# Generic Statistical Information Model (GSIM):

# Overview

(Version 0.4, May 2012)

**DRAFT FOR REVIEW**

**Please note** the development of GSIM is a work in progress. GSIM v0.4 is not intended for official publication.

**Instructions for reviewers** and a **template for providing feedback** is available at **http://www1.unece.org/stat/platform/display/metis/GSIM+Version+0.4**

## About this document

This document presents an overview of GSIM. The document consists of general information about the scope and purpose of GSIM, why it is needed, the expected benefits, as well as an introduction to the structure and content of the model. This document is aimed at top managers. The Overview Layer also helps other readers (e.g. target audiences beyond statistical agencies) understand whether GSIM is relevant for them and whether they should explore the other layers.

This work is licensed under the Creative Commons Attribution 3.0 Unported License. To view a copy of this license, visit <http://creativecommons.org/licenses/by/3.0/>. If you re-use all or part of this work, please attribute it to the United Nations Economic Commission for Europe (UNECE), on behalf of the international statistical community.



# What is GSIM?

1. The Generic Statistical Information Model (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.
2. The Generic Statistical Information Model is an internationally agreed set of definitions, attributes and relationships that describe the pieces of information that are used in the production of official statistics.
3. GSIM provides a common language to describe information that supports the statistical production process from the identification of user needs through to the dissemination of statistical products.
4. GSIM is aligned with relevant standards, such as DDI and SDMX, but it is not directly tied to them, neither to specific technologies nor other concrete implementation details. GSIM provides common semantics that can be used unambiguously across and between different implementations.

# Purpose of GSIM

1. Across the world statistical agencies undertake similar activities albeit with variation in the processes each uses. Each of these activities largely consume and produce the same information (for example all agencies use classifications, create datasets and publish products). Although the information used by statistical agencies is at its core the same, all agencies tend to describe this information slightly differently (and often in different ways within each agency). There is no common means to describe the information we use. This makes it difficult to communicate clearly within and between statistical agencies and without this there is no foundation for in-depth collaboration and nor greater standardisation and the sharing of tools and methods.
2. GSIM defines the information that supports all statistical business processes (especially as defined in the GSBPM) with the following purposes:
* Provide definitions of the information objects supporting all statistical business processes, especially those defined in the GSBPM.
* Improve communication between different disciplines involved in statistical production, within and across statistical organisations; and between users, producers and providers of official statistics.
* Generate economies of scale by enabling intra and inter-agency collaboration, especially through reuse of information, methods or technology.
* Enable configurable, rule-based and modular ways of producing statistics, thus minimizing human intervention in the production process.
* Provide a basis for flexibility and innovation, including support for the easy deployment of new statistical products and the adoption of new types of statistical data sources.
1. GSIM is one of the foundations for modernizing statistical production and moving away from traditional silos. By defining and grouping objects common to all statistical production, regardless of subject matter, GSIM enables statistical organizations to rethink how their business could be organized to generate economies of scale.
2. GSIM 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, GSIM helps to explain significant relationships among the entities involved in statistical production, and supports the development of consistent standards or specifications.
3. A model alone cannot transform an organisation or its processes, but GSIM is modelled to allow for innovative approaches to statistical production to the greatest extent possible; for example, in the area of dissemination, where demands for agility and innovation are increasing. At the same time, GSIM supports more traditional approaches of producing statistics.

# Scope of GSIM

1. GSIM provides the information object framework supporting all statistical production processes as described in the GSBPM, giving the information objects agreed names, defining them, specifying essential properties, and indicating their relationships with other information objects. It does not, however, make assumptions about the standards or technologies used in implementation.
2. The information objects defined include those to allow the specification and introduction of new data sources for more innovative data collection, and also the generation of new statistical products.
3. GSIM does not include models of the information objects supporting business functions within an organisation such as human resources, finance, or legal functions, except to the extent that this information is used directly in statistical production.

# Overview of GSIM v0.4

1. GSIM v0.4 is structured on three levels:
2. **Group level** –Figure 1 shows the four groups which comprise the highest level of the model.



***Figure 1. The Group Level of GSIM v0.4***

* The Activity group contains sets of information objects required to manage the programmes that make up statistical production. This includes Acquisition, Statistical and Dissemination activities.
* The Production group contains sets of information objects that describe the processes, methods and rules that are used in statistical production. Included in this group are Process Control, Process Components and Rules.
* The Conceptual group contains sets of information objects that describe the concepts used and their practical implementation, allowing users to understand what the statistics are measuring. It includes Population Unit, Variable and Classifications.
* The Information group contains sets of information objects that describe the results of the stages of statistical production. This includes Dataset, Products and Services.
1. **Set level** – This level shows the sets of related information objects within each Group.



***Figure 2. The Set Level of GSIM v0.4***

1. **Object level** – In this level the specific information objects that are contained in the sets displayed in Figure 2 are detailed, including definitions, attributes and relationships.
2. Further details about the Set and Object levels of GSIM can be found in the GSIM Communication Layer documents.

