VERSITA

Journal of Official Statistics, Vol. 29, No. 1, 2013, pp. 203–211, DOI: 10.2478/jos-2013-0015

# Discussion

*Siu-Ming Tam*[1] *and Bill Gross*[1]

## 1. Introduction

We congratulate the Editors for the production of this special JOS issue on "Systems and Architecture for High-Quality Statistics Production", and the selection of articles for inclusion in this volume. Statistics organisations, and national statistical organisations (NSOs) in particular, are at a crossroad. Faced with significant statistical challenges that have not been seen before (Pink et al. 2009), they can choose to transform their business model to "survive and thrive" or, ignoring these challenges, to continue to operate the way they have been successfully operating in the past and face the increasing risk 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 are embarking on a course with a view to offering new statistical products and services. They are finding creative ways to harness new data sources (a vision for products) and are actively looking for more efficient and effective ways in the production of statistics (a vision for processes). They also aim to maintain the same or a similar level of quality of the products.

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. Recognising the need to work collaboratively with NSOs in the world, the United Nations Economic Commission for Europe (UNECE) and, in late 2012, the United Nations Economic Commission for Asia and the Pacific (UNECAP) have set up High Level Groups comprising Chief Statisticians and Directors General from national and international statistical offices to work with member countries to develop an international vision and strategy for business transformation (HLG BAS 2011).

In addition, currently led by the Australian Bureau of Statistics (ABS), the NSOs from Canada, Italy, New Zealand, Norway, Sweden, and UK are members of a Statistical Network established to develop and experiment with models for working together to co-design and co-develop statistical frameworks, concepts, and tools. One of the more notable

[1] Methodology and Data Management Division, Australian Bureau of Statistics, Locked Bag 10, Belconnen, ACT 2617, Australia. Emails: Siu-Ming.Tam@abs.gov.au and Bill.Gross@abs.gov.au

**Acknowledgments:** The authors wish to thank Chris Duffy, Alistair Hamilton, Michael Glasson, Brian Studman, and the reviewers for their comments on earlier drafts of this article. The responsibility for the content of this article remains, however, solely with the authors. Disclaimer: the views expressed by the authors do not necessarily reflect those of the Australian Bureau of Statistics.

achievements of the Statistical Network is the development of the Generic Statistical Information Model (GSIM) (Hamilton and Tam 2012a). This is a metadata model for describing the information objects used in the statistical production process, as described by the Generic Statistical Business Production Model (GSBPM) (Vale 2009).

We will restrict our discussion to the Group of Four (GoF) articles by Nealon and Gleaton, Seyb et al., Struijs et al., and Thalji et al. The reason for our choice is principally a reflection of our interests in this challenging and emerging area in the work of NSOs. We will refer to these articles as the NASS (National Agricultural Statistics Service), SNZ (Statistics New Zealand), SN (Statistics Netherlands), and RTI (RTI International) articles respectively. Clearly the four articles cover a lot of ground in describing their journey to transform their business, and it will not be possible to discuss all aspects. Instead we chose to discuss the strategic issues in the business transformation, how they align with the emerging international vision for transformation, and conclude with comments on what we believe are significant challenges facing the NSOs as a statistical community.

## 2. The Products Vision – the GoF Articles

There are more similarities in the visions articulated in these articles than differences. The SNZ and SN articles describe efficiency, responsiveness, and expanding data sources as major aims for their business transformation. On the other hand, the NASS article articulates efficiency as the main aim, and achieves this through economy of scale by centralising their collection and network operations in a small number of offices in the NASS. The RTI article outlines improvement in accuracy and timeliness of their statistical products and services as their main aim, and seeks to achieve it by developing and implementing data collection and management systems to manage the production life cycle of statistics.

Whilst we generally agree with the articulated vision of transformation in the GoF articles, we feel that, as the articles are predominately focussed on the process vision, there is a gap in information about their statistical offerings. We argue that the products vision is the *raison d'etre* for the process vision and, without a vision for the product, the process vision can degenerate into a mere pursuit for efficiency. Whilst efficiency is important, we believe that innovation and creativity in harnessing new data sources to produce richer, more relevant statistics more responsively are the main challenges facing official statisticians now. The four emerging IT mega trends of Cloud, Social, Big Data, and Mobile provide major opportunities.

