

2016



Australian Government
Australian Taxation Office

ATO SUBMISSION TO PRODUCTIVITY COMMISSION'S INQUIRY INTO DATA AVAILABILITY AND USE

December 2016

The Australian Taxation Office (ATO) welcomes the Productivity Commission's draft report on *Data Availability and Use*. The ATO considers that, if adopted, the recommendations will help the ATO make more data available:

- for other government agencies and researchers to build a richer evidence-base to inform policy and administration; and
- for the public service and business to improve service delivery for the community.

This submission provides some insights based on the ATO's experience and outlook that we hope will be worth considering for the final report.

1. Supporting evidence-based policy and administration

The ATO is developing a longitudinal individuals tax file for policy research. This builds on the ATO's history of making data available for researchers, including releasing annual confidentialised unit record income tax sample files for almost a decade and publishing aggregate annual tax statistics tables for around 100 years.

The Australian Longitudinal Individuals File (*aLife*) will enable researchers to develop a deeper evidence-base to inform policy and administration.

The *aLife* file will be made available to researchers in the Sax Institute's Secure Unified Research Environment (the SURE) from mid-2017. Researchers will initially be from approved research institutions including universities, approved not-for-profit institutions and government agencies.

Approved researchers will be able to log into the secure 'trusted access' environment from their remote desktops. They will be able to see de-identified unit record data, but will be prevented from printing or taking data out of the secure environment (or bringing other data into the secure environment without approval). Researchers will be able to publish aggregate results once they have been checked to ensure that those results cannot be used to re-identify taxpayers. This approach represents a shift in the ATO's practice from confidentialising and perturbing data inputs to deidentifying data inputs and then checking confidentialised aggregate outputs leaving the secure environment.

The ATO is developing a household version of the file to enable household and intergenerational analysis, and plans to extend *aLife* to include employee-employer links. However, for the latter to be widely utilised by researchers, it would require law changes to allow ATO to disclose to trusted researchers ATO datasets that include potentially re-identifiable large businesses. This is because the largest businesses in highly concentrated sectors may be easily re-identified using any useful set of combined characteristics.

The ATO also sees additional value in linking *aLife* with other data sources where allowable under the law and is exploring linking non-tax data from partner agencies to *aLife*.

This project has given us several insights into making administrative data available for researchers:

1.1 The role of custodians

Administrative data needs to be curated by the data custodian. The ATO is the largest collector of financial data in Australia. As a data custodian, the ATO is ideally placed to curate data and ensure its data is fit for purpose. In particular, as the data custodian, we have extensive knowledge and experience in the collection methodology, storage, data limitations and system expertise. For example, we are able to identify and resolve issues with administrative data using ATO's internal systems (rather than just data) and sectoral expertise, including knowledge of changes in tax laws and how these flowed through to data holdings. Given this, the ATO considers that custodians are best placed to curate their data holdings, rather a centralised body.

Data custodians should work with researchers. The ATO has identified and resolved issues with administrative data by working with researchers. While we had shared this administrative data with other agencies for several years, it was not until we worked directly with researchers that those issues were identified. This highlights the importance of data custodians working with researchers to ensure

the quality of data. There are benefits from each data custodian keeping their 'door open' to researchers to get insights on the broader use of its data. Otherwise, potentially valuable data will sit in silos or not be collected because it is not seen as a high value asset. In addition, by working with researchers, data custodians can sound out ideas to best address trade-offs between utility for research and protecting privacy. The ATO has benefited from working closely with its Academic Advisory Group and Sir Roland Wilson Scholars. The benefits of these collaborations would not occur if there was only 'one centralised door' to access administrative data for the whole public service.

Data custodians should **curate data for research and integrate it with other datasets through virtual infrastructure, while maintaining some control**. This reduces issues with version control and transferring data to a central repository. For example, the ATO can best manage amendments and late lodgments over time and manage risk of sensitive taxpayers, such as the removal of certain sensitive taxpayer records. Custodian control is also needed to reduce the risk of a cyber-security attack and unintentional disclosure by other agencies, thereby helping maintain community confidence and willing participation in the system.

The ATO considers that there would be benefits from establishing an independent data office, which would provide guidance on confidentiality and disclosure rules and deed templates. This would particularly be useful should agency specific legislation become subordinate to the proposed *Data Sharing and Release Act*. This is one potential role for the new National Data Custodian. Another important role of the National Data Custodian could be to provide a framework, and potentially offer risk assessments, for data released in secure environments for 'trusted users'. These risk assessments should identify risks focusing on what is allowable under the relevant law and help agencies implement mitigation strategies.

