**Enhanced Generic Models to Support the Standardisation of Statistical Production**

**Steven Vale**

**United Nations Economic Commission for Europe**

**Abstract**

Standardisation of statistical production requires generic models to describe the different parts of the production process and the information consumed and created by it. Such models should also provide a common language to facilitate benchmarking and quality assurance between processes and organisations.

The High-Level Group for the Modernisation of Statistical Production and Services (HLG), a global group of chief statisticians from national and international statistical organisations, has launched an ambitious modernisation programme, based on key standards such as the Generic Statistical Business Process Model (GSBPM) and the Generic Statistical Information Model (GSIM). To ensure that these models were fully fit for purpose, they were reviewed and revised during 2013, as part of a HLG project on “Frameworks and Standards for Statistical Modernisation”.

This paper outlines the models and the role they play in modernising and standardising statistical production, as well as presenting the changes and enhancements introduced in the new versions released in December 2013. It also considers the practical implementation and use of the models in different organisations, and the ways in which they have contributed to standardisation and quality assurance.

**1. Introduction**

The High-Level Group for the Modernisation of Statistical Production and Services (HLG)[[1]](#footnote-1) was created in 2010 by the Bureau of the Conference of European Statisticians. It comprises the heads of national and international statistical organizations, and has a mandate to reflect on and guide strategic developments in the ways in which official statistics are produced.

This mandate has been translated into a vision and strategy, both endorsed by the Conference of European Statisticians, which promote the goal of “standards-based modernisation”. The logical first step was therefore to identify the necessary standards, and develop new ones where necessary. The following diagram (Figure 1) was created as part of the development process of the HLG vision and strategy, and shows how the initial focus was on the conceptual models: the Generic Statistical Business Process Model (GSBPM)[[2]](#footnote-2) and the Generic Statistical Information Model (GSIM)[[3]](#footnote-3). More recently the focus has shifted to a more practical approach, applying these models through a Common Statistical Production Architecture[[4]](#footnote-4), and mapping them to key data and metadata management standards.



Figure 1: The four pillars of statistical modernisation

As the models have been applied in more and more contexts around the world (including a few cases outside official statistics[[5]](#footnote-5)), feedback and ideas for improvements were generated. This feedback and implementation experience was used to develop new versions of both the GSBPM and the GSIM during 2013. This work was part of a project on “Frameworks and Standards for Statistical Modernisation”, overseen by the HLG.

**2. The GSBPM**

The GSBPM describes and defines the set of business processes needed to produce official statistics. It provides a standard framework and harmonised terminology to help statistical organisations to modernise their statistical production processes, as well as to share methods and components. The GSBPM can also be used for integrating data and metadata standards, as a template for process documentation, for harmonizing statistical computing infrastructures, and to provide a framework for process quality assessment and improvement.

Part of C the Common Metadata Framework[[6]](#footnote-6), "Metadata and the Statistical Cycle", developed by the UNECE Steering Group on Statistical Metadata, refers to the phases of the statistical business process (also known as the statistical value chain or statistical cycle) and provides generic terms to describe them. As a result, version 4.0 of the GSBPM was released in April 2009. It was subsequently widely adopted by the global official statistics community. Version 5.0 (see Figure 2) was released in December 2013.

Figure 2: Version 5.0 of the GSBPM

The GSBPM is not intended to be a rigid framework in which all steps must be followed in a strict order, instead it identifies the possible steps in the statistical business process, and the inter-dependencies between them. It should be seen more as a matrix, through which there are many possible paths.

The GSBPM aims to be sufficiently generic to be widely applicable, and to encourage a standard view of the statistical business process, without becoming either too restrictive or too abstract and theoretical. It is intended to apply to all activities undertaken by producers of official statistics, at both the national and international levels, which result in data outputs. It is designed to be independent of the data source, so it can be used for the description and quality assessment of processes based on surveys, censuses, administrative records, and other non-statistical or mixed sources.