We therefore believe that readers of the GoF articles would benefit from learning their plans for the creation of new or better statistical offerings through the strategic use of the ever-increasing alternative sources of data: of administrative data available from governments, transactional data from businesses, and Big Data (or "organic data", Groves 2011) from social media, as well as through the opportunities available from the IT mega trends. That said, we felt that SNZ and SN are better positioned than NASS and RTI to do so, given the process transformation being undertaken in these organisations, a point we will return to later in this discussion. An excellent description of the characteristics of Big Data in terms of volume, velocity, variety, and degree of structure can be found in Daas (2012).

In articulating their products vision, the UNECE High Level Group stated:

> "The raw materials, the data we use to create our products, will need to be found in the data that are available throughout society. . .The active pursuit of data and the creation of products that give insight from an impartial perspective, our unique selling point, will be our new mission. It will undoubtedly mean that our organisations will have to leave their comfort zone and will have to question the work that seems so normal at present . . ." (HLG-BAS 2011)

There is a paucity of novel and innovative combinations of these different data sources, both existing and emerging ones, in the creation of new statistics. In the case of Big Data, Glasson (2012, internal ABS report) provides an interesting insight into why this may be the case by noting the differences in the supply chain and value proposition of these data sets between statistical agencies and nongovernmental organisations. There are some examples of a Big Data application for official statistics. Hal Varian, Chief Economist of Google, demonstrated in his lecture to ABS staff in 2011 how using Google search terms and statistical modelling can be used to "nowcast" US unemployment. The methods used in nowcasting are available in the public domain (Varian and Choi 2009).

Kimura et al. (2012) also describe a collaborative project between the National Data Centre of Japan and NTT DOCOMO, a Japanese major network operator, to use the latter's operational data about location and attribute to provide the capacity to estimate the daytime population on a real time basis.

Other examples of uses and issues with Big Data can be found in the Volume of Significance published in August 2012 by the Royal Statistical Society of the UK and in Daas (2012).

We urge statistical offices to release their plans for the creation of such new statistical products and services in the public domain with a view to sharing information and knowledge with their counterparts. In the ABS, our plans include the development and implementation of interdepartmental protocols for the creation of linked datasets combining ABS data with administrative data, subject to appropriate approvals, to provide the information to address complex issues facing contemporary Australia, and the development of facilities for accessing such datasets for secondary data analysis without compromising the confidentiality of respondents' data (Tam et al. 2009; Tam 2011).

## 3. The Process Vision

There is a wealth of information provided in the GoF articles on the transformation of statistical processes, ranging from organisation-wide transformational change to structural/systems change impacting on the collection and processing of information. We are particularly interested to compare their process vision with that of an emerging international vision for statistical transformation.

The process vision as articulated by UNECE High Level Group stated:

> "The production of statistics should be based on common and standardised processing. . .We view this as the industrialisation and standardisation of statistics production. . .The production of official statistical information should have its own industrial standards . . ."

Summarising these aspirations, the ABS recently used the slogan "Industrialisation, Modernisation, Transformation, and Future-Proofing" (IMTP) to galvanise support for its business transformation program across the ABS, and embarked on a strategy based on the following philosophies:

- plug and play – to build modular components (e.g., for imputation or weighting) which are easy to assemble into a reliable processing system, with an aim to supporting a future "assemble-to-order" statistical business model.
- standardisation and corporatisation – to standardise and corporatise concepts, classifications, methods, tools, and processes to reduce cost and lead time in the creation of statistical production processes and to improve timeliness and relevance of ABS statistical outputs. The effort to standardise and corporatise ABS metadata is huge and should not be underestimated.
- metadata-driven processes – to build content-agnostic processes whose metadata-rich outputs are driven entirely by input metadata.
- connectivity – to ensure statistical data and metadata can flow freely from one modular component of a statistical system to another with little, if any, manual intervention.
- capability building – to equip staff with the skills in anticipation of the move from low value-added to high value-added statistical activities; skills of which will be needed following the completion of the transformation, and
- international cooperation – to share or collaboratively build common standards, methods, tools and processes to decrease the ABS costs of their development (if going alone), and to work productively with the Statistical Network to achieve these aims.