The ATO supports the Productivity Commission's model of Accredited Release Authorities for data integration and access platforms. This model puts competitive pressure on entities to reduce costs and improve service offerings. Further, this would improve agency and researcher understanding of the benefits and shortfalls of specific data, and ensure that the agency's understanding of data for its sector is brought to bear on integration issues. The ATO would be willing to serve as an Accredited Release Authority.

1.2 Trust in the system

The ATO agrees that it is important for the public service to build a social licence to increase data sharing. The Productivity Commission should propose a model to support this.

As part of this model, drafters of the proposed *Data Sharing and Release Act* should consider how the new law would interact with existing legislation, because inevitably there will be inconsistencies. The ATO considers that the proposed overarching principle-based law reform is important to enable better use and accessibility of data. Investment in technological solutions by itself would not be sufficient.

To maintain trust in the system, the proposed *Data Sharing and Release Act* should establish penalties for users who breach rules around data use – harmonising the patchwork system of existing laws and extending penalties to trusted users. It is also important to consider cross-jurisdictional issues where the trusted user resides overseas, issues raised by commercial data held in government agencies, and information sharing on non-compliant users.

The Productivity Commission should consider how to address the issue of cumulative disclosure and differencing that may arise from releasing aggregate data from a secure environment. Aggregate data released from a secure environment may not be identifiable by itself, but differencing between aggregate data may result in taxpayers (or other entities) becoming identifiable. This could create 'externalities' whereby initial researchers wanting to take detailed aggregated tables out of secure environments impose constraints on future researchers. To reduce this risk, the ATO's deed with approved researchers will limit researchers from taking large amounts of aggregate data out of the secure environment. The ATO will also enable researchers to collaborate within the secure environment (so only final output is taken out of the secure environment).

1.3 Transition

The Productivity Commission should **set out an implementation timeline**. The Australian Public Service should focus initially on each data custodian improving the quality of data and sharing its data holdings, building the social licence and introducing the *Data Sharing and Release Act*. After this investment in developing quality datasets and law, the second stage could involve investment in technology to allow multi-party computation and integration of data curated by data custodians.

Taking a step-by-step approach may help to manage the considerable resources that may be required to implement some of the recommendations in the report if they were adopted. This problem would be magnified if all Government agencies were seeking relevant resources at the same time.

The Productivity Commission's final report should provide **guidance on the pricing of access to data and ownership**, with an eye to maximizing the public benefit. Full cost recovery on data access limits its use and means that Australia does not get full value for this asset and may also erode public trust. The report could have strong warnings against privatising public data and imposing paywalls. Further, marginal costing may not be appropriate for the curation of data that becomes a valuable public good (marginal costing being an amorphous concept, particularly when data can be reused and one person using it does not diminish someone else's use).

2. Supporting improved service delivery for the community

As the Commission is aware, this review is taking place at the same time as there is a strong focus on investigating ways of improving the efficiency of the public service delivery, based on shared information and shared technology.

The Productivity Commission could consider expanding its recommendations in light of the growing interest in '**real time' data sharing to improve service delivery**. This would present a value proposition for the community because it should only have to 'tell the government once' about changes in circumstances. It would enable the public service to provide better service at a cheaper cost for the community. For example, several agencies across government (including the NSW State Government) are connecting to the Australian Business Register to populate their client relationship systems. This reduces red tape for business and improves the quality and use of data.

In terms of particular recommendations, the ATO is supportive of the Productivity Commission's recommendation that citizens have greater control over the data held about them by government. The ATO agrees that citizens should be enabled to share their government data directly with third parties. Apart from anything else, this enables contestability in service offerings. This value proposition would help build the social licence for increased use and sharing of data under the proposed *Data Sharing and Release Act*. In addition, the systems required to enable real time sharing of data would help improve data management across government.

It is worth noting, however, that there would be significant resources required to implement this access (and some other aspects of the Productivity Commission Proposals for consumer control) for the third parties as well as government. Further, if the Government enables citizens to transfer their personal data from government to third parties, it may be worth considering whether there should be consumer protections on use, retention and destruction of that data. For example, consumer protections to enable citizens to ask third parties to destroy data after use or when requested where it is otherwise not required to be kept by law.

The Productivity Commissioner could **outline a framework for agencies to collect data beyond their immediate needs**. The framework could be that each agency could collect data on behalf of government (potentially across jurisdictions) based on their comparative advantage, in terms of types of clients they interact with.

Thank you again for the opportunity to comment.