Australian Government Data Sharing and Release Legislation

Issues Paper for Consultation

August 2018

Suneel Jethani and Dale Leorke

University of Melbourne

This paper was drafted by the authors as a submission to the Australian Government Department of Prime Minister and Cabinet's Data Sharing and Release Legislation Issues Paper. We have done so in our capacity as researchers currently engaged in the study of Australian Government workers attitudes and behaviours towards data sharing and release. This research has been conducted in partnership with the Melbourne Network Society Institute (MNSI) at the University of Melbourne. The Issues Paper is a document that outlines principles that may guide the development of data sharing and release arrangements between government departments and, in some cases, its release into the public domain via the Australian Government's open data portal, MAGDA. In our submission we consider whether the concepts and principles outlined in the issues paper adequately acknowledge contextual factors relating to purposes for which the data was originally collected and potential issues which may arise through its reuse and/or release into the public domain. It is our view that the acknowledgement of some of the definitional ambiguity around terms such as 'sharing', 'release' and 'data' is an important aspect of risk analysis, legislative interpretation and scalability of any proposed framework which might attend to the attributes and sensitivities of data sharing, analytics and publication by government departments and agencies.

Introduction

The object of a new Data Sharing and Release Bill is to promote sharing and release of data between Commonwealth Entities, Commonwealth Companies, and in some cases the Australian Public in order to better inform government decision-making, foster greater transparency, stimulate growth and innovation, and shed light on complex emerging social, cultural, and economic problems. To this end, the Bill aims to alter the approach to data sharing and release from one which restricts data sharing to one which enables the demand for data sharing and release to be authorised through a frame work guided by a five-part model which considers: (1) data type, (2) the proposed and potential users of the data, (3) the access environment for the data, (4) the format and level of granularity of the data that is made available, and (5) the purpose of use. This model is commonly known as the 'five safes' framework. In what follows, we draw from the preliminary findings of our research which includes fifty semi-structured interviews and an online survey of actors working in or with federal, state and local governments involved in various government data analytics and publishing initiatives to:

- Outline some of the definitional ambiguities within the proposed data sharing and release framework
- Raise a number of questions and issues in relation to the terms of reference and scope of the proposed framework.
- Consider how assumptions about risk and trust might influence the proposed data sharing and release authorisation process.
- Examines some of the social and ethical dimensions of designing the proposed framework

Throughout our paper several modest recommendations are made for ways that might improve or extend the purview, scope and readiness of the proposed legislation and framework.

Definitional Ambiguity around the term Data

Data, information and knowledge are often considered in relation to one another and as being linked by processes of analysis and interpretation. Thinking broadly, data is understood as something that can be collected and analysed to create information which can then be used to make informed decisions. By contrast, knowledge is thought to be derived through *experience* gained through routinely dealing with information about a particular subject or situation (Beynon-Davies, 2002). A problem that many of our research participants have identified is that although data, information, evidence or knowledge are abstract concepts if data is thought of in material terms, as a physical object, then factors which relate to its original purpose, its potential to be mobilised, and the value that is envisaged for it can be considered in ways that help predict how the assessment of data sharing and release requests might become scalable and transferable across government.

Further, in defining data within the framework we recommend that the data is defined in expanded terms that incorporate the extent to which the data has been collected passively as a by-product of certain government activities that are documented and recorded or actively produced with a pre-defined purpose in mind. In order to preserve the integrity of the national data system we recommend that documentation of the processes involved in deriving insights from data being sought occurs and that this be outlined in the legislation and the accompanying authorisation-seeking protocols. We argue that this produces an extended provenance for given datasets that allows for the potential for data re-use to be understood, and for methods and techniques for data use and re-use to be shared between government departments and agencies.

Scope and Purpose of the Proposed Framework

Our research identifies some gaps that might be appropriate to consider in defining the scope of the legislation. The proposed Bill will accredit trusted users and may be extended to encompass States and Territories and/or other non-government actors for pre-specified purposes under certain conditions. Our preliminary research data suggests that in some cases there may be deep interdependencies on intermediaries for the provision or maintenance of certain data assets and we recommend that the legislation be widened to encompass intermediaries in the data supply chain at the organisational level. In our view inclusion or exclusion criteria should be set through open, consultative and documented channels at the level of specific data categories *as well as* by their proposed upstream uses.

We support legislating a purpose test provided there is sufficient plasticity for applications falling outside the purposes of informing government policy making, supporting the design and delivery of government services and policy or when a clear public benefit has been identified. We recommend, however, that government operations be treated as a separate element of the purpose test with distinct compliance mechanisms in place. We feel that by treating government operations as a separate category within the purpose test the Bill will align more closely with its goal of achieving greater transparency around government activities and spending.

Our research also indicates that the purpose test could be used as a mechanism for engaging data custodians and users in dialogue in regards to the use of government data by

academic institutions we recommend that the Bill mandate that when research based on data made available by these provisions that it is published in open access journals with appropriate attribution and acknowledgement of not only the data sharing agreements but the individual data sets used as cited objects (e.g. by URI or DOI).

Assumptions About Risk and Trust

The assumption of neutrality in the analysis of data that is free from human bias is not necessarily acknowledged in the proposed process for authorising data access requests. Neutrality of objectivity of the data may not be inherent with the dataset which makes the unearthing and revealing of patterns and trends particularly fraught. It may be that understandings of analytical methods may be assumed to mirror the tenets of the scientific method though in practice they may be 'more open to using a hybrid combination of additive, indictive and deductive design … that seeks to generate hypotheses and insights' (Kitchin, 2017: 5) that may or may not be reasonably attributable to the data being sought. We see a need to articulate the type of questions being asked of the data at two stages in the authorisation process: at the stage that a request is triggered and then for consideration when the purpose test is being applied.

While the five safe system is widely adopted and recognised as a robust standard– one wonders how detailed will the criteria for these questions be? One issue that has come up in our research is that such guiding models can be vague and open to interpretation, such as the stipulation that data that is not "sensitive' won't be released – "sensitive" can mean many things, including data that might embarrass or reflect poorly on the government agency or policymakers. This can lead to conservative approach to leading data, contradicting the imperative to make data that is useful being made open. We propose that in order to enhance the use of the five safes model as a guiding principle for the Bill that a series of sub-questions and categories be formulated in consultation with relevant stakeholders.

Our research indicates that there is much confusion around Australian Government Workers' perceptions about the risks associated with data sharing and release. The different modalities of risk that our preliminary results suggest are based around: *knowledge* and *experience* of dangers and harms associated directly to government data sharing processes which can often be conflated with other more mainstream accounts of data related social harms such as large data breaches, hackings or misuse (e.g. Cambridge Analytica, celebrity photos). *Perception of government* attitudes towards technology and its associated risks and the struggle over interests pertaining to public good, *economic* benefits as a trade-off for any perceived risk and as viewing data sharing and release through a *principle of precaution* which could be subsequently relaxed incrementally as norms and comfort thresholds are established around data sharing practices.

We appreciate recommendations proposed will require further consideration as to establish how they could be implemented within the proposed framework and which would require significant revisions of the ideas mapped out in the issues paper.

References

Beynon-Davies, P., 2002. *Information systems: An introduction to informatics in organisations*. Palgrave Macmillan.

Kitchin, R., 2014. *The data revolution: Big data, open data, data infrastructures and their consequences*. Sage.