Refer to Hamilton and Tam (2012b) for a more detailed exposition of these ideas.

### *3.1. GoF Transformation Strategies*

Whilst all four statistical organisations plan to standardise tools, SNZ and SN aspire to build "plug and play" capability in their processes and tools with an aim of assembling components to form end-to-end statistical process systems quickly and with little or no IT involvement. In contrast, NASS and RTI have more modest goals and aim to "consolidate and generalize survey applications" (NASS) and for configuration of a particular collection to be "done by a programmer with less experience" (RTI). Building "plug and play" capability is a substantial capital investment and for statistical agencies with a large number of collections, there is ample opportunity to reap the budgetary benefits of such investment. Equally important, particularly for centralised statistical agencies, is that this plug and play capability enables new collections – whether direct, from administrative or transactional data or repurposing of existing holdings – to be developed and made operational quickly so that emerging user needs can be met in the time frames they expect.

While SN's (IT) treatment of metadata is unclear, the other organisations have information about data, with SNZ storing "metadata close to the data" but "in a way that facilitates data integration" and the others having a central store. SN and RTI extend the traditional thinking of metadata covering classification and data items to statistical

processes, but NASS and SNZ handle information about processes and workflows separately. In addition, RTI has metadata about data, processes and workflows in a central system, and this has a number of advantages. It allows a process to update metadata about its outputs automatically and provides a comprehensive store from which to generate metadata for dissemination. It is easily searchable, thereby facilitating repurposing of data, reuse of workflows and use of standard processes. A central system should encourage metadata to be created in a standardised fashion.

We believe that metadata should be extended to cover paradata (Scheuren 2005) and metadata about statistical processes, as they are important to support an assemble-to-order business model as described above. Whilst we are not necessarily wedded to a centralised metadata store, federated data stores need to ensure that the metadata in all the stores are searchable and accessible by all statistical production staff. Furthermore, good metadata registration is needed for maximisation of metadata re-use.

Notably, SNZ makes extensive use of tools developed by other statistical agencies, such as BANFF (Statistics Canada) and Blaise (from SN), which is also used by NASS. The uptake in the use of tools developed by other NSOs is less than one would have liked to see. Part of the reason is that the tools that could be used in an organisation are constrained by the base level language (e.g., R for SN) or statistical package (e.g., SAS for Statistics Canada, SNZ, ABS) and associated data storage. More widespread use of these tools in the "future world" seems to require the developers to recognise the importance, and the provision, of "plug and play", "connectivity", and "metadata-driven" capabilities in the tools.

Summarising the lessons learnt in implementing a similar transformation program undertaken in the Office of National Statistics (UK), Penneck (2009) emphasised the importance of leading and sustaining a cultural change in the "new way of doing things". Whilst some references were made in the GoF articles about engagement with their business counterparts and statistical staff in general, there is not a lot of information, if any, in these articles to assist those who want to learn from their experience. We urge the authors make them available in the public domain.

### *3.2. Enterprise Architecture (EA)*

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). Successful change programs in an NSO will, in our view, require transformation of the business process, of how statistical data and metadata are managed, and of the technology infrastructure and application platforms. Outside the statistical world, it is customary to use an EA (Zachman 1987) as a tool to describe the interplay between, and steer, these essential elements of the change program. We believe that it is a useful tool for NSOs in undertaking their transformation programs.

From The Open Group's (2009) Architecture Framework, "[t]he purpose of enterprise architecture is to optimize across the enterprise the often fragmented legacy of processes (both manual and automated) into an integrated environment that is responsive to change and supportive of the delivery of the business strategy." It is customary to visualise the EA

to comprise a Business Architecture layer, an Information Architecture layer, an Application Architecture layer, and a Technology layer, typically referred to as the Enterprise Domains. At the strategic level, the EA also comprises business segments including statistical leadership, statistical production, statistical support, and business support, which transcend people, products, and performance in the statistical organisation.

It is also customary to accompany the EA with a set of EA principles. These are used as a "guiding light" when it is unclear which one of a number of alternative choices, or pathways in the transformation, should be adopted. The SNZ article provided a set of the principles for their "Economic statistics business architecture", but we believe a broader set covering all official statistics would, for obvious reasons, be preferred.

