Innovations in Australia's social statistics information program

Eric Morris* and Gemma Van Halderen
Population, Labour and Social Statistics Group
Australian Bureau of Statistics, Canberra, Australia

* Corresponding Author, Locked Bag 10, Belconnen, A.C.T., Australia, 2616, email eric.morris@abs.gov.au

Keywords: Transformation; Household Surveys; Web forms; Data integration; address register

Abstract

The demand for access to more detailed information to support informed decision making by Governments, businesses and community continues to increase. For the Australian Bureau of Statistics (ABS), this comes at a time of increasing financial constraints, capability limitations, complex work programs and ageing technical infrastructure. To meet these challenges, the ABS is radically transforming its business processes.

Under the auspices of this transformation environment, the ABS' Population, Labour and Social Statistics (PLaSS) Group has commenced work on a range of initiatives to deliver an integrated social statistics program, maximising coherence across administrative, household survey and Census collections and utilising newly developed corporate infrastructure. This paper describes the broad program of change including the introduction of an address register to support a list-based sampling framework for household surveys; the introduction of web forms; extending the time in sample for the Monthly Population Survey (MPS); data integration activities; and measuring Australia's progress. This paper will also outline some of the challenges that have been encountered and the opportunities that are yet to be realised.

Introduction

The Australian Bureau of Statistics (ABS) is Australia's National Statistical Office. The ABS is responsible for delivering a wide range of statistical information across economic, social and environmental domains. Our business includes household and business based collections, the use of administrative data, account compilation, and statistical information products that focus on particular themes such as wellbeing, progress and gender.

The ABS has embarked on a major transformation in the way population, social, economic and environmental statistics are designed, compiled and disseminated. [Pink et al, 2010]. This transformation, entitled ABS2017, is strongly driven by a need to face emerging challenges to existing statistical collections that will be familiar to most NSOs. These challenges include increasing financial constraints, capability limitations, complex work programs and ageing technical infrastructure. However, it will also be a chance to harness potential statistical advances arising through a range of emerging opportunities that will also be familiar to all statisticians - the increasing accessibility of administrative data, the power of technology in all stages of the collection process and the implicit opportunities in integrated data.

Transformation to leverage opportunities supports a refocussing in line with emerging social policy agendas, including societal progress, wellbeing, human development and sustainable development. [Stiglitz, Sen and Fitoussi (2009); European Commission (2009), UN High Level Panel (2013)]. The effective development and production of social statistics to meet the information needs of the 21st century requires a focus that integrates

themes. A transformed information system, integrating data from multiple sources, can deliver big gains.

This paper is in two parts. The first will consider some of the current initiatives underway or currently being explored to enhance the quality and depth of the ABS social statistics program. The second will look forward to see how these initiatives contribute as building blocks to construct the ABS social statistics program of the future.

1. Current developments in social statistics

1.1 Address register based social collections

While Australia has household and person registers built for specific purposes such as managing the electoral and telephone systems, until now these have not been able to be utilised in any significant manner for statistical purposes. Barriers of coverage, accessibility, development cost, confidentiality and general quality with existing registers have meant that almost all major social collections conducted by the ABS still rely on a multistage area based sampling methodology.

The ABS is now close to developing a national address register for use in the 2016 Census of Population and Housing, and eventually for ABS' program of household-based social surveys. This register is constructed from Australia's Geographic National Address File (G-NAF), a geocoded address index for the whole country, listing all valid physical addresses in Australia. G-NAF is assembled by state and local government authorities and aggregated and cleaned according to an agreed set of business rules.

To adapt this file to build a frame of dwellings , the ABS has gone to significant lengths to understand coverage problems and determine how they can be treated with area-specific business rules. Arrangements are also being established with the custodian of G-NAF, for rules to update the address register to ensure ongoing consistency and temporal coherence. Further advances will emerge through the integration of further dwelling information and spatially located data from other sources to support more advanced sample designs and make the address register a data source in its own right. Of the most significance are the ABS' efforts to access Australia's Integrated Public Number Database (IPND) which is a listing of all valid national telephone numbers. If the challenge of associating a personal telephone number with a household can be solved, the result would be a substantial reduction of enumeration costs for most household-based surveys and a significant increase in sample design options and resulting design efficiency. The capacity to utilise dual frames such as the national address register together with another source where coverage of a single frame is inadequate is also an area flagged for further investigation.

1.2 Collecting social statistics via webform

The collection of high quality social data over the web is a major emerging challenge for survey designers. For the ABS, developing methods to electronically collect data from business still offers the best returns for investment. However there are also compelling drivers pushing the rapid development of electronic/web based options for ABS social surveys, including cost efficiency, respondent amenity and rapid receipt of data.

Several factors need to be considered when determining which surveys are given the highest priority for development of an e-form, particularly given the significant set up costs to design, develop, test and deploy a web based instrument. Collections that are conducted frequently, have relatively simple cognitive requirements or target the general population (against those that target specific subgroups like the elderly, immigrants etc) are considered most suited to an e-form. For the ABS, another consideration is that the full

savings arising from introducing electronic modes are likely to be slow to eventuate with the clustered sample designs widely utilised in many collections requiring a very high e-form take up rate to substantially reduce enumeration costs. Changing the characteristics (especially location and size) of the survey interviewer pool also takes time.

