



Australian Government

Commonwealth Grants Commission

**PRODUCTIVITY COMMISSION INQUIRY INTO
DATA AVAILABILITY AND USE**

**Commonwealth Grants Commission
Submission**

- 1 This submission represents the views of the staff, or the Commission's secretariat, and does not necessarily represent the views of the Commonwealth Grants Commission (CGC).
- 2 The CGC is the agency responsible for recommending the distribution of more than \$60 billion in Commonwealth transfers of GST revenue to the States and Territories (hereafter referred to as States).
- 3 This is a very data intensive process. We have a broad remit, requiring data covering all aspects of State service delivery and revenue raising activity. However, as a small agency, we have limited capacity to develop and maintain our own datasets across the breadth of our areas of interest. As a non-data custodian we rely on data from third party providers that are experts in their field, such as the ABS, AIHW, NCVER, but also from States.

DATA QUALITY

- 4 We consider that in addition to considering and expanding the range of data available to users, the inquiry should also focus on the importance of improving the quality of data available.
- 5 We do not under-estimate the efforts that key data collection agencies such as the ABS and AIHW put into ensuring the quality of their data, but we are concerned by the risk that competing priorities in coming years could be at the expense of validation and quality assurance of data. We would like to emphasise that data are only useful when they are:
 - reliable (it measures what it claims to measure)
 - timely
 - comparable over time
 - comparable across geographic areas (especially States)
 - stable (not subject to significant revisions).

DATA ACCESS

- 6 Concerns with maintaining the confidentiality of respondents are obviously critical both from a right to privacy perspective as well as maintaining the trust and confidence of respondents. The ABS has put significant effort into maintaining this confidence, while improving the access that data users have to vital information. However, we consider that more can be done.

- The limitations imposed in TableBuilder continue to limit the nature of research that can be done. We strongly support any efforts the ABS can make to continue its path of improving access to its data. TableBuilder currently:
 - has a limit of 5 million cells (this can preclude measuring a variable by fine geography, and standardising by more than a few variables)
 - restricts sparsely populated tables (this precludes cross-classifying correlated variables, such as producing journey to work origin-destination tables)
 - does not allow user defined fields.
- Randomising rather than confidentialising microdata means that all information gathered by the ABS is available to researchers. Data are slightly modified to protect confidentiality without changing usability. This is a great improvement over the historical approach of removing large amounts of data from datasets. The ABS has not yet developed an approach where access to survey data could be similarly improved. However, if this is possible it should improve the usefulness to clients as well as reducing the cost and time taken to prepare survey microdata for release.
- AIHW requires State permission before providing data to third party users such as the CGC. The process of obtaining State approvals can be very time consuming. Streamlining this process would be of great benefit. This could be done by any of the following means.
 - Granting ‘format based’ approval, so an annually required data request does not require approval for identical requests each year.
 - Developing guidelines where the AIHW would have authority to approve data requests not deemed to be politically or personally risky.
 - Developing guidelines where the AIHW would have authority to approve data requests to approved agencies. The Independent Hospital Pricing Authority (IHPA) currently has authority to provide data to any agency in the Treasury portfolio.
- Until 2012 the ABS provided the CGC with Government Finance Statistics (GFS) unit record data, under a return to source protocol, as the State and Commonwealth treasuries had endorsed this. Since then, however, the ABS has reinterpreted its Act and now considers that it requires the permission of all individual agencies described in the data, not merely the agency providing data to the ABS. This new interpretation of the Act has created significant barriers to our ability to analyse data without any improvement in the privacy offered to providers.
- To access the Characteristics of Employees survey data (a supplementary survey to the Labour force survey) to examine wage differences between States, the CGC embedded an officer at the ABS. While this approach was satisfactory for the CGC’s purposes, some States have had difficulty replicating these

arrangements and argued a lack of ability to access the data used by the CGC has reduced transparency in its work.

DATA FOR THE FUTURE

- 7 We consider data linkage to be the most effective means for addressing the policy questions facing Australia today and into the future. Data linkage offers:
- the potential to create powerful datasets at low cost, and
 - the best means to link government finances with client need, which is a driver of many of the policy debates in Australia.

