Response to Issues Paper on Data Sharing and Release

The Allens Hub for Technology, Law and Innovation is a community of scholars at UNSW Sydney aiming to add breadth and depth to research on the interactions among law, legal practice and technological change in order to enrich scholarly and policy debates and enhance understanding and engagement among the legal profession, the judiciary, industry, government, civil society and the broader community. The views of those participating in this submission are our own, based on our research, and do not represent the official views of UNSW Sydney or Allens.

We were grateful for the opportunity to present our views at a consultation on 30 July 2018; this document attempts to summarise these in writing, taking account of the subsequent discussion at that event and conferral with colleagues after the event.

Timing

As explained in this submission, the work being undertaken as part of this project provides an opportunity to rationalise the current patchwork of laws dealing with how the government shares information internally and externally. However, given the tight timing, there is a risk that, instead, the government will merely add an additional patch to the existing morass of conflicting policies and inconsistent principles. I urge the Department to consider extending the timeline and taking the opportunity to rationalise the existing legal framework. While the delay may be regrettable, the long-term benefits of a clear, principle-based, risk-based legal framework for handling government data will be invaluable.

Patchwork

The current law around information-sharing is unnecessarily complex. The study of law enforcement information sharing as part of the Data to Decisions CRC (copy provided separately at the consultation), revealed a range of issues. Some of the challenges are definitional. For example, different legislation will use different terms (and different definitions of the same term) to describe the object of analysis – is it data, information, communications, records or documents? And are these physical things or digital signals or both? There are also different terms to describe the relationship between such things and particular agencies responsible for them – data might be held, in the custody of an agency, under the control of agency, in the possession of an agency, in the care of an agency, or an agency might be responsible for it or have acquired or obtained it. Again, each of these terms often comes with conflicting definitions.

In addition to definitional issues, there is an issue with the assumption of much legislation about data (or equivalent term) that it is held (or equivalent term) by one entity. The question is then whether it is given to another entity and in what circumstances this is required, encouraged, permitted, or punished. However, none of this works as well with new ways of storing data – a common data platform through which multiple agencies can access data stored on one or more public or private servers does not fit easily into the existing framework.

All of these issues are discussed in the report, albeit in the specific context of law enforcement information sharing. The advantages of a single Act that resolves the current confusion, dealing with all information sharing questions in a principle-based way according to a coherent set of concepts are great. Such an Act can and should recognise distinctions based on the diversity of data and circumstances, but there is no need for hundreds of separate provisions in different legislation using inconsistent concepts and definitions. For example, only a thorough review, based on existing work of the ALRC, can derive a principles-based understanding of the

circumstances in which secrecy laws are appropriate. Our report included recommendations as to how the legal framework could be reworked in order to improve information sharing for law enforcement purposes. These could be combined with this project in order to improve the current complex patchwork laws rather than being excluded from scope.

Data Diversity and factors considered in the data sharing process

The Issues Paper acknowledges the diversity of data in terms of attributes and sensitivity. There are, however, further diversities that need to be acknowledged and addressed. For example, data is diverse in its *quality*. As an example, consider the extent to which a law enforcement database reflects the actual state of crime in a particular location. It is likely that recorded rates of murder will be broadly accurate, but recorded rates of domestic violence and sexual assault will be misleading. Policy responses based on flawed data can be harmful if they ignore real facts that are not recorded or are recorded with systemic inaccuracies. Diversity in data quality suggests that, together with any sharing of data, meaningful information needs to be provided or made available to recipients and potential users as to conditions under which particular data were collected and methods by which those data were filtered, cleansed and otherwise pre-processed, alongside some support aimed at ensuring recipients or users understand the significance of this information.

Context is another important aspect of diversity. The Issues Paper (p8) talks about the different sensitivity between health data and transport data. However, transport data can be sensitive in some contexts. If a person is being stalked by a member of the public service (or other person with whom data is shared), they may be very concerned that that person has information on their commuting patterns. One cannot assume that all data, even within a particular dataset, has the same practical sensitivity for all data subjects in all circumstances.

Another important diversity to recognise is the diversity of perspectives on data in different segments of the community, including as to the kinds of purposes to which particular kinds of data may legitimately be put. This is no doubt evident from the broad range of different views through the submission process, but particular communities may have different attitudes to data. For example, the data sovereignty movement has emerged from within Australian and international indigenous communities. Again, information and training must be directed towards ensuring recipients or users have insight into and are attentive to these kinds of diversity.

A further aspect of data diversity concerns the amenability of data to reuse, repurposing and representation in multiple formats, either in their own right or in combination with other data. Because of this, the purposes to which data may potentially be put may not be entirely foreseeable at the time of its initial sharing. The purpose test and the Five Safes framework need to be interpreted and implemented with this in view. Appraisal of "safe people" should include whether users are well equipped and worthy of trust in relation to potential reuse of data for purposes subsequently identified. Data recipients' and custodians' "safety" should also be assessed, and accreditation afforded, with regard to use that may extend to combining, reformatting, representing, interpreting and explaining data, not just using it for those immediate purposes identified at the time of the request for sharing.

Conditions built into Data Sharing Agreements will need to take account of these factors if safeguards are to be meaningful and effective. Furthermore, care needs to be taken to ensure that accountability and remediation mechanisms built into Data Sharing Agreements and the National Data Commissioner's oversight role take account of the aforementioned diversity.

