

**UNITED NATIONS  
ECONOMIC COMMISSION FOR EUROPE**

**CONFERENCE OF EUROPEAN STATISTICIANS**

**Seminar on Statistical Data Collection**  
(Geneva, Switzerland, 25-27 September 2013)

**Topic (ii): Centralising data collection**

**ACCELERATING THE CENTRALISATION OF DATA COLLECTION AT  
THE AUSTRALIAN BUREAU OF STATISTICS**

**Working Paper**

Prepared by Jenine Borowik, Adrian Bugg and Bruce Fraser, Australian Bureau of Statistics

**I. Introduction**

1. The Australian Bureau of Statistics (ABS) is Australia's national statistical agency. It was established in 1905 as the Commonwealth Bureau of Census and Statistics and became the Australian Bureau of Statistics in 1975. Each year, the ABS collects data from approximately 400,000 household survey interviews and 350,000 business survey returns. A Census of Population and Housing is conducted every five years, collecting data from a population of approximately 23 million Australians.
2. The ABS has been progressively centralising its data collection activities over the past decade, with an accelerated centralisation effort over the past two years. The drivers for an accelerated centralisation of data collection are discussed in section II. The centralisation effort has allowed the corporate capabilities for data collection to be extended, in particular the capability for electronic data collection. This is discussed in section III. The key strategies employed by the ABS to build its collection capability are discussed in section IV, along with some of the main benefits realised and challenges faced. Current activities and emerging opportunities are discussed in section V.

**II. Key drivers for the centralisation of data collection activity**

3. Prior to the year 2000 it was common for data collection activities to be conducted by individual survey areas. The survey area would design and build their survey forms, mail them out and receive back the completed returns, manage the reminders, follow-up, and other elements of sample management, and perform data entry where it was required. Centralised activity was mainly restricted to managing contracts (e.g. centralised management of a printing contractor to print survey forms, and centralised management of mail despatch) and management of the household interviewer workforce. This decentralised model is no longer feasible for an agency with the scale and diversity of data collection that the ABS has.
4. The key drivers for the centralisation of ABS data collection activity have been:

- Rapid expansion in the availability of data;
  - Increased expectations and requirements from the users of ABS statistics;
  - Community expectations for more convenient and consistent data collection approaches; and
  - Community expectations for productivity improvements from government agencies
5. The rapid growth in information technology over the past two decades has led to a massive increase in the amount of data that is available, and the introduction of new techniques and areas of research such as data mining and big data methods designed to extract information from this wealth of data. A national statistical agency such as the ABS must be well-placed to utilise these sources of data to fulfil its mission of providing a responsive national statistical service, and its legislated responsibility to ensure the maximum possible utilisation, for statistical purposes, of information available to official bodies.
  6. The growth in data availability and in the information provided by the analysis and compilation of these data has in turn led to increased understanding in the user community of what can be done, and has increased the level of expectations on the ABS. There are private sector organizations that are able to harness these non-survey sources of information to produce their own statistics, for example on job vacancies and on retail sales. The users of official statistics require information sooner, and with greater detail (for example, for smaller areas or demographic groups) than previously.
  7. The expectations of survey participants have also changed. Respondents expect to be able to provide data in the manner that is most convenient for them. Electronic modes must themselves meet respondent expectations for usability and functionality, and these expectations increase over time as technology improves and becomes more diverse. In the area of household data collection, householders are becoming increasingly more resistant to interviewers conducting interviews in their home, or to participating in telephone interviews at times that suit the interviewer.
  8. All government agencies, including the ABS, are expected to continually improve their productivity, so that they deliver the same (or improved) outputs with reduced funding. The ABS must meet this challenge by rationalising its IT systems to use common data collection and sample management systems, by using common processes and procedures for data collection, and by realising economies of scale from the centralisation of functions.
  9. The expectations for improved productivity and improved convenience of reporting means that it is not viable for a statistical agency to attempt to meet the increased user expectations by simply increasing the sample size of its surveys or by asking its respondents to provide information earlier. New methods must be developed in order to satisfy the full range of increased expectations. The requirement for more innovative data collection methods is itself a driver for increased centralisation of data collection functions.

