

# Generic Statistical Information Model (GSIM): User Guide

(Version 1.0, December 2012)

## About this document

The document contains information to help users understand what a conceptual model is. It provides guides to help understand GSIM, explains how GSIM can be used as a communication tool and describes the steps to be taken to implement GSIM at a business and technical level. The document also includes information on how some early adopters of GSIM intend to use the model.

This document should be read in conjunction with either the GSIM Communication paper or the GSIM Specification.



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.

**Table of Contents**

Introduction..... 3

A conceptual model ..... 3

Understanding GSIM..... 4

GSIM as a communication tool ..... 6

Implementing GSIM at a business level ..... 9

Implementing GSIM at a technical level ..... 11

Examples of Implementing GSIM ..... 11

    GSIM and Statistics New Zealand..... 12

    GSIM and the Australian Bureau of Statistics ..... 14

    GSIM and the International Monetary Fund Statistics Department ..... 16

Annex A: Scenarios ..... 17

    Scenario 1: GSIM and metadata management..... 17

    Scenario 2: GSIM and the acquisition of data ..... 21

    Scenario 3: GSIM and sample selection and estimation..... 25

    Scenario 4: GSIM and dissemination of statistical information ..... 30

    Scenario 5: GSIM and quality..... 33

Annex B: Template for Case Study ..... 36

## **Introduction**

1. GSIM V1.0 is the first internationally endorsed reference framework for statistical information. This over-arching framework will play an important part in modernizing, streamlining and aligning the production and standards associated with official statistics at both national and international levels.
2. GSIM V1.0 is the result of extensive, multi-disciplinary development and consultation across the international statistical community. Over time it is expected that GSIM will be refined as a result of implementation experience and further development.
3. It is expected that statistical organizations will progressively adopt and implement this reference framework and the ‘common language’ it provides. It is intended that GSIM may be used by organizations to different degrees.
4. This User Guide is part of the first public release of GSIM. The document contains information to help users understand what a conceptual model is. It provides guides to help understand the model, explains how GSIM can be used as a communication tool and describes the steps to be taken to implement GSIM at a business and technical level. The document also includes information on how some early adopters of GSIM intend to use the model.

## **A conceptual model**

5. GSIM is a conceptual model. Being a conceptual model, the focus of GSIM is on high-level concepts and the relationships between them, and not on implementation details.
6. By design, GSIM does not refer to any specific IT setting or tool. Statistical organizations use a wide range of in-house and proprietary hardware and software platforms; this environment also changes over time. GSIM is designed to be platform-independent in order to be relevant to all stakeholders and robust over time.
7. Across the world statistical organizations undertake similar activities albeit with variation in the processes each uses. Each of these activities use and produce similar information (for example all organizations use classifications, create data sets and publish products). Although the information used by statistical organizations is at its core the same, all organizations tend to describe this information slightly differently (and often in different ways within each organization). There has been no common way to describe the information we use.
8. GSIM defines and describes the pieces of information (called information objects) that are important to statistical organizations. It also gives users the relationships between the information objects. By describing statistical information in a consistent way, statistical organizations become able to communicate unequivocally and to collaborate more closely (both at a national and international level).

9. A conceptual model such as GSIM cannot be implemented directly. It requires a further level of detail. The sections on Implementing GSIM at a business level and Implementing GSIM at a technical level provide more detail on how this might be done.

## Understanding GSIM

10. The documentation of GSIM v1.0 has a number of layers. Each layer provides more detail on the model and is targeted at different audiences. There is:

- A number of brochures which provide a high level introduction to GSIM
- A Communication document which gives an overview of the model. It also describes the scope and benefits, what the model means for staff in a statistical organization and relationships to other standards and models.
- A User Guide (this document) which explains how GSIM can be used and implemented.
- A Specification document which provides deeper detail about the model, an explanation of the GSIM Extension Methodology and descriptions of how GSIM relates to other standards and models at a detailed level.
- An Enterprise Architect file which contains the UML descriptions of the model.

11. To help target the efforts of staff in statistical organizations, Table 1 provides some guidance on which documents are most appropriate for different audiences.

*Table 1. Reader Guides*

Audience	Suggested documents
Top & Senior managers	GSIM Brochures
Middle Managers, Subject Matter Statisticians and Methodologists	GSIM Brochure GSIM Communication document GSIM User Guide
Architects, business analysts and metadata specialists	GSIM Communication document GSIM User Guide GSIM Specification (particularly section II A Concepts group, Annex E and Annex G for metadata specialists)
Solution Architects	GSIM Specification GSIM User Guide Enterprise Architect file

12. There are some terms and concepts related to statistical information which are important to a statistical organization which are not represented in GSIM v1.0 as specific information objects. In particular, these include Methodology, Quality and Reference Metadata. The following paragraphs explain why these are not included as separate information objects in GSIM v1.0. If an organization has specific unmet requirements for information relating to methodology, quality or reference metadata in the meantime, the extension mechanism that is provided within GSIM can be used during local implementation to fit those specific requirements.

## *Methodology*

13. Methodology is a crucial consideration in production of official statistics, particularly in terms of determining methods to be used during the statistical business process.
14. Methodology designates the study, science and/or theory of “method”(s). Methods themselves are represented as an information object within GSIM but the science of determining methods is not. Methodology can be seen reflected in several GSIM information objects. For example, the rules and parameters that are defined from a methodological view point. The design of a statistical production process sets out the methods to be followed when the process is run. For example, when running a process such as editing, the *Rules* for the procedures in such a tool are defined by existing best practice methodology. The method used is therefore likely to be fairly specific either to an individual process or to a group of similar processes.
15. "Methodology" is not an information object in itself and therefore not modelled as such by GSIM. GSIM is intended to be a generic model, capable of supporting all current and future methods. An example on how the information in a common process conducted by methodologists (sample selection and estimation) is captured by GSIM can be found in Annex A.

## *Quality*

16. Quality means different things in different settings. There is quality as an organizational aspect, the quality of the processes and the quality of the statistics. While methodology is embedded in the design of the statistical production process, quality is typically linked to the instance (i.e. to production runs) of the process.
17. Quality is relevant at a number of different levels of instances of information objects. For example, as an attribute to an information element (e.g. quality flag), as an attribute to a data set (e.g. status provisional data, final data, revised data). It also appears as process quality information. The product quality is laid down in a quality report, which is itself also a statistical product. It is not appropriate for quality to be represented within GSIM as a single information object, distinct from all other information objects. Various facets/aspects of quality are, instead, described, assessed and/or managed using various information objects within GSIM.
18. Quality reports traditionally mainly refer to the quality of the statistics. Quality information and quality reports can be tied to the production process as a whole and/or to parts of it. Quality is present in the inputs and outputs of process steps in the Generic Statistical Business Process Model (GSBPM), acting as process data to control and define rules for processes. The outputs from a GSBPM process step using GSIM information objects can be used as quality measures (in the form of Process Metrics).
19. Quality related to statistics can have many forms depending on its purpose. Depending on the scope, it will refer to different information objects in relation to relevant processes.
20. An example on how quality information is captured by GSIM can be found in Annex A.

## *Reference metadata*

21. There is currently no globally agreed definition of the scope of reference metadata, and GSIM should remain generic rather than linked to one specific definition.
22. Conceptual metadata are represented in GSIM in the Concepts Group, methodological and procedural aspects of metadata are represented by GSIM Production Group. As described above, various information objects from GSIM can be referenced when describing the quality of statistics.
23. GSIM models connections between data and its associated metadata. For example, from a *Data Set* through its *Data Structure Definition*, relevant information from the Concepts Group (for example about *Variables*, *Classifications*, *Populations*, *Concepts*) can be discovered. If the data were being represented in SDMX then the conceptual information (and process information and quality information) relevant to that data would be presented as Reference Metadata.
24. When defining and managing conceptual content in its own right, such as negotiating the definition of a new standard classification, the content isn't typically referred to as "Reference Metadata". In that case the conceptual metadata isn't being considered in the context of describing the contents of a specific *Data Set*.
25. If the term "Reference Metadata" is considered more generally as the ability to refer from the definition of a particular information object (typically data) to other information which is relevant when considering that object then aspects of metadata may be modelled by means of object attributes. The extension mechanism within GSIM can be used to define additional attributes for information objects if these are required (including use of the extension mechanism).

## **GSIM as a communication tool**

26. A major barrier to effective collaboration within and between statistical organizations has been the lack of common terminology.
27. A 'survey' in Statistics Canada, for example, is a 'survey instance' according to the UNECE, a 'collection cycle' according to Australian Bureau of Statistics, and a 'study' according to the external research community. These examples are just the tip of the iceberg.
28. This has made it difficult to communicate clearly within and between statistical organizations and without a common statistical language, there is no foundation for in-depth collaboration, standardization, or sharing of tools and methods.
29. GSIM, as the unifying 'common language' for official statistics, will enable rationalization, cooperation and collaboration within and between statistical organizations.