## 4. Methodology and Methodology Architecture

Whilst all articles in the GoF mentioned statistical methods in conjunction with tools in the context of standardisation, the focus of the articles is on the tools, rather than the methodology. What role did methodology play in the transformational change carried out in the four statistical organisations, and how did methodologists contribute to the transformation process?

We believe that methodologists play a crucial role in shaping and contributing to the development and implementation of strategies to support a business transformation program. In recognition of this, we believe that the EA for statistical agencies should be extended to include a Methodology Architecture (MA) within the Business Architecture layer. Underpinned by the NSO transformational needs, the MA lays out the vision for methodology, supported by an envisaged future inventory of statistical methods and uses, contrasting with the current inventory, and a transition plan for migrating from the current methodological state to the future state.

In developing the MA, we believe consideration needs to be given to the following issues:

- how the suite of methods should be modularised – what should the scope and boundaries of each module be and what are the interdependencies?
- how to trade off between ease of application (usability), scope of application (flexibility), efficiency, and simplicity?
- the availability and use of process quality measures.
- the soundness and consistency of relevant conceptual frameworks underpinning the methods, and
- the relationship between MA and the Business, Information, Application and Technology layers within the EA.

## 5. Challenges for the NSOs as a Community

Pink et al. (2009) argued that given the enormity of the transformation program, it will be difficult, and indeed not sensible, for NSOs to do it alone. In the transformational space, we believe that NSOs could work (better) together in the development of the EA framework, standards (as was shown in the development of GSBPM and GSIM), products, ways to improve skills, and the processes to manage the transformation. In what follows,

we outline what we believe are key challenges for the NSOs in collaborating as a community.

### *5.1. Collaboration in the Development of Future Statistical Products and Services*

The use of Big Data as shown in Varian and Choi (2009) and Kimura et al. (2012) to create timely statistics relevant to official statisticians demonstrates the potential value of these data sets to contribute to the NSOs' mission to provide timely and relevant statistics and services for more informed decision making. One question faced by NSOs is: "Is this a space that we should be in?" In our view, the answer is obvious if we want to remain relevant in the future. If so, the question is how. Until recently, official statisticians gave little attention to Big Data's potential. However, this is going to be changed with the UNECE signalling its intention to commission a paper to understand what Big Data is and how it can be used in official statistics.

The statistical community will benefit from more sharing of ideas, developments and advances in this space, including the methods for processing and quality assuring the data.

### *5.2. Collaboration in the Development of Business Architecture (BA)*

The GoF articles and our interactions with statistical counterparts in other parts of the world show that there is general convergence in our thinking on improving the efficiency and effectiveness of, and growing, our business. There is also remarkable similarity between the ideas about reusing data, reusing metadata, and reusing statistical methods and processes to improve the NSOs' value proposition to their users. Their vision is in strong accord with the vision of HLG-BAS.

Is the commonality of purposes for transformation in the NSOs of such a high level that it warrants the development of an international "reference" for statistical BA? If so, what work is needed for this to occur, who should be leading this work and how can NSOs contribute? The benefit of such an international collaboration effort is enormous, including the placement and recognition of statistical production standards like GSBPM, GSIM in this architecture. In late 2012, UNECE released GSIM 1.0 (UNECE 2012), including a User Guide and specification for consultation within the international statistical community. The Statistical Network referred to earlier has commenced a project to collaborate on developing a "common" BA for the participating countries in the Network.

### *5.3. New Form of International Collaboration to Create Statistical Processes and Tools?*

Hamilton and Tam (2012b) argued that official statisticians can learn a lot from the model used by some well-known smartphone manufacturers in getting the world's developers to develop applications for their handsets to improve the functionality of the phones and user experience. They argued that this collaborative model should be added to the "stand-alone" model (as used in the development of PCAXIS, Blaise, etc.), the "co-development and co-design" model (as used in developing GSIM), and other existing models for international collaboration.

Under this "platform-based" collaboration model, developers (e.g., staff in the NSOs) can upload applications to an internet-based statistical applications store for sharing. In choosing the processes or tools to plug and play, the official statistician shall no longer be restricting the search to applications within their own statistical organisation, but will have applications from other NSOs available to search and to download. In addition, private developers who want to sell their products to NSOs can also upload them to the same store for showcasing to the NSOs.

