Methodology Architecture – A Roadmap for New Methodological Directions in the Australian Bureau of Statistics

Siu-Ming Tam, Canberra, AUSTRALIA

Corresponding author: Siu-Ming Tam, email: Siu-Ming.Tam@abs.gov.au

Abstract

To maintain its strong brand as a central statistical agency into the 21st century, the Australian Bureau of Statistics (ABS) is transforming the way it acquires, collates, uses, reuses and disseminates statistical information. To support this transformation, innovative, industrialised and contemporised statistical methods and tools will be required. Methodology architecture provides a roadmap for systematically assessing and developing these 21st century methods and tools covering the full spectrum of the statistical production cycle.

1 Introduction

Tam and Gross (2013) argued that national statistical organisations (NSOs) are at a crossroad. Faced with significant statistical challenges that have not previously been seen before (Pink et al., 2010), they can choose to transform their business model to "survive and thrive" or, ignoring these challenges, continue to operate the way they have been successfully operating in the past and face the increasing risks of becoming irrelevant in the future.

In order to sustain the mission of NSOs to provide the statistical evidence for informed decision making, the choice is abundantly clear. Many NSOs have started thinking about, or embarking on a course with a view to, offering new statistical products and services, including creative ways to harness new data sources. They have also been actively looking for more efficient and effective ways in the production of statistics. Whilst striving to improve the timeliness and reduce the cost of official statistics production, NSOs will need to make certain that the same or similar level of quality of the products and services is maintained.

Undertaking business transformation of the statistics production process and international collaboration in the development of concepts, classifications, tools and products are increasingly seen as promising strategies to address these challenges.

2 Vision of the Australian Bureau of Statistics (ABS) for business transformation

The ABS transformation vision comprises a products vision and a process vision.

The ABS' products vision is a 21st century NSO which has the ability and agility to combine ABS data with other strategic sources of data, including administrative data, transactional data, Big Data, and "organic" data (Groves, 2011), to produce more timely and relevant official statistics (Pink et al. (2010)). As well, the process vision sees the ABS fully embracing industrialisation and standardisation in its business model in the production of statistics.

3 Methodology Architecture (MA)

A business transformation program is fundamentally a change program. To be successful in any change programs, it is important to articulate and communicate the vision and create a roadmap for change, including the creation of short term wins (Kotter, 1995). Outside the statistical world, it is customary to use an enterprise architecture (EA) (Zachman,1987) as a tool to describe the interplay between, and to steer, these essential elements of the change program.

Being part of the EA, an MA is a transformation plan for methodology. Underpinned by the NSO transformational needs, the MA lays out the vision for methodology. The MA therefore comprises an envisaged future inventory of statistical methods and tools, covering the full spectrum of the Generic Statistical Business Process Model (GSBPM) (Vale, 2009). Contrasting with the current inventory, the MA will also provide a transition plan for migrating from the current methodological state to the future state.

The MA vision for the ABS is "to provide a set of methods that underpins the products and process visions of the ABS transformation program." Fundamental to achieving

this vision are five key elements in developing the future inventory of statistical methods and tools: innovate, industralise, contemporise, build capability, and build support.

In this paper, we will restrict our discussion to the first three of the five key elements.

4 Innovate

New methods and tools are required to address the new and emerging challenges from the ABS products vision. An example of this is in the linking of data to create fused unit record files or new statistics, and their processing, analysis and dissemination at the micro and macro data level. The goals are to improve relevance, production cycle time, organisational capability and meet legislative requirements to protect the confidentiality of individual personal or business information. The ABS has been working with academics and collaborating with other NSOs in the development of the requisite tools and methods. In 2012, the ABS signed a Memorandum of Understanding with the Australian technology universities to part fund two professorial positions to undertake research and supervise post-graduate students in data confidentiality and data integration. Methods where ABS has made advances are using the EM algorithm for data linking but methods and tools will be needed for statistical disclosure control and quality assurance and assessment for integrated data. Some of the tools developed are Table Builder and Analysis Service under the ABS Remote Execution Environment for Microdata. Other tools such as tau-Argus, mu-Argus, SDCMicro, and other industralised tools yet to be developed will be evaluated.