<b>Box 1. Other industries use information models</b>
---

The Official Statistics industry is not alone in recognizing the need for a ‘common language’ to underpin standards-based modernization.

Travellers, for example, can make their own arrangements to travel anywhere in the world by selecting and booking available flights. Their travel is enabled by a standardized aviation information model, which support streamlined flight planning and in-flight navigation. These models also support air-traffic control to reliably handle increased traffic, reduce fuel use and coordinate on-time arrivals and on-ground services. These standardized information models have promoted the adoption of an industry-standard ‘common language’.

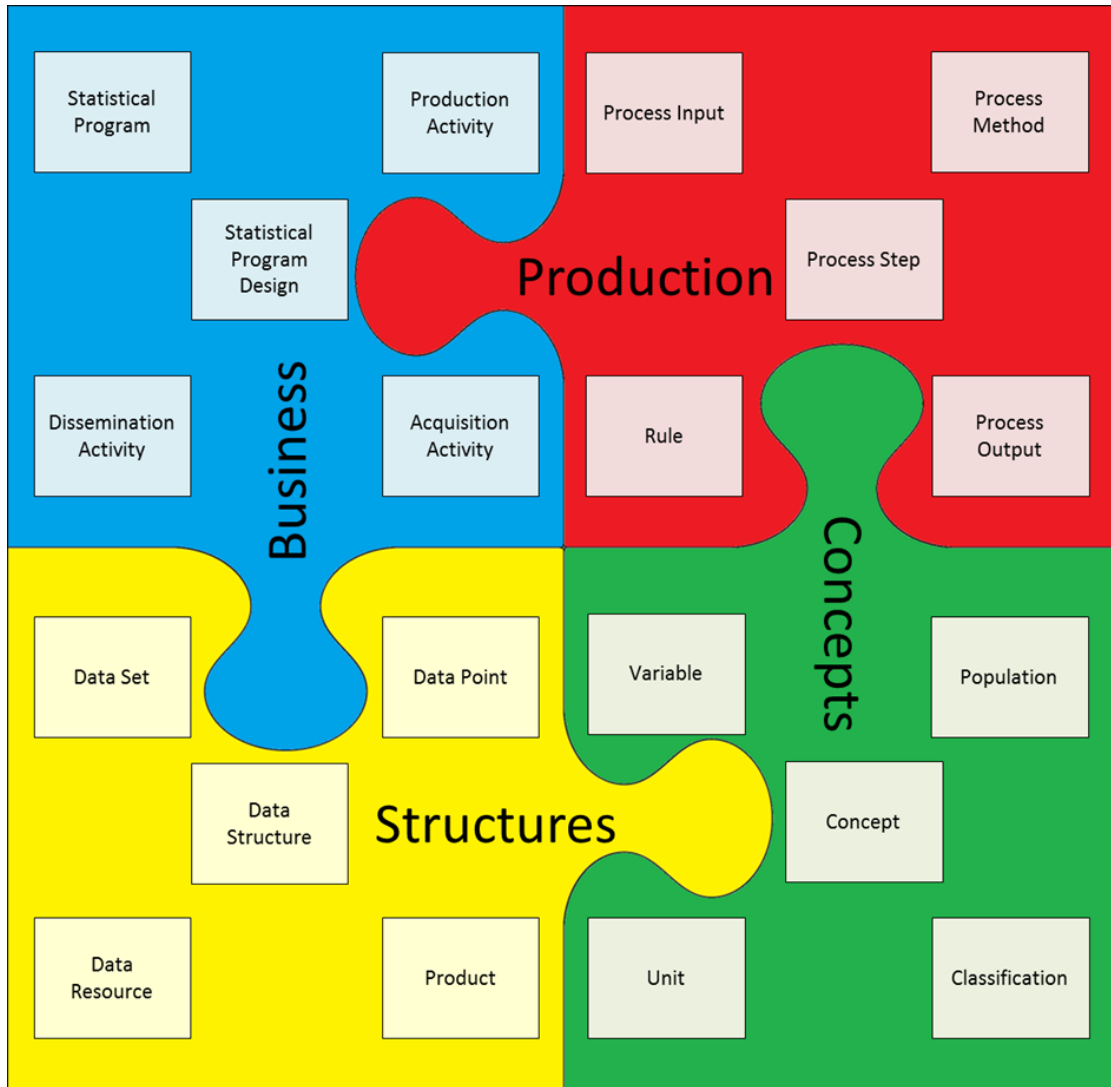
GSIM, as an information model, provides the equivalent common, unifying language for the official statistics industry.

30. GSIM provides staff with simple, easy to understand views of complex information. By describing statistical information in a consistent way, statistical organizations become able to communicate unequivocally and collaborate more closely.

31. Information can be a vague concept. Often when staff are asked to describe the information that is the input to and output from a statistical process, this can be a difficult task. It is difficult because most people don’t know where to start.

32. GSIM can be used to educate staff. It gives them a framework for thinking about information. It provides a common terminology to describe the processes and their inputs and outputs that are used to generate official statistics. At the simplest level, staff could look at Figure 2 (below) and start to identify whether these information objects are relevant to them.

33. The layered documentation of GSIM means that staff can find the appropriate level of detail for them. For some, this level is an overview of the information objects. Others will want to see all the information objects, their attributes and relationships.



*Figure 2. Simplified view of GSIM information objects*

34. By providing these simple views on statistical information, GSIM can be used as a way to help staff understand what technical standards such as Data Documentation Initiative (DDI) and Statistical Data and Metadata eXchange (SDMX) describe. If staff can relate these standards to more familiar terms, they become easier to understand and use. For example, knowing that in DDI a ‘Universe’ is the same as the *Population* information object in GSIM helps.

35. Using GSIM as a common language will increase the ability to compare within and between statistical organizations. All processes that lead to the production of statistics can be described in this one integrated model. This includes the analysis of business needs, the establishment of statistical programs, the development and management of statistical methods, the design of production processes and their cyclical execution.



36. For example, GSIM can be used by staff (at all levels) to share and compare the concepts used in their work. It is also agnostic of subject matter, so it can be used to compare statistical production across subject matter departments.

## **Implementing GSIM at a business level**

37. GSIM can either be mapped to an existing information model or adopted as is by a statistical organization. The steps to adopting GSIM are outlined in the following two sections. Please see the GSIM Specification for further details about the information objects mentioned in these sections.

### *A statistical organization has an existing information model*

38. It should be a straightforward task to map an existing information model onto GSIM. The order in which the mapping is undertaken could be dependent on the orientation of the existing information model.

39. If the information model focuses primarily on metadata, start with the Concepts group of the GSIM. This group comprises information objects such as *Variable*, *Classification*, *Value Domain*.

40. If the information model stresses mainly the management of data sets, the point of departure could be the Structures group. This is where notions such as *Data Set* (concerning data) and *Data Structure* (concerning metadata) can be found

41. The information model could be principally oriented on the design and execution of statistical processes. In this case, start the mapping with the Production group. It offers concepts such as *Process Step Design*, *Input Specification*, *Output Specification*, and provides a link to policy and method information through a *Design Context*.

42. It might, however, be preferable to begin with information objects referring to the main activities of a Statistical Organization. In this case, start by focusing attention on the Business group, which offers information objects referring to the *Statistical Program*, the *Acquisition Activity* and the *Dissemination Activity*. It is also in this group that concepts related to high-level management of the statistical process, such as *Statistical Need*, *Assessment*, *Business Case* are found.

43. These four groups are interrelated through relations linking information objects across the group borders, so you will easily find a path from one group to the other.

### *A statistical organization adopts GSIM as its information model*

44. Most statistical organizations have some information models distributed over one or more repositories (catalogues, data bases, etc.) to manage statistical methods, statistical metadata, architectural principles, policy provisions and similar things. In many cases these information model may be implicit rather than explicit. Typically these models will represent a subset of the

information objects in GSIM – containing just those which are relevant to the purpose of the particular model.

45. A statistical organization may choose to:

- adopt one of the existing models as organization's preferred information model, map this model to GSIM and adopt the sections of GSIM which are needed to address any gaps in coverage or
- adopt GSIM to bring all this information into one consistent model.

46. The order in which the different bits of information are brought together under the GSIM will be dependent on the relative importance of the collections of information.

### *GSIM and GSBPM*

47. Although GSIM can be used independently, it has been designed to work in conjunction with the GSBPM. It supports GSBPM and covers the whole statistical process. It is assumed in this section that an organization either uses GSBPM or uses another business process model (which can be mapped to GSBPM).

48. Adopting GSIM at a business level involves an analysis of the information being used, managed and processed when designing and producing statistics. When designing a new process or redesigning an existing process, the process should be mapped to GSBPM and the information objects should be mapped to GSIM. In Annex A, there are a number of examples of how this can be done.

49. This work is useful, because GSIM makes business processes and methods “visible”, where other design approaches keep them buried in application code and documentation. This opens up a range of technical possibilities. For example:

- Comparing IT solutions
- Sharing IT solutions across subject matter domains, or even between statistical organizations

50. After undertaking this exercise, it is possible that there will be information objects that your organization needs to describe, but that are not accounted for in GSIM.

