GSIM and more

InKyung Choi United Nations Economic Commission for Europe (UNECE)



@SESRIC Workshop July 10th 2023



Agenda

- 1. Introduction to GSIM
- 2. Linking GSBPM and GSIM
- 3. SDMX-DDI-GSBPM







Part 1: Introduction to Generic Statistical Information Model (GSIM)*





Modernisation

- Modernisation of statistics can be supported by re-use and sharing of methods, components, processes, metadata and data
- Generic Statistical Information Model (GSIM) is the first internationally endorsed reference framework for statistical information
- GSIM is a **conceptual** framework playing an important role in modernising, streamlining and aligning the standards and production associated with official statistics, at both national and international levels.







What is GSIM?

• It is a reference framework of internationally agreed definitions, attributes and relationships that describe the pieces of information used in the production of official statistics (called "information class").



• This framework enables generic descriptions of the definition, management and use of data and metadata throughout the statistical production process.







Development of GSIM

WHO WAS INVOLVED?



17 statistical organisations





- Developed by statistical organisations and for statistical organisations
- GSIM version 1.0 (2012)
- GSIM version 1.1 (2013)
- GSIM version 1.2 (2019)
- GSIM version 2.0 (2023)





Supporting standards





GSIM Groups

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







Some GSIM information classes



127 information classes exist in GSIM (version 2.0)











GSIM information classes that tell a story







What is the relationship between GSIM & GSBPM?

- GSIM and GSBPM are complementary models for the production and management of statistical information. Greater value can be obtained when both models are applied together.
- GSBPM models the statistical production process and identifies the activities undertaken by producers of official statistics that result in information outputs.
- GSIM helps describe GSBPM sub-processes by defining the information classes that are used by them, that flow between them, and are created in them in order to produce official statistics.





GSIM experiences



Statistics Canada from GSIM e-training (2019)

The Generic Statistical Information Model at Istat

- GSIM applied at Istat for the conceptual representation of metadata in metadata system
- Most of the GSIM Information Objects from Concept and Structure Groups already implemented in the Istat Structural Metadata System SUM-MS
- Business, Exchange and Structure Groups Information Objects are supporting the redesign of Istat System for Reference Metadata and Quality of Statistical Processes SIDI/SIQual
- Business Group Information Objects are also being considered for modelling the metadata-driven system for managing the data processing in shortterm surveys (SINTESI)



Istat (Italy) from GSIM e-training (2019)

UNECE

Functions, GSIM



Hungary from ModernStats World Workshop (2022)



Sweden from ModernStats Community of Practice (2023)

How GSIM-objects are currently being implemented in Statistics Finland



Finland from ModernStats World Workshop (2019)

GSIM objects pinpointed

- · GSIM is rich enough to provide useful objects.
- Statics dimension: to charaterise the main steps or objects of the process
 Business services, Process Steps, Business Functions. All of them tagged with attributes
 and administrative details
- · Dynamic dimension: to describe the sequence of steps
- Objects around the process design
- Input (core, process support, parameter)
- output (core, process execution log)
- · Methods (specification of the technique used in the process)
- Rules (mathematical or logical expression)

France from ModernStats World Workshop (2022)





Benefits of GSIM experienced

- Provides a common language to improve communication between statistical organisations at national and international levels
- GSIM in combination with GSBPM creates an environment prepared for reuse and sharing of methods, components and processes
- Enables statistical organisations to rethink how their business could be more efficiently organised by defining information classes common to all statistical production, regardless of subject matter
- Provides a common language to improve communication between the different roles in statistical production (business and information technology experts)







Part 2: Linking GSBPM-GSIM*







Linking GSBPM and GSIM Task Team (2019-21)

- 20 members from 11 participating institutions (Australia, Canada, Egypt, Hungary, Italy, Mexico, Poland, South Korea, Sweden, ILO, DDI)
- Derived a robust set of GSIM classes that could be used as inputs and outputs based on use cases commonly taking place in many statistical organisations.
- Made it easier to design systems to track information flow through statistical business processes
- Contributed to building a "de facto" integrated view of the ModernStats models (now expanded by the Core Ontology for Official Statistics work)
- Made it easier to create implementations, in particular CSPA services (now continued by the SDMX-DDI-GSBPM mapping work)







