Distr. GENERAL

30 October 2012

WP. 30
ENGLISH ONLY

UNITED NATIONS
ECONOMIC COMMISSION FOR EUROPE

CONFERENCE OF EUROPEAN STATISTICIANS

Seminar on New Frontiers for Statistical Data Collection
(Geneva, Switzerland, 31 October-2 November 2012)

Topic (v): Economies of scale from using common tools and methods

MOVING TO COMMON SURVEY TOOLS AND PROCESSES – THE ABS EXPERIENCE

Contributed Paper

Prepared by Jenine Borowik, Adrian Bugg, Brent Ross, Australian Bureau of Statistics, Australia

I. Introduction

1. The Australian Bureau of Statistics (ABS) is positioning itself to take advantage of new technologies and data sources within the developing international information industry. With other National Statistical Institutions (NSIs) the ABS is a member of a significant cohort within this industry, those with responsibility for official national statistics. Through use of standard methods and common tools this cohort has the potential to engage and extend economies of scale within agencies, across NSIs, and as participants in the global information industry.

2. Building on previous organisational change and progressive process improvement the ABS is now working towards fundamental transformation of its information management processes. This opens a way to further efficiencies and improved capabilities which are increasingly difficult to find within existing methods of statistical production.

3. The ABS recognises the value of international collaboration to magnify the benefits of improved information management using the leverage of economies of scale. The work of the Conference of European Statisticians within the UNECE, and particularly the High Level Group for Business Architecture in Statistics (HLG-BAS), has established the strategic vision for this collaboration. The theme of industrialisation and standardisation is central to our focus on economies of scale.

4. The term "economies of scale" will be applied broadly to the benefits of industrialising statistical production, although some of these benefits may relate to process improvement and synergies which are not entirely due to increasing scale. While adopting new methods and technologies has the potential to reduce costs at any scale, the benefits can be enhanced by widespread use within and across NSIs.

5. This paper describes ABS experience with common tools and processes, and elaborates the opportunities and initiatives which will help to realise the strategic vision for the industrialisation of statistical production.

II. A brief history of data collection in the ABS

A. Standardisation and integration

6. The broad sweep of ABS data collection over several decades can be characterised as moving from specialised tools and methods for individual organisational units, towards standardised tools and methods shared by clusters of organisational units. The three main clusters of statistical programs are household surveys, business surveys and the population census, although there are collections which remain substantially self-contained and do not fit well with any of the established clusters.

7. A common thread for household surveys has been the adoption of Computer Assisted Personal Interviewing (CAPI) and a suite of common tools for data collection by field officers. Common elements in survey methodology and centralised interviewer management formed a foundation for process integration. More recent development of tools for metadata authoring, instrument design and data processing has further consolidated household survey processes.

8. Business surveys have typically been based on paper forms for self-completion by data providers, with limited use of electronic forms. Many business surveys were drawn together by integration of economic statistics based on the ABS business register, including standard survey frame maintenance procedures and routine creation of a common frame to improve sample management. There has also been centralisation of forms design, collection methodology and standard question wording which has improved the quality and standardisation of survey instruments. Establishment of an Economic Statistics Data Centre (ESDC) has resulted in better coordination of interactions with data providers based on a shared provider management platform. More broadly, a defined end-to-end framework for statistical production includes a shared input data warehouse with integrated input editing processes.

9. The population census has historically stood apart from other data collection activity with purpose built systems to serve its 5 year cycle, extensive field operations and requirements for high volume processing over a short period. The census is primarily based on paper forms delivered by field officers, although about 30% of forms submitted in the 2011 census used the optional on-line electronic form (eCensus).

10. Use of administrative datasets has steadily increased, taking advantage of tax data established with the introduction of the Goods and Services Tax, and a wide range of other data sources. A single organisational unit has been established to centralise the handling of administrative data sets, although the diversity of data sources makes it difficult to standardise processes. Ongoing efforts to improve these processes are framed in terms of managing data sources rather than collecting individual data sets.

11. Overall there have been pay-offs in better coherence of statistical outputs, reduced duplication of processes, improved quality assurance, and savings in design work. Use of established software products (such as SAS, Oracle and Blaise) integrated with ABS infrastructure represents some progress in using common tools and platforms. Operational efficiencies have been realised in the adoption of computer based data collection technologies, including CAPI, spreadsheet based electronic forms for off-line completion, and more recently on-line web data collection. Each of these resulted in incremental cost reductions.