Broader focus

- 8 We consider that where possible data producers should endeavour to produce datasets of broad appeal to possible research needs, rather than for specific research projects.
- 9 Some linkage projects are defined narrowly to focus on a specific policy question, rather than a broad linkage which would enable other researchers to explore alternative issues. For example, the mental health services project involved linking a small subset of Medicare Benefits Schedule (MBS) and Pharmaceutical Benefits Scheme (PBS) records. If the full set of MBS and PBS records had been matched, it would have obviously created a much more powerful dataset that could be used to answer questions on the socio-demographic users of health services other than just mental health services, including questions on the fiscal incidence of Medicare funding (e.g. who gets bulk billed?)
- 10 Similarly, the linkage of death and census records has an exclusive focus on Indigenous life expectancy estimates. This dataset could also be used to measure death rates, and cause of death patterns by:
- labour force status
 - household income
 - level of education
 - family status
 - language spoken at home.
- 11 The focus of the CGC's work on State services means that we would not have any immediate need for either a census-death registration or census-MBS dataset. However, they do provide a good illustration of the principles that should apply.
- 12 As a small agency that is not a data custodian, we will never be involved in the development of data linkage. Agencies like ours rely on high quality data coming from the ABS and other data providers. If new datasets are too narrowly defined for

specific research projects, they are unlikely to meet our needs, or the needs of the broader research community.

- 13 The ABS obviously has to ensure that its statistics are relevant to the community. It cannot merely build a data set and users will come. It does, rightly, ensure that the data enhancement projects do have policy needs which justify the resources used to produce them. However, research questions do arise quickly, and where the additional cost of making these projects more broadly applicable is relatively low, the ABS should attempt to meet all potential users needs, not merely those identified at the start of the project.

Hospitals

- 14 A hospitals-census linked dataset would significantly expand the ability of the CGC to accurately measure the profile of people using hospitals, and the relative needs of different States for hospital funding.
- 15 Spending on admitted patients represents about 15% of the State budgets and understanding the patterns of hospital use is of very significant policy interest. It is this interest that has driven statistical work at the AIHW and elsewhere to create a nationally comparable dataset of hospital use.
- 16 There is significant interest in, and data on, hospital separations. However, there is currently no data on patients. We do not know how many people go to hospital each year, just how many hospitalisations there are. This data gap, along with information on the attributes of patients can be best met by linking the National hospital morbidity database with the census.
- 17 Associated with such a data linkage project are questions of methodology and statistical need.
 - Is a single year's hospitalisation sufficient, or should multiple years' data be linked?
 - Should this also be linked to death registrations and/or MBS and PBS records to further enhance its value to health researchers? (Linking census to both hospital and MBS data would enable much needed analysis into the substitutability of different types of health care).
- 18 However, the most difficult questions are those relating to the sensitivity of the data. The confidentiality of patient and hospital data, as well as census data is important. The respective data custodians need to protect the confidentiality of their data. As States have proven willing to provide their data to the AIHW, it seems reasonable that they could also provide it to the ABS.

The future of censuses

- 19 In 2015 the ABS and the federal government canvassed a proposal to cancel the 2016 Census. In light of this, we consider that whether Australia needs a census program every 5 or 10 years or not at all is worth considering in this inquiry.
- 20 A less frequent, or abolition of, the census program would be possible if the data needs currently met by the census program could be met by either a large sample household survey program, or an administrative population register. It is important to prove that such an alternative is viable for the broad range of users before the census is discontinued.

Administrative census

- 21 Administrative population registers have the attraction of greater frequency than the current census program at lower financial cost. However it is worth considering whether these alternatives can actually produce data of the quality produced by the census.
- 22 While Australian and overseas experience may suggest that administrative systems may be able to produce a population register, the range of data available on such a register is likely to be significantly less than that available from the census, and the match rates achieved in linking different administrative systems do not appear to be high enough to guarantee that an accurate population register can be created.
- 23 As such, the CGC would strongly suggest that any move to reduce the frequency or content of Australian censuses be done only after an alternative system has been proven successful, not merely on the prospect that one might be possible.

Australian Population Survey

- 24 As part of the proposal to cancel the 2016 census, the ABS proposed an annual survey of 400 000 households. While this, and the revised, smaller program offer much to be excited about, they do not replace the need for a census. The diversity of Australian society means that small population groups and small area data are of significant interest to policy makers.