The GSBPM recognises over-arching processes that apply throughout statistical business processes. These can be grouped into two categories, those that have a statistical component, such as quality and metadata management, and those that are more general, and could apply to any sort of organisation, such as human resource and financial management.

**3. The GSIM**

The GSIM is a reference framework of information objects, which enables generic descriptions of the definition, management and use of data and metadata throughout the statistical production process. It provides a set of standardized, consistently described information objects, which are the inputs and outputs in the design and production of statistics. As a reference framework, the GSIM helps to explain significant relationships among the entities involved in statistical production, and can be used to guide the development and use of consistent implementation standards or specifications.

The GSIM provides the information object framework supporting all statistical production processes such as those described in the GSBPM , giving the information objects agreed names, defining them, specifying their essential properties, and indicating their relationships with other information objects. GSIM identifies around 110 information objects. Examples include data and metadata (such as classifications) as well as the rules and parameters needed for production processes to run (for example, data editing rules).

Information objects are grouped into four top-level groups, as follows:

* The **Business** group is used to capture the designs and plans of statistical programs, and the processes undertaken to deliver those programs. This includes the identification of a Statistical Need, the Business Processes that comprise the Statistical Program and the evaluations of them.
* The **Exchange** group is used to catalogue the information that comes in and out of a statistical organization via Exchange Channels. It includes objects that describe the collection and dissemination of information.
* The **Concepts** group is used to define the meaning of data, providing an understanding of what the data are measuring.
* The **Structures** group is used to describe and define the terms used in relation to information and its structure.

Figure 3 gives examples of the objects that are in each of the four top-level groups.



Figure 3: Examples of GSIM information objects

The GSIM and the GSBPM are complementary models for the production and management of statistical information. As shown in Figure 4, the GSIM helps to describe GSBPM sub-processes by defining the information objects that flow between them, that are created in them, and that are used by them to produce official statistics.



Figure 4. GSIM and GSBPM

Version 1.0 of the GSIM was released in December 2012, followed by version 1.1 in December 2013.

**4. Revising the models**

As mentioned above, the international collaboration project “Frameworks and Standards for Statistical Modernisation” included the revision of the GSBPM and the GSIM. This project had four main objectives:

* To ensure that the international statistical community has access to the standards needed to support the modernisation of statistical production and services
* To increase coherence between these standards
* To provide support mechanisms for the practical implementation of these standards within national and international statistical organisations
* To ensure effective promotion and maintenance of the GSBPM and the GSIM, including the release of new versions as appropriate.

Some of the work packages within the project were:

* Work Package 1: Support for the implementation of the GSIM - This work package included providing support for a community of GSIM “early adopters”, by creating a “GSIM Implementation Group” which met regularly by web conference, and shared ideas and questions through a wiki-based discussion forum.
* Work Package 2: Enhancement of the GSIM - This work package gathered feedback from the users of the GSIM, including those with experiences of practical implementations, and potential user groups within and outside official statistics.
* Work Package 3: Update of GSBPM - This work package gathered feedback from users on potential changes to GSBPM. As GSBPM version 4.0 was adopted in 2009, and already implemented by over 50 statistical organisations around the world, changes were deliberately kept to a minimum. They had to be supported by a good business case, and a majority of the participating organisations, to ensure that the impact of the new version on the user community was kept to a minimum.
* Work Package 4: GSIM / DDI / SDMX mapping - This work package included detailed mappings between the information objects in GSIM and those in the information models of DDI and SDMX. The aim was to identify any issues affecting the coherence of these standards, and propose solutions where possible.
* Work Package 8: Mapping GSBPM to the Fundamental Principles of Official Statistics - This requirement was identified during a high-level seminar for leaders of statistical organisations in SPECA (Special Programme for the Economies of Central Asia) countries.

