25th Meeting of the Wiesbaden Group on Business Registers
- International Roundtable on Business Survey Frames

Tokyo, 8 – 11 November 2016

Luisa Ryan and Linda Bencic
Australian Bureau of Statistics
Session No. 4

Quality and Coverage

Quality Improvements to the ABS Business Register

Introduction

Statistical business registers are foundational infrastructure in the collection and production of economic statistics and as such it is critical that they are high quality. A high quality business register is the product of quality (1) administrative data and (2) profiling information as these are the key inputs to both establishing and maintaining such a register. The ABS Business Register has a two population units model. Large or complex units are captured in the profiled population via direct contact with the business. The remainder of units (the non-profiled population) are assumed to have simple structures. These are updated primarily through use of administrative data from the Australian Taxation Office (ATO). Both populations are regularly updated as part of the ongoing ABS Business Register maintenance program. To continue to improve quality the ABS also collaborates with key stakeholders to test and introduce new initiatives. This paper provides two case studies of where the ABS Business Register Unit is working with key stakeholders to ensure and improve ABS Business Register data quality. The first case study outlines the use of point of contact coding by the ATO to code businesses to an industry, the impact of this on ABS industry coding and how ABS have been working with the ATO to improve this process. The second case study outlines a joint pilot project that is currently in progress with the ABS Annual Industry Statistics team to combine profiling and editing functions in order to inform optimal business structures, better manage respondent burden and identify ways of integrating these functions in the future.

ABS Business Register maintenance

1. In Australia, the ABS Business Register is the source of the majority of economic survey frames, business demography statistics and increasingly, it is being used to enable linking / integration with other datasets. There is an expectation within the ABS that the ABS Business Register will reflect real

world business structures and that the contact details, industry and size information will be accurate and updated in a timely manner. The register maintenance program is managed by the ABS Business Register Unit (BRU). There are dedicated maintenance teams responsible for each of the administrative data and profiling streams.

2. A major ABS transformation program is currently in progress. This will deliver changes to the conceptual model underpinning the ABS Business Register, plus enable maintenance approaches to further evolve, for instance through the use of more efficient systems and a consolidated business survey instrument. The ABS Business Register case studies in this paper support this broader ABS transformation program.

Administrative data overview

3. The maintenance of administrative data for the non-profiled population on the ABS Business Register relies on monthly data updates from the Australian Business Register (ABR) that is maintained by the ATO. The monthly updates include new businesses (births), ceased businesses (deaths), businesses that have recommenced (resurrections) as well as changes to name, address (location and postal), contact details, and industry. In addition, updates in terms of tax obligation roles applied for and cancelled (e.g. Goods and Services Tax (GST), Income Tax Withholding (ITW)) are received from the ATO. Updates to size are captured via the receipt of turnover data from the quarterly Business Activity Statements and biannual payees data also from the ATO. Survey specific information (e.g. for tourist accommodation and international trade in services) is maintained by the survey area but appended to the ABS Business Register by the BRU for use in the production of survey frames. ABS employs a series of quality gates to ensure that the ATO data quality and ABS derivations are within expectations. There are standard checks to identify missing data, remove duplicates and suppressed units, and analyse changes to the ATO dataset. Where there are significant or unexpected changes for particular data items, these are discussed with the ATO.

4. The coding of industry data has been undertaken by the ATO since the early 1990s. Industry coding on the ABR commenced with the introduction of the Australian Business Number (ABN) in July 2000. Units were initially coded to the Australian and New Zealand Standard Industry Classification (ANZSIC) via an automated coder provided and supported by the ABS. Units that failed to be assigned a code by the auto coder were manually coded by ATO staff including contacting providers to determine the appropriate industry code (where necessary). This arrangement was supported by

ongoing ABS training of ATO staff, the establishment of quality expectations (e.g. 90% accuracy for employing units) and ABS peer review of a sample of units each month. In December 2013, the ATO introduced point of contact coding. The following case study discusses this change to industry coding and how ABS and the ATO have been working together to maintain and where possible improve the quality of information produced regarding industry.

Case study – coding of industry data

5. Prior to December 2013, the online ABN registration form had two data fields to enable industry coding: the main industry in which the business operates (drop down box) and the main activity of the business (free text). Around half of all registrations were able to be coded using the autocoder. Despite ATO investing significant resources in manual coding, the number of legal entities without a valid industry code grew to exceed 1 million legal entities, with around 140,000 of these in scope of the ABS common frame. To address this situation, ATO worked with ABS to develop a more effective process for allocating industry code during the registration process.

6. The new approach is referred to as point of contact coding (POCC). It involves the applicant entering their main business activity as free text in an on-line form. The tokens of the free text are then parsed through an index of terms used in all ANZSIC category descriptions. Using a predetermined set of rules and weighting algorithms, the most relevant category descriptions are returned as a dropdown list in order of relevance for the applicant to select. Each category description has an associated industry code (e.g. 3242 Carpentry Services) and main industry code (e.g. CON (Construction)). The coder has in-built logic to cater for common grammatical variations (eg. plumbing, plumber), common misspellings (eg. accommodate, accomodate) and 'noise' words (eg. after, about). Registrations cannot proceed without a category being selected. This approach has improved industry coding via online registration to 100%. There are still a small number of papers forms that are submitted and these are still manually coded. ABN holders have an obligation to update details with the ABR within 28 days of a change occurring, including any changes to industry activity.