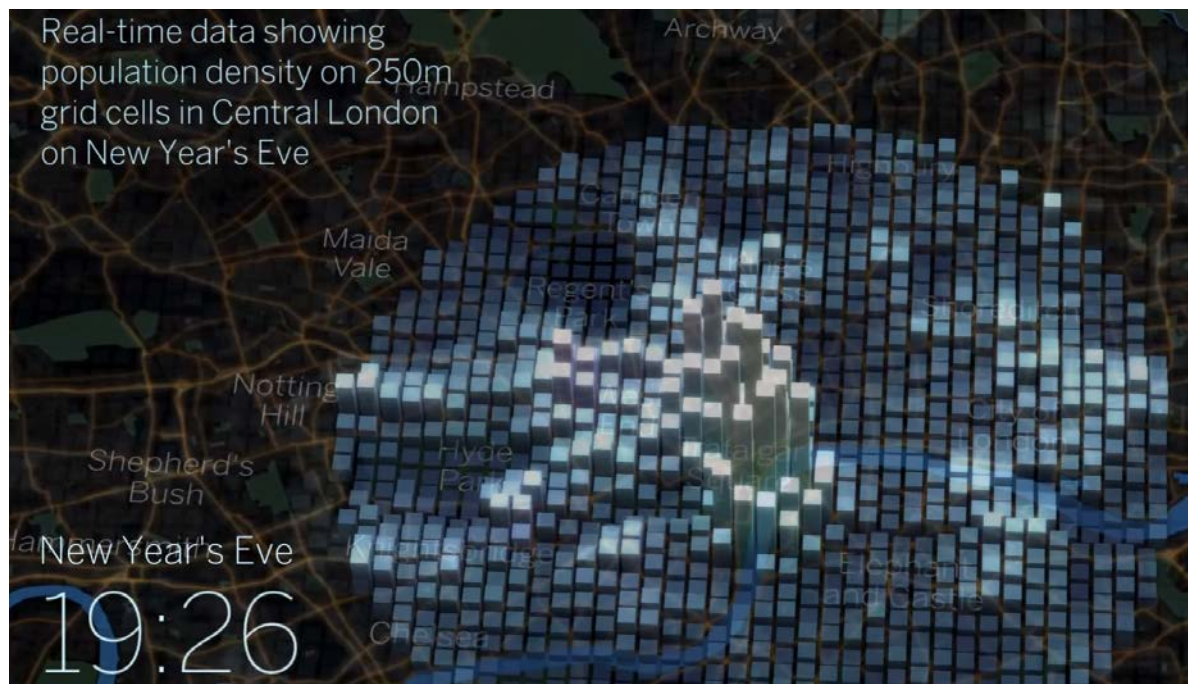
Mobile phone metadata

- 25 The inquiry asks “What private sector datasets should be considered high-value data?” Our GPS-equipped phone and other technology contain a massive amount of data about where the population is and when, and how it moves.
- 26 In considering creating a publicly available dataset with some or all of these data, we first need to consider the data needs.
 - Do we need a real time dataset?

- knowing and predicting crowd sizes for police response
- knowing traffic volumes and speeds to improve the responsiveness of the road and public transport network.
- Or a non-real time dataset for
 - Transport planning
 - Estimating service (non-residential populations) for planning delivery of services. The time dimension of this can be:
 - i very short, measuring daytime populations of employment or entertainment districts
 - ii longer: measuring seasonal populations, and Fly in-Fly out work patterns.
- Do we need any demographic information associated with these data?
 - Phone companies know at least the date of birth and gender of their customers.

- 27 Services are not always delivered near where people live. While governments and private sector providers rely on estimated resident population to plan service delivery, this planning could be improved if we had better data on where people are, not just where they live.
- 28 Again, the CGC is a very minor player in the need for these data, and we have not yet even engaged with existing data providers in our modelling of State road expenditure. However, from our little research, it appears that Australia has several companies producing traffic data, including Google, Waze, Here and Suna. We have identified one company willing to sell it: Here.com.
- 29 In Europe and America, data are available on real time population densities, as illustrated in Figure 1. As we understand it, Australian data such as these are not available to the public or to governments at any price. It may or may not be produced in house by telecommunications companies. We consider there is a demand for these data, and the Productivity Commission should recommend a strategy for their creation and use.

Figure 1 Population density in Central London on New Year's Eve.



Source: Inrix population analytics.

Video of full day: <http://inrix.com/products/inrix-insights-volume/>