In developing electronic collection solutions, a range of challenges have to be addressed. Much of this work covers new ground and the ABS is collaborating with a range of stakeholders including other National Statistical Offices in addressing these issues:

i. Technical and cognitive issues emerge in building solutions for multiple devices and operating systems especially given the rapid rate of change in the range of communications media coming to the market.

ii. To seamlessly integrate electronic data capture with the other parts of the statistical production process often requires substantial reworking of the whole process. This stretches financial resources, puts business as usual work at some risk and requires substantial investment in human capital to conceptualise and build the solutions.

iii. It is important to develop standards for collection of key data items, not just within the one collection framework, but also across the ABS social statistics program and in harmony with international approaches. This will take time and also may require compromises to be made for individual designs for specific data items.

iv. Accessibility and respondent preference dictate that a large number of respondents are adverse to web based collection, thus requiring multiple response modes to be offered for the same collection. This creates a challenge where differences can emerge in responses via different modes which impacts on the collection design. Do you design a web based form to most closely mimic the effects and output from other modes or do you design the most effective e-form which might produce higher quality electronic data at the expense of comparability between modes and may require a total collection redesign? In considering where mode effects are most significant, some early observations and research suggest those data that require high cognitive effort, data requiring coding or classifying, and sensitive data are likely to be the main areas where difference appears. Self-enumeration also potentially threatens the coherence of survey scope and coverage across modes as the burden for determining key survey attributes switches from the interviewer to the respondent.

v. The new technology raises issues of accessibility, particularly for people with impairments. Usability for some populations, particularly the visually impaired, can be significantly enhanced with good design. However, this also requires substantial effort in development.

1.3 Getting better value from existing survey arrangements

The ABS has a well-established set of social surveys. There is significant investment in the design and conduct of these collections in their current form which tends to work as a barrier against major change. Notwithstanding this, options for substantial re-engineering of this program starting with new designs from scratch, are being explored with a view to evaluate and cost transformational change.

In addition, a range of options for change requiring minimal to moderate survey and collection redesign and restructuring are being explored in the current period. These options include:

<u>Longer use of samples once recruited</u>

A major component of the cost of ABS surveys is the recruitment phase. Until an address register with contact details is fully available, staff will continue to be deployed in the field to recruit survey participants in the majority of selected dwellings in most collections. This is a feature of the ABS Monthly Population Survey (containing the Labour force collection and a range of supplementary surveys), which sees participants recruited for eight months. To take advantage of this expensively recruited sample, the ABS is starting to experiment with using the selected sample beyond eight months (months 9-12) to conduct other collections and for testing.

Split forms to optimise contact of specific populations

The ABS goes to great expense to screen for rare populations in certain surveys (e.g. Aboriginal and Torres Strait Islander peoples, population who are disabled or caring for a disabled person). Screening is often a simple matter of door-knocking many households until one with the desired characteristics of the specific survey is discovered. The Census of Population and Housing can assist in this process by giving information about the prevalence of specific populations within small areas. However a more cost effective approach being explored is to couple a range of social collections into one survey vehicle, with the specific collection chosen after the application of screening questions that determine key characteristics of the population unit.

Better sample targeting

Until recently, the selected survey samples for ABS special supplementary surveys (SSS) were selected in parallel with the sample for our Monthly Population Survey. Although this is generally more efficient for enumeration as sample co-location allows for more optimal field interviewer location, it resulted in a less optimal design where a SSS was attempting to target a specific population (generally lower socio economic locations). Under the latest, 5 yearly, design of our survey samples, the urban portion of the SSS sample was selected independently of the Monthly Population Survey sample, using design information obtained from the Census of Population and Housing that identified the prevalence of low socio-economic households in specified small areas.

Tailoring enumeration

Certain subpopulations present various enumeration opportunities and challenges that can reduce costs of survey enumeration while maintaining quality. These opportunities are generally being explored with regard to our largest social collection the Monthly Population Survey. As an example, persons living in certain types of dwellings such as prisons, mining camps, hospitals etc, may tend to change their social circumstances (e.g. particularly employment status) less frequently than other population members. Accordingly enumeration can also be less frequent, presenting a large cost saving for a small quality impact to monthly statistics.

1.4 Working across domains to produce new statistics

In other parts of this paper, we have already touched on the importance of developing social statistics that span multiple statistical themes. Several very specific work streams have this objective as a primary goal.

Data integration

Data integration is the joining of data from different sources at the unit record or small area level to develop new datasets. The potential value to the National Statistical System from combining high value data sources, previously used in isolation, to generate an expanded range of statistics for policy formulation and to pursue new research opportunities is immense [Conn, 2010]. For this reason, data integration is an area of significant growth across the ABS.