7. ATO are the owners of the POCC. However, to assist with the POCC development four ABS staff were outposted to the ATO in 2013 to review the functional documentation, improve the index list and undertake testing of the POCC, including a comparison with the manual coding methodology. The ABS also had opportunity to input into the directions being taken with the development and implementation of the POCC via membership on the Board that oversees ABR directions.

8. Both the ABS and ATO have a strong interest in the maintenance of high quality industry data, noting the ABR is used by around 500 government agencies for both unit record and aggregate data analysis. A joint work program has been established to understand the quality of industry data delivered by the POCC.

9. The ATO provide ABS with quarterly snapshots of the main business activity free text description and the chosen category description from the POCC in addition to the associated 5 digit industry code and main industry. Using this information, the ABS has been able to undertake the following exercises:

    - Analysis of the coder's effectiveness is undertaken a couple of times a year. This involves checking how well the main business activity description provided matches the chosen industry code description. The POCC coder is achieving an 80% match rate with the ABS coder for those units that can be coded. A number of areas have been identified for improvement to the POCC. These include terms covering on line selling, cleaners, and mortgage brokers. The ATO have built a facility to test proposed changes to the industry weightings and updates to the POCC weightings will be implemented by the ATO in due course.

    - An analysis of the impact of the change in the distribution of ABN industry coding across division and subdivision levels one year after the introduction of the POCC was conducted. The overall conclusion of this exercise was that there has not been a significant impact on the quality of industry coding for units survey frames. All subdivisions had less than 1% absolute proportional change except for other store based retailing (-1.29%), construction services (-1.25%) and professional, scientific and technical services (1.06%) for employers, and building construction (1.54%), property owners and real estate services (-1.2%), and construction services (-1.06%) for non-employers. Taking into account real world events, there were three subdivisions identified for further investigation: Commission based Wholesaling, Construction Services, and Professional and Technical Services.

10. Currently the only way to confirm that the selected industry code actually reflects the main business activity undertaken is via contact with the ABN holder. The ATO run an annual integrity survey with a sample of ABN holders. The Survey conducted in 2016 found that 85% of ABNs had the correct industry category descriptions recorded on the ABR, increased from 80% in 2015. The ATO are committed to undertaking the following activities to improve the integrity of the ABR:

- Direct contact with ABN holders where there is evidence to suggest details are incorrect or out of date
- Cancellation of ABNs where there is no evidence of activity in the last two years
- Synchronisation of ABN information with other government agency databases where those agencies are the source of truth (and the information can be released).

11. Moving industry coding to the POCC has reduced the accumulation of invalid industry codes and the quality has been within acceptable limits. Historically however, where a significant unit in the non-profiled population is considered to have an incorrect industry code on the ABR, these units have been moved into the profiled population. This is because information on industry coding collected by ABS cannot be released outside of ABS. In the future, an ABS industry flag may be used for the non-profiled population.

12. The coding of industry data by the ATO has been a good initiative with whole-of-government benefits as ABR industry data can be released to other government agencies. This has resulted in a more consistent use of industry data across the Australian Government for a broad range of purposes including analysis, policy making, service delivery and emergency management.

13. The ATO have plans to expand ABN registration options. In the future, in order to streamline whole of government operations, ABN registrations will be available from other government websites and via accounting software packages. This change will present a new set of challenges in relation to industry quality coding and future updates to industry classifications. The partnership between ABS and ATO will provide an important basis to work through these and find solutions.

Profiling Overview

14. Profiling is a method for analysing the legal, operational and accounting structures of large Enterpriseroups in order to build the statistical units with the most efficient structure for the collection of statistical data. Profiling has been used by the ABS since the late 1980s to ensure the ABS Business Register accurately reflects the structure of units undertaking major economic activity in the Australian economy.

15. The ABS profiling program records the structure and activities of Enterprise Groups according to statistical standards and classifications, through constant monitoring of economic media and other sources such as major projects, mergers and acquisitions. ABS profiling achieves strong outcomes by

building the trust and cooperation of providers as well as representing the needs of ABS stakeholders to ensure user requirements are satisfied.

16. Profilers have a strong understanding of statistical and economic concepts and the ability to effectively liaise with data providers. Each cycle, profilers are allocated to industry teams that are established to develop expertise within particular industries. Ongoing training and peer review is also employed to ensure data quality is maintained. Regular audits and validations are conducted to identify any key issues with data or areas for investigation. Management information is maintained on a weekly basis to monitor progress with profiling allocations across industries and for each profiler.

17. There are two methods used by ABS for profiling:

    - Personal profiling: The largest and most complex businesses are profiled using the 'personal profile method' requiring the profiler to interview the data provider; and
    - Mail profiling: where the business is still significant enough to be profiled, however are less complex and can be profiled using the Update of Business Structures (UBS) survey form. Profilers then conduct any necessary follow up to ensure that information is both received and correct.

