Department of Infrastructure and Regional Development submission to Data Availability and Use inquiry

Executive Summary

The Department of Infrastructure and Regional Development is responsible for the design and implementation of the Australian Government's infrastructure, transport and regional development policies and programmes. To support these roles, The Department relies on datasets from a range of sources, both internal and external to the Department, and where necessary, from targeted statistical collections made by the Department's research arm, the Bureau of Infrastructure, Transport and Regional Economics (BITRE).

As both a heavy user of a wide range of datasets and a producer of several key datasets, the Department is in a unique position to provide input to this inquiry.

The Department's use of datasets sourced from other public sector agencies is needlessly hindered by legislated privacy and confidentiality barriers specific to each agency. The aim of the legislation is to protect the confidentiality of individuals' information, yet the focus is on the organisations that statisticians work for, rather than on how trustworthy they may be.

This same legislative framework restricts the Department's ability to share datasets with analysts outside the Department. Checking datasets for breaches of confidentiality consumes significant resources for little benefit, however the principle of maintaining the confidentiality of individuals' information is vital to maintaining the trust of our survey participants and the honesty of their responses.

The Department submits that it is possible to carefully share confidential survey information with selected analysts that hold appropriate security clearances.

Early work on developing partnerships with private sector data custodians show great potential for the use of "Big Data", subject to strong confidentiality protection.

This submission notes that privacy and confidentiality issues are not the only barriers to data sharing. The development and use of common standards and classification is an important prerequisite for data matching and the analysis of multiple datasets. Access to unit record information may allow classifications to be reverse-engineered, however such exercises would be very resource-intensive.

Finally, supporting documentation for the inquiry touches only lightly on data accessibility issues that may restrict data use. The Department has found that while there is a significant volume of information presented on the Department's website in downloadable spreadsheets, less technical users can find themselves "drowning in a sea of numbers". The Department is currently investigating a number of graphical and data visualisation options to improve the accessibility of our datasets.

Datasets owned by the Public Sector

The Department is a significant user of datasets owned by other public sector agencies, including the Australian Bureau of Statistics (ABS). However, the Department is also the custodian of several important datasets representing the output of surveys conducted in-house by BITRE.

Regardless of their origin, public sector datasets provide a solid evidentiary base for the Department's support of the Australian Government's infrastructure, transport and regional development policies and programs. While some parts of the Department's work have a focus on national issues (such as aviation policy), much of the Department's work is location-specific (infrastructure placement, transport bottlenecks, regional economics, etc.). Access to spatial data is therefore very important to the Department, at as fine a level of detail as is available.

Most public sector datasets fall under some form of privacy or confidentiality legislation specific to the agency holding the data. Legislation usually only permits access to information that has been aggregated to a level where it is no longer possible to identify individual responses. This severely restricts the spatial detail available to the Department's analysts and presents a significant barrier to small area or "last mile" analysis.

Restricting access to external staff not only reduces the utility of the datasets, but also makes it difficult for industry experts to assist with survey design problems. Case Study 1 (below) provides a summary of relevant issues that arose when the ABS undertook a survey of road freight movements jointly funded by Australian, state and territory governments. ABS confidentiality legislation restricted the level of geographic detail available to clients to Statistical Area Level 3 (SA3), which was useful in informing large scale freight corridor issues, but did not have the resolution to inform local infrastructure issues.

Case Study 1: ABS data access -- Road Freight Movements Survey, 2013-14

In 2013–14, the Australian Bureau of Statistics (ABS) undertook a survey of road freight movements in Australia, the first such survey since 2000–01. Jointly funded by the Australian, state and territory governments, at a total cost of $3.2 million, the survey targeted approximately 16,000 freight vehicles, from a total population of over 370 thousand freight vehicles, achieved a response rate of approximately 73 per cent, yielding information for around 120,000 separate freight trips.

The survey results provide valuable information about the size, commodity composition and geographic distribution of the road freight task to help inform freight policy and infrastructure planning, particularly around major freight corridors. However, data confidentiality and reliability necessarily limit the published geographic detail to Statistical Area Level 3 (SA3). While this level of disaggregation is quite useful for information about key freight corridors, it provides only limited information on more local policy and planning issues, such as 'last-mile' access and freight volumes on local roads. Nonetheless, the data is the best available nationally-consistent data to inform freight planning.