Risk of uneven data availability and misleading policy

There are a variety of ways in which the evidence-based policy drawing on data analysis can prove deeply flawed. An example from Australia was an attempt a few years ago to integrate federal welfare data with state-based data on children in foster care. A correlation was identified between children who move homes more often and higher welfare needs as adults. The policy proposal (presented at a workshop in Canberra) was to encourage states to keep children in their initial foster home. However, as a statistician in the audience pointed out, there is also a strong correlation between the number of fire engines in attendance and the damage caused by a fire. The reason is a common causal factor (severe fire), so reducing the number of fire engines would have the opposite effect to that intended. Similarly, there may be a common cause for the foster care scenario (such as child with greater needs that fewer foster carers can meet). Similarly, there are a variety of other scenarios, such as child abuse in the first foster placement, that causes greater difficulties settling children into a new home and greater difficulties as adults. Alternatively, there may be worse outcomes than relying on welfare, such as imprisonment, homelessness or death. Faced with this response, I hope that the proposed

policy was never implemented. However, it highlights that developing policy from data analysis requires teams with expertise in policy development, statistics and data science who are able to work constructively together. The recent challenges with Robodebt further highlight the need to consider how policy is implemented given that data science and automated decision-making systems often rely on probabilities rather than certainties, in that case the probability that an individual's income does not fluctuate over the course of a year. Given the risks of bad policy derived from inappropriate inferences drawn from data, or inattention to the uneven distribution of data and reasons for it (i.e., the fact that there will always be more data available on those who have historically engaged with or been the focus of government programs), it is important that policy projects are properly constructed, evaluated and reviewed. The maintenance of social trust in data sharing will be contingent on ongoing demonstration of responsible data use. For these reasons, reporting by the National Data Commissioner ought not to be confined to the promotion of "good news stories", as the Issues Paper suggests, but should also extend to reflexive analysis of, and learning from, "bad news stories", including those arising from lawful, well-intentioned data use (i.e., not just from intentional misuse or unauthorised accessing of data).

Ethics

One concept that receives little attention in the Issues Paper is ethics. When university researchers use individuals' data in research, there is a formal ethics review conducted by a committee. Similar arrangements are in place in hospitals and some government departments, but it is not universal in government and it is rare in the private sector. It is important that ALL uses of personal data, particularly for research, are based on a similar process.

In the United Kingdom, there is a data ethics framework that applies to the whole of government. This was based on local consultation and so there is no reason to believe that an Australian framework would be identical. However, it is worth considering the benefits here of some of the categories dealt with in the UK including proportionality, understanding the limitations of data, training and skills, transparency and accountability as to tools, data and algorithms, as well as responsible use and evaluation. In the UK, there is a "workbook" that government actors are encouraged to work through. Data ethics principles, data impact assessment guides, data innovation risk assessment tools, and related accountability mechanisms, have also been developed, or are in the process of development, in many other national jurisdictions and international organisations.

In Australia, it may make more sense to develop training courses and/or an ethics review board process in sectors where this does not yet exist. In addition, some kinds of data access should be fully recorded and audited with random checks for compliance with protocols. In other words, the specific ethics framework in the UK may not be optimal in Australia but having *an* ethics framework and related guidance and accountability mechanisms developed through community consultation is essential. It is also likely that particular variants of, or elaborations from, any such framework would need to be developed for different kinds of public sector decision-making and policy contexts.

Privacy

The relationship between the proposed legislation and the existing data protection laws (in the *Privacy Act* and elsewhere) needs deeper thought than that which can be conducted in the short time available for this process. In particular, there is a philosophical gap between a consent-based model of data use and a model based on the assumption that more data use is "good for society." The current law has many problems, particularly its patchwork nature, but more extensive community consultation would seem essential in moving away from a consent-based model. In particular, we endorse the views of the submission from Melanie Marks et al on the significance and risks of this shift, suggesting that further analysis and consultation is required. We also agree with what was said there about the need to consider the issue of liability.

Culture

The Issues Paper identifies culture as one challenge preventing beneficial information sharing. The Data to Decisions CRC report (provided separately) sets out findings in relation to culture in the context of law enforcement information sharing. It concluded that pointing the finger at culture is not necessarily helpful. In addition to specific recommendations on improving culture, it pointed out that the complexity of the existing legal framework combined with strong negative consequences of breaching secrecy provisions is an

important element in understanding the rationality of risk aversion. It is also important to appreciate that certain kinds of risk aversion have very specific historical provenance and rationales: risk aversion among Indigenous Australians, for example.

Language

There are some terms used in the Issues Paper that seem inappropriate given the context. For example, where the Australian government uses or shares data it holds, those people should be described as "individuals" or "subjects" rather than "consumers". The government's use or sharing of that data should benefit them *as citizens* and not *as consumers*, even if the ultimate benefit is economic.

The Issues Paper also uses the word "release" (meaning open data release) alongside "sharing" relatively frequently. These are very different things. Disclosing unit record level personal data in the context of the Five Safes for a particular project is very different from open release on the Internet. The re-identification of previously released MBS/PBS data demonstrates the risks of data "release". Once "released", there is no going back, and the probability of re-identification increases over time due to growth in computing power (eg quantum computing) and increase in the quantity of data about individuals available for cross-matching. At the same time, there is no significant advantage in data "release" at unit record level – anyone with a serious interest in research can create circumstances that are higher up the scale of the Five Safes than the open Internet.

Rule of Law

There are a variety of important questions that come into play where data is used by government to make decisions affecting individuals. The requirement of equality before the law, for example, renders problematic discriminatory uses. Should an individual be treated differently by their government because people "like" them (for example, demographically) have behaved in a particular way in the past? These are important questions, and the subject of existing research within the Hub.

Participants: Lyria Bennett Moses, Ross Buckley, Fleur Johns, Graham Greenleaf, Katharine Kemp, Marc de Leeuw, Kayleen Manwaring, Alana Maurushat, Monika Zalnieriute