51. Most organizations have legacy systems and administrative practices that will require an extension of GSIM to meet organization-specific implementation needs. In particular, processes relating to corporate management are outside the scope of both GSIM and GSBPM.

52. GSIM is robust, but can readily be adapted and extended to meet users' needs. In order to implement GSIM, you will need to identify the organization-specific information that needs to be integrated into your own extension of GSIM. Examples include preferred platform and standards, standard documents to be produced when developing a new statistical program, etc.

53. In order to extend GSIM usefully, it is important to use the mechanism provided within GSIM, and to document every extension carefully. The quality of this documentation is fundamental for a successful use of the extensions for communication between all participants in the activities of the organization. Moreover, extensions to GSIM are not for internal use only. They should be submitted to the UNECE Standards Steering Group ([support.stat@unece.org](mailto:support.stat@unece.org)), which will keep a record of existing GSIM extensions. Some of them might well be approved as something that should be added to the agreed model.

## **Implementing GSIM at a technical level**

54. In order to effectively implement GSIM at a technical level, it should be implemented at a business level first.

55. An organization may choose to implement GSIM as the information model that defines their operating environment. In order to implement GSIM into a system, it is necessary to map it to an organization's implementation models. Preliminary mappings to SDMX, DDI and other standards are given in GSIM v1.0 Specification. This more technical use of GSIM requires an analysis of gaps between GSIM and IT implementation systems at a detailed level.

56. The design of GSIM takes into account the possibility to map to implementation models, such as SDMX or DDI. Such a mapping can be used to establish a link between GSIM and its technical implementation. GSIM thus helps to:

- Compare IT approaches;
- Detect double work in legacy systems
- Avoid double work in new systems.

## **Examples of Implementing GSIM**

57. GSIM v1.0 is the first public release of GSIM. As such, statistical organizations have not yet had the opportunity to implement it. However, a number of organizations have started to plan how they will use GSIM.

58. This section contains short statements from some of these organizations on how they plan to use GSIM. At the moment, these statements represent intentions, not results. As GSIM is used by more organizations, there will be further examples of its use.

59. Annex B provides a template which statistical organizations are encouraged to complete regarding local information models and the use of GSIM.

## **GSIM and Statistics New Zealand**

60. Developing GSIM has been a large, collaborative effort from agencies across the globe. The aim was to deliver a product that could be used to benefit the international statistical community, as well as individual agencies, as we seek to modernize the production of official statistics.

61. Statistics New Zealand has identified three initial uses for GSIM. These do not depend on the shared adoption of GSIM. They represent the value that GSIM offers individual agencies. Using GSIM allows an agency to benefit from the collective expertise of the international community, to ensure that terminology, processes, and systems design fit with best practice. In future there will be collective benefits, but in the meantime GSIM can be put to practical use to help Statistics New Zealand transform the way it delivers official statistics.

### *To collaborate within the organization and beyond*

62. GSIM will inform our terminology as we work towards standardisation across the organization. Using the same language will help our teams, including IT and statistical professionals, to collaborate better. We will use GSIM to help new staff understand the basic concepts in the production of official statistics. This will help build our staff capability in understanding the information needs of the statistical production process, and ensure they can more easily rotate around roles in the organization.

63. Using a common model and terminology between agencies will be a strong foundation for international collaboration. The time currently required for projects to get up and running hampers the collaboration effort and reduces the time we spend working towards the project goals. A common model will reduce any miscommunication and misunderstanding.

64. Reducing the barriers to working with other agencies will ensure we better align our thinking with others, and allow us to consolidate our joint vision for the future of statistics.

### *To underpin and understand existing standards*

65. GSIM provides a complete picture of the information that underpins the statistical production process. Existing international standards provide only a partial picture and do not meet the specific needs of statistical agencies. Use of standards such as DDI, SDMX and others such as geospatial standards will be enhanced when looking through the lens of GSIM. We will use the model to support our understanding and help us to determine where these standards meet our needs, where gaps exist, and where future development should be focused.

66. As a reference model, GSIM will help to determine how existing standards complement each other. For example, contributing to analysis of where needs can be best met by using DDI, areas where needs can be best met by use of SDMX, and areas that can be best met by using other standards.

67. GSIM will support the need to develop guidelines and standards that are independent of specific implementation standards. The terminology and definitions in GSIM are meaningful and familiar to statisticians, and a single model covering the entire statistical production process will enable a more consistent overall view of our data and information.

*To inform system design*

68. We will use GSIM to inform information model designs behind our systems and as a benchmark and starting point for design. It will help provide assurance that our own systems are developed to a high standard.

69. Using the model will improve the efficiency of our development projects by avoiding re-analysing our information needs for each and every project. For example, we are already using the model to inform our development of metadata management systems. As we develop our new classification management system we will aim to be aligned with GSIM.

70. With over 100 projects currently underway to transform the organization, GSIM has a huge potential to improve our processes. Using GSIM during the development of several new systems will be an opportunity to identify successes and areas for improvement.

71. GSIM will help us identify our most high-value and critical data so we can appropriately resource them.

*To lay the foundations for the future*

72. Many of the initial uses for GSIM we're planning are based on the assumption that at present the model is not yet mature enough, and without sufficient detail, to implement into systems. GSIM is a conceptual model.

73. We hope that as GSIM is developed in the coming years, the information model will be able to be directly implemented and systems built upon it. GSIM will begin to be implemented into our statistical infrastructure systems to enable standardized information exchange across Statistics New Zealand.

74. In the future we will be able to reduce the time it takes to develop new systems – the base information model will already exist – and where other agencies have developed systems based on the same model, these will be able to be repurposed for our own needs.

75. We hope to build our capability by enabling our use of solutions from other statistical agencies and contribute to enable others to benefit from the flexibility of our own new solutions.

## **GSIM and the Australian Bureau of Statistics**

### *ABS 2017*

76. The Australian Bureau of Statistics (ABS) considers GSIM to be a key enabler of practical collaboration and sharing across a number of levels. This includes at the international level, and also for guiding and facilitating organization programs such as the ABS2017 transformation program.

77. Announced in February 2010, the ABS2017 program aims to transform the way the ABS collects, collates, manages, uses, reuses and disseminates statistical information. This will improve the usability, value and timeliness of statistics for government and the community. The approach to adopting and applying GSIM as a common reference framework across the ABS is expected to have many analogies with the approach used with GSBPM<sup>1</sup>.

78. GSIM includes a level of detail (the Specification Layer) which is not present in GSBPM. This will support a number of uses of GSIM such as the ability to map GSIM to existing implementation standards like DDI and SDMX.

79. It is expected that GSIM will continue to evolve. ABS will take into account the fact GSIM isn't yet as mature as GSBPM. It will take into consideration the extension guidelines defined in GSIM and determine the impact of GSIM on ABS business requirements.

80. A small task team has been established within ABS 2017 to plan and co-ordinate adoption of GSIM in the first half of 2013.

### *Enterprise Architecture*

81. The task team will formalize the inclusion of GSIM as a reference framework in the ABS Enterprise Architecture. GSIM will be a core artifact within Information Architecture but also influence the description of Business and Applications Architecture.

### *Using GSIM at a business level*

82. The task team will gather feedback from various ABS initiatives which seek to apply GSIM as a reference framework.

83. One such initiative is currently reviewing the diversity of terms and concepts currently used by subject matter statisticians across the ABS to refer to Statistical Metadata. The project aims to harmonize the terminology used within the ABS in future.

---

1 GSBPM is used as part of the ABS Enterprise Architecture (EA). It provides a common point of reference when planning, specifying, managing and coordinating individual projects and initiatives across the ABS. GSBPM is used as a common set of terminology and common basis for categorization of statistical production activities.

84. Wherever possible the reference terms to be used in future will be drawn directly from GSIM. Where there is an agreed ABS business need to diverge, the relationship between ABS usage and GSIM usage will be documented. Some, but perhaps not all, of these divergences may result in a proposal to update GSIM terms and/or definitions in the next version.

#### *Implementing GSIM in the ABS*

85. GSIM is also interlinked with the ABS' development of metadata infrastructure, in particular the Metadata Registry and Repository (MRR) and the ABS Transitional Metadata Model (ATMM).

86. The Metadata Registry and Repository is a core part of the ABS infrastructure, and will make capturing and reusing metadata much easier and more efficient. The MRR is made up of two parts:

- a repository that will store statistical information (including metadata, data, process definitions, etc).
- a registry or catalogue that will allow us to register metadata, easily search for that metadata, and discover and retrieve metadata held in the 'store' part of the MRR.

87. The ATMM is the model that underlies the MRR. It defines the way statistical information is registered for discovery. The model combines the influences of a number of international standards (including GSIM) with the core requirements of ABS processes and gives a core set of information objects for use across the ABS.

88. The ATMM could be described as the 'operationalization' of GSIM (or the ABS specific version of GSIM). It has enabled the ABS to bring forward the development of the MRR, and to provide valuable input into the GSIM development process – experts working on the ATMM and MRR projects have contributed to GSIM development.

