

Generic Statistical Information Model: An innovative collaboration which facilitates international collaboration

Abstract

Jenine Borowik¹, Carrie Ashley¹, Aurito Rivera¹ and Steven Vale²

¹ Australian Bureau of Statistics, Canberra, Australia

² United Nations Economic Commission for Europe (UNECE) Statistical Division

Contact author: Jenine Borowik, email: jenine.borowik@abs.gov.au

The Statistical Network was initiated in June 2010 with the following objective:

Collaboration in practical small steps to industrialise methods and processes to quickly and effectively benefit all participating National Statistical Offices.

The Generic Statistical Business Process Model (GSBPM) was used within the Statistical Network to identify the sub-processes within statistical production where there was greatest interest and opportunity to develop collaborative solutions.

At that time, however, there was no common reference framework to describe the statistical information (data and metadata) which would be input to, and output from, each sub-process. The Statistical Network therefore initiated development of the Generic Statistical Information Model (GSIM).

GSBPM and GSIM aim, between them, to provide common terminology, improving communication about the production of statistics, within and between organizations. This, in turn, facilitates collaboration and exchange of good practices, leading to greater efficiency.

During 2011, the value and importance of GSIM became recognised more widely beyond the Statistical Network. In 2012, an unprecedented international collaboration effort was initiated by HLG (High Level Group for the Modernisation of Statistical Production and Services) to complete development of GSIM V1.0.

This paper will summarise the development of GSIM and provide an overview of the content of GSIM V1.0 which was released in December 2012. The paper will include consideration of the strategic context and benefits associated with GSIM.

Now GSIM has been developed, it will facilitate further international collaboration to develop practical, sharable solutions that address specific business needs.

The paper outlines these next steps. This includes outlining how GSIM can be used as a framework to guide consistent use of implementation standards such as SDMX and DDI in a manner which supports the business practices and needs of statisticians.

The paper also outlines plans to further evolve GSIM, to more fully meet business needs, based on experience from applying the current version in practice.

Key words: common reference framework; information model; High-Level Group

1. Introduction

Official statistics must stay relevant to the modern world. National Statistical Offices (NSOs) face the shared challenges of limited funding, aging infrastructure, capability shortages, ambitious work programs, big data and a rapidly evolving environment. In addition, official statisticians need a flexible and agile statistical environment that supports the production and analysis of coherent statistical information from multiple sources in a cost effective manner.

The ability to engage in effective collaborations between producers of official statistics – underpinned by common frameworks and standards – is essential for being able to respond to changing needs and opportunities in an agile manner and to remain relevant to governments’ and societies’ needs for statistical information to support analysis and decision making.

With every challenge comes opportunity. The body now known as the High-Level Group for the Modernisation of Statistical Production and Services (HLG)¹ has identified the Generic Statistical Information Model (GSIM) as a key standard, in partnership with the Generic Statistical Business Process Model (GSBPM)², to drive the modernization of official statistics.

GSIM V1.0, released in December 2012, is the first internationally endorsed comprehensive reference framework for statistical information that can be applied regardless of statistical subject matter domain. GSIM is the result of extensive, multidisciplinary development and collaboration across the international statistical community. As a result of innovative techniques and outputs, including creative design processes and communication, GSIM was developed in less than two years. The bulk of the development occurred over ten months.

Statistical agencies have a long history of working together to establish conceptual frameworks for statistical domains. One well-known and broadly influential example is the System of National Accounts (SNA). The role of GSIM when defining and exchanging statistical information can be seen as in some ways analogous to the role of SNA when compiling and comparing national accounts data. Exact implementation practices for collecting and compiling data to produce National Accounts are decided at a national level. However, the SNA provides a common reference framework to promote consistency of definitions and practices and to promote comparability of outputs.

This paper is an abridged form of the full paper of the same name, located on the UNECE website.³ The paper outlines how, and why, GSIM was developed. It identifies learnings which might be relevant to future national and international collaborative developments. It also discusses strategic benefits associated with implementing GSIM, along with a range of “next step” activities which are making use of GSIM V1.0.

2. What is GSIM and how is it used?

GSIM is a reference framework of internationally agreed definitions, attributes and relationships that describe the pieces of information that are used in the production of official statistics (information objects). Examples of information objects in scope for GSIM include data (held in datasets) and metadata (e.g. classifications, variables, questions and populations).

¹<http://www1.unece.org/stat/platform/display/hlgbas/High-Level+Group+for+the+Modernisation+of+Statistical+Production+and+Services>

²<http://www.unece.org/stats/gsbpm>

³<http://www1.unece.org/stat/platform/download/attachments/75564198/Developing+GSIM+V1.0+-+and+further+developments.docx?version=2&modificationDate=1366029857271>

GSIM provides a common language to describe information that supports the whole statistical production process. Official statisticians follow a set of processes, such as those described in the GSBPM, to produce official statistics. GSIM defines and describes the information (data, metadata, rules and parameters etc.) that flows between the steps in these processes, from the identification of user needs through to the dissemination of statistical products.

