

## **Meeting of the Group of Experts on Business Registers**

Brussels, 21 – 23 September 2015

*Name of author(s): Luisa Ryan, Jenny Foster and John Machin*

*Organization: Australian Bureau of Statistics*

*Session No. IV*

### ***Transforming the ABS Business Register***

#### **Executive Summary**

Statistical business registers have traditionally been designed and used to support statistical surveys and censuses, via the provision of survey frames, and as a source for broad demography outputs. While these objectives will remain as primary roles of the ABS Statistical Business Register, as the ABS modernises its strategies and systems, its roles are being expanded to include data integration and micro data analysis as an input to policy decision making.

This paper outlines the key directions for the ABS Business Register during organisational transformation, including the vision of an expanded and outwardly facing ABS Business Register that is the integrating spine for firm level data.

Part 1 of this paper provides a brief history of the ABS Business Register and sets out the key directions for the ABS Business Register under transformation, including the benefit of joint activities with administrative data providers. Part 2 provides an outline of the semantic analytical techniques that are being explored in relation to statistical register developments and big data in the ABS.

## **Part One: History and Directions of the ABS Business Register**

### **Introduction**

1. The ABS is embarking on an organisational transformation program that covers the transformation of ABS environment, strategy, governance, infrastructure, people and culture. The outcomes of the program are improved efficiency and timeliness, an expanded range of ABS activity, a reduced risk of errors and a reduction in red tape. The necessity for the transformation program is driven by the need to ensure a sustainable model for the organisation in a tight budget environment whilst also responding to evolving user needs. The aim is to change the focus from collections to solutions. The transformation program commenced on 1 July 2015 and will take five years to complete.
2. At the heart of the transformation is the development of a number of enterprise wide services that will support the loading, processing and dissemination of ABS statistics. These services will be used by all ABS statistical areas, replacing the current stovepipe approach. The ABS's approach to information management is being modernised to include:
  - an Enterprise Data Management Environment (EDME) which will be the single source of truth for all ABS input and output data;
  - a standards based corporate metadata registry where all metadata will be stored;
  - a statisticians' workbench which will be the interface for staff in statistical production areas to interact and manage their operations in the new environment.
3. At the centre of the strategy to transform the economic statistics program is the redevelopment of the ABS Business Register to provide an expanded and more outwardly focussed capability. The economic statistics program will leverage off this development, as the vision is for the ABS Business Register to be the integrating spine for all ABS economic collections.

### **History of the ABS Business Register**

4. In 1999 the ABS Business Register system was introduced following a 5 year development project. It was developed using ODBII, the object oriented database management system developed by Fujitsu.
5. Australia introduced the New Taxation System during the period 2000 to 2002. There were two main parts to the taxation reform. The first covered the removal of a range of wholesale sales taxes and the introduction of a Value Added tax known as the Goods and Services Tax (GST). The second covered the introduction of the Australian Business Register (ABR). Organisations register with the ABR and receive an Australian Business Number (ABN). Only registered organisations are eligible to participate in the GST system and able to claim GST credits.
6. In response to the New Taxation System, the ABS adopted a new units model and a two population approach to the maintenance strategy and the inclusion of non-employing units. The ABR became the major updating source for the vast majority of units. These units with relatively simple structures are known as the Non-Profiled Population. The

other units (the Profiled Population) are significant and/or have complex structures so are profiled by the ABS.

7. A further ABS Business Register redevelopment was undertaken in 2007 when it was advised that the OBDII platform would no longer be supported. This was predominantly a technical project.
8. While the ABS Business Register successfully fulfils its requirements in the production of survey frames and business demography, this comes at a considerable cost. The introduction of any change such as additional data items or derivations to the ABS Business Register takes considerable (up to 9 months) lead time. The underlying design of the ABS Business Register is 21 years old and systems changes have predominantly needed to be low cost (eg. switching off functionality rather than a fundamental revision to the system). Although the Economic Units Model was changed in 2013, the design of the ABS Business Register was not changed.
9. The ABS has been slowly migrating economic statistical collections to the ABS Business Register. The efficiency of this work and the pace of change has been hampered by the need to ‘bolt on’ additional functionality to hold collection specific variables. This work is particularly labour intensive under the current ABS Business Register design.
10. Under the ABS transformation program the ABS Business Register will have a more central role in the production of economic statistics. The ABS Business Register will provide the frame for most economic statistical collections and will play a key role in the integration of economic datasets.

