Linking GSBPM and GSIM: Hands-on Activity

Background

The Generic Statistical Business Process Model (GSBPM) provides a standard framework and terminology to help statistical organisations to harmonise their statistical production processes. The Generic Statistical Information Model (GSIM) is a reference framework for statistical information that provides a set of standardised, consistently described information classes. Conceptually the two models are closely related and complementary as GSIM describes information classes that can be used as inputs and outputs in the statistical production process.

Instruction

The Linking GSBPM-GSIM task team went through 44 GSBPM sub-processes and identified what information is needed as inputs to carry out activities in each sub-process and what outputs are produced as a result of the activities; and represented them in GSIM classes. In this hands-on activity, participants will go through a similar exercise, but for a few selected GSBPM sub-process.

Please read instruction below (2 min), conduct the exercise (18 min; use the last blank page if needed) and once you are done with the exercise, explain to colleagues in your group the result and compare it with those who choose the same sub-process (10 min)[1]:

1. Think of an example of any statistical production process (e.g., labour force survey, tourism statistics production), preferably one that you have already worked on or are familiar with.
2. Choose GSBPM sub-process from List 1 and think how the production process from (1) is conducted in this sub-process
3. What are the important information inputs (e.g., data, user needs) needed for you to carry out activities described in this sub-process for the production process example from (1)?
4. What are the core information outputs generated when you finish this sub-process for the production process example from (1)?
5. Do you see any GSIM classes from List 2[2] that seem to be able to represent the information from (4)?

---

[1] Check also Linking GSBPM-GSIM task team report to see how the team did as a reference
[2] Have a question about a GSIM class? Ask ModernStats Clinic (slido.com # 6158786)!

List 1: GSBPM sub-process description

GSBPM Sub-process 1.5 Check data availability

This sub-process checks whether current sources of data could meet user requirements and the conditions under which they would be available including any restrictions on their use. An assessment of possible alternatives would normally include research into potential administrative or other non-statistical sources of data, to:

- *Determine whether they would be suitable for use for statistical purposes (e.g. the extent to which administrative concepts match data requirements, timeliness and quality of the data, security and continuity of data supply);*
- *Assess the division of responsibilities between data providers and the statistical organisation;*
- *Check necessary ICT resources (e.g. data storage, technology required to handle incoming data and data processing) as well as any formal agreements with data providers for accessing and sharing the data (e.g. formats, delivery, accompanying metadata and quality check).*

When existing sources have been assessed, a strategy for filling any remaining gaps in the data requirement is prepared. This may include identifying possible partnerships with data holders. This sub-process also includes a more general assessment of the legal framework in which data would be collected and used, and may therefore identify proposals for changes to existing legislation or the introduction of a new legal framework.

GSBPM Sub-process 2.2 Design variable description

This sub-process defines the variables to be collected via the collection instrument, as well as any other variables that will be derived from them in sub-process 5.5 (Derive new variables and units), and any statistical or geospatial classifications that will be used. It is expected that existing national and international standards will be followed wherever possible.

This sub-process may need to run in parallel with sub-process 2.3 (Design collection), as the definition of the variables to be collected, and the choice of collection instruments may be inter-dependent to some degree. Preparation of metadata descriptions of collected and derived variables, statistical and geospatial classification is a necessary precondition for subsequent phases.

GSBPM Sub-process 5.2 Classify and code

This sub-process classifies and codes the input data. For example, automatic (or clerical) coding routines may assign numeric codes to text responses according to a pre-determined statistical classification to facilitate data capture and processing. Some questions have coded response categories on the questionnaires or administrative source of data, others are coded after collection using an automated process (which may apply machine learning techniques) or an interactive, manual process.

GSBPM Sub-process 6.1 Prepare draft outputs

This sub-process is where the data from sub-processes 5.7 (Calculate aggregates) and 5.8 (Finalise data files) are transformed into statistical outputs such as indexes, seasonally adjusted statistics, e.g. trend, cycle, seasonal and irregular components, accessibility measures, etc., as well as the recording of quality characteristics such as coefficients of variation. The preparation of maps, GIS outputs and geo-statistical services can be included to maximise the value and capacity to analyse the statistical information.

List 2: Selected GSIM objects[3]

GSIM class	Definition
Assessment	result of the analysis of the quality and effectiveness of any activity undertaken by a statistical organization and recommendations on how these can be improved
Business Case	proposal for a body of work that will deliver outputs designed to achieve outcomes
Code List	list of Categories where each Category has a predefined Code assigned to it
Data Set	organized collection of data
Data Structure	structure of an organized collection of data (Data Set)
Information Resource	abstract notion that is any organized collection of information
Information Set	organized collections of statistical content
Referential Metadata Set	organized collection of referential metadata for a given Referential Metadata Subject
Represented Variable	combination of a characteristic of a population to be measured and how that measure will be represented
Rule	mathematical or logical expression which can be evaluated to determine specific behavior
Process Execution Log	Process Output listing events generated by a Process Step Instance
Process Method	specification of the methodology which will be used to perform the work
Process Metric	Process Output summarising some aspect or property of the execution
Provision Agreement	legal or other basis by which two parties agree to exchange data
Statistical Classification	set of Categories which may be assigned to one or more variables registered in statistical surveys or administrative files, and used in the production and dissemination of statistics
Value Domain	set of permissible values for a Variable
Variable	use of a Concept as a characteristic of Unit Type intended to be observed

---

[3] The definitions are from the new version of GSIM that is currently being updated so they may look slightly different from the current version of GSIM