### **III. Corporate capabilities for data collection**

10. Centralisation of data collection functions means there is a management structure that has a strong focus on data collection methods and processes. This provides ownership for ongoing innovation and leadership in data collection, and allows investment in research and development and in systems to be better coordinated. This approach allows new corporate capabilities to be developed relatively quickly.

#### **A. Electronic collection modes**

11. Within the ABS, there has been a focus on electronic methods of data collection over the past decade. There is a clear expectation from all stakeholders – including survey participants, users of statistics, and governments - that electronic collection should be offered to survey participants. ABS developments have included offline electronic forms such as Excel spreadsheet forms, XBRL-based Standard Business Reporting, and online survey forms.

12. The earliest modes of electronic reporting offered to ABS survey participants were offline electronic forms. The first offline electronic forms were offered in the late 1980s on floppy disks. Development of offline electronic forms was accelerated with the passing of the Australian Electronic Transactions Act (1999), which required the ABS to accept electronic responses to requests for written information. Excel versions of ABS survey forms were produced and an electronic Secure Deposit Box facility was implemented to allow secure submission of these forms.
13. In 2006 the Australian government established a whole-of-government Standard Business Reporting (SBR) program modelled after the Netherlands approach of XBRL taxonomies that enable direct transmission of business data to government agencies. The ABS was a key contributor to the development of the Australian SBR taxonomy.
14. Online forms were first offered by the ABS for the 2006 Census of Population and Housing, and the first online business survey form was offered in 2009. Online forms were a significant part of the data collection strategy for the 2011 Census of Population and Housing and the 2011 Agricultural Census. In 2012 an online form was deployed for a proportion of the Monthly Population Survey which is the vehicle for the collection of Australian labour force data. In 2013, online forms are also available for 8 key business surveys with an ongoing program rolling out online forms for the remaining business surveys.
15. Emerging modes of electronic data collection for surveys are the use of smartphone apps that can be used to automatically collect data on behalf of a survey participant (e.g. collection of travel or location information) or to provide electronic diaries or log books that allow survey participants to record key information at the time it becomes available (e.g. time use data).
16. The ABS is also investing in systems to collect regular streams of transactional data, such as supermarket scanner data.
17. Electronic reporting offers survey participants a choice in how they report data to the ABS. There are a small number of surveys in the ABS, for example financial industries surveys, where there has been 100% take-up of the electronic mode, as it is a natural choice for businesses in this industry to report through this mode. For many ABS surveys operations there are significant numbers of respondents who prefer to respond by traditional modes and it remains important to maintain corporate capability in these collection modes.

## **B. Traditional collection modes**

18. The ABS currently maintains separate workforces for household and business survey data collection operations. Each workforce is centrally managed, although the workforce is split geographically across a number of ABS sites.
19. The use of a small number of operational sites that are geographically spread and located in the largest cities of Australia provide the ABS with good access to labour markets across Australia. If labour markets tighten in one part of the country there is still opportunity to recruit good staff from other parts of the country.
20. Telephone collection operations for business surveys have been established in three locations, in New South Wales, Queensland and Western Australia. This model has operations on the east coast and the west coast of Australia, with the time zone differences allowing business survey providers to be contacted across a period of 12 hours or longer each working day, without requiring staff to work outside normal office hours.
21. Centralised management of the data collection workforce has also given the ABS the opportunity to negotiate separate conditions of work for its interviewer workforce. Household interviewers have always worked under different conditions to office-based staff due to the need for them to work outside of normal office hours, and to work from their own home (in order to collect data from households in all regions of Australia). Office-based collection staff are currently employed under

the same Enterprise Agreement as other office staff, but there is the option of negotiating a separate Enterprise Agreement with data collection staff in the future if required (e.g. to increase the span of working hours).

### **C. Despatch and data capture**

22. The ABS has a centralised Despatch and Collection Unit that provides the capability for large scale mail or electronic despatch of collection material. This includes surveys forms, approach letters, reminders and brochures.
23. All collection instruments that are returned to the ABS through mail or through email are received by this unit. It maintains the scanning equipment that captures the data reported on hardcopy survey forms, and the systems that extract data from offline electronic collection instruments such as Excel spreadsheet forms.
24. This unit also manages the acquisition of large administrative datasets and is currently looking at options to extend the ABS corporate capability to include central receipt of transactional data.