#### **Key directions for the ABS Business Register under transformation**

11. The ABS Business Register redevelopment will be an early ABS transformation project, as it will support transformation across the economics work program. The ABS Business Register redevelopment aims to provide an expanded and more outwardly focused capability. The vision is that the ABS Business Register will:
  - Become the integrating spine for all firm level data, providing a common link to all data about a particular firm to which ABS has access.
  - Provide the infrastructure for unit and small area statistical products, including the Business Longitudinal Database, and by using geospatial data, regional statistics.
  - Become a data source in its own right, as a tool for unit level analysis for ABS analysts, and available to non-ABS analysts for purposes such as unit level analysis and micro-simulation.
12. To improve the current performance of the ABS Business Register, an exploration of an alternate design is being undertaken. The requirements that have emerged are that the ABS Business Register should be simplified and rely on a small core of essential business structural information with a linking variable (eg. ABN) to integrate the other items currently included on the ABS Business Register. Standard Business Register outputs will be produced through linking the core Business Register to items on other key data files. Examples of other files could include economy wide transaction data, survey specific

data, profiling information or provider management information. Links to other registers will also be facilitated (eg. the ABS Address Register which predominantly support social and population statistics). The ABS Business Register, the source data files and the resulting outputs would all be stored on the EDME.

13. This approach would support the production of linking ABS and non-ABS source data files to produce both current (eg. frames) and new statistical series and products. For instance, while there are no plans to change the ABS Economic Units Model, this approach will facilitate both production and institutional sector views of the economy from the underlying legal entity information, as well as industry specific views. In this sense the redeveloped ABS Business Register will become the spine for the economic statistics work program and allow for an expanded use of the ABS Business Register.
14. It is envisaged that a more flexible ABS Business Register will better meet user requirements (including both aggregate and microdata analysis input to policy related questions) with lower cost output production. It will also facilitate a simpler way to migrate economic collections to the ABS Business Register infrastructure, which in turn will open up further potential for data linking opportunities and expanded outputs.
15. Profiling of large and complex businesses will continue to be a critical activity to ensure accurate business structures on the ABS Business Register. As part of the transformation, ABS will also consolidate the profiling of register units with the editing of survey data into a single team, to ensure data coherence across collections and effective stakeholder management outcomes.
16. The ABS Business Register redevelopment will be conducted using an agile approach. The ABS is developing a prototype of the ABS Business Register using semantic technology. The semantic approach is described in Part Two of the paper.

#### **Joint activities with administrative data providers**

17. One of the objectives of ABS transformation is more open and engaged partnerships that are driven by mutual respect and benefit. This is a critical element in the modernisation of a statistical system. Administrative data providers are key relationships that need to be fostered, particularly where they have the potential to replace the direct collection of data and increase opportunities for expanding statistical products.
18. The ABS has been using Australian Business Number (ABN) registration data and tax role data to maintain the ABS Business Register since 2002, when ABS requirements were included in the initial design of the Australian Business Register (ABR). This data will continue as the source of truth for maintaining the quality of the legal entity information on the ABS Business Register. The ABR has been registered as an essential statistical asset for Australia.
19. The ABS has worked with the ABR (through engagement at all levels) to further maximise the quality of the ABN information on the ABR as a whole of government resource. Recent initiatives by the ABR that have been directly supported by the ABS

have included the introduction of point of contact industry coding at the point of registration and the introduction of business location information.

20. Point of contact industry coding was introduced in December 2013. During the registration process the registrant is asked to type in a main activity. The registrant is then presented with a list of industry index entries from which they choose the one that best represents their business. The ABS assisted in the design of the coder index via the outposting of four staff to the ABR, and is currently working with the ABR to further improve the index weightings. This development has significantly reduced the number of units that are excluded from ABS survey frames because they have not been able to be either automatically or manually coded to an industry.
21. While there is significant demand for regional statistics, the ABS Business Register is not currently geospatially enabled. The location statistical unit has not been added to the ABS Business Register. Location data including address, industry and geo-codes has recently been added to the ABR for all new registrations and through a special exercise for some large businesses. The ABS plans to incorporate the ABR location data onto the ABS Business Register, and the ABS and ABR are working together to explore additional options to source and maintain location level data for both registers.
22. There are other administrative data sources that can be used to supplement the ABS Business Register. For example, data on company cancellations and the Ultimate Holding Company from the Australian Securities and Investment Commission (ASIC) has been used to update groups in the profiled population. Potentially the ABS may also supplement the ABR population with some ASIC registration information in order to add investment vehicles which are currently out of scope of the ABR.
23. Where administrative data are built into statistical processes there is risk that administrative data could be ceased or changed in the future. This needs to be managed through high level and operational engagement and building an understanding of the statistical value of the administrative datasets.
24. The proposed design of the ABS Business Register will enable changes to administrative data sources to be incorporated at a relatively low cost and enable the production of multiple views to be maintained.

