**Meeting of Group of Experts on Business Registers
Geneva, 30 Sep – 2 Oct 2019**

Luisa Ryan and Andrew Naughton

Australian Bureau of Statistics (ABS)

**Session: Use of Administrative Data Sources for Statistical Business Registers**

**Using Administrative Data to improve data quality in the ABS Business Register**

**Introduction**

The Australian Bureau of Statistics (ABS) Business Register was established in the early 1970s using information from the 1969 Economic Census. Following this, registrations data from the Australian Taxation Office (ATO) have been the primary source for updating the ABS Business Register, supplemented by the introduction of profiling of large businesses in the 1980s.

The use of administrative data in the ABS Business Register has delivered many benefits to ABS and the Australian community. Administrative data has replaced the expensive and time consuming exercise of running an integrated economic census. It has reduced respondent burden, and it has improved the quality of data on businesses in Australia through improved coverage and a coherent use of classifications and units.

There is a strong nexus between the ABS Business Register and the taxation system, and the ATO is a strategically important relationship for the ABS. The ABS has worked with the ATO to take advantage of ATO developments to improve and enhance the ABS Business Register. More recently, ABS has also been exploring if other data available within the ATO and other government agencies could be used to improve the data quality of the ABS Business Register and expand its scope. This has been managed taking into account that administrative data is not usually designed for statistical purposes so it needs to be understood and carefully managed to leverage the use benefits (eg. increased coverage, improved classifications, reduced provider burden and growth in the business).

This paper will present: (1) a short history of the usage of administrative data in the ABS Business Register (2) the critical importance of relationship management with the administrative data provider and some strategies ABS has employed (3) the exploration of other administrative data sources to support emerging requirements; and (4) ABS strategies to handle data quality and risk issues when using administrative data.

**History of usage of administrative data in the ABS Business Register**

In the 1960s it became evident that to measure the economy comprehensively broader statistical infrastructure was required. The strategy was to develop an integrated economic census to improve the quality of industry statistics and provide data for use in the National Accounts. The 1969 ABS Economic Census integrated manufacturing, mining, electricity and gas, retail trade and selected services, and wholesale trade data. It ensured comprehensive coverage, consistency in scope across collections, standardization of concepts, and consistency in the reporting units via an economics unit model. From the 1969 Economic Census the first ABS integrated register of businesses was borne.

Following the 1969 Economic Census, regular updates of ATO Group Employer Registrations were used to maintain the ABS Business Register, supplemented by an ABS group employer census to collect additional data items. Group Employers were businesses registered to make income tax deductions from the salaries they pay to their employees for transmission to the ATO. This was facilitated by special legislation enabling the ATO to provide ABS with these records.

As the ATO data holdings have evolved, ABS has been able to leverage these changes to implement the following developments to the ABS Business Register and frame products:

* In 1983 ABS and ATO established a joint project to examine the options for expanding the use of taxation data. As a result, in 1986 the *Income Tax Assessment Act 1936* was amended to give the Statistician access to identified taxation records of businesses for Administering the *Census and Statistics Act 1905*. Income tax data was subsequently implemented as a source to supplement and improve coverage in the late 1980s.
* In 1989 the ABS received the first electronic file of group employers. This was used to update the ABS Business Register with improved timeliness and it reduced the coverage leakage that resulted from the use of the paper based process.
* In 2001, significant taxation reform was introduced, including a single business registration process and a goods and services tax:
	+ The Australian Business Register (ABR) was introduced as a registry for all Australian businesses who were issued with an Australian Business Number (ABN). The ABR is enacted via the *A New Tax System (Australian Business Number) Act 1999 (ABN Act)* and enables data to be shared with other Agency Heads for the purpose of carrying out the functions of the Agency. The ABR gave ABS access to a list of all registered businesses (economy wide coverage) with a unique identification number. In addition, registrations of non-employers also become available. The trade-off with the use of this source was the loss of establishment and location data.
	+ The introduction of the *A New Tax System (Goods and Services Act) 1999* resulted in the introduction of Business Activity Statements (BAS) that in most instances are reported quarterly (but can also be monthly or annual). The BAS include data on turnover and wages and salaries, and these data items enabled the development of new size indicators for ABS survey frames. BAS data are also used for estimation and imputation in various ABS surveys. The Economic Activity Survey also uses BAS data to (a) produce synthetic estimates at the 4 digit industry class level and (b) impute the contribution of micro non-employing units.
* The current extract from the ATO was designed in 2001 and combines registrations data from the ABR, and data on tax role and employment from the ATO. It is received on a monthly basis.
* In 2017 a process to modernise administrative business registries in Australia commenced. While the development work for this is still in progress, ABS have been an active participant in the consultation process to ensure statistical objectives will be met. This includes the review of the relevant legislation.