89. As GSIM itself has become more fully developed over time (particularly from v0.8 onwards), it has begun to have a greater impact on the continuing development of the ATMM, further bolstered by the learnings and experiences of the team members who have directly worked on GSIM.

90. Using GSIM together with other implementation standards to create the ATMM and MRR is just one example of local implementation, but it is proving to be an important one.

## **GSIM and the International Monetary Fund Statistics Department**

91. The IMF Statistics Department intends to use GSIM (and GSBPM) at the point in time when we are looking to upgrade/re-engineer existing processes in the context of the Department's on-going "Streamline, Standardize, Automate" initiative. These opportunities will be used to build a holistic view of the end-to-end statistical process.

92. The first activity will identify the sub-processes and information objects used by the IMF in specific instances of the statistical process.

1. Create an "as is" inventory of sub-processes.
2. Identify core characteristics of each sub-process, such as description and owner.
3. For each sub-process, identify its inputs and outputs.
4. Match each IMF sub-process to a sub-process in the GSBPM.
5. Match each input and output to an information object in the GSIM.

93. The second activity will utilize GSBPM and GSIM to look for opportunities to improve the effectiveness and efficiency of the IMF's work. Analysis that we may undertake to identify such opportunities may include:

1. Studying organizational distribution of activities where IMF sub-processes occur in different work areas, but relate to the same GSBPM process.
2. Investigating situations where the same GSIM information objects are used as inputs and outputs to different IMF sub-processes, to identify potential areas for standardization and automation.
3. Exploring the extent to which existing standards and technologies (e.g. SDMX) can be leveraged to provide the basis for such standardization and automation.

94. These two activities will be conducted iteratively, starting off with a pilot of the sub-processes being worked on as part of an existing automation project, with the aim of organically spreading to a wider range of sub-processes as further automation projects are initiated. This will allow the "bottom-up" description of existing work processes and information objects (the "as is" inventory) to be analyzed within the framework of the "top-down" generic models. At the same time, the understanding and use of the generic business process and information models can be gradually widened and strengthened across the Department.



## Annex A: Scenarios

95. The examples offered in this annex are not taken from concrete implementations of GSIM (as these don't currently exist). They have been devised with a teaching goal in mind. They illustrate how some of the main activities of a statistical organization can be described and managed by using GSIM.

96. Each of the scenarios in this section are common for statistical organizations. They are described first in simple everyday language that will be familiar to a large number of staff in statistical organizations. The scenarios are then described, for a second time, in terms of the information objects in GSIM and the sub-processes in GSBPM. The following scenarios are included:

- Metadata management
- Acquiring data
- Sample selection and estimation
- Disseminating data
- Quality

### Scenario 1: GSIM and metadata management

#### *Managing definitions of variables*

97. Each organization will have its own specific set of processes to manage definitions of variables, but the simplest and most common process would be to update a variable for an established statistic and a more complicated, but still common, process would be to create a new variable for a new statistic. The process of updating or creating variables would be carried out within the specify needs and design phases of the established or new statistic.

98. Other processes are conceivable, and supported by GSIM, such as creating a new variable for an established statistic, reusing existing variables from one statistic in another new or established statistic, but these processes will not be further detailed in this scenario.

99. Part 1: Updating a variable for an established statistic

- a) Update variable – Change the validity period of the outdated variable to reflect the fact that it is no longer valid. Copy the outdated variable and change the definition and validity period for the updated variable. Remember to also update documentation of variables that will be derived from this one. Connect the updated variable to relevant concepts (populations, categories {e.g. male, female and other}, other variables) for the established statistic, if these connections were not copied from the outdated variable.
- b) Update/create or re-use value domain – If necessary update/create the value domain (allowed values of codes [e.g. m, f, o] for categories) for the updated variable and connect the updated variable to the updated/created value domain. Otherwise connect the updated variable to the existing value domain.

- c) Change or re-use data source – check whether the value of the updated variable can still be obtained from the same data source. If not, then update the collection method.
- d) Update/create or re-use question – if the value of the updated variable is to be obtained from a question in a questionnaire, then check whether an existing question can be re-used, possibly updated or one or more new questions need to be designed. Connect the appropriate question(s) to the updated variable.
- e) Update the design of statistical outputs – Connect the appropriate unit/table/cube designs to the updated variables.

100. Part 2: Creating a new variable for a new statistic

- a) Create variable - Document definition of new variable. Remember to also document variables that will be derived from this one.
- b) Identify concepts – check whether appropriate concepts (populations and categories) are already documented in a concept management system. Document new concepts for the new statistic, if necessary. Connect the variables to the relevant concepts.
- c) Identify value domain - check whether an appropriate value domain is already documented in a value domain management system. Document a new, re-use or update an existing value domain for the new variable. Connect the variable to the value domain.
- d) Identify data source – check whether the value of the new variable can be obtained from an existing data source or whether it needs to be collected e.g. using a questionnaire.
- e) Identify question – if the value of the new variable is to be obtained from a questionnaire, then check whether a relevant question/question group is already documented in a question bank and/or questionnaire. Document a new question/question group, if necessary. Connect the question to the new variable.
- f) Designing statistical products – Connect the unit/table/cube designs to the new variables.

*Mapping of example description to GSIM information objects*

101. GSIM can be used by those people involved in metadata management to identify the pieces of information that they require to undertake their roles and in the design of systems for metadata management.

102. Part 1: Updating a variable for an established statistic

- a) Update Variable – Change the validity period of the outdated *Variable* to reflect that it is no longer valid. Copy the outdated *Variable* and change the definition and validity period for the updated *Variable*. Remember to also update documentation of *Variables* that will be derived

from this one. Connect the updated *Variable* to the relevant *Concepts* (*Population*, *Categories* and other *Variables*) for the established statistic, if these connections were not already copied from the outdated *Variable*.

- b) Update/create or re-use value domain – If necessary update/create the *Value Domain*. The *Represented Variable* is then the association of the updated *Variable* with the updated/created or re-used *Value Domain*.
- c) Change or re-use Data Set – check whether the value of the updated *Variable* can still be obtained from the same *Data Set*. If not, then update the *Acquisition Design*.
- d) Update/create or re-use Question – if the value of the updated *Variable* is to be obtained from a *Question* in a *Survey Instrument* (questionnaire), then check whether an existing *Question* can be re-used or one or more new *Questions* need to be designed. Connect the appropriate *Question(s)* to the updated *Variable*.
- e) Update the design of *Representations* – Connect the appropriate *Unit/Dimensional Data Structures* (unit/table/cube designs) to the updated *Variables*.

**Table 2. Part 1: Updating a variable for an established statistic**

<b>Activity Steps</b>	<b>Applicable GSBPM sub-process</b>	<b>Applicable GSIM Objects</b>
a) Update Variable (Re-use Population, Category)	2.2 Design variable descriptions	<ul style="list-style-type: none"> <li>• <i>Concept</i></li> <li>• <i>Population</i></li> <li>• <i>Conceptual Domain</i></li> <li>• <i>Variable</i></li> </ul>
b) Update/create or re-use Value Domain	2.2 Design variable descriptions	<ul style="list-style-type: none"> <li>• <i>Represented Variable</i></li> <li>• <i>Value Domain</i></li> </ul>
c) Change or re-use data source	1.5. Check data availability 2.3 Design data collection methodology	<ul style="list-style-type: none"> <li>• <i>Variable</i></li> <li>• <i>Data Set</i></li> <li>• <i>Acquisition Design</i></li> </ul>
d) Update/create or re-use Question	2.3 Design data collection methodology	<ul style="list-style-type: none"> <li>• <i>Variable</i></li> <li>• <i>Question</i></li> <li>• <i>Survey Instrument</i></li> </ul>
e) Update the design of Representations	2.1 Design outputs	<ul style="list-style-type: none"> <li>• <i>Unit Data Structure</i></li> <li>• <i>Dimensional Data Structure</i></li> <li>• <i>Variable</i></li> <li>• <i>Representation</i></li> </ul>

103. Part 2: Creating a new variable for a new statistic

- a) Create Variable - Document definition of new *Variable*. Remember to also document *Variables* that will be derived from this one.
- b) Identify Concepts – check whether these are already documented in a *Concept (Population and Categories)* management system. Document new *Concepts* for the new statistic, if necessary. Connect the *Variables* to the relevant *Concepts*.
- c) Identify Value Domain - check whether this is already documented in a *Value Domain* management system. Document new, re-use or update an existing *Value Domain*. The *Represented Variable* is then the association of the *Variable* with the *Value Domain*.
- d) Identify Data Set – check whether the value of the *Variable* can be obtained from an existing *Data Set* or whether it needs to be collected e.g. using a *Survey Instrument* (questionnaire).
- e) Identify Question – if the value of the *Variable* is to be obtained from a *Survey Instrument*, then check whether a relevant *Question/Multiple Question Item* is already documented in a question bank and/or *Survey Instrument*. Document a new *Question/Multiple Question Item*, if necessary. Connect the *Question* to the new *Variable*.
- f) Design Representations – Connect the new *Unit/Dimensional Data Structure* (unit/table/cube designs) to the new *Variables*.