Linking GSBPM and GSIM - Context









Linking GSBPM and GSIM - Specification









Linking GSBPM and GSIM - Specification



Specification example: GSBPM sub-process 5.3 Review and Validate

Process Input Specification	Process Design	Process Output Specification
Core Input type	Process Method	Core Output type
Data Sets (unit Data Sets) to be reviewed and validated	Review Data Sets and Process Methods	Data Sets (unit Data Sets): updated Data Set
Data Structures associated with Data Sets to understand Data Sets	Apply Process Methods and Rules to review Data Sets	Data Structure associated with Data Set
Represented Variables to be reviewed and validated	Apply Process Methods and Rules to validate Data Sets	Referential Metadata Set: descriptions of the Process Methods used,
Process Methods that specifies methodology for review and validation	Calculate quality measures specified by Process Methods	relevant information to be passed along with Data Sets
Rules, as designed in Phase 2	Update Data Sets and associated element in Data Structure with results from review and validation	Process Metric type
Parameter Input type		Quality measures related to review and validation such as:
Parameter values to be used for review and validation methodologies as specified in Process Method such as:		 Number of validations conducted Number of outliers detected
 Limit value for edit Rule (interval for valid values) Threshold for checking outlier 		Quality measures of Process Step such as:
Process Support Input type		Time spent to complete the Process Step (derived from Process Execution Log)
Auxiliary Data Sets or any Information Resource to be used for review		Cost spent to complete the Process Step
and validation, e.g., historic comparison, macro-level comparison		Process Execution Log type
Technical / methodological handbooks, policies or guidelines to be followed regarding data validation as well as guality management		Execution log such as
		Time that Process Step started
		Time that Process Step ended
		 Any message or event log generated from software used for review and validation







Linking GSBPM and GSIM - Information Flows









Scenario: NSO and a sewage treatment company recently established Memorandum of Understanding (MoU) that agrees on the principle of data sharing. The company commits to make their data available via API for NSO to use for statistical purpose.

This diagram aims to demonstrate how GSIM classes in Exchange Group can be used to represent information in GSBPM sub-processes relevant to exchange (while the focus of this diagram is on Exchange Group, some of non-Exchange Group classes were also added to make the story flow more smooth

GSIM Exchange Group in Action



GSIM Business Group in Action









GSIM classes for all GSBPM sub-processes

GSBPM Phase 1

GSBPM Phase 2...7

GSBPM Phase 8







Available on the Linking GSBPM-GSIM "information flow" wiki page







Part 3: GSBPM-SDMX-DDI



* This presentation materials are based on "Supporting Standards Group" by Zoltán Vereczkei (Hungary), Chair of the Supporting Standards Group, from <u>Modernisation Workshop 2022</u>



GSBPM-SDMX-DDI Task Team Context



Statistical Data and Metadata eXchange

M2 .	

Data Documentation Initiative



Benefits

- International / global standards
- Free / open
- Helps reuse tools and concepts
- Increases interoperability
- Improves quality



Difficulties

- Which standard to use for which use case?
- Technical, people get lost
- Which artefact to use for which stage of production process?







GSBPM-SDMX-DDI Task Team (2022-)

• 15 members from 11 participating institutions (Canada, Hungary, Italy (NSO), Italy (Bank), Mexico, Kingdom of the Netherlands, USA, BIS, DDI, ILO, OECD)









GSBPM-SDMX-DDI Task Team Objectives



Source: Steven Vale: Exploring the relationship between DDI, SDMX and the Generic Statistical Business Process Model







GSBPM-SDMX-DDI Task Team Objectives

Sub-process 1.4 Identify Concepts

SDMX

- SDMX provides the means of capturing Concepts that can then be organized (and managed) in Concept Schemes.
- Relevant SDMX artifacts/instrument: Concepts, ConceptScheme; (optional) SDMX modeling guideline, SDMX Glossary

DDI

- DDI provides the means of capturing Concepts that can then be organized (and managed) in ConceptSchemes.
- Relevant DDI artifacts: Concept, ConceptScheme, DDI-C (optional)

Example for GSBPM sub-process 1.4

Objectives

- Provide short description why/how SDMX/DDI helps as an entry point
- List relevant SDMX/DDI artefacts under each sub-process
- Map relevant SDMX/DDI artefacts under each sub-process to GSIM class

Results and (some preliminary) findings

- By using GSBPM as common linkage points, identifies where SDMX and DDI work together, their strengths
- With mapping between SDMX and DDI made easier, more chances for interoperability between the two standards







Resource

- <u>GSIM Wiki</u>
- <u>GSIM Resource Repository</u>
- GSIM E-training (2019)
- Linking GSBPM-GSIM
- <u>GSIM implementation roadmap and tools</u>







Thank you!

Feel free to contact me for more information (choii at un dot org)





