



The changing role of the Census in Australia's integrated data landscape

James Mowle and Ross Watmuff

Paper prepared for the 16th Conference of IAOS
OECD Headquarters, Paris, France, 19-21 September 2018

Session 5.B, Friday, 21/09, 10:30: Use of administrative data for compiling statistics

James Mowle
James.mowle@abs.gov.au
Australian Bureau of Statistics

Ross Watmuff
Ross.watmuff@abs.gov.au
Australian Bureau of Statistics

The changing role of the Census in Australia's integrated data landscape

Prepared for the 16th Conference of the
International Association of Official Statisticians (IAOS)
OECD Headquarters, Paris, France, 19-21 September 2018

Note:

This Working Paper should not be reported as representing the views of the Australian Bureau of Statistics. The views expressed are those of the authors.

ABSTRACT

Over the last two decades the role and nature of the traditional Census of Population and Housing has been increasingly challenged by the growing potential of administrative and new big data sources. Many non-register-based countries are pursuing Census approaches which seek to improve data and reduce costs by favouring integrated administrative data over directly collected data.

At the same time, these countries are beginning to harness increasing statistical power and insight by integrating alternative data sources to their traditional Census and survey data assets. Rich longitudinal views of the population can be built up as sources covering different time periods are drawn together. In this context, Australia's five yearly Census data collection is taking on new and increasing statistical value for government policy development and the research community.

This paper uses the Australian context to examine the changing role and value of traditional Census data within increasingly integrated data environments and reflects upon the implications for future Census models.

Keywords: Census, Administrative data, integrated data

1. TRADITIONAL ROLE OF CENSUS DATA

The Australian Bureau of Statistics (ABS) has been conducting a national Census of Population and Housing since 1911, which is the largest and highest profile of its statistical collections. It uses a ‘traditional’ full-enumeration long-form Census model with a comprehensive snapshot covering everyone in the country at a single point in time. The Australian Census is currently conducted every five years, as required by national legislation, with planning well-underway for the next Census in August 2021.

The primary purpose of running the Australian Census has been to provide a population count base for deriving official population estimates. This base measure of the number and key characteristics (age, sex, usual residence) of people in Australia on Census night provides a reliable basis to estimate the population for each state and territory and local government area.

A second, growing purpose of the Australian Census has been to collect a set of key socio-economic data to output information for small geographic areas and population groups, a level of detail which is not supported through sample surveys. No other single information source in Australia can provide the range of socio-economic data for small populations that are required for informing government policy and business decisions.

Despite the Census currently meeting these key purposes, the ABS is facing increasing demand for data to better inform Australia’s increasingly complex economic and social environment. Expectations are growing for more frequent data, greater levels of detail, longitudinal insights and socio-economic connections to inform complex policy issues.

2. AUSTRALIA’S INTEGRATED DATA LANDSCAPE

The ABS has traditionally made extensive use of administrative data, especially from the public sector, to produce national statistics. Over the last decade administrative data has become increasingly available, accessible and viewed as a national resource to be leveraged for improved decision making (Australian Department of Prime Minister and Cabinet (2015)¹. Combined with better technology and improved methods of combining and analysing large datasets, this has resulted in more opportunities for using integrated administrative data for insights not previously available from traditional statistical collections.