The ABS also receives a range of other ATO datasets on a regular basis for use in other ABS economic and social statistical production programs.

**Relationship management with administrative data providers**

Relationship management with key administrative data providers is critical to obtaining and sustaining high quality data, including leveraging new opportunities and problem solving if issues do arise.

ATO data provided to the ABS is considered strategic (as it reduces ABS direct collect data) and critical (significant impact if not available). As a result, the ABS has a long history of high level engagement with the ATO and the strength of the relationship has resulted in ABS being able to improve the scope and quality of data received over time. The relationship is further supported by a Memorandum of Understanding signed by the Australian Statistician and Taxation Commissioner that has an emphasis on mutual benefit. There is also supporting subsidiary documentation that it updated more frequently.

The supply and quality risks of using ATO data are minimised through strong engagement and building relationships across all levels of both organisations, covering strategic, operational and technical aspects. It is through these interactions that a shared understanding and accountability to manage risk is built, including ensuring the delivery of ongoing data requirements and maintenance of quality.

Some joint strategies that ABS and ATO have employed to foster the joint relationship include:

* Senior Executive ABS / ATO meetings to develop strategy and ensure the health of the relationship.
* The ABS Deputy Statistician of the Statistical Services Group is a Member of the ATO Business Registry Strategic Advisory Council. Participation on other Boards in the past has ensured good statistical outcomes. For instance, when the BAS reporting was introduced ABS was able to contribute to the business case to include some data items that would support economic measurement (eg. turnover, wages and salaries, capital / non-capital expenses).
* Ongoing and regular operational meetings at the dataset level. The Business Register Unit have bi-monthly meetings with the ABR /ATO, plus more regular meetings on specific project work.
* A permanent ABS Outposted officer (relationship manager) to the ATO. The role of this person is to manage the overall relationship, including acting as a liaison point between the two agencies and identifying new opportunities for both agencies.
* Outpostings to support specific initiatives as required. Recent examples include the placement of four ABS staff to help develop an ABR industry coder in 2013, and the outposting of ABS staff in 2014 and 2015 to help develop employment and turnover indicators.
* Regular reporting to ATO on how ATO data are being used by ABS, including for data integration projects and its data quality.
* The organisation of hackathons to explore potential uses of ATO data.
* As part of the MOU, ABS provide an agreed quantum of advisory work to the ATO each year, usually on statistical methods.

Note: ABS legislation does not enable the sharing of unit record data with the ATO.

**Exploration of other administrative data sources to support emerging ABS Business Register requirements**

There are a number of emerging and contemporary statistical requirements that need to be supported by the ABS Business Register. These include the measurement of the finance sector, and international / globalisation measurement. The use of additional administrative data sources can support improved data quality, the development of new data items, and enable new uses of the ABS Business Register. Such developments support both regular statistical production and new innovative statistical solutions.

To support these emerging statistical requirements, the ABS has developed the ABS Business Register Spine concept. The spine contains the minimum information needed to integrate individual datasets. The integration of different datasets happens by linking the input datasets and spine to produce the outputs required. This approach enables the derivation of a more diverse range of outputs and presents greater analytical opportunities through the linking of administrative data. The legal entity (ABN) is the key linkage variable. Refer to Attachment 1.

The ABS has a range of datasets that are already available internally. Some of these have been explored to improve quality and develop new indicators, in addition to some other external datasets not previously utilised by the ABS. The case studies below present two such examples.

Case study 1: Improved Data Quality and coherence across collections for the finance sector

In Australia, financial and international collections use a sector based statistical unit rather than a production based unit. This has led to challenges in developing ABS Business Register based frames. To support this objective, the ABS Economic Units model (Attachment 2) has been expanded to include an institutional unit. A significant exercise was undertaken to improve the sector coding for these institutional sector units using a range of additional administrative data sources:

* Business Income Tax (BIT) from the ATO – used to define statistical units and apply sector codes.
* Managed Funds data from the Australian Securities Investment Commission (ASIC) Companies Register – used to define statistical units and apply sector codes.
* Financial entity sector information from the Australian Prudential Regulatory Authority (APRA) – used to cross check outcomes.