As a reference framework, one of GSIM's key purposes is to improve communication between different disciplines involved in statistical production, within and between statistical organizations; and between users and producers of official statistics. By supporting improved collaboration, GSIM holds the potential to generate economies of scale through reuse of information, methods or technology. It provides a basis for flexibility and innovation, and will make it easier and faster to collaborate and exchange tools and ideas. GSIM has been designed to support greater automation of the statistical production process. It can be used as a teaching aid to build staff capability, and also allows NSOs to validate existing information systems and compare with best practice in other organizations.

GSIM is aligned with relevant data management and exchange standards, such as SDMX⁴ and DDI-L⁵, but it is not directly tied to them, or to any specific technology. As an overarching framework, GSIM will play an important part in modernising, streamlining and aligning the production and implementation standards associated with official statistics.

This paper does not introduce, in detail, GSIM as a product. Information for understanding GSIM as a product is available through the GSIM Home Page⁶.

3. GSIM as a cornerstone of the HLG Strategic Vision

The need for a common framework in regard to statistical information was identified in June 2010, when the Statistical Network (SN) formed in order to collaborate in practical small steps to industrialise methods and processes. At the time, the SN included NSOs from Australia, Canada, New Zealand, Norway, Sweden and the United Kingdom (Italy joined in 2012). The project team responsible for the initial GSIM collaboration, which included members from each of the six NSOs within the SN, first met (by teleconference) in late November 2010.

Development of GSIM became of global interest, and was accelerated, as HLG's Strategic Vision became more widely understood and supported during 2011.

HLG was established by the Bureau of the Conference of European Statisticians late in 2010⁷. HLG's mission is to oversee development of frameworks, and sharing of information, tools and methods, which support the modernisation of statistical organisations. Their aim is to improve the efficiency of the statistical production process, and the ability to produce outputs that better meet user needs.⁸ There are currently ten members of HLG. Each is the head of a NSO or the chief statistician of an international organisation (UNECE, OECD, Eurostat).

In March 2011, HLG released its Strategic Vision⁹, identifying that producers of official statistics should be able to work together as an "industry", including defining and applying shared "industry standards and frameworks" which will facilitate collaborative development and sharing of processes, methods, IT components and statistical information.

⁴Statistical Data and Metadata eXchange. See <http://sdmx.org/>

⁵DDI (Data Documentation Initiative) Lifecycle Specification. See <http://www.ddialliance.org/Specification/>

⁶<http://www1.unece.org/stat/platform/pages/viewpage.action?pageId=59703371>

⁷The body was referred to as HLG-BAS until the end of 2012.

⁸<http://www1.unece.org/stat/platform/download/attachments/58492100/HLG+ToR+2013+to+2015.doc>

⁹<http://www1.unece.org/stat/platform/display/hlgbas/Strategic+Vision>

In this vision, GSIM is referenced as both a cornerstone and pre-requisite for industrialising statistics, highlighting its critical role in the global agenda for pursuing modernisation.

4. GSIM development and communication

In a number of ways, GSIM development became an exemplar of the approach to collaboration that HLG is championing for the future. This included:

- mobilising resources internationally, and defining and managing the project, through collective agreement and commitment by heads of national and international agencies rather than through a single international agency assuming governance
- adopting an “agile” development framework, focusing on timely initial delivery to be refined through application and using methods such as “Sprints” to supplement – and improve the effectiveness of – more traditional approaches to collaboration between agencies. Two Sprints were held, the first hosted by the Statistical Office of the Republic of Slovenia, and the second hosted by Statistics Korea. These Sprints were defined as intensive collaboration exercises, bringing together experts from different disciplines in a single location for a fixed period of time to achieve a given set of outputs.

HLG made a strategic decision to prioritise timely delivery of GSIM V1.0 over more extensive trialling and refinement prior to initial release. This enabled timelier realisation of initial benefits and progress on subsequent steps. It also allowed GSIM to be refined further based on experience with practical application and testing rather than more theoretical forms of evaluation. This in turn will lead to rapid and business needs driven refinement of the definition of the model.

GSIM development was innovative not only in terms of process, but also in terms of the suite of information made available as communication products. Unlike conventional frameworks and standards in statistical subject matter domains, GSIM was delivered as a comprehensive package of material for different audiences. The formal specification is available in two main forms. One form (in Microsoft Word or PDF) is designed to be read directly by people. The other form (in Unified Modelling Language) is designed to be used systematically by modelling tools. There are also two Communication Brochures, one Communication Paper, a User Guide, a Readers’ Guide, and other supporting material. These outputs reflect the multidisciplinary design process, and are essential for making the content readily accessible to different users such as top managers, statisticians, methodologists and IT experts.

The use of multidisciplinary Sprints, along with input from communication specialists and subject-matter experts to produce a range of documents targeting different audiences, seems to be an approach that could be considered as a good practice, and used more widely.

Another key feature of the Sprints and the wider GSIM development work, strongly encouraged by the HLG, was openness. This took several forms, including inviting a representative of the data archives community to join the second sprint and subsequent activities, using social media to publish updates on progress during the sprints, and generally encouraging input and feedback from as many people and organisations as possible, not just those who were physically present.