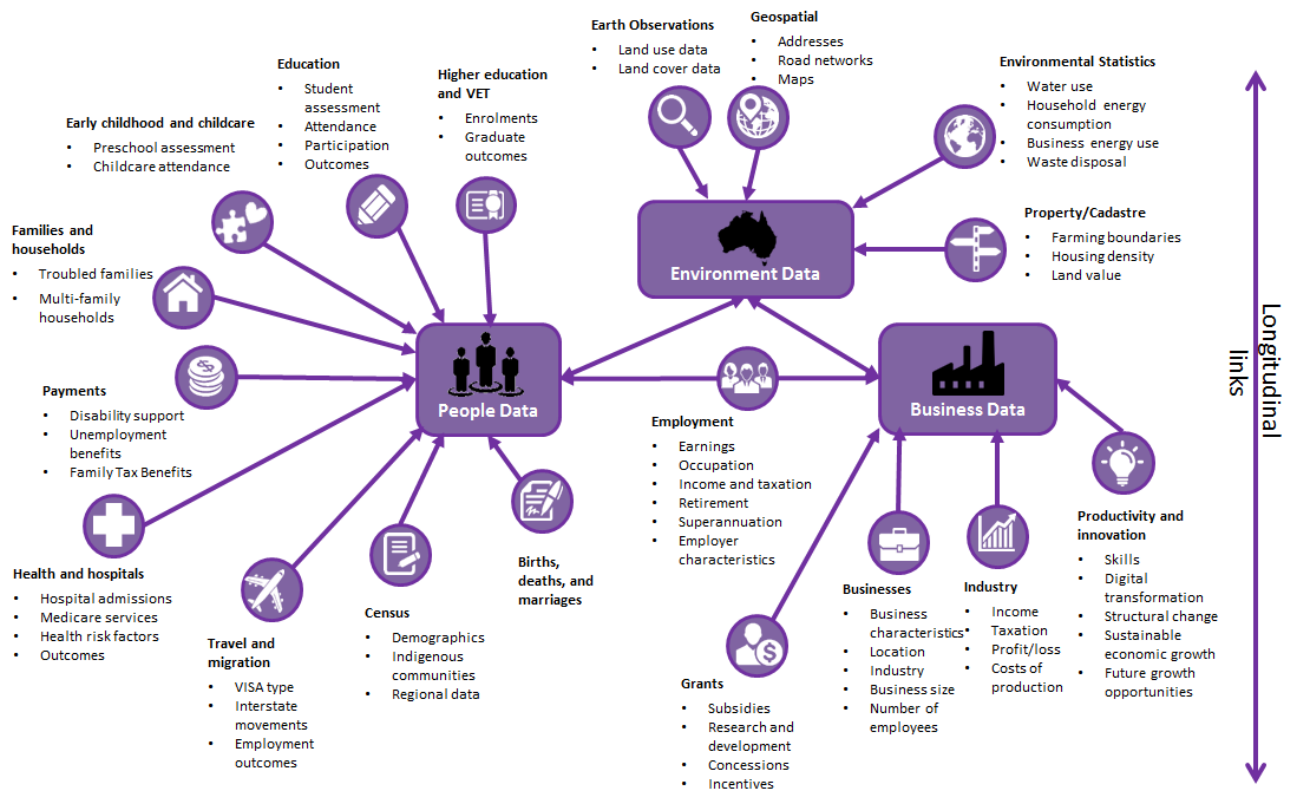
A key focus of the ABS’s data integration program since 2005 has been on augmenting the Census with administrative data to enhance topic coverage, provide longitudinal insights and improve official statistics. This was initially done through a series of discrete projects under the Census Data Enhancement (CDE) program.

The success of the ABS’s data integration work program is set against a backdrop of increasing government interest, and more importantly investment, in public data initiatives. In 2017, the Australian Government committed \$131 million over three years to a number of government agencies to maximise the

¹ On 7 December 2015, the Australian Government released its Public Data Policy Statement as part of the National Innovation and Science Agenda. The Policy Statement commits Commonwealth Government entities to: specific actions designed to optimise the use and reuse of public data; release non-sensitive data as open by default; and collaborate with the private and research sectors to extend the value of public data for the benefit of the Australian public.

use and value of the Government's data assets through a [Data Integration Partnership for Australia \(DIPA\)](#). The DIPA builds on earlier data integration progress and aims to create high-value, enduring integrated national assets to build a comprehensive picture of Australian populations, business and the environment over time (Figure 1).