The BIT and APRA data were already available within the ABS. The ASIC data was purchased after discussion with the ASIC Registry team.

These sources also indirectly helped improve the industry coding across all ABS economic units where there are known relationships between industry and sector coding, particularly in the finance sector.

The ability of the finance and institutional sector surveys to use ABS Business Register based survey frames will also improve the coherence of economic statistics across the production and sector based collections. This opens up opportunities to link finance sector data to the ABS BR spine to enable data confrontation and statistical solutions via data integration or modelling. In the future there will also be opportunities to replace direct collect data with administrative data due to the link between production and institutional units.

Case study 2: Development of new indicators for globalisation

In our globally connected world it is critical to understand the activity of multi-national enterprises. Statistical Business Registers are well placed to support this need via the inclusion of foreign ownership and activity indicators. The ABS Economic Research Hub and Business Register Unit have undertaken an exercise to attach a foreign ownership indicator to Australian companies as a general resource and for use in analytical projects relating to multi-national enterprises.

The experimental indicator has been developed by using foreign shareholding information reported on the Business Income Tax (BIT) form to identify the level of foreign ownership reported by companies to the ATO at the ABN level. This data is then mapped to the ABS statistical unit (Type of Activity Unit) to create a Level of Foreign Ownership indicator at Type of Activity Unit (ABS statistical unit). A data source for partnerships and trusts still needs to be identified. A data source and method for analysing the economic activity of Australian businesses investing abroad (outward investment) also needs to be determined.

The prototype foreign ownership indicator is now available internally for matching to other survey results to provide new views on international economic activity occurring in Australia and will be incorporated into the ABS Business Register in the future.

In the future the dataset could also be used as an analytical data-set for micro economic analysis and data modelling or as a frame to run targeted benchmark surveys (or undertake case studies) relating to foreign investment

**Strategies to handle data quality issues when using administrative data**

There are advantages and disadvantages to using administrative data. The ABS has used a range of strategies to mitigate the risks and difficulties with the use of both new and ongoing administrative data.

Administrative data is not usually designed for statistical purposes so it needs to be understood and carefully managed to leverage the use benefits. This section uses the ABS Statistical Production Activity Model (SPAM – ABS equivalent of the GSBPM) (see Attachment 3) to provide an outline of issues that ABS has faced in relation to using administrative data in the ABS Business Register and how these have been managed.

Scope, Design and Assemble

*Is the data fit for purpose?*

Large administrative files are often provided to a NSO in an initial exercise for understanding and assessment. Depending on the format, these can sometimes require specialised code to research and interrogate (SAS, SQL, R). Understanding the specific formats and how to use the content can take time and may require ongoing discussion with the administrative data provider. When assessing a new administrative dataset time needs to be taken to understand the benefits and limitations of how data can be used, particularly where datasets are large or complex. A time series is required to understand the “normal” level and any seasonality. It is also important to analyse downstream impacts on survey data and the National Accounts, including any back casting requirements if a new data source is being introduced.

|  |
| --- |
| **ABS case study** The recent changes to the sector coding methodology on the ABS Business Register using the new data sources were developed over a period of 2 years in order to ensure the new data sources were understood appropriately and the impacts to ABS outputs could be managed. A number of papers were presented to the ABS Economic Statistics Methods Board to promote the changes and obtain endorsement.  |

*Alignment of scope, coverage and concepts*

The conceptual scope of the administrative dataset needs to be determined, including whether it includes the full set of units. Where the scope is deficient secondary administrative data sources or some direct collection may be required.

|  |
| --- |
| **ABS case study**When developing the foreign ownership indicators there were some entity types (partnerships and trusts) for which no appropriate data sources could be identified. While this is a deficiency, it has not prevented progress for the entity types where these are available. The incomplete scope will remain on the work program.  |

The coverage is the proportion of activity that the units represent. While a dataset may have conceptual coverage of the economy, there may be some units missing. It is critical to understand the level of coverage and be transparent about this.

|  |
| --- |
| **ABS case study**The ABR includes a location unit. This has been populated at the point of business registration since 2013. As the requirement is not retrospective or enforceable for the stock population, only some large businesses have provided locations data at points in time. Given the data are not consistently maintained (due to the volume and the impost on businesses), the coverage of the stock population is incomplete. ABS have commenced work on a project to supplement the ABR locations data for use on the ABS Business Register. In the longer term, ABS will also continue to work with the ABR to improve the administrative source.  |