### **D. Workforce and sample management**

25. As data collection has been progressively centralised within the ABS, a large number of separate workforce and sample management systems have been consolidated into corporate systems that provide this capability.
26. Currently, separate systems are maintained for household and business survey operations. The ABS intends to move onto a single integrated corporate system for workforce and sample management within the next few years.

### **E. Collection instruments and tools**

27. The ABS has centralised data collection instrument design and production capability. It has also centralised the capability for design and production of other respondent materials, such as approach and reminder letters.
28. Collection instrument design is a shared responsibility between the centralised instrument design area and the statistical area responsible for an individual survey. The survey area has the role of specifying the content requirements for the instruments – the data items that are required at the end of the collection process. The centralised instrument design area has the role of designing an instrument that will deliver these data requirements while meeting ABS standards for collection instruments and minimising the reporting burden on respondents through well-designed instruments.
29. Similar skills are used to provide a corporate capability for respondent correspondence. Standard reminder and follow-up strategies are in place, with the capability to modify the standard approaches for individual surveys as required (e.g. to produce additional reminders if response rates are lower than expected).
30. ABS efforts to improve the consistency of respondent communication are illustrated by the development of an ABS Surveys Charter. A survey charter for business survey respondents was first produced in 2003 with a household surveys charter following in 2006. The two separate charters were replaced with a single ABS Surveys Charter in 2010.
31. Written questions and complaints from respondents are received and handled centrally to ensure a consistent approach based on standard templates and responses, and to meet government expectations for how agencies communicate with citizens and business. The area handling respondent correspondence also manages respondent requests for exemptions from ABS surveys, and prepares the cases for the issuing of Notices of Direction (a formal direction that places a legal requirement on a respondent to provide the survey information).

## **F. Collection methods and procedures**

32. The ABS has a centralised Data Collection Methodology section that provides coordinated research into new and improved collection methods. This area also produces corporate standards for collection materials, such as the ABS Forms Design standards.
33. The ABS uses standard Sample and Frame Maintenance Procedures for business surveys to help ensure coherence in its suite of economic statistics. This means standard treatment of non-response and non-contact across surveys, and standard rules for dealing with businesses that cease operations, change ownership or otherwise change their structure or characteristics.
34. Standardisation of processes is also achieved through common training of collection staff and using collection staff across a range of survey operations.

## **IV. Strategies used to build and consolidate corporate capability**

35. There have been strong benefits to the ABS from the centralisation of data collection activity and the development of new corporate capabilities for data collection, but also a range of challenges. A number of strategies have been adopted by the ABS to meet the challenges and maximise the benefits.

## **G. Organizational structure**

36. One of the first challenges was to implement the new corporate structures that incorporated a centralised data collection area. There were three key organizational changes in the past decade that have produced the current centralised structure. These were:
  - The Business Statistics Innovation Program (BSIP)
  - Establishment of a Data Acquisition and Collection Branch
  - Integration of Census of Population and Housing data collection
37. The Business Statistics Innovation Program (BSIP) was a transformation program that impacted on the full range of business survey operations. The aims of the program were to free up resources to give a greater focus to statistical leadership, analysis and coherence of outputs; to improve and rationalise systems and processes; and to improve the management of ABS survey respondents. These aims were achieved by centralising a range of business survey functions. Data collection functions were centralised into the Economic Statistics Data Centre (ESDC). This centre was responsible for the production of business survey frames (including a common frame for most business surveys), collection instrument design and production, workforce and sample management, despatch and data capture, administrative data acquisition, and survey respondent correspondence and follow-up. The ESDC began operation in December 2002.
38. To ensure a seamless transition, it was necessary to identify the resources currently used in each survey area for these collection activities and to transfer these resources to the ESDC to enable it to continue the existing processes for each survey. After the functions had been successfully transitioned to the new environment, the ESDC would then be able to standardise and rationalise operations across survey to make productivity improvements. Standardisation also improved the coherence of the ABS's economic statistics output.
39. The process of identifying and transferring collection processes was referred to as 'podding and docking'. In the first stage, the survey area would continue to undertake their own collection operations, but would do so by establishing a separate 'pod' within their area to undertake these functions. This ensured each survey area had accurately identified the level of resources required to undertake their collection functions. For the second stage of 'docking', the entire pod was moved out of the survey area and into the ESDC. As each survey area pod was docked the ESDC was able to re-

structure the work to form work areas of functional specialties – in forms design, in follow-up, in data capture, and so on.

