Submission in response to:
Data Sharing and Release Legislative Reforms Discussion Paper & Privacy Impact Assessment

Dr Luke Heemsbergen, Deakin University.
Dr Alexia Maddox, Deakin University.
Dr. Robbie Fordyce, Monash University.

We are thankful for the opportunity to respond to the data sharing discussion paper and impact assessment. Their dissemination and request for response aligns to practices of good government, which we are happy to be part of.

Further, the need for review of data sharing and release is real. The optimism of an inherent public good found in open and shared data has been continually eroded since the early 2000s. The post-war development of 'Open Government' policies in the US was specifically aligned as a capitalist-democratic 'good' in opposition to the Soviet Union's information control system of censorship and propaganda. Yet experts (eg. Charalabidis et al 2012; and Orlandi et al 2015) note increasing and significant skepticism regarding the value of open data, particularly with regards to mythologies of positive impact and utility without sufficient assessment of context-specific impacts.

This is to say that the same data is used by different actors for different purposes, most of which custodians of data are *unable* to foresee; data use is generative by definition. As a result, Public Data as a process and resource needs to plan for both socially beneficial and socially detrimental use cases in order to build resilient data sharing. In the current environment, the mere existence of data stores - released, shared, or kept secret - can be as much a liability as an asset on the balance sheet of any public or private organisation; see, for instance, Equifax (Equifax 2019), ANU (ANU 2019), the Australian Government (Packham 2019), or the password leak monitoring by Microsoft security researcher Troy Hunt (Hunt 2009, 2019).

Empirical cases such as these falsify again and again the inherent good of ever more data. Instead, we have learned that data is a resource that can be and will be exploited in all senses of the term, regardless of the intentions of its distributors or 'accredited custodians'. Capacity to re-identify data or its use in aggregate for future discriminatory scenarios only increases with every additional dataset available; there is no delete key, there is no undo, and - with divergent and variable data literacies across populations - many individuals have little specific awareness of what data exists about them or how it is used.

From this revised understanding of how society puts data to use, we respond via four points to the discussion paper and commissioned privacy impact assessment.

1. We applaud moving past a binary of open/closed to consider that, how, when, and with whom data is shared matters. **To enable operationalisation of the policy goals enunciated within this spectrum we propose cryptographic Data Sharing Agreements** (DSA) to effectively and efficiently govern as follows:

- Data Sharing Agreements (DSA) should, like Creative Commons licensing, be articulated on three levels: a 'human-readable' level for the layperson, a 'lawyer readable' level for legal contexts, and a 'machine readable' level to enable automated data management in line with legal and humanistic obligations. More than providing 'comprehensive' and 'basic' enunciations, a DSA should in itself computationally govern how attached content can be utilised. One might turn to encryption (and/or digital rights management) for those instances where DSAs promote specific ways of sharing vs. public release. Further, any product made via a DSA should offer its DSA license as part of that product; the publication of DSAs should be required to record any uses or modification (eg. with combination of other datasets) of relevant data that creates new data.

- To implement this requirement, each DSA's computer readable layer could be cryptographically unique as they are issued to organisations, with any modifications noted on a distributed ledger. As partners' data use evolves, so to must their disclosure of that use. The utilisation of public resources (by public or private actors) should be accounted for and visible for scrutiny. Such a decentralised ledger of the DSA mechanism would allow DSAs to function as a powerful and equitable form of governance to data sharing.

- Such a system would enable and enforce auditing forensics: DSAs should also be versioned and timestamped to include both their origins and transactions between parties, including information about when and how data was obtained both by the issuer of the DSA and by those that obtain data from an open data provider. This would include information about historical DSAs that a particular data packet may have been released under, the date of issuance, the terms of use, references to any memoranda of understanding, the legal authority for notification in the event of misuse, the accredited receiver, and be cryptographically signed on release.

- Relatedly, we also applaud the need expressed in the Impact Assessment for third parties (accredited or service providers) to include a 'user friendly public information resource [on] Core data sharing and data release activities; Data sources; and a register of data sharing agreements' and believe the above mentioned DSA implementation offers a way forward to operationalise these requirements in an effective, efficient and standardised manner that codes legislation into possible actions.

2. We note the impetus for considering 'a purpose test to maximise public benefits while meeting community expectations.' We understand the sentiment (academics too must now meet this requirement for public resources), **but point out the limits in framing data sharing policy on the knowns versus the unknowns**. It is the unknowns that are of concern and that sharing policy needs to address. We raise the following points in relation to this:

- While a holistic purpose test is indicative of the profile of a specific project, it is not indicative of the profile of the data that enables it. Data is generative and we should assume it can and will be

used in ways unintended by the actor/project seeking maximum public benefit, and protect it as such.

- While evaluating DSA on appropriateness of project, authorisations and environments of sharing, protections, and outputs is necessary, it is not sufficient. Evaluation efforts need to focus substantively on the exceptions - that which happens outside the appropriate and advertised uses of the specific data.