Figure 1. Data Integration Partnership for Australia – long term vision for data assets



The current focus of the ABS's contribution to DIPA is to expand the utility of, and enable new public policy insights from, current integrated data assets that bring together rich cross-portfolio data for research and policy purposes:

- [Multi-Agency Data Integration Project](#) (People Data), that combines information on healthcare, education, government payments, personal income tax, and the Census;
- [Business Longitudinal Analytical Database Environment](#) (Business Data), that 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;
- [Linked Employer-Employee Database](#) (Employment data), that combines person and business level taxation data.

Privacy, confidentiality and social license are critical aspects for the ABS in navigating developments in the integrated data landscape. The Australian community has complex and diverse perspectives towards privacy of information. Progress made by the ABS in data integration has been done alongside Privacy Impact Assessments to ensure compliance with the Australian Privacy Principles and that privacy risks are identified and mitigated (Australian Bureau of Statistics, 2018a).

As these integrated data assets are developed by the ABS over time, linked administrative data has the potential to provide more frequent and timely information on the size, structure and location of the Australian population, as well as a rich source of detail on socio-economic characteristics and outcomes.

At the same time it should be noted that integrated administrative data currently present some critical quality challenges impacting its usefulness. Two key challenges are:

1. incomplete coverage of important small populations such as recent arrivals and lower socio-economic groups; and over coverage and duplication of some population groups; and
2. up-to-date location information for individuals, which particularly impacts data for mobile populations.

3. EMERGING VALUE OF THE CENSUS IN INTEGRATED DATA ASSETS

With the linkage of administrative data increasing in Australia, Census data is taking on a key emerging role in deriving value from linked data assets. The Census plays a central role in many of Australia's current data integration projects, particularly those focused on People Data for socio-economic insights. In particular, these assets are providing opportunities for longitudinal analysis that were previously sought through panel studies, since they draw together unit record data from sources over multiple time periods including the periodic "snapshot" provided by a Census.

As new administrative and big datasets emerge, the Census is also taking on an increasingly important role in assessing the quality of these in respect to the quality challenges noted above. The Census is therefore becoming increasingly valuable as a 'ground truth' source to understand and anchor integrated administrative data assets.

The characteristics of Census data that support its traditional objectives provide unique value in integrated data assets:

- rich source of socio-economic data for longitudinal analysis
 - factors that help predict socio-economic outcomes on other datasets
 - outcomes given historic patterns and factors provided from other sources;
- rich source of attributes that depend on personal reporting e.g. family and household relationship information, proficiency in spoken English, unpaid work.
- reliable geographic residence location information of individuals – something that is often incomplete or lagged in administrative datasets;
- high population coverage at a single point in time;

The examples below highlight some policy and longitudinal insights being made through incorporating the Australian Census into integrated administrative data assets.

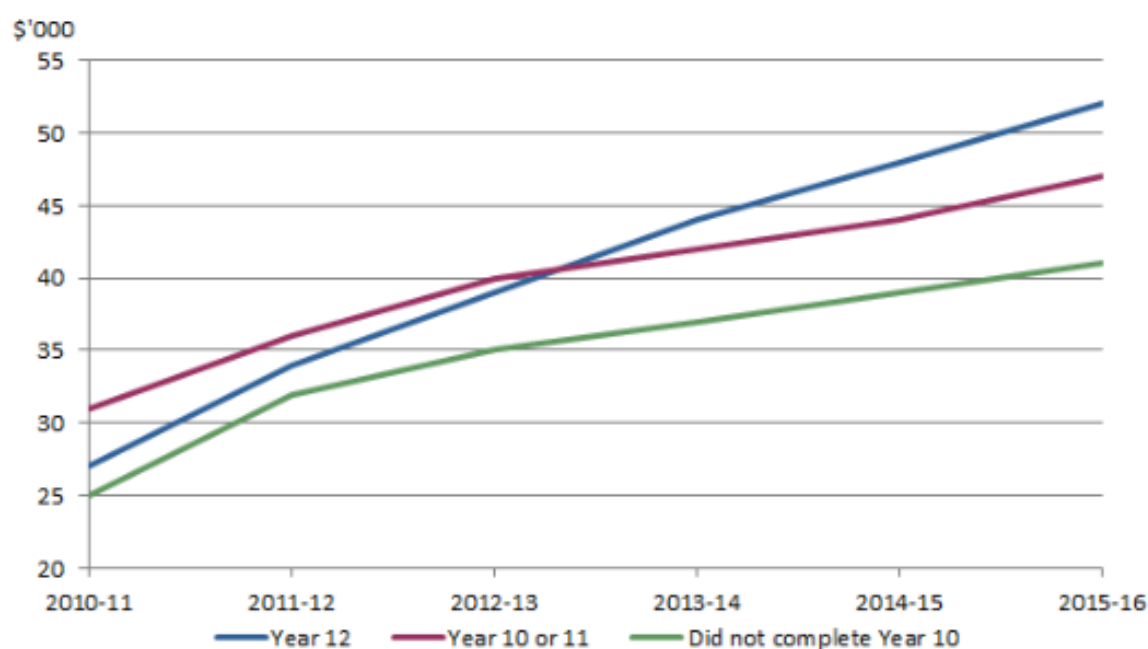
3.1 Changes in personal income for young people who completed different levels of schooling

