

**Plug and play statistical components – the cornerstone for future proofing  
Australia’s statistical systems**

by

Gillian Nicoll, Frank Yu and Michael Meagher

Australian Bureau of Statistics

ABS House 45 Benjamin Way Belconnen ACT 2617 Australia

Gillian Nicoll, PH: +61 2 6252 5533, F: +61 2 6252 7511

E: Gillian.Nicoll@abs.gov.au

**Abstract**

National statistical offices are facing challenges on a number of fronts. Users of statistical information want better quality data, and they want it faster, at a cheaper price, and in a format that suits their own systems. Providers want easier ways to provide their data so that the burden and cost of providing their information is reduced. At the same time, NSIs are facing reducing budgets and increasing costs.

Australia’s national statistics office, the Australian Bureau of Statistics, has set out to meet these challenges head-on. Through large scale innovation across the ABS, we are radically transforming the way we acquire, collate, use, reuse and disseminate statistical information. Our aim is to have a ‘plug and play’ system where statistical processes and technology can be assembled to order with minimal IT involvement. This new world will enable seamless reuse of skills, business processes, services, information and data models, tools and technologies across all collections in the ABS.

The paper describes our transformation activities. Our statistical processes are being reengineered, using international standards. We are building foundation infrastructure to support the transformation, including a Metadata Registry and Repository, an Enterprise Data Warehouse and Statistical Workflow Management.

**Keywords:** Transformation; GSIM; GSBPM; assemble-to-order.

**1. Introduction**

In 2012, the Australian Bureau of Statistics (ABS) initiated a transformation program that will change the way ABS collects, collates, manages, uses, reuses and disseminates statistical information.

Through large scale innovation across the ABS, we are radically transforming the way we work. This transformation is aimed at reducing the cost and time of doing business; growing our business through new statistical products and services; and delivering a large scale digital Census. Our aim is to have a ‘plug and play’ system where statistical processes and technology can be assembled to order with minimal IT involvement. This new world will enable seamless reuse of skills, business processes, services, information and data models, tools and technologies across all collections in the ABS.

This paper describes the ABS’s transformation activities including: our strategic goals, new enterprise architecture; new information frameworks; new information infrastructure; and our new ways of acquiring and disseminating statistical data.

It also describes some of the challenges the ABS has encountered along its journey, and will explain how its work could help the international statistical industry to embrace a plug and play culture.

## 2. Strategic Goals

The ABS is facing challenges on a number of fronts:

- demand for higher quality, cheaper and more timely data;
- expectations of providers for easier ways to provide their data;
- reductions in our budget;
- escalating costs, especially in difficulty to contact households ;
- difficulty in attracting new staff due to increased competition ;
- increased complexity of work programs, including measurement of global changes; and
- increased expense of maintaining ABS's siloed and aging statistical processes and infrastructure.

The ABS has decided to attack these challenges head on.

At the heart of the transformation is the ABS Enterprise Architecture, that will support the re-engineering of business processes to bring about a leap in re-use and automation of end-to-end statistical processes. To achieve this, it is essential to manage and standardise both metadata and processes across all household and business statistical collections run in the ABS.

The foundation information infrastructure consists of three components, all based on a common information framework:

- a standards based Metadata Registry and Repository to enable discovery and re-use of metadata;
- a Statistical Workflow Management (SWM) system to increase the transparency, automation and re-use of business processes at the application level
- an Enterprise Data Warehouse (EDW) to manage statistical data across the end-to-end statistical process, using metadata from the MRR to locate and link datasets

To maximise the potential for automation of business processes, there are strong linkages between the SWM, MRR and EDW; so that processes are highly configurable and driven by metadata. A service-orientation approach to systems development is being used, where re-usable well defined, self-contained functional modules ('services') are connected together to run statistical business processes

## 3. Enterprise architecture

The ABS 2017 transformation program will provide us with a future world where architectural standards have been defined in sufficient detail to enable the development of statistical production solutions, which can be re-used easily across all household and business collections in the ABS.

The new ABS Enterprise Architecture is the plan that translates the ABS's vision for the future into the change required to move the ABS into the new world. It will facilitate business and information transformation through the provision of a conceptual view of the ABS's 'to-be' architecture.

The EA guides ABS planning and investment decisions at both the strategic and tactical levels, ensuring all decisions support the ABS's plug and play vision.