**Table 3. Part 2: Creating a new variable for a new statistic**

<b>Activity Steps</b>	<b>Applicable GSBPM sub-process</b>	<b>Applicable GSIM Objects</b>
a) Create Variable	2.2 Design variable descriptions	<ul style="list-style-type: none"> <li>• <i>Variable</i></li> </ul>
b) Identify Concepts	1.4. Identify concepts	<ul style="list-style-type: none"> <li>• <i>Concept</i></li> <li>• <i>Population</i></li> <li>• <i>Categories</i></li> </ul>
c) Identify Value Domain	2.2 Design variable descriptions	<ul style="list-style-type: none"> <li>• <i>Value Domain</i></li> <li>• <i>Represented Variable</i></li> </ul>
d) Identify data source	1.5. Check data availability 2.3 Design data collection methodology	<ul style="list-style-type: none"> <li>• <i>Variable</i></li> <li>• <i>Data Set</i></li> <li>• <i>Survey Instrument</i></li> </ul>
e) Identify Question	2.3 Design data collection methodology	<ul style="list-style-type: none"> <li>• <i>Variable</i></li> <li>• <i>Question</i></li> </ul>

		<ul style="list-style-type: none"> <li>• <i>Multiple Question Item</i></li> <li>• <i>Survey Instrument</i></li> </ul>
f) Design Representations	2.1 Design outputs	<ul style="list-style-type: none"> <li>• <i>Unit Data Structure</i></li> <li>• <i>Dimensional Data Structure</i></li> <li>• <i>Variable</i></li> <li>• <i>Representation</i></li> </ul>

**Scenario 2: GSIM and the acquisition of data**

*Acquiring data*

104. The majority of statistical organizations collect data in one form or another. The collection or acquisition of data begins with identifying the need for data and results in the statistical organization having a resource of data to process, analyses and disseminate.

105. Each organization will have its own specific set of processes to collect or acquire data. Generally, this process will consist of these following steps:

106. Part 1: A need for data is identified

- a) Statistical organization determines need for new data
- b) Decide on the concepts that are to be measured
- c) Check what data and sources are already available.
- d) Decide whether the data will be acquired and how

107. Part 2a: An administrative data source is available

- a) An agreement is made with a register owner.
- b) Administrative data are delivered from the register owner
- c) The data is forwarded to an environment for pre-processing

108. Part 2b: A survey is needed to collect the data

- a) Decide the variables that measure concepts and the applicable classifications
- b) Decide on questions to ask and the question-sequence
- c) Build the physical instrument
- d) Collect the data
- e) Finalize the collection

*Mapping of example description to GSIM information objects*

109. GSIM can be used by those people involved in the collection process to identify the pieces of information that they require to undertake their roles and in the design of systems for collection purposes.

110. Part 1: A need for data

- a) Statistical organization determines need for new survey - A statistical organization will determine that there is a new *Statistical Need*. An example of this need might be an unemployment figure. This *Statistical Need* will usually be expressed in terms of a *Subject Field*, like Labour, and a *Population*, like Australian citizens.
- b) Decide on the concept that is to be measured - The statistical organization will need to do conceptual work to establish exactly what it is trying to be measured –to determine exactly what the required *Concepts* are. In the context of an unemployment figure, one of the *Concepts* would be unemployment.
- c) Check what data and sources are already available - The statistical organization will make an *Assessment* of what data is already available to them. This may involve searching the organization’s existing *Data Resources* to check whether relevant *Data Sets* are already held, which could be reused. It could also involve reviewing the organization’s *Provision Agreements* with *Data Providers* to see what administrative data could be accessed.
- d) Decide whether the data will be acquired and how - The *Process Outputs* from the above two processes will result in a *Change Definition* - a formalized statement of how the organization should react to the *Statistical Need*. This *Change Definition* will feed into a *Business Case*. Based on *Assessments* made by the statistical organization, a particular *Acquisition Activity* will be proposed. This may include collecting data using a survey or an administrative source. This activity will be described by a *Collection Description*. If the *Business Case* for the *Statistical Need* is accepted, a *Statistical Program* (for example, a Labour Force Survey) will be initiated.

**Table 4. Part 1: A need for data**

<b>Activity Steps</b>	<b>Applicable GSBPM sub-process</b>	<b>Applicable GSIM Objects</b>
a) Statistical organization determines need for new data	1.1 Determine Needs	<ul style="list-style-type: none"> <li>• <i>Statistical Need</i></li> <li>• <i>Subject Field</i></li> <li>• <i>Population</i></li> </ul>
b) Decide on the	1.4 Identify Concepts	<ul style="list-style-type: none"> <li>• <i>Concept</i></li> </ul>

concept that is to be measured		
c) Check what data and sources are already available	1.5 Check Data Availability	<ul style="list-style-type: none"> <li>• <i>Data Resource</i></li> <li>• <i>Data Sets</i></li> <li>• <i>Provision Agreement</i></li> <li>• <i>Data Provider</i></li> <li>• <i>Assessment</i></li> </ul>
d) Decide whether the data will be acquired and how	1.6 Prepare Business Case	<ul style="list-style-type: none"> <li>• <i>Change Definition</i></li> <li>• <i>Collection Description</i></li> <li>• <i>Acquisition Activity</i></li> <li>• <i>Business Case</i></li> <li>• <i>Statistical Program</i></li> </ul>

111. Part 2a: An administrative data source is available

- a) An agreement is made with a register owner - As part of the *Acquisition Design*, the statistical organization makes a *Provision Agreement* with a *Data Provider* (the owner of the register). The *Provision Agreement* will outline the *Data Location* (where the *Data Set* can be retrieved from) and a *Data Flow*. The *Data Flow* could be a link to a specific *Data Set* file or to a *Business Service* which will consume a query and return a *Data Set*.
- b) Administrative data are delivered from the register owner - In the *Acquisition Activity*, the *Data Provider* will make the *Data Set* available at a specific *Data Location* via a *Data Flow*. The *Data Set* will be structured according to the agreed *Data Structure*.
- c) The data are forwarded to an environment for pre-processing - The *Data Set* is fed into the *Data Resource* for pre-processing. Pre-processing means that *Instance Variables* are created by extraction and derivation from the *Instance Variables* that have been received.

**Table 5. Part 2a: An administrative data source is available**

<b>Activity Steps</b>	<b>Applicable GSBPM sub-process</b>	<b>Applicable GSIM Objects</b>
a) An agreement is made with a register owner	2.3 Data Collection Methodology	<ul style="list-style-type: none"> <li>• <i>Acquisition Design</i></li> <li>• <i>Provision Agreement</i></li> <li>• <i>Data Provider</i></li> <li>• <i>Data Location</i></li> <li>• <i>Data Set</i></li> <li>• <i>Data Flow</i></li> <li>• <i>Business Service</i></li> </ul>

b) Administrative data are delivered from the register owner	4.3 Run Collection	<ul style="list-style-type: none"> <li>• <i>Acquisition Activity</i></li> <li>• <i>Data Provider</i></li> <li>• <i>Data Set</i></li> <li>• <i>Data Location</i></li> <li>• <i>Data Flow</i></li> <li>• <i>Data Structure</i></li> </ul>
c) The data are forwarded to an environment for pre-processing	4.4 Finalize Collection	<ul style="list-style-type: none"> <li>• <i>Data Set</i></li> <li>• <i>Data Resource</i></li> </ul>

112. Part 2b: A survey is needed to collect the data

- a) Decide the variables that measure concepts and the applicable classifications - The statistical organization will need to decide on the *Acquisition Design*. One of the first inputs for this is to define the *Variables* (for example, Unemployment Status) and *Classifications* (for example, industry classification) which will be collected via the *Survey Instrument*.
- b) Decide on questions to ask and the question-sequence - The statistical organization will then design the *Survey Instrument* (e.g. questionnaire for the Labour Force Survey). The design of the *Survey Instrument* will depend of the *Mode(s)* (CATI interview) and the *Data Channel(s)* (phone) that will be used. The *Questions* ('Last week, did you do any work at all in a job, business or farm?') and *Value Domains* (Yes, No) and *Units of Measure* (dollars) used in the response options and designed. The *Questions* will be grouped into *Question Blocks* and the *Control Transition* (flow logic or question sequence) between the *Questions* will be determined.
- c) Build the physical instrument - Once the *Survey Instrument* has described what is to be collected, an *Instrument Implementation* (for example, a Blaise Program) is created.
- d) Collect the data - An *Acquisition Activity* takes place. The *Acquisition Activity* executes a number of processes required to collect the data via the specified *Data Channel* (phone).
- e) Finalize the collection - The collected data is loaded into a *Data Resource* for further processing.