Throughout the conduct of the survey, including the survey design and planning phases, the ABS consulted widely and kept funding agencies well informed of survey progress. However, at some points during the survey, ABS legislative requirements necessarily restricted ABS officers from sharing some details about survey details with client agencies and sometimes inhibited a fully shared understanding of all the issues involved. In particular, ABS officers were conscious of divulging information that might inadvertently reveal the survey weightings, and thereby respondent information.

Additionally, client agencies will sometimes seek more detailed information from the survey than is available from the published survey outputs to better help inform policy development and planning. While additional unpublished data is available on request, for client agencies existing processes appear overly bureaucratic, require significant lead times to arrange approvals and access, and so are less conducive to the often short timeframes and quick turn-around requirements of a modern and responsive public sector.

Joint inter-agency data initiatives conducted with the ABS can also involve significant administrative overhead, including tripartite type agreements between data providers, the ABS and the other agencies involved. While not in themselves a barrier to collection and sharing of data, such processes nonetheless impose additional administrative overhead for the ABS and other agencies involved, are often different enough to typical private sector non-disclosure agreements that potential providers find it necessary to seek separate legal advice and present to the private sector an image of separate government agencies, rather than one joined-up APS. It would be preferable for all parties concerned in such ventures if the APS could appear as a single unified entity, the data protections and privileges applying to any one agency, equally applying to all participating agencies that have certified and compliant data systems and processes.

The Department has also explored the possibility of alternative sources for restricted public sector data to access the level of detail required for spatial analysis. In Case Study 2, the Department identified a need to determine and measure Australia's key international maritime freight routes. The ABS provides a subscription service for International Cargo Statistics (ICS) based on information they receive from the Department of Immigration and Border Protection (DIBP). However, confidentiality legislation prevents the ABS from providing a subscription service with the level of detail necessary to match with ship movements databases to calculate routes. As an alternative, the Department approached DIBP directly to establish a data sharing arrangement for the necessary level of detail. While both agencies are committed to data sharing principles when "there are no insurmountable

legislative barriers", significant resources continue to be devoted to legal assessments of the relevant privacy legislation and the *Australian Border Force Act* (2015).

Case Study 2: Matching datasets to identify key maritime freight routes

The Department's research arm, the Bureau of Infrastructure, Transport and Regional Economics subscribes to two key data sources for its maritime reports. First, *Lloyd's List Intelligence* is a private sector dataset which provides data on ship by ship movements to and from Australian ports. Second, *International Cargo Statistics* (ICS) is a public sector dataset provided by the ABS, based on Customs data on the value and volume of international trade received by the ABS from the Department of Immigration and Border Protection (DIBP).

The two datasets are complementary in that the former provides information on the movement of ships, but not the cargo that they carry. The latter provides information on international cargo carried by ships to and from Australia, but not how the cargo is moved from origin to destination.

While both datasets are individually valuable, in their present form they cannot be linked because ABS confidentiality standards restrict the release of sufficient detail. This means BITRE is unable to analyse and report on fundamental economic and policy issues for our portfolio, including:
- Supply chain mapping of goods and commodities, from origin to destination. This deficiency impairs the Department's ability to:
    - forecast freight moving through our ports to better understand future demand for landside infrastructure that supports our international trade;
    - provide an improved evidence base for our national freight and port strategies;
    - understand international trade routes connecting Australia.
- Mode of transport of specific goods and commodities (e.g. 'what was the value of containerised international trade through Australia's ports last quarter?').
- Timing of movement of goods and commodities. Currently, the above data sources allow BITRE to infer monthly total movements; however finer time resolution is lacking. This renders it difficult for BITRE to research port congestion issues.