**5. The results**

New versions of both the GSBPM and the GSIM were released at the end of 2013. Mappings between the GSIM and DDI / SDMX, as well as between the GSBPM and the Fundamental Principles of Official Statistics, were also released. Information about the standards and models needed for statistical modernisation was consolidated in one place, a new “Virtual Standards Helpdesk”[[7]](#footnote-7).

**5.1 GSBPM v5.0**

The main changes to the GSBPM between versions 4.0 and 5.0 are as follows:

* Phase 8 (Archive) has been removed, and incorporated into the over-arching process of data and metadata management, to reflect the view that archiving can happen at any stage in the statistical production process.
* A new sub-process: "Build or enhance dissemination components" has been added within the "Build" phase to reflect the growing importance of having a range of dissemination options.
* Within the “Process” phase, a clearer distinction has been made between the detection of potential errors and inconsistencies, and their subsequent treatment
* Several sub-processes have been re-named to improve clarity
* The descriptions of the sub-processes have been updated and expanded where necessary
* The terminology used has been changed to be less survey-centric, in recognition of the growing use of non-survey sources (administrative data, big data etc.).

**5.2 GSIM v1.1**

Version 1.1 of the GSIM has simplified some areas of the model, where complexity was a potential barrier to implementation, and increased the alignment of GSIM with other models and good practices.

A review of the Neuchâtel Model of classification terminology was taking place independently as a joint activity between the UN Expert Group on Classifications, and the Conference of European Statisticians Steering Group on Statistical Metadata. The reviewers and the project team realised that there was considerable overlap in terms of what the two models were trying to cover. The result is that the revised Neuchâtel model has been incorporated as the classification part of GSIM .

Work has started in 2014 to develop a new generation of the DDI standard. This draws heavily on the modelling work done in the context of the GSIM revision. As a result, coherence between these standards, which is already good, is likely to improve further.

**6. The Future – Governance and change management**

Further revisions to the GSBPM and the GSIM are not expected in the short-term (3-5 years). It is important to balance the benefits of stability against any requirements for change. However wider use and implementation of these models is certain to identify areas where they could be improved. The Virtual Standards Helpdesk includes discussion forums, where users are invited to contribute ideas. These ideas will be dealt with in two ways:

* Cases where issues can be resolved by adding or enhancing documentation will be dealt with as they arise
* Cases requiring changes to the models themselves will be stored until they reach a “critical mass”, when they will be considered as inputs to revisions of the models. This is likely to happen no more than once every five years, and the frequency can be expected to reduce as the models become more mature.

The “owner” of the GSBPM and the GSIM is the High-Level Group for the Modernisation of Statistical Production and Services[[8]](#footnote-8). This is the group that will decide when new versions are needed. However, for practical purposes, since January 2014, maintenance of the models has been delegated to the Modernisation Committee on Standards. This committee is responsible for managing feedback, overseeing the sharing of implementation experiences, and enhancing supporting documentation. An example of the work to be undertaken by this Modernisation Committee is the recently-launched work to develop recommended quality indicators for each GSBPM sub-process.

1. http://www1.unece.org/stat/platform/display/hlgbas [↑](#footnote-ref-1)
2. http://www1.unece.org/stat/platform/display/GSBPM [↑](#footnote-ref-2)
3. http://www1.unece.org/stat/platform/display/gsim [↑](#footnote-ref-3)
4. http://www1.unece.org/stat/platform/display/CSPA [↑](#footnote-ref-4)
5. For example: http://dx.doi.org/10.3886/DDILongitudinal05 [↑](#footnote-ref-5)
6. http://www1.unece.org/stat/platform/display/metis/The+Common+Metadata+Framework [↑](#footnote-ref-6)
7. http://www1.unece.org/stat/platform/display/VSH [↑](#footnote-ref-7)
8. http://www1.unece.org/stat/platform/display/hlgbas/Modernisation+Committee+on+Standards [↑](#footnote-ref-8)