12. However, the structural separation of household surveys, business surveys and population census has remained in place. Processes are often similar in principle across collections, although historical factors result in substantially different implementations. Variations in processes can persist even where common tools are used on a shared application platform.

13. The ABS has historically developed discrete system components, whether for individual collections or integrated clusters. These systems have been purpose built to requirements at a point in time and with defined relationships to other ABS infrastructure. These have pushed the

boundaries of what was possible at the time and served the ABS well. However, the inflexibility and high cost of change has resulted in some monolithic developments being regarded as legacy applications from the time they were fully operational. They were not built to support the pace of change which is now a constant.

14. The development of new tools and methods has typically been slow, costly and high risk, and generally systems have not been able to keep pace with statistical ambitions. Large scale development projects have enabled occasional leaps forward, although this generational change then establishes a new status quo with limited flexibility.

15. Previous initiatives in metadata development and reuse have attempted to standardise tools and processes, with limited success. This has resulted in a patchwork of data and metadata repositories, and associated process control applications, connected by occasional threads of standardisation. The existing suite of infrastructure serves as an effective statistical assembly line, but does not offer a clear path of convergence towards a single framework for ABS data collection.

B. Organisational change, governance and culture

16. Organisationally the ABS model for data collection evolved as a loose federation of cottage industries. Individual responsibility for statistical outputs and objectives required the autonomy to develop specialised systems and processes tailored to local requirements. Corporate governance arrangements aided the consolidation of integrated clusters of collections, but could not go further without structural change.

17. The Business Statistics Innovation Program (BSIP) in the early 2000s implemented the ESDC to deliver shared data collection services for business surveys, a significant step towards standardisations of processes. Some of these services have subsequently been extended to household surveys.

18. Since the commencement of the Information Management Transformation Program (IMTP) in 2010 the ABS has explored the operational implications of fundamental change. Several pathfinder projects have taken forward proof-of-concept work for specific statistical projects using components of infrastructure to support shared metadata and processes. In 2012 a new round of organisational change has established the ABS 2017 program, combining programs for infrastructure development and data collection operations, and including the population census.

19. While cultural change can be slow and difficult, it does tend to follow organisational change as projects reconsider their corporate responsibilities and identity. Following BSIP, the new data collection units in the ESDC were more conscious of their role as collection specialists contributing to the public face of the ABS. Individual statistical programs recognised this expertise and supported the process re-engineering necessary for progressive migration to new standardised processes. Similarly, organisational change associated with ABS 2017 is promoting new directions in corporate standardisation of tools and processes, and encouraging engagement of staff with a new paradigm for statistical production.

III. Future directions in data collection

A. Recognising challenges and opportunities

20. The ABS is facing challenges shared to some extent by all NSIs, including budget constraints, pressure for relevance, expectations of greater responsiveness, declining survey response rates, and direct competition from other organisations and businesses involved in the information industry. Also, there has been a fundamental change in the information environment with an exponential increase in the volume of data and the number of potential data sources, which has come to be known as “big data”.

21. Previous initiatives in standardisation and common tools have helped to progress integration and cost reduction by forming clusters within the portfolio of statistical programs. However, they tend to reach a limit where cost factors or lack of flexibility prevent further consolidation. Some data collection activity cannot be aligned with standard infrastructure without compromising short term statistical objectives. This points to the need for a new paradigm which allows the necessary rigidity of standardisation to coexist with flexibility in data collection processes and statistical outputs.

22. Economies of scale and the broader benefits of industrialisation should not be interpreted simplistically as "bigger is better". Diseconomies of scale also exist in some circumstances, so there is a need to consider the notion of an appropriate or optimum scale. Increasing scale and other features of industrialisation need to be adopted intelligently where they can add most value.

23. Related to economies of scale we might also recognise "economies of scope", for example where there are synergies in accessing a wider range of diverse data sources to achieve strategic objectives. This would require flexible methods to access data in different environments, to accept data in a wide variety of formats, and to validate the content and provenance of datasets. The challenge is to manage this data diversity within a common conceptual framework using a common toolset and deliver it to shared processing infrastructure.

24. A high capital assembly line model optimised for a standard manufactured product can demonstrate economies of scale, but may also constrain flexibility. This is not necessarily a good analogy for industrialised statistical production. Information industries depend heavily on ubiquitous standards, such as the protocols which support the internet, digital communications, optical disks and image processing formats. The industrialisation of statistical production should take advantage of scale and widespread standardisation only where it facilitates flexibility, innovation and connectivity.

25. A more appropriate analogy would be the use of "apps" on mobile devices, or traditional computer applications on established operating systems. These examples have shown that strong standards for connectivity to ubiquitous infrastructure can support enormous creativity and diversity.