The original DIBP dataset includes international cargo information classified by individual ship, which would allow linkage of the two datasets; however, following lengthy engagement with DIBP, a data sharing arrangement is not yet in place. As at the date of this submission, DIBP is seeking legal advice regarding whether the *Australian Border Force Act* (2015) permits sharing of the relevant data. Noting the recentness of this Act, DIBP is understandable cautious in setting precedent interpretations of its new legislation. The Department of the Prime Minister and Cabinet's recently released 'Guidance on Data Sharing for Australian Government Entities' provides encouragement to agencies to err on the side of data sharing when 'there are no insurmountable legislative barriers'. However, data sharing in the foregoing circumstance still clearly requires a legal assessment of risk; and these assessments can be lengthy. In this case, the absence of a data sharing arrangement constitutes a barrier to BITRE undertaking work of fundamental importance to its portfolio.

Datasets owned by the Private Sector
The Department subscribes to a number of regularly published private sector datasets compiled by commercial or semi-commercial organisations including:
- Lloyd's List Intelligence;
- OAG Aviation Solutions; and
- Centre for Transport Energy & Environment.

These datasets often provide packaged information from a wide range of publicly available sources, representing a cost-effective way of accessing public data.

Recently, the Department has approached private sector owners of large, "Big Data" datasets to investigate their suitability for detailed analysis. Following on from Case Study 1, where ABS road freight movements statistics lacked the detail for "final mile" analysis, Case Study 3 summarises collaboration between ABS, BITRE and private sector custodians of freight telematics data.

Case Study 3: Freight Performance Measurement Project

Australian governments—Commonwealth, state, territory and local—each year invest large amounts in transport infrastructure, spending over $23 billion in 2013-14. Much of the spending is on primary arterial routes used heavily by the transport and logistics industry, which the Australian Logistics Council (ALC) has estimated accounts for around 8.5 per cent of total gross domestic product (GDP) (ALC 2014).

However, due to the high cost of collecting freight-related data by traditional survey means and the size and diversity of the freight transport sector, there is a comparative lack of detailed and up-to-date data on road freight movements in Australia. This hinders effective planning and decision-making about where and when to invest in infrastructure.

Noting significant advances in the penetration of information and communications technology (ICT) within the transport industry, the Australian Bureau of Statistics (ABS) and the BITRE are currently undertaking the joint Freight Performance Measurement Project. This project is intended to assess the technical feasibility, level of industry cooperation and utility of producing road freight statistics using industry-sourced, in-vehicle telematics and administrative data (e.g. GPS, in-vehicle telematics, freight/logistics management systems).

While the study is still ongoing, a number of key themes are emerging:
- There is broad general support across industry for the project, with most industry stakeholders recognising the potential benefits of this information to help better inform planning and investment.
- Confidentiality is industry's primary concern around the provision and use of vehicle telematics data, and that the data be used only for aggregate statistical purposes.
- There are challenges relating to industry coverage in terms of technology penetration, with the uptake of technology more prevalent in larger transport companies.
- The majority of freight industry operators utilise third-party provided telemetry equipment and services. While data ownership resides with the client, telemetry service providers generally collect and store all data.
- Provision of vehicle telemetry data can be automated, reducing the administrative burden on data providers and simplifying collection and production of aggregate statistics. For the current project, some operators have provided permission to their telematics service provider(s) to provide sample data directly to the project team. In most cases, this can be done via a telematics service provider-supplied Application Program Interface (API), which facilitates machine-to-machine transfer of data.

While industry coverage of vehicle telematics equipment is not currently sufficient to use this information to estimate total freight volumes, work to date suggests that it may provide valuable insights into preferred freight routes and freight vehicle travel times, which inform key freight policy and infrastructure planning and investment priorities, including:
- Identifying key freight-related infrastructure bottlenecks and priorities;
- Providing information on road freight productivity trends; and

Origin-destination freight movement data, which would help inform forecasts of likely future growth in freight vehicle traffic.

So far, this project shows great potential for Big Data to provide opportunities for governments to partner with business to improve the information available to governments for better policy making in a very non-intrusive and low-cost way. However, maintaining the confidentiality of individuals' and businesses' information is an important pre-condition for private sector cooperation.