The ABS has recently released a longitudinal extract of the Multi-Agency Data Integration Project, which integrates data from the 2011 Census with longitudinal administrative data relating to personal taxation, social security and Medicare benefits between 2011 and 2016.

Including the 2011 Census information in the longitudinal product enables comparisons between groups with different socio-demographic characteristics, such as educational attainment, occupation and industry of employment, country of birth, and marital status.

Using highest year of schooling completed data from the 2011 Census enables analysis of administrative income data for young people and how this varies over time for different levels of educational attainment. In 2010-11, young persons (aged 20-24 in 2011) whose highest year of school completed was Year 10 or 11 had a median income of \$31,000, higher than the median income for persons that had completed Year 12 (\$27,000) (Figure 2). Four years later, in 2015-16, young persons whose highest year of school completed had been Year 12 had the highest median income (\$52,000) compared to that of young persons whose highest year of school completed had been Year 10 or 11 (\$47,000) or who had not completed Year 10 (\$41,000).

Figure 2. Median income(a)(b) of persons aged 20-24 years in 2011, by highest year of school completed(c)



(a) Data weighted to resident Census 2011 population

(b) Total income from Personal Income Tax in Australian Dollars (\$AUD). Medians are calculated on income data that has been categorised into \$1000 ranges.

(c) Highest year of school completed as reported in Census 2011.

Source: Australian Bureau of Statistics, 2018b.

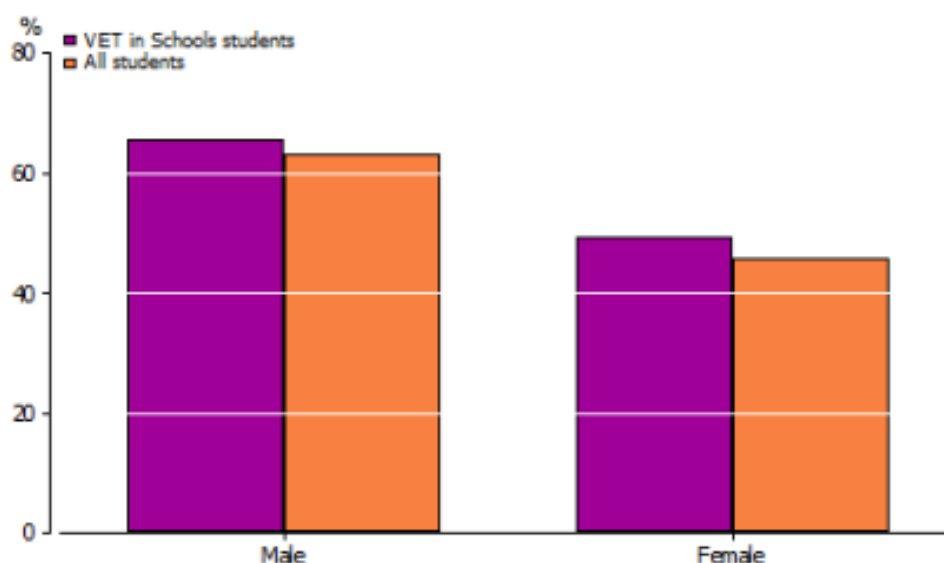
3.2 Employment outcomes from school-based vocational education and training

Integrating vocational training data from 2006 with 2011 Census data has provided insights into the post-school pathways and destinations of students participating in Vocational Education and Training (VET) programs in schools and addressed a critical gap in understanding their longer term employment outcomes.