**Table 6. Part 2b: A survey is needed to collect the data**

<b>Activity Steps</b>	<b>Applicable GSBPM sub-process</b>	<b>Applicable GSIM Objects</b>
a) Decide the	2.2 Design Variable Descriptions	<ul style="list-style-type: none"> <li>• <i>Variable</i></li> </ul>



variables that measure concepts and the applicable classification		<ul style="list-style-type: none"> <li>• <i>Classification</i></li> </ul>
b) Decide on questions to ask and the question-sequence	2.3 Data Collection Methodology	<ul style="list-style-type: none"> <li>• <i>Mode</i></li> <li>• <i>Data Channel</i></li> <li>• <i>Question</i></li> <li>• <i>Unit of Measure</i></li> <li>• <i>Value Domains</i></li> <li>• <i>Question Block</i></li> <li>• <i>Control Transition</i></li> <li>• <i>Survey Instrument</i></li> </ul>
c) Build the physical instrument	3.1 Build Data Collection Instrument	<ul style="list-style-type: none"> <li>• <i>Survey Instrument</i></li> <li>• <i>Instrument Implementation</i></li> </ul>
d) Collect the data	4.3 Run Collection	<ul style="list-style-type: none"> <li>• <i>Acquisition Activity</i></li> <li>• <i>Data Channel</i></li> </ul>
e) Finalize the collection	4.4 Finalize Collection	<ul style="list-style-type: none"> <li>• <i>Data Resource</i></li> </ul>

### Scenario 3: GSIM and sample selection and estimation

#### *Sample selection and estimation*

113. The majority of statistical organizations select samples and compute estimates. This example illustrates the application of a few statistical methods, so it is well suited to show how methodology is modelling in GSIM.

114. This example was used in the CORE project (COMmon Reference Environment). Each organization will have its own specific set of processes to do sample selection and estimation. However, generally, this process will consist of these following steps:

#### 115. Part 1: Sampling

- a) Establish the population.
- b) Determine sampling method.
- c) Compute strata statistics
- d) Allocate the sample
- e) Select the sample

116. Part 2: Collection

- a) Collect survey data
- b) Check which methodology to use
- c) Check and correct survey data

117. Part 3: estimation

- a) Check which methodology to use
- b) Calibrate survey data
- c) Compute estimates

*Mapping of example description to GSIM information objects*

118. GSIM can be used by those people involved in the methodological processes to identify the pieces of information that they require to undertake their roles and in the design of systems.

119. A central role is played by the *Design Context*, an information object representing a repository of principles, best practices and proven solutions supporting the production of coherent statistics in a transparent and reproducible way.

120. Part 1: Sampling

- a) Establish the population - the *Target Population* is Banks in the Netherlands
- b) Determine sampling method - access the *Design Context* of sampling, to retrieve a *Process Method* suited for sampling banks. The *Process Method* to be applied is a stratified random sample because the *Population* of banks is skewed on the *Represented Variable* Turnover. This *Process Method* comprises the following three steps (c - e):
- c) Compute strata statistics – establish the *Frame Population* (a subset of the *Target Population* that is available for surveying) and apply a stratification *Rule* to classify it according to the *Instance Variable* Turnover into a number of strata, and compute, for each stratum, the mean and standard deviation of a set of auxiliary *Represented Variables*. Store this information in a *Data Set* whose *Data Structure* specifies a record for each stratum and a set of two *Represented Variables* (Mean and SDev) for every auxiliary *Represented Variable*.
- d) Allocate the sample – apply a *Rule* supplied by the *Process Method* to find the optimal sample allocation across strata. The output of this *Rule* is a *Data Set* whose *Data Structure* specifies a record for each stratum and one *Represented Variable* for the allocation value.

- e) Select the sample – draw a stratified random sample of *Units* from the *Frame Population*, according to the previously computed optimal allocation.

**Table 7. Part 1: Sampling**

<b>Activity Steps</b>	<b>Applicable GSBPM sub-process</b>	<b>Applicable GSIM Objects</b>
a) Establish the population	1.1 Determine needs for information	<ul style="list-style-type: none"> <li>• <i>Target Population</i></li> </ul>
b) Determine sampling method	2.4 Design frame & sample methodology	<ul style="list-style-type: none"> <li>• <i>Design Context</i></li> <li>• <i>Process Method</i></li> <li>• <i>Population</i></li> <li>• <i>Represented Variable</i></li> </ul>
c) Compute strata statistics	2.4 Design frame & sample methodology	<ul style="list-style-type: none"> <li>• <i>Target Population</i></li> <li>• <i>Instance Variable</i></li> <li>• <i>Represented Variable</i></li> <li>• <i>Data Set</i></li> <li>• <i>Data Structure</i></li> </ul>
d) Allocate the sample	2.4 Design frame & sample methodology	<ul style="list-style-type: none"> <li>• <i>Rule</i></li> <li>• <i>Process Method</i></li> <li>• <i>Frame Population</i></li> <li>• <i>Represented Variable</i></li> </ul>
e) Select the sample	4.1 Select sample	<ul style="list-style-type: none"> <li>• <i>Unit</i></li> <li>• <i>Frame Population</i></li> </ul>

121. Part 2: Collection

- a) Collect survey data – approach one *Unit* at a time and apply to it the *Survey Instrument*. The *Process Output* of this step is a *Unit Data Set*, described by a *Unit Data Structure*.
- b) Check which methodology to use - Access the *Design Context* of data editing to retrieve the *Process Methods* applicable to the *Represented Variables* of the *Unit Data Set*.
- c) Check and correct survey data – the validation *Rules* specified by the *Process Methods* for a subset of the *Represented Variables* of the *Unit Data Structure* are applied to the *Instance Variables* of the *Unit Data Set*, and the correction *Rules* are applied to the *Instance Variables* that fail validation.

**Table 8. Part 2: Collection**

<b>Activity Steps</b>	<b>Applicable GSBPM sub-process</b>	<b>Applicable GSIM Objects</b>
a) Collect survey data	4.2 Run collection	<ul style="list-style-type: none"> <li>• <i>Unit</i></li> <li>• <i>Survey Instrument</i></li> <li>• <i>Unit Data Set</i></li> <li>• <i>Unit Data Structure</i></li> </ul>
b) Check which methodology to use	2.4 Design frame & sample methodology	<ul style="list-style-type: none"> <li>• <i>Design Context</i></li> <li>• <i>Process Method</i></li> <li>• <i>Unit Data Set</i></li> <li>• <i>Represented Variable</i></li> </ul>
c) Check and correct survey data	5.3 Review, validate & edit	<ul style="list-style-type: none"> <li>• <i>Rule</i></li> <li>• <i>Process Method</i></li> <li>• <i>Represented Variable</i></li> <li>• <i>Instance Variable</i></li> <li>• <i>Unit Data Structure</i></li> <li>• <i>Unit Data Set</i></li> </ul>

122. Part 3: Estimation

- a) Check which methodology to use - Access the *Design Context* of estimation of stratified data and retrieve the *Process Methods* suited for the *Population* of banks. Two *Process Methods* will be applied in sequence (steps b-c).
- b) Calibrate survey data – the weights computed in Part 1 are no longer valid, due to survey errors such as non-response. The calibration *Process Method* specifies a *Rule* to compute new weights. The weights will be stored in the *Instance Variables* described by the *Represented Variables* designed for this purpose.
- c) Compute estimates – select the aggregation *Process Method*, which specifies the following steps:
  - i. Select an aggregation *Rule* (sum, average, median, etc.)
  - ii. Design a *Dimensional Data Structure* to describe the aggregates to be produced
  - iii. Design an aggregation frame, specifying how *Unit Measure Components* of the *Unit Data Structure* of the *Process Input* contribute either to *Dimensional Identifier Components*, or in combination with the calibrated weights, to *Dimensional Measure Components* of the *Dimensional Data Structure* of the *Process Output*.

- iv. Apply the selected aggregation *Rule* and the aggregation frame of the previous step to the *Unit Data Set* (used as *Process Input*) to produce the *Dimensional Data Set* specified in the *Dimensional Data Structure* (the *Process Output*) of step 3b.