Standardising the collection, sharing and release of data: Classifications

While improvements to the availability and sharing of public and private sector data holdings would provide significant benefits to analysts, researchers and policy makers, there remain other barriers to linking datasets at the unit level or at a low level of aggregation. Analysis would benefit from some level of coordination across data custodians in terms of standards, definitions and classifications so that analysts can compare "like with like". In Case Study 4, the Department provides a recent example where the lack of a common geographical classification presents a significant challenge to regional economists.

Case Study 4: Profiling Regions

The Department of Infrastructure and Regional Development regularly publishes *Progress in Australian Regions Yearbook*, which is designed to help answer the question of how our regions are progressing against social, economic, environmental and governance indicators. The development of this publication highlighted the availability limitations for regional data. Some indicators included in the publication were only available with a limited level of geographic detail, or were available with various customised geographic classifications that did not allow for easy comparison.

In developing the Yearbook, regional researchers were faced with the challenge of trying to integrate data sets that are based on different geographic classifications. While the ABS tends to apply a consistent geographic classification (the Australian Statistical Geography Standard), other government agencies and research bodies use a variety of geographies that vary according to the way that information is collected. This means that different data sets are not directly comparable, and pose challenges for making inferences about how different statistical indicators relate to one another. Examples include:

- NAPLAN data on the literacy and numeracy of Australian school students is published on a geographic scale that classifies school locations into Metropolitan, Provincial, Remote and Very Remote areas. This is similar, but not directly comparable to the ABS Remoteness Areas that classify locations into Major Cities, Inner Regional, Outer Regional, Remote and Very Remote areas.
- Tourism Research Australia provides data on tourism demand and supply according to the boundaries of tourism regions. This is a custom geography designed to reflect tourism markets, but which is not directly comparable to the more widely used ABS classification system.
- Health system data is published according to Medicare Local Service Delivery Areas, another custom geography based on the catchment areas of local Medicare providers.
- The Australian Electoral Commission publishes electorate statistics, including enrolment and voting distribution, according to electoral division boundaries.

Datasets are usually compiled with a geography that is well suited to their immediate purpose or the way source data is arranged, but this is often poorly suited to wider comparison and analysis. With access to greater detail (such as street address), it may be possible to reverse-engineer datasets to a different geographical classification, however a more efficient solution would be for agencies to collaborate on a single classification to be used across the board.

Enhancing business confidence and trust

The Department's preferred approach when seeking data of any kind (from responses to surveys we conduct to accessing "Big Data"), is to base data sharing on productive, negotiated collaboration where rules are agreed up front, collaboration starts off small and very manageable, growing to more ambitious levels as trust is established.

Case Study 5 illustrates that with all data collection exercises, there needs to be a strong narrative justifying respondents' effort, with clear benefits for survey participants. While issues of concern may be voiced by industry associations and other representative bodies, it is important that the Department has access to statistics to inform its decision making and to monitor the effectiveness of its policies, programmes and regulations. It is of the utmost importance that businesses trust the Department enough to share sensitive information, and for the Department, the foundation of this trust lies in our legislated guarantee of confidentiality.

Case Study 5: Release of detailed aviation statistics to CASA

The role of the Civil Aviation Safety Authority (CASA) is to maintain, enhance and promote the safety of civil aviation in Australia, with particular emphasis on preventing aviation accidents and incidents. As well as safety regulation, CASA also provides comprehensive safety education and training programmes. To this end, CASA prepares regular Sector Risk Profiles which analyse the key risks facing various aviation sectors. Appropriate treatments for the risks are proposed and their implementation is monitored.

BITRE's aviation statistical datasets are an important input to these profiles, providing information on the number of hours flown by various aircraft types undertaking a range of tasks. This information allows CASA to calculate accident rates to compare risks associated with various flying activities and various aircraft types. Clearly, the more detailed the data provided, the more detail available to CASA researchers, however the detail available is restricted by legislation.

Section 10 of Air Navigation Regulation 2016 states that persons in receipt of statistical information collected under the authority of the Regulation "must not disclose the information in a way that allows individual responses to be identified, except:
   a) In accordance with section 9 [to provide statistics on international airlines to the International Civil Aviation Organisation]; or
   b) In the course of performing a function under the Act or this instrument; or
   c) Where the person who provided the information has given their consent to the disclosure of the information."