# Benefits of GSIM

1. Statistical organisations are confronted with shrinking budgets and pressure to respond to increasing information needs. The current approach to producing statistics is characterised by stove-pipe (or silo) production processes, both at the national and at the international level. Limited integration of processes leads to inefficiencies, both within statistical organisations, and in the international statistical community.
2. Opportunities for common development and sharing of tools, methods and processes are largely unexplored. Although statistical organisations have the experience and methodology to deal with the data deluge created by growing demands and advances in information technology, they do not have the resources to fully explore new possibilities. Statistical production still requires a great deal of manual intervention, which is not only resource intensive, but introduces the potential for human error.

## A common language

1. A significant advantage of using GSIM is to improve communication at different levels:
* between the different roles in statistical production (statisticians, methodologists and information technology experts);
* between the statistical subject matter domains;
* between statistical organisations at national and international levels.
1. Improving communication will result in a more efficient exchange of data and metadata within and between statistical organisations, and also with external clients and suppliers.

## Supporting tradition and innovation

1. As a common reference framework for information objects, GSIM will support current production processes and facilitate the modernisation of statistical production. Implementation of GSIM, in combination with GSBPM, will lead to more important advantages. GSIM will:
* create an environment prepared for reuse and sharing of methods, components and processes;
* provide the opportunity to implement rule based process control, thus minimizing human intervention in the production process;
* generate economies of scale through development of common tools by the community of statistical organisations.
1. At a strategic level, GSIM could be used to direct future investment towards areas of statistical production where the common need is greatest. It could also lead to some degree of specialisation within the international statistical community. For example, some organisations could specialize in seasonal adjustment, time series analysis or data validation, and other organisations could take advantage of this expertise.

## Strategic vision for the future of statistics

1. GSIM forms a crucial part of the strategic vision for the future of statistics production[[1]](#footnote-1) and statistical organizations began to collaborate on GSIM’s development in late 2010. As many statistical organizations are currently planning or investing in modernization programs, they have identified the development of GSIM as a key priority.
2. Conceptual models relating to statistical processes and information - GSBPM and GSIM - form two cornerstones of the vision for the future of statistics. Both must be in place for this vision to be realized. GSBPM already exists and GSIM is now an urgent priority for the statistical community.



***Figure 3. The HLG BAS diagram on modernizing statistics***

# Information versus processes: GSIM and GSBPM

1. GSIM and GSBPM are complementary models for the production and management of statistical information. GSIM focuses on the **information objects** used and/or produced in a **statistical business process**. The GSBPM, as a model of the **statistical business process**, identifies the activities undertaken by producers of official statistics that result in **information outputs**.

GSBPM is a business process model

GSIM models information about processes and their inputs and outputs



***Figure 4. Relationship between GSIM and GSBPM***

1. Greater value will be obtained from GSIM if it is applied in conjunction with GSBPM. Likewise, greater value will be obtained from GSBPM if it is applied in conjunction with GSIM. For example, harmonization of metadata is necessary in order to achieve standardization of production processes. Nevertheless, just as GSBPM has been applied to date without GSIM, it is possible (although usually less than ideal) to apply GSIM without GSBPM. For example, an agency may currently be using a local variation on GSBPM to model their statistical business processes, rather than using GSBPM itself. This decision in regard to modelling statistical business processes should not necessarily prevent them deciding to apply GSIM as a reference framework for statistical information.
2. In the context of GSBPM, GSIM can be harnessed as a tool to help describe and define the interrelated set of sub-processes within a statistical business process and the types of information used in those processes to produce official statistics.
3. GSIM may be used directly and implemented through systems or it may be used as a reference framework to which agencies can map their existing information models and communicate with other agencies using a common language.

# How Agencies could use GSIM

1. In the shorter term GSIM can be used by agencies to:
* Build capability among staff by using GSIM as a teaching aid that provides a simple easy to understand view of complex information and clear definitions
* Validate existing information systems and compare with emerging international best practice and where appropriate leverage off international expertise
* Guide development or update of local or international standards to ensure they meet the broadest needs of the international statistical community
1. It is intended that GSIM may be used by agencies to the degree that they find it of use. It may be used in some cases only as a model to which agencies refer when communicating with other agencies to clarify discussion, in other cases an agency may choose to implement the final GSIM as the information model that defines the agency’s operating environment. The modular nature of GSIM means that it is also possible for an agency to choose to use part but not all of GSIM (e.g. information described by the Production group but not that described by the Conceptual group). All these scenarios for the use of GSIM are valid, although those agencies that make use of GSIM to its fullest extent may expect to realise the greatest benefits.
1. Strategic Vision of the High-level Group for Strategic Developments in Business Architecture in Statistics (HLG-BAS). Available at [www1.unece.org/stat/platform/display/hlgbas/Strategic+vision+of+the+High-level+group+for+strategic+developments+in+business+architecture+in+statistics](http://www1.unece.org/stat/platform/display/hlgbas/Strategic%2Bvision%2Bof%2Bthe%2BHigh-level%2Bgroup%2Bfor%2Bstrategic%2Bdevelopments%2Bin%2Bbusiness%2Barchitecture%2Bin%2Bstatistics). [↑](#footnote-ref-1)