Data integration is also a challenging area to pursue as many datasets being utilised are complex and can have significant quality issues such as poor metadata or incomplete coverage of the population of interest etc. Data access and use require a very high level of public trust and data custodians are often reluctant to relinquish control over datasets and the statistics that they can generate. To address these concerns, all heads of Commonwealth Government agencies in Australia have endorsed a set of principles to govern the integration of Commonwealth data for statistical and research purposes, as well as a set of governance and institutional arrangements to support these principles.

Significant linkage work is proceeding under these arrangements, with datasets integrated covering a range of diverse domains including education, migrants, income tax, Census, Indigenous mortality, early childhood and more. The progression of linkage work is itself leading to significant statistical development work across different stages of the Generic Statistical Business Process Model. This work includes the development of quality measures for integrated data, the development of more powerful linkage methodologies, and the refinement of dissemination methods to appropriately release sensitive datasets.

Broad measurement of societal progress

In order to address an emerging and strengthening global interest in progress being made across multiple facets of society, in recent years the ABS has continued to update its flagship statistical release 'Measures of Australia's Progress' (MAP) [ABS, 2012]. This increasing interest at both national and international levels is manifested in further understanding what societies really care about, how to account for the social and environmental dimensions of economic development, and what dimensions of progress are important within specific communities and local areas.

Accordingly, the ABS has recently undertaken a broad-ranging consultation that asked Australians: 'What is important to you for national progress?' This consultation will lead to the development of a refreshed set of indicators for the next generation of MAP [ABS, 2012]. A new version of MAP, based on the consultation results, will be released in the second half of 2013.

2 The social statistics program of the future

The innovations noted to date are all pieces in a big puzzle that, when assembled, will form the ABS social statistics program of the future. Even as the ABS pursues many smaller statistical developments, the complete framework for the program of the future is also being visualised through both exploring the future in individual statistical theme areas and population groups, and through alternative 'whole of program' models being developed by a small dedicated team. This work is progressing in 2013-14 with a view to have some alternative program models to test and evaluate shortly thereafter. Not only is this work essential in its own right to future proof the ABS social statistics program, but it is also vital support to the transformation facing the whole ABS as it will determine the infrastructure required to support the ABS' population, labour and social statistics program going forward.

The Census of Population and Housing will be the key building block around which the future social statistics program will be shaped. The program itself will be strongly dependent on opportunities the Census affords in enhancing survey design, sampling and data integration. Other possibilities include strengthening the scheduling of the whole social statistics program around the Census to improve collection efficiency and output range and quality. The potential also exists to vary Census content according to the characteristics of the household or to replace some directly collected Census data by administratively sourced data (e.g. income reported to the Australian Taxation Office).

Most models proposed at this stage also feature a greater role for continuous and omnibus style collections to leverage the benefits of longitudinal and cross sectoral data and for the purpose of more efficient collection. Such collections, augmented with data integrated from other sources, will go a long way to informing the complex stories that cut across a number of themes. Accordingly, it is also certain that administrative data will play an ever increasing role as a data source.

The key challenges of this work can probably be summarised into a few related themes. The first is ensuring that change is pursued effectively while not harming business as usual. This is a difficult assignment with high quality staff needed to cover both angles and with limited resources available to pursue transformation activities. The second issue is the practical requirement to build parts of the solution now while the broad plan for complete program transformation is not yet agreed. The challenge is akin to laying bricks for a house before the architect has finalised the plans.

There are a number of other significant areas of exploration worth mentioning. These particularly relate to statistical methodology or infrastructure areas and are likely to warrant significant attention in the future program. These include the potential (depending on the characteristics of address registers that may be built) to move to sample persons rather than households in many collections, the effective use of metadata to drive the whole statistical cycle and the development of better dissemination methods preserving confidentiality (particularly for administrative and integrated data).

3 Conclusion

Visualising and executing a model for the future is hard. Building a framework to provide a future population, labour and social statistics program that is relevant, responsive and trusted requires clever thinking, ambition and a high tolerance of risk in large measures. As the ABS embarks on this journey in the social statistics space, we expect to come out with a model that secures us for the future. In this process, we encourage other NSOs to share similar experiences of transformation to the benefit of us all.

References

Australian Bureau of Statistics, (2012) *Measures of Australia's Progress. Aspirations for our Nation: A Conversation with Australians about Progress*, cat. no. 1370.0.00.002.

Conn, S. (2010) Widening the Data Net: NSO Leadership role, http://isi.cbs.nl/iaos/YoungStatisticians/papers/2011-2nd-Conn.pdf

Pink, B, Borowik, J, Lee, G (2010) *The case for an International Statistical Innovation Program* http://www.abs.gov.au/websitedbs/d3310114.nsf/4a256353001af3ed4b2562bb00121564/064584f68877204fca2576c0001a0fa8!OpenDocument,

European Commission (2009). *GDP and beyond: measuring progress in a changing world* 'GDP and beyond: measuring progress in a changing world' ,

Stiglitz, J, Sen, A, Fitoussi, J-P (2009) *Report by the Commission on the Measurement of Economic Performance and Social Progress,* Commission on the Measurement of Economic Performance and Social Progress.

United Nations High Level Panel of Eminent Persons on the Post-2015 Development Agenda (2013), *A New Global Partnership: Eradicate Poverty and Transform Economies through Sustainable Development*