

Friends of the Chair Paper

5 The Institutional Transformation of NSOs – New methods, services and roles

(Moderator Mariana Kotzeva)

b) New services including microdata access and linking. ([Eurostat and Australia to provide discussion document from Global/National perspectives](#))

Introduction – The Broader Context

The Australian Government recognises that the data it holds is a strategic national resource and that greater use and sharing of public data facilitates opportunities for enhanced productivity, increased competition, improved delivery of government services and research outcomes.

The Secretaries Data Group (SDG) and Deputy Secretaries Data Group (DSDG) (both chaired by the Department of Prime Minister and Cabinet) provide governance of public data initiatives across Australian Government entities. A network of senior Commonwealth officials, referred to as the Data Champions, promote the use, sharing and reuse of data across entities.

In the Australian context, the Office of the National Data Commissioner is responsible for implementing a simpler data sharing and release framework across government. The framework aims to break down the barriers preventing the efficient use and reuse of public data, while maintaining the strong security and privacy protections that the community expects. The ABS works closely with the Office on data sharing issues.

Long standing data sharing arrangements across the public service are complex and typically hinder the efficient and effective use of data.

Barriers to greater sharing of data within government include:

- a dense web of legislative requirements which lack consistency
- a culture of risk aversion, leading to overly cautious legislative interpretation and approval process complexity
- the lack of a whole-of-government approach.

The Australian Government's Public Data Policy Statement was released on 7 December 2015. This statement provides a clear mandate for Australian Government entities to optimise the use and reuse of public data; to release non-sensitive data as open by default; and to collaborate with the private and research sectors to extend the value of public data for the benefit of the Australian public.

The ABS's Role in Data Sharing and Integration

As Australia's official data agency and a prominent member of the SDG, DSDG and Data Champions fora, the ABS has a key role in the whole of government approach to the sharing and reuse of data. This is inherent to our role in maximising the value of public data and statistics.

Combining datasets is an efficient and effective way of creating new insights about communities, families, industry and the economy and potentially improves the development and delivery of government services in areas such as health, education, infrastructure, and other community services.

Data integration is also more efficient and less costly than undertaking surveys, and can help to reduce the burden on Australian households and businesses to provide data that is already available.

The ABS began to build its data integration capabilities and assets in the early 2000's. In 2017-18, the ABS received additional funding under the whole-of-government Data Integration Partnership for Australia (DIPA) to expand its data integration capacity and assets. DIPA is an initiative to make better use of existing public data.

A number of Commonwealth agencies contribute data and policy analysis skills to the program, while others provide technical expertise and infrastructure. The partnership enables statistics to be produced in a more streamlined and cost-effective manner, and improves our ability to analyse data.

The creation of DIPA is a further maturity and centralisation of existing data integration projects:

- The Multi-Agency Data Integration Project (MADIP) is a partnership among Australian Government agencies to develop a secure and enduring approach for combining information on healthcare, education, government payments, personal income tax, and population demographics (including the Census) to create a comprehensive picture of Australia over time.
- Business Longitudinal Analysis Data Environment (BLADE) combines business tax data and information from ABS surveys with data about the use of government programs, to provide a better understanding of Australian businesses and the economy.
- Australian Census Longitudinal Dataset (ACL D) uses data from the three most recent Censuses to create a research tool for exploring how Australian society is changing over time.

Further details regarding these project is provided as an attachment to this discussion note.

Microdata Access and Linking

Expanding safe access to microdata

In October 2018, new Australian legislation expanded the ABS' ability to release firm-level data for statistical purposes. The legislation prohibits the release of direct identifiers (such as name and address) but otherwise permits the release of information where there are risks of re-identification, with the intention that those risks are managed in the context of a 5 Safes Framework.

The 5 Safes Framework is a multi-dimensional approach to managing disclosure risk. Each 'safe' refers to an independent but related aspect of disclosure risk.

Currently, firm-level data is made available by the ABS through its DataLab system. The DataLab system is a secure system wherein users can analyse information where there is a risk of re-identification. The DataLab includes built in protections to detect and prevent users from intentionally or unknowingly taking re-identifiable information out of the system.

When supported by strong administrative processes, access to firm-level data within DataLab can be achieved in accordance with a 5 Safes Framework by:

- requiring users to register, sign legal commitments binding their use of the information, and attend mandatory training ('safe people')
- requiring users to provide justification for accessing the data ('safe projects')
- assessing that the user's systems are appropriate in preventing unauthorised access ('safe settings')

- assessing the disclosure risk of the underlying dataset prior to it being made available to users in DataLab, and employing manual and automated protections to reduce unacceptable risks ('safe data')
- managing the disclosure risk of a user's output by requiring all outputs to be vetted by ABS staff prior to being released from the DataLab system ('safe output')

Discussion Points

The ABS has had to negotiate a range of issues in arriving at a point where we now are significantly better able to provide a service to the Australian government and its associated research communities.

The issues the ABS continues to address include:

- The flow of information within government
- The contested space in the role of statistical institutions as data integrators
- Social Licence & Trust – the balance between privacy and maximising the use of data
- Meeting the demands of researchers and analysts in providing reasonable access to sensitive data
- Growing skills within our own organisation as well as across government
- Moving into the Cloud as a means to meet an exponentially growing demand
- How do these issues align with the experience and expectation of others?

Attachment

The Multi-Agency Data Integration Project (MADIP) is a partnership among Australian Government agencies to develop a secure and enduring approach for combining information on healthcare, education, government payments, personal income tax, and population demographics (including the Census) to create a comprehensive picture of Australia over time.

Authorised researchers can use unidentified MADIP data to look at patterns and trends in the Australian population, and provide new insights into the development and evaluation of government policies, programs, and services (such as healthcare) to ensure they are delivering value to the people and communities who need them.

Examples of approved MAPID research projects include:

- Measuring the social return on investment from education and training (2018)
- Use of government services by older Australians (2018)
- The prevalence and impact of mental health disorders on income support and student outcomes (2018)
- Integrated analysis for sustainable regional development Northern Australia (2018)
- Pathways for workers affected by industry downturn (2018)
- Examination of the gender wage gap (2018)

The Business Longitudinal Analysis Data Environment (BLADE) BLADE contains data on all active businesses from 2001-02 to 2015-16, sourced from:

- Department of Industry, Innovation and Science (DIIS) programs
- the Australian Taxation Office
- Intellectual Property Government Open Data, produced by IP Australia
- ABS surveys, including the Business Characteristics Survey, Economic Activity Survey and the Survey of Research and Experimental Development – Businesses.

Approved researchers and analysts can use BLADE to study how businesses fare over time and the factors that drive performance, innovation, job creation, competitiveness and productivity.

Examples of BLADE projects include these research papers:

- Export behaviour and business performance
- Entrepreneurship dynamics in Australia
- Business dynamics of a clean energy policy

The **Australian Census Longitudinal Dataset (ACLD)** uses data from the three most recent Censuses to create a research tool for exploring how Australian society is changing over time.

The 2011-2016 ACLD brings together a representative five per cent sample from the 2011 Census with corresponding records from the 2016 Census, producing a dataset with over 1.2 million records. Similarly, the 2006-2011 ACLD brings together a five per cent sample - approximately 1 million records - from the 2006 Census with corresponding records from the 2011 Census.

In March 2019, the three wave ACLD was released which brings together a five per cent sample from the 2006 Census with corresponding records from the 2011 and 2016 Censuses.