While BITRE is committed to meeting CASA's needs for detailed statistics, the Regulation requires BITRE statisticians to check all datasets for breaches of confidentiality before releasing them to CASA. These confidentiality checks become more and more complicated as the size and level of detail increases in a dataset. In addition, customised detailed subsets need to not only be checked for confidentiality breaches themselves, but also compared to other datasets previously released to ensure that it is not possible to compare datasets to expose an individual's confidential information. As the number of customised datasets increases and the detail within them increases, the resources and time required to check for breaches of confidentiality increases exponentially.

In short; BITRE expends a significant amount of resources to still fall short of the detail that our main clients would like.

Any move to allow greater sharing of confidential data (even between public sector agencies where officers have been rated as sufficiently trustworthy by the AGSVA) would require changes to Air Navigation Regulation 2016. To maintain the trust of survey respondents, the confidentiality provisions in the Regulation would need to be expanded to include all people that have access to unit record information.

It is important to recognise that in this example, providing CASA officers with unit record information may reveal individuals undertaking unsafe flying activity. In such instances, CASA officers would be forced to make a decision between their responsibility to maintain the confidentiality of survey responses and their responsibility to prevent aviation accidents.

Any action taken against operators based on their confidential response would be considered a breach of trust and may have a significant impact on the level of response to future surveys as well as the veracity of responses.

Any change to the Regulation must include a strong statement regarding the primacy of protection for confidential information.

Suggestions for a way forward

In discussing the availability and use of public sector datasets, the key barrier to greater sharing is not just that the information is protected by privacy and confidentiality legislation, but that the legislation is usually specific to each agency. In effect, it assumes that for each agency, only agency staff may be trusted with confidential information and that external public sector officers may not be trusted, regardless of how trustworthy they are rated by the Australian Government Security Vetting Agency (AGSVA – the central agency for the processing and granting of security clearances for the majority of Australian Government agencies and state and territory agencies).

It is ironic that with increased staff mobility across the Australian Public Service, statistical officers that are entrusted with confidential statistics one day, may be deemed to be no longer worthy of this trust if they transfer departments the next day.

It is this concept of trust that could usefully be explored by the inquiry.

> The Department **suggests** that access to confidential data should be open (for statistical purposes only and on a needs basis) to any statistician within the Australian Public Service with an appropriate security clearance as determined by the AGSVA. Any statistician receiving confidential data would in-turn agree to confidentiality and non-dissemination responsibilities, specified in relevant legislation.

Any expansion in the availability and use of public sector datasets should be accompanied by rigorous legislative protection for the confidentiality of individual responses. It is vital that businesses and individuals are free to provide frank and honest information to Departments without fear of negative legal or regulatory outcomes.

> The Department **suggests** that any changes to legislation provide clear guidance that confidential data may be used for statistical purposes only and that information provided for statistical use must not be used for regulation compliance or litigation.

As technology penetrates business activity, Big Data provide more opportunities for government to partner with businesses to improve the information available to governments for better policy making in a very non-intrusive and low cost way.

> The Department **suggests** that increased emphasis be put on working cooperatively with third-party business metrics providers to embed information exchange requirements in standard software products. Confidentiality legislation should be expanded to include data feeds from private sources of Big Data.

While organisations compiling datasets usually do so in order to meet specific requirements and often have specific metrics, the use of specific standards and classifications can be a barrier to sharing of these datasets, particularly spatial datasets.

> The Department **suggests** that data custodians maximise the use of common spatial standards and classifications to better enable data matching and comparison exercises.

Finally, data use is not only restricted by the availability of datasets, but also by how accessible datasets are to potential users. Recent initiatives such as Data.Gov have significantly increased the volume of public sector datasets available to the public. While this is of great benefit to data technicians, other potential users of this information may find themselves swamped in a "sea of numbers".

> The Department **suggests** that appropriate graphical or data visualisation interfaces should be developed and made publicly available with datasets. This would be a valuable addition to the Data.Gov.au initiative.