Details of the unprecedented international collaboration effort, including GSIM Sprint 1 and 2 objectives and achievements; the June 2012 Business Plan¹⁰ for the GSIM Development Project; the work of the Specification Layer Task Teams and GSIM Integration Team; and the process of web conferences and communications, are outlined in the full version of this paper.

¹⁰<http://www1.unece.org/stat/platform/download/attachments/59703371/19-+GSIM+Business+Plan.doc?version=1>

5. Next Steps

Next steps involve applying GSIM to underpin subsequent steps in the HLG Implementation Strategy for Modernization of Statistical Production. Another key set of next steps is to gather, analyse and act on experiences with implementation of GSIM V1.0 so an enhanced version of GSIM can be designed, agreed and released for use in a timely manner.

a. Outcome realisation at a national level

A number of NSOs have already developed, and are currently implementing, plans for corporate adoption of GSIM. These implementers are working together through the GSIM Implementation Group to share plans, experiences, issues encountered, presentation material and other supporting documentation. The GSIM Implementation Group will be instrumental in gathering and prioritising feedback to allow the next version of GSIM to more fully and more consistently meet the needs of implementers.

b. CSPA Project

Having accelerated and completed development of GSIM V1.0 during 2012, HLG initiated the Common Statistical Production Architecture (CSPA) Project¹¹ for 2013. This project aims to:

- Create a standardised architecture for statistical production solutions, including processes, information and systems, and to allow specifications and ultimately applications to be re-used easily within and between statistical organisations.
- Enable and advance the sharing of production processes or components, thus reducing costs.
- Provide the basis for a central inventory or repository with life cycle management of sharable production processes and components.

CSPA is sometimes referred to as “Plug & Play” Architecture because its aim is to enable a solution developed by one NSO, which is compliant with the standardised architecture, to be “plugged in” (and “play correctly”) to support relevant business processes undertaken by another NSO.

The CSPA Project includes defining, and undertaking a practical “proof of concept” application of the standardised architecture. Alignment with the information model specified in GSIM is one of the defining characteristics for the standardised architecture.

c. Informal taskforce on metadata flows

This taskforce is mapping the typical flow of metadata through the nine phases of the GSBPM, using GSIM V1.0 to characterise the metadata. The work recognises that while GSIM is not dependent on GSBPM, or vice versa, much of the value for NSOs comes from how these two frameworks related to Statistical Information and Statistical Business Processes fit together to characterise the business activity of producing official statistics.

The output of this work will be a generic characterisation of where in the statistical business processes various types of metadata are typically created, reused and updated.

The taskforce’s application of the frameworks to practical (but generalised) business analysis is identifying potential gaps and areas of ambiguity in both GSIM V1.0 and the current version of

¹¹<http://www1.unece.org/stat/platform/download/attachments/58492100/Plug+and+Play+project+outline.docx?version=1&modificationDate=1360670755446>

GSBPM (V4.0). As GSBPM V4.0 was released prior to the development of GSIM, it describes some information objects from GSIM using out of date, or nonaligned, terminology. An update and alignment process for GSBPM which will commence later in 2013.

d. GSIM/ SDMX / DDI

In January 2013, HLG initiated a work package to determine the relationship between the information objects in GSIM (the business oriented conceptual framework), and objects defined in the information models associated with SDMX and DDI-L (two commonly used standards for representing and exchanging statistical data and metadata).

This work will gauge how fully, and how effectively, the existing standards support the full range of information objects defined in GSIM. Change proposals may then be initiated which seek to bring the standards into fuller, and more consistent, alignment with GSIM.

The aim, which will be achieved progressively, is to ensure that agencies which choose to adopt GSIM as their conceptual framework for statistical information will have simple, consistent and standards based means available to them for applying the framework in practice when defining, representing and exchanging data and metadata.

Aligning use of the standards with GSIM means that the solution chosen, where multiple options exist, should best address the business uses associated with that information rather than necessarily being the most technically “satisfying” or “least demanding” option. The solution should also be in common, or at least fully consistent, with the way other implementers decide to address the same need.

e. Designing and agreeing the next version of GSIM

A GSIM V1.0 Discussion Forum was established once GSIM V1.0 was released. Any implementer, or reviewer, of GSIM V1.0 is welcome to contribute issues. The GSIM Implementers Group reviews open issues collectively and negotiates recommendations. Based on the feedback gathered by September 2013, a recommendation will be made to HLG as to whether a revised version of the GSIM is needed, and if so, whether this would be a minor revision or major change. The volume and significance of feedback gathered so far would seem to suggest some form of revision of GSIM will be agreed in September 2013.

6. Conclusion

HLG members have assessed the approach to developing and delivering GSIM V1.0 as highly successful. HLG’s strategy of increasing GSIM’s fitness for purpose through statistical agencies starting to apply V1.0 in practice, and identifying future refinements and extensions, appears at this early stage to have momentum and to be succeeding.

HLG focuses on collaborations between agencies in regard to business processes, statistical methods, IT components and statistical metadata and data repositories. In other contexts, however, the positive experiences with developing and implementing GSIM as a common reference framework may also have relevance for approaches used to develop and evolve common conceptual frameworks for specific domains of statistics.