Within the EA, the Business Architecture looks in detail at the functions that the ABS undertakes and identifies discrete, non-overlapping activities. The Business

Architecture takes the enterprise domain and progressively breaks it into finer level business functions. This progressive disaggregation, and the relationships between the activities, forms the basis for establishing the capability requirements to transform the organisation.

The ABS defines capability in any activity to be the combination of people, methods, processes, knowledge and systems that enable the organization to carry out that activity to achieve a business outcome. Careful analysis of the existing capability, compared with the future requirements, will provide a roadmap for investment in each of the dimensions of capability.

By understanding the future Business Architecture and the required capability, an organisation can develop a roadmap to transform and modernise.

## **5. Information frameworks and standards**

The ABS will use two key information frameworks to underpin all of its transformation work. These frameworks provide a common language for statistical production and, when used together, will enable the ABS to promote reuse of statistical information resources, not just in the ABS, but across the world.

The first framework is the Generic Statistical Business Process Model (GSBPM) which provides standard terminology to describe statistical production processes. The GSBPM is well established and has been integrated into the language of the ABS.

The second framework is the new Generic Statistical Information Model (GSIM) which defines the information required to run a statistical process (such as the structure of the input datasets; the variables to be defined; the different levels of aggregation of these variables), and the information outputs (such as output datasets; process metrics) that are released after that process has run.

The ABS is now operationalizing GSIM, as part of our Statistical Information Management Framework. In addition to GSIM, this framework includes:

- A governance framework, including processes for standardising metadata, versioning, and clarifying ownership and custodianship; and
- A capability framework, which will identify any gaps in the skills required within the ABS to use the new information model, and the strategies to fill those gaps.

The Statistical Information Framework is critical to ensure that the ABS information infrastructure is the source of authoritative metadata appropriate to drive the plug and play environment of the future.

## **6. Information Management Infrastructure**

There are three critical pieces of new infrastructure being developed within the ABS that will provide the foundation to our new world: the Statistical Workflow Management System; the Metadata registry and repository; and the Enterprise Data Warehouse. Once linked, the foundation infrastructure will drive metadata driven systems, and will provide the ABS with the ability to keep information on all processes, quality measures, transformations, etc throughout the lifecycle of the data path.

## 6.1. Statistical Workflow Management System

The first new part of the foundation infrastructure is the Statistical Workflow Management System (SWMS). SWMS provides the ABS with the capability to run its business process in an automated and managed manner:

- controlling and storing the common business processes used by SMAs,
- retrieving data and metadata from the metadata registry and repository, and from the enterprise data warehouse.

The SWMS provides the basis for building, automating, and re-using processes from high level workflow through to low level technical. It provides the engine for metadata driven processes making ABS operations more efficient.

Without strong governance, there is a high likelihood of duplicating processes and/or building inappropriate components using SWM, thus achieving little re-use or automation. As such, the concept of SWM goes beyond the use of BPMS software, and encompasses governance, including role and responsibilities around building and implementing new processes.

It is envisaged that capturing processes in SWM will require three layers of modelling:

- A high level business process model – built by business analysts/and process re-engineers with statistical staff (as part of the multidisciplinary approach discussed earlier)
- A detailed business process model – built by business analysts, using modelling tools
- The executable business process in ActiveVos – built by technologists.

In the longer term, it is envisaged that SWM will provide the basis for assembling high level processes for new collections. The execution of the new end-to-end collection process will require a significant amount of configuration of both statistical metadata and process metadata or paradata. This represents a significant change from current business practices and requires a high degree of maturity in operating SWM and SOA.

## 6.2. Metadata registry and repository

Another critical part of the new information infrastructure is the Metadata Registry and Repository. The Metadata Registry and Repository is designed to be the centralised, standards based, authoritative store of statistical metadata for the ABS. It consists of:

- the Repository – the centralised store for standards based statistical metadata, and
- the Registry – the catalogue that tells a user what's in the repository.

The MRR has three main functions, to register metadata, search for metadata, and retrieve metadata.

The split of the MRR into two components facilitates its three main functions. Upon registering a piece of metadata to the MRR, selected elements of the metadata package are recorded in the registry, while the metadata package is saved in its entirety in the repository. The registered elements provide an index for the complete metadata package and are the elements most likely to be used in searching. Only metadata conforming to approved standards will be able to be registered (in the first instance using specific versions of DDI and SDMX).