18. Each year the ABS profiling cycle commences in late February and concludes in December. At the beginning of the cycle the Update of Business Structures (UBS) survey form is dispatched for all mail profiles. A Microsoft excel version of the paper form is also available for personal profiling or where the number of activities undertaken by a unit exceeds the capacity of the paper form.

19. Data collected during ABS profiling includes ABNs, main activity, wages and salaries, business activities, employment and contact information. Coverage information relating to activity in trade in services, international investment and research and development is also sought. The profiling data obtained is entered into the ABS Business Register database, known as BRIMS (Business Register Integrated Management System). Each quarter a Common Frame is extracted from BRIMS forming the basis for the selection of units in ABS economic surveys.

20. Throughout the profiling cycle there are a number of activities occurring simultaneously which may also trigger the need to profile or re-profile a unit. These include identified 'new-ons' (triggered through admin data), sample frame maintenance procedure (SFMP) queries (notification of change initiated by a provider or a survey area), 'drops' (using size benchmarks to determine which

businesses can be moved to the non-profiled population), and 'floating ABNs' (a quarterly process identifying units likely to belong to profiled groups). Real world events such as privatisations are also monitored and structures updated as appropriate.

21. The benefits of profiling are that it allows the ABS Business Register to be responsive to changes in the economy in order to support the broader ABS economic statistics work program. Profiling builds relationships with data providers, it produces better response rates (as a result of data providers better understanding the purpose of data collection), it results in more accurate data (as legal entity administrative data is not always available or accurate) and it reduces workloads for businesses.

Case study – combining profiling with other statistical functions

22. In early 2016, the BRU and the Annual Industry Statistics (AIS) areas within the ABS embarked on a pilot to test the potential for integrating previously siloed functions with a view to improving efficiencies (process), improving data quality (outcomes) and optimising the provider experience (process and outcome). The AIS produces information about the structure, financial operations, performance and production of Australian industries. This information is used to monitor the economy and the business cycle, and support the compilation of the Australian National Accounts and Environmental Accounts.

23. This pilot was established in response to the broader ABS transformation program that aims to reduce cost, reduce statistical risk, reduce red tape, reduce time to market of the statistics produced and to grow the business. The pilot aims to support achievement of the goals of this transformation program by providing intelligence and evidence to guide best practice for business survey consolidation, explore the potential to utilize existing or alternate data sources and inform critical infrastructure requirements**.**

24. Large businesses are critical to understanding changes to the economy over time and are the main contributors to the completely enumerated streams of annual, sub-annual and ad hoc business surveys. However, current ABS processes for the profiling and editing of survey data for large and/or complex businesses are undertaken in isolation from each other which can result in process duplication and inconsistent provider management.

25. In the editing and validation phases of the statistical production work program, AIS also contacts providers to query where there are movements that are contrary to expectations. Reasons may

include real changes in business conditions, dynamic changes to business structures such as acquisitions, take-overs or mergers that are not reflected on the sample frame, as well as a misalignment between what providers believe they can report and the requirements of the survey collection. The pilot was established to design and test a standardised approach to profiling, collecting and editing of large businesses and a centralized provider management and survey processing function for large businesses.

26. The pilot team was established with three staff from each of BRU and AIS sections. The pilot is concentrated on the following Industry divisions: Mining, Manufacturing, Utilities, and Construction.

27. The first phase of the pilot was conducted between March and August 2016 and focussed on profiling outcomes. Some preliminary observations from this first phase are as follows:

- Some profiling and annual survey editing processes are duplicated with limited integration. For instance, intelligence collected via annual surveys is not used by profilers to assist in setting optimal business structures.
- The provider management of businesses across ABS is not co-ordinated.
- Balancing provider reporting and frame requirements is not easy. In particular, in industries where structures change frequently (more than annually) businesses may end up reporting against structures that were established through the previous years' profiling.
- Businesses often have a poor understanding of the importance of the business reporting structure and how data are used to broadly inform the national economic interest. Compounded with this, providers often have difficulties with particular data reporting requirements; this is usually related to a mismatch between the timing of surveys and the availability of their own internal reports.

28. The second phase of the trial will be conducted between September 2016 and March 2017. In this phase the pilot team will undertake joint profiling and editing activities, and the pilot will continue to inform and develop strategies to achieve the following key objectives:

- Integrated processes – identify duplication of processes; expand the sharing of information to inform profiling and editing decisions; inform future infrastructure requirements
- Provider management – articulate reporting issues; assess provider burden; identify different applications of policy and processes; enhance the provider experience
- Coherence and data quality – identify different reporting units impacting on coherence; explore treatment of split units

- Capabilities – identify key new staff capabilities that will be required to support integrated profiling and editing in the future
- Improvements to profiling procedures and outcomes.

Conclusion

29. The ABS is continually looking to improve the quality of the ABS Business Register in conjunction with key stakeholders. This paper has provided an overview of the ABS Business Register maintenance program and outlined two case studies highlighting how the BRU has been working with an external stakeholder to improve administrative data industry coding and an internal stakeholder to improve economic statistic program outcomes via the integration of profiling and editing functions.