40. The next major organizational change was to combine business survey and household survey collection operations into the one organizational structure. This grouping is currently called the Data Acquisition and Collection Branch, and was established in July 2007.
41. The key challenge faced by this new branch was to integrate household and business survey operations to realise further benefits for the ABS and for its survey participants, while maintaining or improving the quality of collected data.
42. The centralisation has been further accelerated by moving the Census of Population and Housing branch into the same organizational grouping as the Data Acquisition and Collection Branch. These two branches were placed into the Program Delivery Division of the ABS 2017 Group in February 2012. The ABS 2017 group is leading a major transformation program within the ABS, and the placement of the Census branch and the Data Collection branch within this group recognises the key role these areas will have in implementing the transformation. Borowik et al (2012) discusses the ABS 2017 transformation program in more detail.

## **H. Electronic reporting capability**

43. The economies of scale delivered by centralisation and the productivity gains from standardisation of processes and rationalisation of systems have freed up resources to invest in developing new corporate capabilities. Initial attempts to build new capability were reflected in a Multi-Modal Data Collection project (which produced the ABS's first spreadsheet-based survey forms) and in support for the Australian government's Standard Business Reporting project (XBRL-based reporting). These projects were successful in developing some corporate capability but did not form a complete strategy.
44. The challenge facing the ABS was to articulate a holistic strategy to modernise data collection for the demands of the 21<sup>st</sup> century. A strategy was developed over the course of 2011. The first step was an articulation of a vision as to what data collection in the ABS would be like in five years' time. This vision statement helped to focus attention on what we wanted to achieve, and in getting people's attention and securing their support. A detailed business case for a 4 year transformation program to achieve the vision was then put together over a period of several months and submitted to the senior management of the organization for approval and resourcing. The transformation program was called Acquire@ABS. This branding helped to emphasise that the program was not just about better ways to collect survey data from respondents, but also other methods of acquiring data - from administrative data sources, from transactional data streams, or from direct machine-to-machine transmission of data from survey respondents to the ABS.
45. The Acquire@ABS project gained strong levels of support across the organization, but produced a new challenge. How would the ABS fund the transformation given the other pressing demands for improvement and new capability it was facing in other areas of the organization? The strategy chosen was to make an early investment in online survey forms, and to transition existing paper form and face-to-face household interview modes of collection to the internet mode sooner rather than later. The productivity savings produced by take-up of the internet mode could then be invested in further transformation activities.
46. To drive quick take-up, the initial strategy adopted was a "translate now, transform later" strategy. What this means is that the initial deployment of online survey forms were primarily a translation of the existing paper survey forms into an online form. While there were many ideas for how the ABS could take advantage of the functionality offered by online forms to improve the collection process, these were largely put to one side for the initial deployment of online forms. Respondent convenience was also a strong factor in the initial development, to ensure that the take-up rate of the online forms was high, and that the timeliness and quality of data reported by respondents would be maintained or improved.

## **V. Current activities and future opportunities**

### **I. International collaboration**

47. The transformation of data collection methods and processes to meet the increased expectations of users and respondents is a significant undertaking. The changes in the external environment are not unique to Australia, but part of a global pattern. There is a strong case to be made for national statistics agencies across the world to collaborate and achieve together more than one agency can achieve on its own.
48. Our international collaboration efforts are being coordinated by a Statistical Network Steering Committee with representation from Australia, New Zealand, Sweden, Norway, Canada, Italy and the UK, with links to the strategy and work of the High-Level Group for the Modernisation of Statistical Production and Services that was established by the Conference of European Statisticians in 2010.
49. These collaboration efforts provide the opportunity for accelerated development of corporate capabilities within agencies. The initial focus has been establishing common frameworks and standards, such as the Generic Statistical Business Process Model (GSBPM) (UNECE, 2009), the Generic Statistical Information Model (GSIM) (UNECE, 2012a), and Common Statistical Production Architecture (CSPA) (UNECE, 2013). Sutton (2011) and Hamilton (2010) give details of how GSBPM has been applied within the ABS, and UNECE (2012b) discusses the implementation of GSIM within the ABS.
50. The ABS is working with Statistics Netherlands on the ongoing development of the Blaise software for survey questionnaires. The Blaise software is used for ABS online surveys, as well as for CAPI and CATI survey interviews. An important reason for choosing Blaise was to allow us to accelerate a move to electronic collection rather than developing our own solution - which would have taken a number of years and high expense.