- As such, we can employ models similar to those used by security researchers to evaluate risk and impact. Security researchers have long used 'penetration testing' to evaluate risks of systems and a corollary 'inequity testing' methodology creates a useful parallel for sharing data. Whether 'privacy' bounties, prizes, or other incentives, we should welcome 'hacking the data' for public good and institutionalise these practices. In addition, intended uses may create unintended social logics that, instead of benefiting people, further subject them to forms of social inequality. 'Hacking the social' for public good has the capacity to make unintended consequences of data applications and services visible that may reinforce existing social inequalities or reify new forms of social exclusion. This approach may take the form of probes derived from upon open data products to test for social backlash and logics of exclusion and inequity.

3. We applaud the vigorous debate on consent. **It is of our opinion that consent is untenable considering the generative nature of data use over time and space.** Relatedly, we acknowledge a 'Government purposes spectrum' of data use from productive to coercive, which includes 'assurance, compliance and national security and law enforcement activities'. However, the focus on data custodians rather than external agencies inhibits meaningful policy and brings into question any efficacy of governance for these concerns. As such, we pose two questions and a suggestion that need to be more adequately addressed:

- What does it mean to sharing practices when we must treat each data product as if consent was not given by those whose data went into it? Citizens and public institutions cannot consent to every conceivable permutation and combination of their data, and thus we should not pretend anyone has the agency to do so. Researchers have very strict rules around data where consent has not been granted, what of the government?

- How will directives that state 'Data Custodians will not be able to share data for compliance and assurance activities under the Data Sharing and Release legislation [or] national security and law enforcement purposes under the Data Sharing and Release legislation' be implemented if the relevant departments are not required to disclose their potentially related activities nor where they leveraged their data from? How do we effectively legislate that, for instance ATO/ASIS cannot use this data for assurance, compliance and/or national security and law enforcement activities once 'shared' with third parties or more widely released?

- Our initial suggestion to these rhetorical questions is that all shared data must be considered as available for assurance, compliance and national security, and thus released only at levels of

fidelity that directly consider these concerns. With regards to notices about release, we would also encourage the development of a database of notices regarding state and/or federal open data to be held by the relevant authority in order to facilitate awareness of the diverse contexts of data release.

4. **Clouds are not control.**

Cloud services, are virtual, dynamic, and potentially stateless, which should trigger concerns over data sovereignty (Irion 2012). Within Australia, this context is complicated by the cross-jurisdictional nature of many data stores affording insufficient or multiple oversight depending on the data warehousing decisions of multiple private entities. This is particularly nuanced in Australia, where state and federal entities have no restriction on the use of extra-national cloud hosting services, unlike the US/UK (DTA 2017). These complications of the Australian context are not well matched by the relevant existing, mostly US-based research materials while, equally, the use cases in Australia are also not well researched. Simply put, Australian Public Data is poorly prepared for detrimental use cases and needs circumspect research and planning before go-live exposure risks.

In conclusion, our suggestions are to facilitate translation across contexts, to provide better mechanisms for forensics of use and potential abuse, and to emphasize the multi-contextual nature of open data that provides specific Australian particularities in regards to release, sharing and security.

Bibliography

ANU 2019. Data Breach. Australian National University. 4 June 2019. Available at: https://www.anu.edu.au/news/all-news/data-breach

Charalabidis, Yannis, Zuiderwijk, Anneke & Janssen, Marijn, 2012. Benefits, Adoption Barriers and Myths of Open Data and Open Government. Information systems management, 29(4), pp.258–268.

DTA 2019. Secure Cloud Strategy. Digital Transformations Agency. https://dta-www-drupal-20180130215411153400000001.s3.ap-southeast-2.amazonaws.com/s3fs-public/files/cloud/secure-cloud-strategy.pdf

Equifax, 2019. Equifax Breach Settlement. Available at https://www.equifaxbreachsettlement.com/

Hunt, Troy. 2009. Troy Hunt.com

Hunt, Troy. 2019 Have I Been Pwned. Available at: https://haveibeenpwned.com/

Ølnes, Svein , Ubacht, Jolien , & Janssen, Marijn. 2017. Blockchain in government: Benefits and implications of distributed ledger technology for information sharing. *Government Information Quarterly*. 34, pp.355-364.

Orlandi, Fabrizio et al., 2015. A systematic review of open government data initiatives. Government information quarterly., 32(4), pp.399–418.

Packham, Colin. 2019. Exclusive: Australia concluded China was behind hack on parliament, political parties – sources. Reuters. 16 September, 2019. Available at: https://www.reuters.com/article/us-australia-china-cyber-exclusive/exclusive-australia-concluded-china-was-behind-hack-on-parliament-political-parties-sources-idUSKBN1W00VF