## **Summary**

25. The ABS organisational transformation program, including the ABS Business Register redevelopment, is an opportunity to review existing practices and introduce approaches and infrastructure that will put the ABS in a strong place to meet current and future Government and community expectations.

## **Part Two: Semantic Techniques**

### **Semantics for statistical registers**

26. ABS is currently exploring the use of semantic techniques, including how they might be applied in the context of statistical registers. Part Two provides an overview of the semantic approach.
27. Using semantic methods to represent the concepts and data that constitute a statistical register of business has several advantages for National Statistical Organisations. The need to address the challenges of increasingly complex organisational structures, volatile legislative environments, and a proliferation of more readily available sources of data regarding activities of economic interest requires a new approach to concept and data management [1].
28. Semantic methods of information management, typically seen in the “Semantic Web”, are based on reusable, robust, and logically rigorous patterns of representation that have been developed and promoted under the auspices of the World-wide Web Consortium (W3C) [2]. These standards – which include the Resource Description Framework (RDF), the Web Ontology Language (OWL), and the SPARQL Protocol and RDF Query Language (SPARQL) – play different roles: RDF connects entities and characteristics through different types of relationships to create a rich information network, OWL describes these relationships in a logically precise and machine-interpretable format, and SPARQL is a language for retrieving facts from the information network.
29. The Semantic Web approach enables complex organisational structures to be meaningfully represented without having to transform them into the arbitrary hierarchies imposed by table-based data storage systems. This avoids the loss of important contextual meaning. An associated ‘additive’ model of information update supports the use of longitudinal queries across the register of business and allows for comparatively efficient changes to both the data and the conceptual models to reflect changes in taxation law or the regulatory environment. Finally the logical rigour of the representation makes the register comprehensible to users – both human and machine.

### **The semantic ‘multi-register’ approach**

30. Underpinning the design of statistical registers with a semantically strong conceptual model allows the purpose-specific meaning of economic statistical units and relationships to be captured more effectively. It also supports better management of sets of units across multiple statistical registers. The ‘multi-register’ approach is intended to localise information about a particular unit type to a specific conceptual class of register (e.g. business, person, and place). This enforces a clear separation of concerns for data management and other custodial responsibilities, and prevents the loss of inconsistency that arises when statistical units are fragmented across disparate register classes. Attachment 1 provides a representation of the semantic approach.

31. Maintaining a logically distinct but semantically related register of place allows a ‘virtually connected’ register of business to benefit from precise and up-to-date information about physical location for its production units. However, this kind of connection does not impose arbitrary constraints on the structure and physical implementation of the register of place. Other register classes – such as registers of person or registers of household – can also be virtually connected without imposing conceptual and data management constraints.
32. Statistical agencies with a ‘multi-register’ system of statistical registers would be able to improve their ability to relate units present in data obtained from diverse sources. The reconciliation of directly collected data, administrative, and more exotic forms of ‘big data’ is supported. In addition, the rich network of relations between a system of registers involving business, place, and person would enable increasingly efficient exploratory analysis to be performed on the sources of data transitively linked by the units in the multi-register system.

### **Current plans and progress**

33. The ABS is currently evaluating a prototype ontology (semantic information model) and infrastructure stack to trial this new approach to a statistical business register. The prototype semantic statistical business register draws on existing administrative data inputs from the Australian Taxation Office (ATO) and augments these with previously unused sources of data in order to obtain a more detailed picture of both institutional and production views of business activity. It is also expected that the methods and tools developed as part of this project will be able to be applied more widely to other registers of statistical entities, and thereby support a multi-entity representation of statistical data.

### **End notes**

[1] Tam and Clarke: *Big Data, Statistical Inference and Official Statistics*; Australian Bureau of Statistics, 2015.

[2] *World-Wide Web Consortium*; <http://www.w3.org/>

## Attachment 1: Semantics for Statistical Registers

