

We would like to present the following feedback in response to the Office of the National Data Commissioner's Discussion Paper, *Data Sharing and Release Legislative Reforms*.

The paper represents a timely and much needed step towards a more cohesive and manageable data sharing environment. There are, however, some general inconsistencies as well as areas that are not sufficiently lucid. Overall, it seems that the paper and the accompanying legislation is attempting to do too much, too soon.

Data Sharing or Data Linkage?

The paper seems to use the concepts of data sharing and data linkage interchangeably, even though these two could be considered conceptually distinct, each with its own practical and legal concerns. Each has its own particular benefits, data sharing supports the concept of better services and better government, whereas data linkage supports better decisions.

For example, the paper describes an example on page 24, of "better service delivery" based on citizens only having to provide a medical certificate once to either the NDIA (for assessment of NDIS participation) or Centrelink (DHS) (for Disability Support Pension eligibility) and then asking the government to share that "data" rather than submitting the information twice. This relies on the agencies involved being able to map identities across each system with 100% accuracy. In the absence of a shared unique personal identifier this is challenging. As pointed out on page 24, pre-filling of myTax is achieved by the ATO but this is presumably through the use of the person's TFN. It is not clear that NDIA and DHS would share a unique personal identifier.

Conversely, the "better informed policy" example, in the tri-fold glossy, involves data linkage of ABS and DET data. There is no identifier allowing ABS Census data to be shared with the DET. A project like this – which is an element of the The Multi-Agency Data Integration Project – relies on probabilistic linkage. This is not 100% accurate but sufficient for population-level insights as per this example and the "better informed research" example, in the tri-fold glossy, using PBS data.

Data sharing uses real-time data and requires a perfect match at the person level, whereas data linkage uses point-in-time extracts of data sets and does not achieve a perfect match.

Data sharing and data linkage are very different concepts, particularly with respect to the interests of citizens. In data sharing, governments are sharing data with the citizen's consent in a transparent manner and would be accountable for data accuracy and security.

In data linkage, the citizen has no knowledge of the link or whether their data was included in the linkage. Furthermore, the average citizen would potentially have little interest in this scenario because the agencies involved in linking would be accountable for proper de-identification and data separation and there is less scope for ramifications at the individual level.

As such, data sharing and data linkage should be dealt with separately (in separate Discussion papers, with separate legislation).

It appears that section 6.4 introduces another term for data linkage – integration. It would be useful if terminology is consistently used throughout the paper.

Private sector use for commercial applications

Section 3.4 states the NDC will champion the greater release of open data, including for commercial uses. Will the government receive payment from the private sector for open data?

On page 9 it states that the legislation will never authorise the release of personal information as open data but on page 7 the risks of re-identifying ‘de-identified’ data are acknowledged. Is enough being done to protect the privacy of citizens in open data?

On page 27 the purpose test is mentioned. How will the purpose test be applied and what prevents private sector entities from stating one purpose in their application but then using the data for another, unacceptable purpose?

Consent

Section 4.6 explicitly states that consent will not be required in all instances. Will citizens have the capacity to remove consent of the sharing of their data, from government? Importantly, will citizens have the capacity to remove consent of the sharing of their data with researchers and/or the private sector? If consent is revoked how will the legislation dictate the removal of that individual’s records from shared data sets?

Increasing Transparency and Data Sharing Agreements

The paper mentions, in many places, the idea of increasing transparency without actually providing any tangible insights into what this would mean for Australian citizens. True transparency would mean that a citizen could log onto their MyGov portal and see all the data that the government has and control which departments have access to that information. It would also allow the citizen to decide whether researchers and the private sector can access their data.

Merely maintaining a register of Data Sharing Agreements and data linkage projects does not promote transparency or accountability. Will the register give citizens the ability to know exactly what data of theirs is being held and linked (or shared) and give citizens the option to not participate? This could be achieved in MyGov, as described above.

A National System

An overarching legislative environment, as outlined in Figure 4, would make it easier to link data both at the state-level and with the Commonwealth. This may not, however, be very relevant for data sharing – without a unique person identifier the scope for a citizen asking for, or allowing, the sharing of state-level data with the Commonwealth is limited because 100% accuracy cannot be guaranteed.

Strict safeguards need to be in place to ensure that any data sharing is validated and 100% accurate; failure to do so may result in more Robodebt scenarios.

Data burden?

Will this legislation create a data sharing “regime”, with citizens required to provide documentation or authorise greater sharing of data for each and every service/transaction with government? The Discussion paper does not touch on how government will operate in a data-sharing environment and if the supposed efficiencies of not having to supply information more than once are designed primarily to benefit the citizen (who in actuality retains more control of their data by having the opportunity to supply updated or different data in each transaction) or benefit the government since stricter controls can be placed on a citizen’s access to government services. Particularly, can this data sharing be used to create an even more onerous regime for welfare recipients and users of government services?

More importantly, would a data sharing environment allow or even encourage departments to shift the burden of collecting data from citizens around, forcing citizens to engage with multiple departments under a scenario where a department chooses not to collect certain data and therefore will not proceed until the citizen provides these data items to any number of other departments who will then share these data?

Will a data sharing regime force all citizens to use MyGov, essentially coercing them to agree to its terms and conditions? Will citizens be forced to data share to access a service or, as in the NDIA/Centrelink example (on page 24), can they still provide separate documentation to two departments if they prefer?

What tolerances will a data sharing regime have in place to support people across the digital divide? Will people with disabilities, people with limited access, experience and familiarity with digital systems, people with incomplete documentation (the homeless, refugees, victims of domestic violence) or people who lack English-language literacy, still be able to access services without sharing the required data?

Data Sharing Not Authorised

This is a critical area where the Discussion Paper does not provide enough clarity. It is unclear how data sharing will relate to assurance activities and service provision (Figure 5). The assurance

clause states that government agencies cannot use data sharing to determine eligibility for government programs and services. However, the example given on page 24 relating to the NDIA and DHS and data sharing of a medical certificate, suggests the opposite. In this example, data sharing is being used to determine eligibility for both NDIS and the Disability Support Pension.