5 Industralise

In the post transformed world, we see significant improvement in the production cycle time through re-use of data available within and without the ABS, as well as through re-use of statistical processes, statistical methods and tools. For this to occur, the post transformation methods and tools need to comply with the following architectural elements:

- plug and play – statistical methods and tools are to be provided in the form of "modules" (with input and output capabilities) which can readily be assembled to form reliable statistical processing systems;
- standardisation and corporatisation – a standard but finite set of statistical methods and tools will be provided to support the business needs of the ABS. These will become the corporate methods and tools. Non-standard methods or tools may only be used subject to compelling business cases;
- connectible and metadata driven – tools will need to be content agnostic, driven entirely by input data and metadata, and readily flow statistical data and metadata into other tools without manual intervention; and
- user-driven – the selected tools can be used reliably by subject matter statisticians with little or no intervention by methodologists. This requires a well crafted set of diagnostics which are easy to interpret and alert subject matter statisticians to issues which need expert assessment.

As an example, SEASABS (SEASonal analysis, ABS standards) is an ABS-developed package with a core processing system based on X11 and X12ARIMA. SEASABS keeps records of the previous analysis of a series so it can compare diagnostics over time and "knows" whether they are consistent with the parameters set at the last analysis. It identifies and corrects trend and seasonable breaks as well as extreme outliers, corrects for trading days if necessary, and allows for moving holiday corrections. Whilst this system is content agnostic, is a metadata driven corporate system and used by subject matter statisticians, it currently cannot be easily plugged and played, and data and metadata cannot flow in and out of SEASABS in a standardised fashion.

6 Contemporise

The survey methods used by the ABS are predominately design based, or model-assisted based. Stratified random sampling and the Generalised Regression (GREG) estimator is the principal sampling strategy for ABS surveys. As well, the Jacknife and

Bootstrap are the principal technique in estimating the design based variance of the estimator. In time series analysis, predominately filter-based (X11, X12ARIMA) methods are used. Whilst model-based methods are used in small domain estimation and in analytical studies in the ABS, model-based methods (e.g. Chambers and Clark (2012); Harvey (1990)) are seldom used in the collection, processing and analysis phases of the GSBPM, primarily because of concerns about quality of the estimates with model mis-specification, and the effort required to develop, test and evaluate models. In survey sampling, the multinomial model/Dirichlet prior for Bayesian inference for survey sampling, as outlined in Aitkin (2008), has the potential as a future sampling strategy. Multinomial models are realistic models in the sampling context, and the use of zero weights in the Dirichlet prior is akin to model-free assumption in Bayesian bootstrapping for the variance of the survey estimators.

7 Challenges

To be successful, the MA must address all the 5 key elements satisfactorily. Each of the 5 elements poses its own challenges, and the ABS will work with academics, experts in the field, other NSOs as well as ABS methodologists to address them.

In our view the most challenging of all the 5 elements is in the building of support, not only amongst the business people who will be consumers of the chosen methods and tools, but importantly amongst our methodologists. We believe that there are three strategies to build this support: (1) regular and consistent messages from senior management on why the changes are necessary; (2) actively seeking the input from methodologists in developing the MA, and (3) skilling our methodologists in new frameworks.

8 Conclusion

The MA is a plan to transform the ABS statistical methods and tools to support the products and process vision of the ABS. The main output of an MA is an inventory of

methods and tools to support the transformed ABS business. As such, the inventory must cover the full statistical production cycle, as articulated by the GSBPM. When developing this inventory, it is important to systematically go through the 5 key elements of innovation, industrialisation, contemporisation, capability and support building. There are significant challenges in building and implementing an MA.

References

Aitkin, M. (2008) Applications of the Bayesian Boostrap in Finite Population Inference. *Journal of Official Statistics*. **24**, Number 1, p21-51.

Chambers, R. and Clark, R. (2012). An introduction to model-based survey sampling with application. Oxford Statistical Science Series

Groves, R. (2011). Three eras of survey research. Public Opinion Quarterly, Vol 75, No. 5, p861 – 871.

Harvey, A (1990). Forecasting, Structural Time Series Models and the Kalman Filter. Cambridge University Press.

Kotter, J. (1995) Leading change: Why transformation efforts fail. Harvard Business Review, Vol. 73, Issue 32, p59-67.

Pink, B., Borowik J. and Lee G. (2010). The case for an international statistical innovation program –Transforming national and international statistics systems. *Statistical Journal of the International Association for Official Statistics*. **26**, p125-133.

Tam, S, and Gross, B. (2013). Transforming the business of official statistics – vision, strategies and challenges. *Journal of Official Statistics*. **29**, pxx-xx.

Vale, S. (2009). Generic Statistics Business Process Model. UNECE

Zachman, J. (1987) A framework for information system architecture. IBM System Journal, Vol 26, No. 3 p276-292.