The project enabled researchers and policy analysts to see whether the students went on to any further study, the highest level of qualification they attained, whether they were employed in 2011 or still studying, and the level of income they had attained.

By including Census data on educational and employment outcomes of those students five years down the track, the project showed that students who do not go on to higher education have better engagement and employment outcomes if they do VET programs in school. Those that did VET programs in school were more likely to be employed full-time five years after completing Year 11. In this group, the level of full-time employment for males five years after Year 11 was 66%, compared to 63% for all students. Similarly, for females the level of full-time employment was 49%, compared to 46% for all students (Figure 3).

Figure 3. Proportion of students that did not go on to higher education(a), and were employed full time in 2011



(a) Year 11 students in 2006 who had not completed a Bachelor degree or above by 2011 and were not studying at a tertiary institution during 2011

Source: Australian Bureau of Statistics, 2014.

3.3 Outcomes for permanent migrants to Australia

Integrating recent immigration data with 2016 Census data has provided insights into the settlement outcomes for permanent migrants to Australia. The linkage of these two data sources allows for valuable insights into settlement patterns of overseas born persons by visa stream (humanitarian, skilled, family) as well as information on citizenship, housing, income, labour force characteristics, changing occupations, educational pathways and family characteristics.

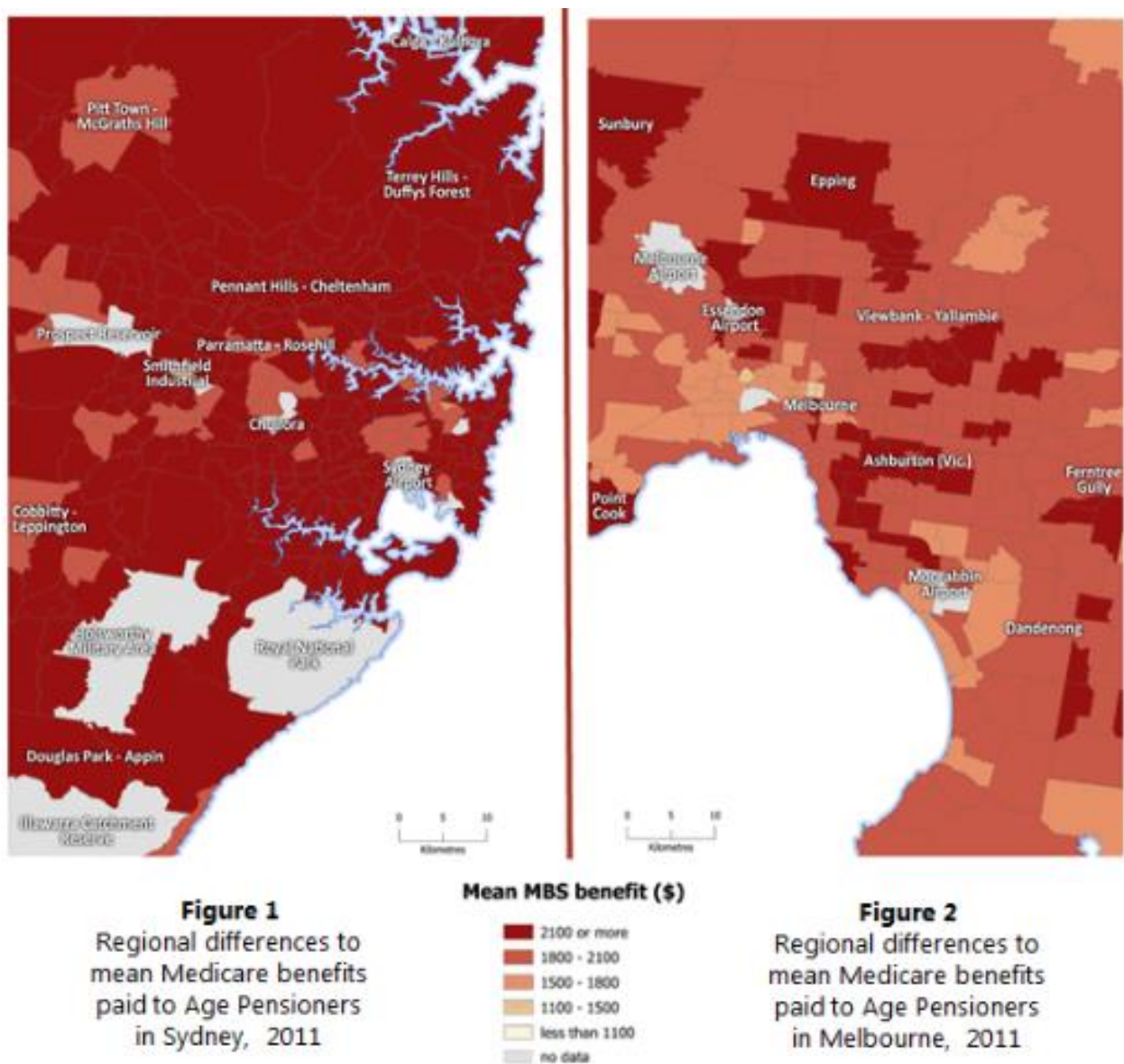
The project showed the geographic mobility of permanent migrants within Australia and their movements between population centres. One in five (20%) permanent migrants reported that they had changed their address within Australia in the year prior to the 2016 Census, with the majority of movements taking place within and to Sydney (Australia's largest city) and Melbourne (Australia's second largest city). Melbourne was the most common destination gaining 12,431 movers, including 2,769 migrants from Sydney. Sydney was the second most common destination gaining 8,780 movers, including 1,798 migrants from other parts of New South Wales (Australian Bureau of Statistics, 2016).