26. In recent years there has been a growing awareness that metadata driven processes could become a reality as service oriented architectures have matured. The ABS now aspires to build tools and processes which are inherently flexible, made possible by a clearly defined backbone of standards and common infrastructure. This can be described as the "plug and play" approach where modular components can be added without unintended impacts on other components. It represents a move away from generational systems deployment based on fixed requirements, towards a dynamic environment where the primary requirement is to support constant change.

B. A change of paradigm

27. The current paradigm is reflected in the history of ABS data collection outlined above. It is characterised by purpose built infrastructure serving clusters of related collections within the organisation. The focus is on direct collection of scarce data rather than engaging with "big data" to take advantage of what is available. The inherent limitations of this paradigm cannot meet the future needs of the ABS in the global information industry.

28. The new paradigm involves design of data collection solutions from a toolkit of standard methods, processes and system components. The Generic Statistical Business Process Model (GSBPM) and Generic Statistical Information Model (GSIM) will provide the conceptual frameworks, and the metadata standards established by the Data Documentation Initiative (DDI) will be used as a protocol to ensure connectivity between processes, services and other system components.

29. The focus of development for statistical processes will be metadata authoring and associated process execution capability. Metadata will include process definitions, variables, and pointers to all required data sources and target repositories. The ABS is developing a Metadata Repository and Registry (MRR) and an environment for Statistical Workflow Management (SWM) which will provide the controlled flexibility required for processes which are both robust and adaptable. This strong backbone of standard technologies, methods and processes will be a platform for modular components which will then allow the high degree of flexibility required.

30. Adapting to the "big data" environment will require a new generation of tools and methods which take full advantage of administrative and transactional data sources. Nominated datasets will be complemented by directly collected data in the enterprise data warehouse where standard analytical tools and methods can be applied on a common platform.

31. Another aspect of this paradigm change is the opportunity to improve testing and quality assurance. Standardisation of process metadata (or paradata) and the associated economies of scale present an opportunity to establish more realistic simulation environments and effective quality checkpoints. If testing can be done more quickly and reliably this will support greater flexibility in design and more robust production processes.

32. Organisational culture is also a critical factor in the new paradigm. Infrastructure can present an opportunity for change, but it will be the myriad of management decisions across the organisation which will ultimately determine how well the outcomes are realised. It will be important to promote the strategic directions widely and effectively within the organisation, and support this with an effective governance model. Success will ultimately depend on the engagement and commitment of individuals to apply the new paradigm to the details of design, development and operations.

C. Towards ABS 2017

33. The ABS 2017 transformation program will industrialise ABS statistical production by adopting this new paradigm.

34. Building the infrastructure backbone for standardised metadata driven processes will facilitate widespread automation and re-use of system components. New systems will conform to design principles and governance to ensure the target outcomes are realised, and will include design elements and system components established by international collaboration. Existing systems may be linked to new infrastructure through interim service components, but the goal will be progressive transformation of legacy systems to adopt the new paradigm. Previous support for clusters of collections will be incorporated into a collective GSBPM view of statistical production.

35. Electronic data collection is a priority because of its potential to reduce operational costs at any scale through process improvement and automation. Some existing processes will be entirely avoided, particularly where manual intervention was previously necessary to support diverse business processes across multiple platforms.

36. ABS 2017 will integrate data collection infrastructure for the population census, household surveys and business surveys. Investment in the population census will contribute to enduring capability rather than delivering single use solutions. Implementation of a comprehensive address register will allow household surveys to move away from clustered samples of nominated locations. It will also enable a shift in the business model for the population census incorporating web data collection, reducing the need for field staff and aligning with other data collection using self-completion survey instruments.

37. Instrument design and production will be based on metadata definitions, automated processes, standard protocols and a single hosting environment for web data collection. Off-line self-completion eforms and paper forms will continue in small numbers, with data capture and

transformation capability linked to the centralised instrument metadata. Administrative and transactional datasets will be defined and managed within the same framework as direct data collection instruments. Other structured provider contact, such as routine notifications, reminders and standard responses, can also be defined as "instruments" with metadata and process definitions.

38. Similarities between instrument design and interface design will allow the same techniques and tools to be used for internal interface components as well as external data collection vehicles. This has the potential to streamline the development of operational interfaces for internal use by further standardising the systems environment and extending the economies of scale.