The MRR will allow for a range of metadata to be registered, for example:

- Content metadata: concepts; classifications; populations; units etc
- Process metadata: type of process used for the collection (eg type of editing or imputation), version used, where it is stored, the datasets and metadata inputs and outputs used for that process
- Data descriptors: physical location; ownership, content.

The registry provides a functionality which is analogous to advanced searches such as the one provided by Amazon (which allows users to search for books by keywords, author, title, ISBN and various other attributes). For example, staff can search the MRR based on the name of a Classification Version or a keyword in the definition of a Concept.

### **6.3. The Enterprise Data Warehouse**

The final piece of the foundation infrastructure is the Enterprise Data Warehouse.

The EDW will be a metadata-driven store to manage statistical data throughout the end-to-end statistical process. Metadata-driven data services will be developed to:

- Respond to the creation or modification of metadata appropriately, e.g. through the creation or modification of appropriately formatted data stores;
- Provide data to processing systems in appropriate formats and on time;
- Capture and store the results of processing in documented versions of data;
- Provide mechanisms to update metadata with the results of processing;
- If necessary, perform metadata driven processing tasks internally;
- Support the management of access.

The EDW will hold all data, including raw unit record data, edited data, aggregated data, and disseminated data. Staff will use metadata from the MRR to locate relevant datasets and link datasets.

The EDW will also provide user-friendly business intelligence & analysis tools to select, manipulate, aggregate and visualise data. This includes reviewing the current toolset and developing adaptors to allow tools to support metadata standards, such as SDMX and DDI. This will provide an enterprise approach and allow for increased automation.

## **7. Service Oriented Architecture**

The ABS is looking to create an environment where processing systems for new collections can be quickly assembled to order by subject matter statisticians.

To achieve this, the ABS has introduced service-oriented-architecture (SOA) to make it easier and faster to build and deploy IT systems that directly serve the goals of a business. Software modules (Services) will be developed to provide standard business functionality which aligns to the capabilities articulated in the ABS's Enterprise Architecture.

The Services will be listed in a catalogue and described in terms that statistical business areas understand. Each Service will be developed at a low enough level to that they are reusable by all statistical areas. Statistical areas can rapidly combine/recombine services to provide end-to-end solutions for new collections. This architecture allows the business to keep its focus on business, and lets ABS strategically invest in new services to keep pace in a rapidly changing world.

## **9. Plug and play internationally**

The challenges faced by the ABS in modernising our infrastructure have been mirrored across the world. Many NSOs have developed statistical systems of in silos, with tightly integrated and localised storage. This tight integration causes difficulty in decoupling components, making it harder to update one part of a process without updating the entire system.

The ABS is working with other national statistical organisations to develop a flexible architecture which will facilitate easy sharing of our infrastructure.

With internationally agreed Plug and Play architecture, technology can be easily swapped and shared between National Statistical Offices. The cost of developing new systems will be dramatically reduced, and updating old systems will be a simple process of plugging in the new component.

## **10. Learnings and Challenges**

To successfully achieve this ambitious change program the ABS must overcome a series of significant challenges. These include:

- Maintaining a core skill base for information modeling and management.
- Maintaining corporate support for the change initiative, in spite of the long lead times for the development of the infrastructure.
- Harmonizing information flows and our business processes, such as formal linkages between GSBPM and GSIM will be essential in the development of robust, metadata driven and sustainable infrastructure.
- Changing the culture to be 'hard-nosed' our investments and our business processes.
- Building an organization which values reusing metadata, rather than creating.
- Aligning our legacy software as part of an effectively transition strategy to the new world.

## **11. Summary**

The environment in which the ABS operates has changed dramatically – requiring richer and timelier data at a lower cost. The ABS is responding to this change by modernizing and industrializing the way in which we undertake our businesses. We will be analyzing our future business requirements through application of our enterprise architecture, and identifying new capabilities that we must develop.

We have commenced this modernization journey by developing core foundational information infrastructure. These are necessary, but not sufficient elements to complete the modernization of the ABS business practices.

While there are a range of challenges, The ABS is confident that the modernization program is achievable by developing our infrastructure in line with international standards (such as GSIM) and collaborating and sharing with infrastructure with other National Statistical Offices.