3.4 Geographic differences in Medicare use by Aged Pension Recipients

Integrating 2011 Census, social security and Medicare data has provided new insights into the relationship between receipt of Medicare benefits and receipt of government payments such as the Aged Pension and how this varies geographically across Australia.

The project found Aged Pension recipients living in the major population centres of each state and territory had a higher rate of use of Medicare services than those living in regional and remote areas of Australia. Mapping the data showed further variations in the use of Medicare services within capital cities (Figure 4). In Sydney, older Australians on the Aged Pension make high use of Medicare services. By contrast, in Melbourne, the use of Medicare services by older Australians on the Age Pension varies considerably from suburb to suburb (Australian Bureau of Statistics, 2017).

Figure 4. Regional differences of mean Medicare benefits paid to Aged Pensioners in Sydney and Melbourne



Source: Australian Bureau of Statistics, 2017

4. IMPLICATIONS FOR NEW CENSUS MODELS BEING CONSIDERED INTERNATIONALLY

The pace and trajectory of international progress in using alternative sources and integrated administrative data continues to show strong potential for National Statistical Organisations (NSOs) to reduce reliance on directly collected data for achieving the traditional aims of a Census, particularly as a population count base for official population estimates.

Recognising this, and responding to funding pressures to reduce costs, many non-register-based NSOs are looking to transition to a Census model that seeks to utilise integrated administrative data over directly collected data.

While Australia is not actively considering alternative Census models at this time, it is worth contemplating the potential impact of two common elements of future models being considered internationally:

1. reducing the frequency of the Census, for example from five-yearly to ten-yearly.
2. shifting resources from a point-in-time traditional Census to a large-scale rolling sample survey.