**Table 9. Part 3: Estimation**

<b>Activity Steps</b>	<b>Applicable GSBPM sub-process</b>	<b>Applicable GSIM Objects</b>
a) Check which methodology to use	2.5 Design statistical processing methodology	<ul style="list-style-type: none"> <li>• <i>Design Context</i></li> <li>• <i>Process Method</i></li> <li>• <i>Population</i></li> </ul>
b) Calibrate survey data	5.6 Calculate weights	<ul style="list-style-type: none"> <li>• <i>Process Method</i></li> <li>• <i>Rule</i></li> <li>• <i>Instance Variable</i></li> <li>• <i>Represented Variable</i></li> </ul>
c) Compute estimates	5.7 Calculate aggregates	<ul style="list-style-type: none"> <li>• <i>Process Method</i></li> </ul>
i. Select aggregation rule	2.5 Design statistical processing methodology	<ul style="list-style-type: none"> <li>• <i>Rule</i></li> </ul>
ii. Design cube data structure	2.1 Design outputs	<ul style="list-style-type: none"> <li>• <i>Dimensional Data Structure</i></li> </ul>
iii. Design aggregation scheme	2.5 Design statistical processing methodology	<ul style="list-style-type: none"> <li>• <i>Unit Measure Component</i></li> <li>• <i>Unit Data Structure</i></li> <li>• <i>Dimensional Identifier Component</i></li> <li>• <i>Dimensional Measure Component</i></li> <li>• <i>Dimensional Data Structure</i></li> </ul>
iv. Apply aggregation rule	5.7 Calculate aggregates	<ul style="list-style-type: none"> <li>• <i>Rule</i></li> <li>• <i>Unit Data Set</i></li> <li>• <i>Process Input</i></li> <li>• <i>Dimensional Data Set</i></li> <li>• <i>Dimensional Data Structure</i></li> <li>• <i>Process Output</i></li> </ul>

## Scenario 4: GSIM and dissemination of statistical information

### *Dissemination of statistical information*

123. The majority of statistical organizations disseminate information in one form or another. Generally dissemination of information begins with the process of designing the outputs (static products or interactive services) to meet a set of user's information needs and results in information being made publically available on a website or other dissemination channel.

124. Each organization will have its own specific set of processes to disseminate information but generally this will consist of:

- a) Selecting the data and information to be disseminated
- b) Setting up output systems to receive data and information
- c) Loading data and information into the output system
- d) Producing products or implementing services to present information to users
- e) Reviewing, editing and approving information for release

125. The above scenario is applicable across a range of types of dissemination including:

- Publication of data and products associated with a single iteration of an ongoing statistical output – for example a survey is conducted, the results are analyzed and findings released on a recurring basis
- Dissemination of data through a service queried by users – for example a data browser or table builder tool where users are able to choose the data they require
- Publication of products including data from multiple statistical outputs – for example a compendium or yearbook that is a separate activity but uses data from multiple sources

### *Mapping of example description to GSIM information objects*

126. GSIM is able to support the dissemination process across multiple types of dissemination including those identified above. A key need of statistical organizations is to disseminate products that include data from multiple outputs.

127. GSIM can be used by those people involved in dissemination processes to identify the pieces of information they require to undertake their roles and in the design of systems for dissemination purposes.

128. Steps involved in the dissemination of statistical information:

- a) Selecting the data and information to be disseminated - The dissemination of information takes place within the context of a *Statistical Program* (the overarching activity or ongoing series, e.g. Employment Survey) as part of a *Statistical Program Cycle* (an iteration of the ongoing activity, e.g. March 2012 Employment Survey) and specifically a *Dissemination Activity*. The first step in this activity is to identify the data and/or information to be

disseminated. The *Variables* required by the intended user audience are identified and the particular *Represented Variables* selected depending on the requirements (as identified by users in an *Information Request*) for data about a particular *Population* or according to a particular *Classification*.

- b) Setting up output systems to receive data and information - Once the data and information for dissemination have been selected the systems used for the process will be configured. This is done according to the *Dissemination Design* which identifies the attributes of the *Dissemination Activity*, such as the *Data Structure*. In many dissemination processes a *Dimensional Data Structure* will be defined. A *Dimensional Data Structure* describes the structure of an aggregate, multi-dimensional table (macro data) by means of *Dimensional Identifier Components*, *Dimensional Attribute Components* and *Dimensional Measure Components*. Both are *Represented Variables* with specific roles in such a table. Dimensions typically refer to *Variables* with coded *Value Domains*, measures to *Variables* with uncoded *Value Domains*. An example for a type of *Data Set* defined by a *Dissemination Data Structure* is a Time Series. It has specific attributes such as frequency and type of temporal aggregation and specific methods, e.g. seasonal adjustment, and must contain a temporal variable.
- c) Loading data and information into the output system - When dissemination systems have been set up and structures created according to the appropriate *Data Structure*, data and information are loaded. *Data Points*, defined by *Instance Variables*, are selected from source *Data Sets*. A check is undertaken to ensure that the required selection of data and metadata from the source *Information Resource* (a collection of *Data Sets* including data and/or metadata) has been correctly loaded.
- d) Producing products or implementing services to present information to users - After data have been loaded to the output system, the type of *Dissemination Activity* determines how the data are presented to users. A *Dissemination Activity* includes either a *Publication Activity* or a *Dissemination Service* as a method to create and disseminate *Representations* to consumers. *Representations* may contain any type of information, for instance statistical data (as a *Data Set* or visualization) or structural or conceptual metadata like a *Data Structure*, a *Code List* or a description of a *Concept*.

A *Publication Activity* results in the creation of *Products* which may be made up of one or more *Representations* and stored to be delivered to users. Examples of *Products* are publications, press releases, etc.

A *Dissemination Service* is the mechanism to create and disseminate *Representations* to users. These *Representations* are created dynamically on the specific request and according to the specific needs of the consumer. These exposes *Data Sets* that may be included in

*Products and Representations*, either as *Data Sets* (e.g. when providing access to public-use micro data) or as a visualization (e.g. a table in a report or an interactive chart on a website).

- e) Reviewing, editing and approving information for release - A completed *Product* or *Dissemination Service* will be reviewed against the requirements of the original *Information Request* to ensure user needs are being met and against the organizations policies and procedures to ensure the outputs meet the required level of quality.

**Table 9. Dissemination of statistical information**

<b>Activity Steps</b>	<b>Applicable GSBPM sub-process</b>	<b>Applicable GSIM Objects</b>
a) Selecting the data and information to be disseminated	6.5 Finalize outputs	<ul style="list-style-type: none"> <li>• <i>Statistical Program Cycle</i></li> <li>• <i>Dissemination Activity</i></li> <li>• <i>Represented Variable</i></li> <li>• <i>Variable</i></li> <li>• <i>Classification</i></li> <li>• <i>Information Request</i></li> <li>• <i>Population</i></li> </ul>
b) Setting up output systems to receive data and information	7.1 Update output systems	<ul style="list-style-type: none"> <li>• <i>Dissemination Design</i></li> <li>• <i>Data Structure</i></li> <li>• <i>Dimensional Data Structure</i></li> <li>• <i>Dimensional Identifier Component</i></li> <li>• <i>Dimensional Measure Component</i></li> <li>• <i>Dimensional Attribute Components</i></li> <li>• <i>Represented Variable</i></li> <li>• <i>Value Domain</i></li> </ul>
c) Loading data and information into the output system	7.1 Update output systems	<ul style="list-style-type: none"> <li>• <i>Data Set</i></li> <li>• <i>Data Point</i></li> <li>• <i>Instance Variable</i></li> <li>• <i>Dimensional Data Structure</i></li> </ul>
d) Producing products or implementing services to present	7.2 Produce dissemination products	<ul style="list-style-type: none"> <li>• <i>Publication Activity</i></li> <li>• <i>Dissemination Service</i></li> <li>• <i>Product</i></li> </ul>



information to users		<ul style="list-style-type: none"> <li>• <i>Output Specification</i></li> <li>• <i>Representation</i></li> <li>• <i>Information Resource</i></li> </ul>
e) Reviewing, editing and approving information for release	7.3 Manage the release of dissemination products	<ul style="list-style-type: none"> <li>• <i>Dissemination Service</i></li> <li>• <i>Product</i></li> </ul>

**Scenario 5: GSIM and quality**

*Quality measures and reports*

129. The majority of statistical organizations will calculate quality measures, and include quality information in a report.

130. The decision was made that quality would not be defined as a separate object (or group of objects) in GSIM, as it could not be adequately defined in this manner. Quality itself can have many forms depending on the purpose that it is there for. It could, for example, be representing the quality of the organization, the quality of the process used, or the quality of the statistics produced. Quality information and reports could be relevant to a number of different levels of a particular information object, or tied to the production process (or parts of the process) as a whole. It is present in the inputs and outputs of all process steps within the GSBPM, and can act as the rules and control of processes.

131. Quality means different things in different settings, and so depending on the scope, it will refer to the different information objects in relation to relevant processes. Quality is therefore not seen as an explicit object in GSIM.

132. This scenario has two parts and looks a little different to the previous four scenarios. The first part describes a process that could be used to calculate response rates (i.e. a quality measure), and the second part how these rates could be included in a quality report in the form of a qualitative assessment. Refer to Figure 1 for a representation of the scenario.

133. Part 1: Calculate the response rates needed for a set of data

- a) A process is setup to calculate the response rates for a *Data Set*.
- b) A quality check is included as a part of this process

134. Part 2: A qualitative assessment statement included in a quality report.

- a) A quality report is prepared

*Mapping of example description to GSIM information objects*

135. GSIM can be used by those people involved to identify the pieces of information that they require to undertake their roles and in the design of systems.

136. Part 1: Calculate the response rates needed for a set of data.

- a) A process is setup to calculate the response rates for a *Data Set* - First, a *Process Step Design* is created to outline the specification of the *Process Step*, in this case the calculation of the response rates for a particular *Data Set*. The design includes the *Process Input Specification* and *Process Output Specification* – the required inputs, and expected outputs of the *Process Step*.