39. There will be a single platform for managing all interactions with data providers, whether they are households, individuals, businesses or other entities. Interactions with providers of administrative and transactional data will be managed on the same platform. This platform may also be suitable for other ABS relationship management requirements associated with output dissemination, community interactions or strategic engagement.

40. Provider approach materials will be consolidated to minimise the reporting burden, complexity and duplication where possible. Data providers will have access to online self-help tools through a provider portal which is integrated with the ABS website. Direct contact with data providers will be supported by a comprehensive view of provider interactions and a suite of functions for service operators to assist interactively.

41. Organisational arrangements will allow for integration of field interviewers (currently supporting household surveys) and office interviewers (currently supporting business surveys) as a combined workforce which can support both household and business surveys. The interviewer workforce will be supported by an online toolset, including access to survey instruments enhanced for personal interviewing where necessary. Standardised online services will support interactive workload assignment and other workforce administration.

42. Collection management will be automated where possible and include a standardised dashboard to ensure a high level of visibility of collection processes. Quality gates and standard reports will be supplemented by a flexible interface to all available process metrics. Web data receivals will be passed through to the enterprise data warehouse in real time. The processes for delivery of administrative and transactional datasets will be aligned with direct collection processes as far as possible, including process management and loading to the enterprise data warehouse.

43. Process control metadata (paradata), will be designed into data collection processes from the early stages. This will improve capability to identify operational adjustments which can maximise response rates and data quality in the course of data collection.

44. The ABS 2017 program aspires to the industrialisation of statistical production. Economies of scale will be realised as standard data collection tools and methods are applied consistently across all ABS data collection activity. International collaboration can magnify these economies of scale, further reducing costs and increasing capability across the international community of NSIs.

D. Avenues to economies of scale

45. As the ABS 2017 program proceeds the ABS expects to realise economies of scale by:

(a) reducing duplication of effort in development and testing;
(b) reducing operational costs as diverse data collection processes are standardised and automated;

(c) reducing the cost of documentation, training and skills development previously duplicated across different technologies and applications;
(d) increasing staff mobility and capability to quickly respond to staff movements.

46. There is potential for international collaboration to extend these economies of scale where there is sharing of development effort for specific application components and processes, including the costs related to testing, documentation and training. The high level standardisation of statistical processes (GSBPM) and information objects (GSIM) will facilitate sharing of specific tools and methods.

47. Standards for information exchange, particularly DDI and SDMX, will assist modularisation and reuse within organisations and across NSIs. These standards also open the way to improvements in information flows and analytical capability with the wider information industry, and with data providers and statistical users across communities and economies.

III. Risks and issues

48. There are significant issues and risks inherent in the new paradigm. It is not an easy matter to transform statistical processes which are tightly coupled to exiting technologies. Residual vulnerabilities and dependencies will need to be identified and managed.

49. There will be a need to establish effective standards and governance arrangements to ensure that we do not simply continue to build specialised incompatible processes using new infrastructure. The infrastructure will make processes more visible and easier to change, but it needs to be supported by effective governance and a corporate culture which embraces the new paradigm.

50. There is a trade-off between standardisation and flexibility. If standardisation is not applied intelligently within the business architecture it can be experienced as reduced flexibility. Similarly, short term flexibility can erode standardisation and the ability to re-use processes.

51. There will be workforce impacts when increasing scale and reducing costs. A positive impact is that standardisation of methods and processes will support skills transfer, as well as effective investment in high quality documentation and training in key technologies. A potentially negative impact is that these efficiencies may also reduce the need for some jobs and skill sets. Often the change will be gradual enough to manage through routine patterns of recruitment, retirement, job mobility and skills development. However, in some areas of activity the changes may be rapid and disruptive, leading to potential for redundancy and displacement of staff in the short term. In the long run the industrialisation of statistics will free resources and skilled staff for work that is analytical, innovative, and design oriented.

52. There are risks in waiting for finalisation of conceptual frameworks and their implementation in components of systems infrastructure which are highly interdependent. There is a tendency for everything to wait for everything which may result in dependency gridlock. This will be a test of the new paradigm to manage constant incremental change by balancing operational realities with the objectives of the transformative change program. A large number of small steps will need a degree of independence to progressively converge on the long term goal without waiting for parallel developments.

IV. Conclusion

53. The ABS supports an ambitious vision for data collection where the use of standards, tools and methods drives down costs while allowing flexibility and promoting innovation. The realisation of this vision requires a paradigm shift in the design of data collection processes, as well as internal cultural change to take advantage of opportunities at a local level. International collaboration will magnify the economies of scale and associated benefits of industrialised statistical production.