Some administrative data items do not align well conceptually with the statistical requirements of the National Accounts. In some instances, it is possible to use administrative data as a starting point and adjust it to derive a more correct economic concept.

|  |
| --- |
| **ABS case study**The ATO and ABS concepts of sector, public / private indicator, not for profit indicator and type of legal entity have some differences. ABS has handled this by using the ATO data as a starting point and applying other information using an ‘on balance’ algorithm to apply codes at the ABN level. For instance, the ABS public / private indicator also takes into account Government Finance Statistics information, and is manually reviewed in a small number of instances.  |

*Quality of coding*

Where possible if statistical coding can be applied in administrative data then this can be a broader resource across government (as administrative data can often be shared at the unit record level) and it can also be a more efficient arrangement. The key consideration from the NSO perspective is the coding quality. This requires ongoing investment by the NSO.

|  |
| --- |
| **ABS case study** Over many years ABS has worked closely with the ATO to embed the *Australian and New Zealand Standard Industrial Classification (ANZSIC)* into ATO business industry codes, as industry is critical to taxation objectives. The benefit to ABS is that ATO have been able to code the industry of an ABN at the point of registration and also apply subsequent updates. ABS has assisted by providing ANZSIC training, aggregate feedback on the quality of ATO coding via audits and investigations, and assistance in the development of point of contact industry coders. This has proved a beneficial whole of government arrangement that reduces overall respondent burden.  |

Acquire

*Unexpected changes to administrative data*

As the primary purpose of administrative data is generally not to support the compilation of statistics, sometimes changes can be made to the administrative data for other purposes or without warning. Administrative data providers may only maintain a dataset at regular intervals rather than in real time, and the resulting bulk updates can have statistical impacts if not understood and managed by the NSO. Statisticians need to be aware of the maintenance methodologies of the administrative data supplier and have contingencies when these change.

|  |
| --- |
| **ABS case study** ABS have regular discussions with the ABR and ATO to understand their change management program and the impacts of any campaigns that they may be conducting. This helps ABS to manage any potential statistical impacts from the known changes. On the occasions when an unexpected impact does occur, ABS will discuss with ATO to understand the cause. When there are a large number of unexpected classification changes the strategies applied have included the resupply of files to ABS or ABS rolling back to a historical classification until the impact is better understood and phased in.  |

*Administrative data ceases*

Sometimes the administrative data provider no longer needs to collect either a dataset or specific data items. A strong relationship with the data provider can ensure prior warning around such change, and can potentially influence outcomes. Often the statistical outcomes are weighed against other objectives such as respondent burden. In these situations it is helpful to take a whole of government approach as if the administrative data source ceases, this may lead to increased data collection by the NSO.

|  |
| --- |
| **ABS case study**In 2016, when the ATO planned to minimise the number of data items on the quarterly Business Activity Statement (BAS) form to reduce respondent burden, ABS were given advance notice. After further discussions, ATO agreed to retain the key data items of interest to ABS and the implementation period meant that ABS could develop other sources to replace the lost data items. Recently, some data was removed from the BAS dataset. Once this was identified ABS and ATO worked together to understand the issue and establish a way forward, including some alternate datasets. The Business Register Unit have a business continuity plan that guides the response if administrative data are not received as expected or significant quality issues are identified.  |

Compile, validate and disseminate

Once implemented, administrative data should be subject to quality gates and measures, as well as clearance processes to ensure data are fit for purpose each cycle.

|  |
| --- |
| **ABS case study**ABS apply quality gates throughout the statistical production process. Quality gates are checks to ensure that expectations have been met at critical points. The ABS Business Register applies quality gates when administrative data are received into ABS, when data are loaded into the Information Data Warehouse, when data are loaded to the ABS Business Register, and when outputs are extracted. The quality gates include quality measures (criteria and tolerances) using a traffic light system and have accountabilities around sign off. While there are many of these, some examples include:* Number of administrative records received
* Counts of active ABN births on ABS Business Register
* Movements in counts by industry (ANZSIC)

Each quarter the administrative data on the ABS Business Register are reviewed through the Common Frame Clearance process. The Clearance report includes a summary of the key movements in the data and highlights any issues or changes. The clearance process includes staff from the Business Register Unit, Methodology, the Tax Data Section and Business Statistics representatives from annual and quarterly collections. Where known changes (eg. remediation, new data sources) are implemented, risk management plans are used and impact analysis are also presented.  |