For the response rate calculation example, the inputs include the *Instance Variables* that describe the original *Data Set* for which the response rates are to be calculated, and the *Parameter Input*, (the calculation formulas used to obtain the Unit and Item response rates). The expected outputs of the described *Process Step* will include a new *Data Set* of the calculated response rate data, and new *Instance Variables* to represent the data.

The execution of the *Process Step* is recorded by the *Process Step Execution Record*, which identifies the (actual) inputs and outputs at the time of the process execution. In this example the response rate calculation process is itself made up to two lower level *Process Steps*, to calculate the unit response rate, and item response rate respectively. Each of these *Process Steps* was defined through the *Process Step Design*, and then when used, captured by the *Process Step Execution Record*.

- b) A quality check is included as a part of this process - As part of the overall *Process Step Design*, a *Process Control* was included in the *Process Step* as a quality check. This quality check stage allows for a check point in the procedure to ensure the calculated response rates are within an acceptable range.

If the calculated rates are within an acceptable range, the output of the process (both the calculated *Instance Variables* and the *Process Metric*) can be used in a subsequent *Process Step* (scenario part 2). If the calculated rates are not accepted, then the process ends, and other *Process Steps* may need to be employed (such as a review of the collection *Process Steps* to allow for more acceptable response rates).

137. Part 2: A qualitative assessment statement included in the quality report.

- a) A quality report is prepared - First, *Process Step Designs* are created for the *Process Steps*, in this case the preparation of the quality report which includes a qualitative assessment (based on the calculated response rates).

There are two *Process Steps* involved. The first is the *Process Step* of preparing the qualitative assessment of the calculated response rates. The *Process Inputs* are the *Data Set*

and *Process Metric* created from the *Process Step* in Part 1 and the *Process Output* is a *Representation* (the qualitative assessment statement).

This *Representation* in turn becomes the *Process Input* of the second *Process Step* which is the creation of the quality report. The *Process Output* is a *Product*.

The execution of the overall *Process* is recorded in two *Process Step Execution Records*.

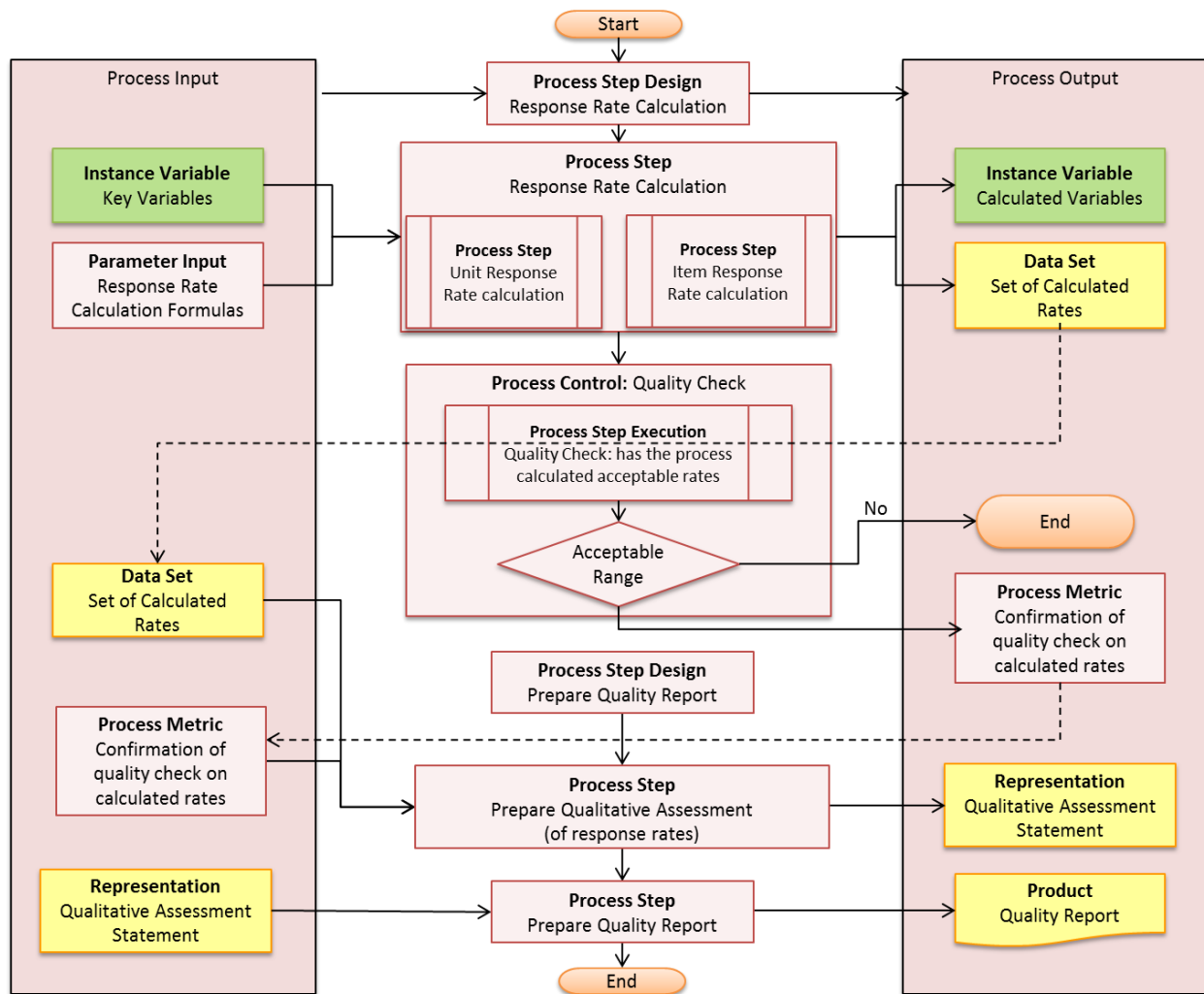


Figure 3. Response Rate Calculation for quality report

## Annex B: Template for Case Study

### GENERIC STATISTICAL INFORMATION MODEL (GSIM) IMPLEMENTATION CASE STUDY < COUNTRY / ORGANIZATION NAME >

This template was created to guide the content and format of case studies on the implementation of the Generic Statistical Information Model (GSIM) in statistical organizations. It is intended to assist organizations in sharing their experiences, and to encourage the update of information as projects develop. From a reader's perspective, a consistent approach to presenting case studies can facilitate better understanding and comparability between organizations.

#### **Instructions for completion:**

The template is divided into three sections. Each section comprises a table with headings in the left hand column (in bold). On the right is a description of the suggested content, which should be replaced with the appropriate text.

The template is flexible. If a particular heading is not relevant, it can be modified to something more appropriate, or deleted. If there appears to be no place for certain information, new headings/sections should be added.

**Questions and feedback** should be directed to the UNECE secretariat: email [support.stat@unece.org](mailto:support.stat@unece.org),

## 1. INTRODUCTION

### **Organization Name**

**Number of staff** The total number of staff in your organization. It would be good to distinguish between central/regional offices where applicable.

**Organization structure** A diagram showing high-level structure of the organization.

**Contact person  
(for Information  
Management)** Name  
Job title / Division (as shown in Organization structure)  
Email  
Phone

**Information Management strategy** Explanation of the overall strategy for developing and maintaining statistical information across the organization. For example, the mandate/program providing framework for information management projects, and basic information management principles used.

**Current situation** Outline the current information management project(s) being planned or

implemented. This can be a short overview. The detail about the project will be provided in the following sections.

## 2. MODELLING THE INFORMATION OF A STATISTICAL ORGANIZATION

- 2.1 Statistical information model** Provide a diagram or explanation of the statistical information model used in your organization. Explain how/when it was developed.
- If you don't have an integrated model of all information managed in your organization, you probably have partial models to manage independently statistical metadata, statistical methods, architectural principles, policy provisions and similar things. In this case, you can provide separate descriptions for these partial models.
- 2.2 Information management system(s)** Describe the information management system, or systems that are used in your organization. Explain how they fit within other organizational systems and clarify which point they are at (i.e. scoping, planning, implementation, etc).
- Explain how the system(s) are interrelated. Descriptions of planned changes and additions would also be helpful.
- This section is intended as an overview of the system.
- 2.3 Costs and Benefits** Describe the costs and benefits of the information management project(s). Costs should preferably be given in terms of human resources rather than money.
- 2.4 Implementation strategy** Explain if the information management project(s) is/are being implemented with a step-wise or 'big-bang' approach. Provide a timeline of project milestones if possible.

## 3. ADOPTING GSIM

- 3.1 Adoption strategy** Have you adopted GSIM? If yes, do you intend to implement it as is, with possible extensions, or to map your own model(s) onto it?
- 3.2 Relation to other models** Do you intend to use GSIM in conjunction with GSBPM, or with your own statistical business process model?
- Do you intend to use GSIM in conjunction with an international implementation standard (for example, DDI and/or SDMX) or your own implementation standard?

\*\*\* END \*\*\*