### **J. Metadata-driven systems**

51. In collaboration and consultation with other national statistical agencies and stakeholders, the ABS is using the agreed Generic Statistical Information Model (GSIM) to create DDI metadata objects that can be used to create online survey forms. This work is still exploratory. The aim is to automatically create Blaise code for an online survey form from the metadata objects that describe the survey questions, and the layout and presentation of the final form.
52. The long-term goal is to create a central repository of metadata that will be accessed by metadata-driven collection and processing systems. Creating a stock of metadata objects that can be re-used across surveys and across iterations of repeated surveys will achieve many of the same benefits as centralisation of functions – it will deliver productivity benefits and it will lead to improved coherence and consistency in statistical outputs. It will also face the same challenges in moving from survey-specific methods and processes to more generic and standardised methods and processes.
53. To effectively use metadata-driven systems, each system must easily ‘talk’ and work with the systems that are upstream and downstream of it. Systems should be modular and ‘plug and play’, so that individual modules can be removed and replaced in future without requiring extensive system redesign work. Currently this work is being progressed internationally by the Common Statistical Production Architecture (CSPA) project, and the ABS is contributing to and watching this work with great interest.
54. If the international efforts on CSPA and metadata objects are successful, it may be possible to create an international repository of metadata objects. Any such repository would of course cater for different presentation standards in different countries, and cater for cultural differences, for example in the way sensitive questions are phrased. A common repository of metadata would allow countries to readily provide multi-lingual survey questionnaires.

## **K. Strategic partnerships**

55. A challenge with these ambitious aspirations is finding the resources to deliver them, especially in an environment where many national statistical agencies face significant budgetary pressures. One strategy for meeting this challenge is to actively seek strategic partnerships with other organizations. The international statistical networks are one example of such a strategic partnership.
56. There may be scope to form further strategic partnerships with industry. Metadata-driven systems would not just of benefit to statistical agencies, there would be wider applications that could make these projects attractive to industry partners.
57. Similarly, there is scope to use these ideas for whole-of-government solutions that extend beyond government statistical agencies.

## **VI. Conclusion**

58. It is currently a very exciting time to be involved in statistical data collections. There are many opportunities for future transformation and the rate of change is rapid.
59. Centralisation of data collection functions is an important enabler for the innovation that is required to take advantages of these opportunities. The acceleration of data collection centralisation within the ABS has left the agency well-placed to respond to the changing data collection expectations of the Australian community, researchers and policy-makers, and to contribute to international collaboration efforts that will strongly influence the future data collection environment for national statistical agencies.

## **References**

Borowik, J, Henden, M and Fraser, B (2012) “Riding the Big Data Wave to Streamline Acquiring and Managing Data”, Meeting on the Management of Statistical Information Systems (MSIS), May 2012

Hamilton, Alistair (2010) “Applying the GSBPM within an NSI: Experiences and Examples from Australia”, Work Session on Statistical Metadata (METSIS), March 2010

Sutton, Trevor (2011) “Generic Statistical Business Process Model (GSBPM) and its contribution to modelling business processes”, Tenth Management Seminar for the Heads of National Statistical Offices in Asia and the Pacific, December 2011

UNECE Secretariat (2009) “Generic Statistical Process Model Version 4.0 – April 2009”, Work Session on Statistical Metadata (METSIS), March 2010

UNECE (2012a) “Generic Statistical Information Model (GSIM): Communication Paper for a General Statistical Audience”,  
<http://www1.unece.org/stat/platform/display/metis/GSIM+Communication+Paper>

UNECE (2012b) “Generic Statistical Information Model (GSIM): User Guide (Version 1.0, December 2012”, <http://www1.unece.org/stat/platform/display/metis/GSIM+User+Guide>

UNECE (2013) “Fostering Interoperability in Official Statistics: Common Statistical Production Architecture”,  
<http://www1.unece.org/stat/platform/download/attachments/78677240/CSPA+v0.1.docx?version=1&modificationDate=1366642833960>