If this new form of international collaboration is to work, a number of prerequisites have to occur, including a reference EA, with internationally agreed protocols for developing applications under the EA, standards for plug and play, connectivity, and metadata-driven systems. Whilst some work has been undertaken on this front, more needs to be done and particularly more involving the great many NSOs outside Europe, Asia, and Pacific, and beyond the participating countries in the Statistical Network.

In regard to defining standards for "plug and play", we are pleased to note that, in November 2012, HLG BAS commissioned a major collaborative work program to be undertaken in 2013.

We thank the authors for their thought-provoking articles.

## 6. References

Daas, P. (2012). Big Data and Official Statistics. Sharing Advisory Board Newsletter, Issue 7. 2–3. Available at: http://www1.unece.org/stat/platform/download/attachments/22478904/issue+7.pdf?version=1&modificationDate=1354288238441 (accessed February 12, 2013).

Groves, R. (2011). Three Eras of Survey Research. Public Opinion Quarterly, 75, 861–871.

Hamilton, A. and Tam, S.-M. (2012a). Toward GSIM V1.0 as a Cornerstone for Common Reference Architecture. Meeting on the Management of Statistical Information Systems (MSIS 2012) Washington, DC, 21–23 May 2012. Available at: http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.50/2012/02_Australia.pdf (accessed February 12, 2013).

Hamilton, A. and Tam, S.-M. (2012b). Platform for International Collaboration – Preliminary Thoughts from the ABS. Meeting on the Management of Statistical Information Systems (MSIS 2012) Washington, DC, 21–23 May 2012. Available at: http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.50/2012/20_Australia.pdf (accessed February 12, 2013).

HLG-BAS (2011). Strategic Vision of the High-Level Group for Strategic Developments in Business Architecture in Statistics. UNECE. Available at: http://www1.unece.org/stat/platform/display/hlgbas/Strategic+vision+of+the+High-level+group+for+strategic+developments+in+business+architecture+in+statistics (accessed February 12, 2013).

Kimura, M., Makita, N., Terada, M., and Kobay, M. (2012). Developing Value-Added Small Area Statistics with Operational Data of Mobile Network. Paper presented at the Conference of the International Association of Official Statistics, Kiev, 12–14 September 2012.

Kotter, J. (1995). Leading Change: Why Transformation Efforts Fail. Harvard Business Review, 73, 59–67.

Penneck, S. (2009). The Office for National Statistics Statistical Modernisation Programme: What Went Right? What Went Wrong? In Proceedings of Modernisation of Statistics Production, International conference organised by Statistics Sweden, Stockholm 2–4 November 2009.

Pink, B., Borowik, J., and Lee, G. (2009). The Case for an International Statistical Innovation Program – Transforming National and International Statistics Systems. Statistical Journal of the International Association for Official Statistics, 26, 125–133.

Scheuren, F. (2005). Paradata from Concept to Completion. In Proceedings of the Statistics Canada Symposium 2005. Ottawa: Statistics Canada.

Tam, S., Farley-Lamour, K., and Gare, M. (2009). Supporting Research and Protecting Confidentiality – ABS Micro Data Challenges and Future Directions. Statistical Journal of the International Association of Official Statistics, 26, 65–74.

Tam, S. (2011). On-line Access of Micro-data at the Australian Bureau of Statistics – Challenges and Future Directions. Proceedings of the 58th World Statistics Congress of the International Statistical Institute.

The Open Group (2009). TOGAF Version 9. Available at: http://www.opengroup.org/togaf/ (accessed February 12, 2013).

UNECE (2012). GSIM Version 1.0. Available at: http://www1.unece.org/stat/platform/pages/viewpage.action?pageId=59703371 (accessed February 12, 2013).

Vale, S. (2009). The Generic Statistical Business Process Model, Version 4.0. Available at: http://www1.unece.org/stat/platform/display/metis/The+Generic+Statistical+Business+Process+Model (accessed February 12, 2013).

Varian, H R. and Choi, H. (2009). Predicting the Present with Google Trends. Google Research Blog. Available at: http://googleresearch.blogspot.com/2009/04/predicting-present-with-google-trends.html (accessed February 12, 2013).

Zachman, J. (1987). A Framework for Information System Architecture. IBM System Journal, 26, 276–292.