Evaluate

*Review of the datasets*

While it is critical for administrative datasets to be stable, in time they may need to be reviewed to ensure both technical and scope currency. With advances in systems and technology the data formats and manipulation of the administrative data may be able to be improved. It is also possible that administrative data sources may expand over time or new collections may be introduced. The NSO should keep a watching brief on this to maximise any data opportunities.

|  |
| --- |
| **ABS case study**In 2014, to reduce data security risk and increase efficiency, ABS and ATO worked together to move the receipt of data from a disc that was couriered to the ABS to an electronic file that could be securely transferred between the two agencies. ABS has been receiving the current tailored ATO business extract file formats (that include registration data, tax roles and employment data) since the early 2000s. While this arrangement has worked very well, ABS and ATO now have a joint collaboration underway to move the ATO business extract to a more contemporary data format and to remove redundant data items. The new approach will also make it easier for the ATO to provide additional data items to ABS in the future.  |

**Conclusion**

Administrative data have been the key input to the ABS Business Register for many years. ABS will continue to investigate new administrative data sources that might become available to improve the quality of the ABS Business Register and expand its content and uses. While the supply and quality of administrative data cannot always be guaranteed, a sustained relationship with the source data provider can help influence outcomes. This paper has presented some examples of how the ABS / ATO relationship has grown over time and how ABS has managed the use of administrative data in the ABS Business Register.

**Attachment 1: ABSBR as an Integrating Spine**



**Attachment 2: ABS Economic Unit Model**

The ABS Economic Units Model is used to determine the structure of businesses is consistent with Australia’s Corporations Law and with the definition of institutional units outlined in the 2008 System of National Accounts (SNA). The model consists of: the Enterprise Group (EG), one or more Legal Entities (LE), one or more Type of Activity units (TAU), and one or more locations.



Note: the presentation of the ISU in this diagram has not been finalised but is indicative.

The **Enterprise Group (EG)** is an institutional unit covering all the operations within Australia's economic territory of legal entities under common control. Control is defined in Corporations legislation. Majority ownership is not required for control to be exercised.

The **Legal Entity (LE)** statistical unit is defined as a unit covering all the operations in Australia of an entity which possesses some or all of the rights and obligations of individual persons or corporations, or which behaves as such in respect of those matters of concern for economic statistics. Examples of legal entities include companies, partnerships, trusts, sole (business) proprietorships, government departments and statutory authorities. Legal entities are institutional units.

The **Institutional Sector Unit (ISU)** is a derived statistical unit. It includes all legal entities within an Enterprise Group that are classified to the same institutional sector class.

The **Type of Activity Unit (TAU)** is a producing unit comprising one or more legal entities, sub-entities or branches of a legal entity that can report productive and employment activities via a minimum set of data items.

A **Location** is a single, unbroken physical area, occupied by an organisation, at which or from which, the organisation is engaged in productive activity on a relatively permanent basis, or at which the organisation is undertaking capital expenditure with the intention of commencing productive activity on a relatively permanent basis at some time in the future.

**Attachment 3: ABS Statistical Production Activity Modal (SPAM)**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Scope**  | **Design** | **Assemble** | **Acquire** | **Compile**  | **Validate** | **Disseminate** | **Evaluate** |
| 1.1 Understand the problem and confirm statistical needs | 2.1 Design solution  | 3.1 Configure Solution  | 4.1 Prepare data acquisition | 5.1 Link datasets  | 6.1 Validate data outputs  | 7.1 Produce dissemination products and services | 8.1 Gather evaluative inputs  |
| 1.2 Identify options and estimate workload commitment | 2.2 Design Workflow | 3.2 Test and finalise solution  | 4.2 Acquire and load data | 5.2 Classify and code | 6.2 Confidentialise outputs  | 7.2 Release dissemination products and services | 8.2 Conduct evaluation |
|  | 2.3 Identify quality assurance |  | 4.3 Validate integrity of acquired data  | 5.3 Derive variable  | 6.3 Explain outputs  |  |  |
|  | 2.4 Package and endorse design |  |  | 5.4 Validate compilation inputs  | 6.4 Clear outputs |  |  |
|  |  |  |  | 5.5 Generate data outputs |  |  |  |