4.1 Reduced frequency full enumeration Census model

With increasing reliance upon large administrative datasets to support official population estimates and enable analysis for small population groups and areas, a traditional full enumeration Census model could be retained but with a longer interval between Censuses, for example every ten years. This is a common Census frequency internationally amongst non-register based NSOs.

A ten-yearly Census model would still provide significant value in integrated administrative data assets as a quality benchmark for integrated data and longitudinal statistical insights, particularly through its rich source of socio-economic factors and attributes that depend on personal reporting. However, the value of the Census in this context would be diminished through the Census data being generally more out of date in the intervening years. Some products and insights may have a corresponding reduced frequency from five-yearly to ten-yearly, or would rely on older Census data.

4.2 Large-scale rolling sample survey

In the future, it is possible that population counts from integrated administrative data could provide comparable or improved quality to the Census as a base for population estimates. In this case, there could be increased incentive to consider shifting Census resources toward a large-scale rolling sample survey model that collects the detailed socio-economic information not otherwise available from administrative data, providing statistics on a more frequent and flexible basis.

The French rolling Census and the American Community Survey both offer examples of rolling collection models. Because data collection is continuous rather than at a single point in time, the data must be pooled over time to produce estimates. Once the rolling collection is established, it has the advantage of providing updated estimates each year by replacing the earliest year of data with the most recent year.

A large-scale rolling sample survey would provide value in integrated data assets through its rich source of socio-economic factors and attributes that rely on personal reporting, and enable more frequent analysis of

linked data. However, size of the sample would limit the level of detail for small areas and population groups without pooling over long periods.

The pooling of data over time may also have a particular impact on how the rolling survey data is used in longitudinal analysis, since any survey-measured factors or outcomes would themselves be multi-year averages compared to the current snapshot Census data. Longitudinal outcomes may therefore be masked over shorter time periods, and only discernible in the longer term.

5. CONCLUSIONS

Many statistical agencies across the developed world are increasingly using data integration for policy and research purposes, with integrated administrative data enabling new statistical and policy insights. The pace and trajectory of recent progress is showing the potential of administrative data to challenge the traditional role of the Census in official statistics in the future.

Although administrative data could potentially provide the main basis for achieving the aims of a traditional Census in the future, the characteristics of Census data that support its traditional objectives also provide unique value in integrated data assets. Many recent applications of integrated data assets in Australian socio-economic policy research have demonstrated the value that Australia's Census data is providing in enabling statistical and policy insights from integrated administrative data. There is therefore a risk that a transition to an administrative-based Census model may constrain its value in an integrated data context.

The ABS is not considering alternatives to the five-yearly traditional Census model at this time. However, as Australia's integrated data landscape further develops and matures, the role of the Census, including its role in integrated administrative data, will need to be considered. It is possible that the growing importance of longitudinal data integration and analysis provides a long-term driver for retaining more traditional models such as the five-yearly Australian Census.

REFERENCES

Australian Bureau of Statistics (2014) Outcomes from Vocational Education and Training in Schools, experimental estimates, Australia, 2006-2011 cat. no 4260.0, ABS, Canberra.

Australian Bureau of Statistics (2016) Understanding Migrant Outcomes - Insights from the Australian Census and Migrants Integrated Dataset, Australia, 2016 cat. no 3417.0, ABS, Canberra.

Australian Bureau of Statistics (2017) MADIP Case Studies viewed at <http://www.abs.gov.au/websitedbs/D3310114.nsf/home/Statistical+Data+Integration+-+MADIP+Case+Studies>

Australian Bureau of Statistics (2018a) ABS Privacy Impact Assessments. Viewed at <http://www.abs.gov.au/websitedbs/D3310114.nsf/home/ABS+Privacy+Impact+Assessments>

Australian Bureau of Statistics (2018b) Microdata: Multi-Agency Data Integration Project, Australia, cat. no 1700.0, ABS, Canberra.

Australian Department of Prime Minister and Cabinet (2015) Australian government Public data policy statement.viewed at <https://www.pmc.gov.au/resource-centre/public-data/australian-government-public-data-policy-statement>