

# Case study: Italian National Institute of Statistics - Istat

Case study: Italian National Institute of Statistics - Istat Case study: Italian National Institute of Statistics - Istat: use of GSBPM istat: use of GSBPM istat: use of GSIM

<b>Contact person*</b>	Marina Signore
<b>Job title</b>	Director of Research - Chief Division "Metadata, Quality and R&D projects"
<b>Email</b>	signore@istat.it
<b>Telephone</b>	+39 06 46733359

## Summary\*

The Istat Unified Metadata System (Sistema Unitario di Metadati, SUM) is based on the following key concepts and objectives: data retrieval and usability (by associating proper meaning to data as well as, methodological and quality information); metadata reuse (in order to harmonise concepts and reduce documentation burden); traceability (with the objective of statistical process transparency and process automation); integration (with the aim of making the different sectors within a statistical institute speak with one voice and support standardisation).

The SUM system jointly manages the three following metadata categories:

- Metadata related to data structure and content include all that is necessary to give a definition and a meaning to statistical data, roughly corresponding to structural metadata;
- Process-related metadata describe the statistical business process in terms of methods (e.g. sampling, collection methods, editing processes) and quality (e.g. timeliness, accuracy), largely corresponding to reference metadata;
- Business-related metadata are useful for the management of an NSI, in planning, executing and assessing both statistical and support activities. These metadata allow stakeholders to connect data and processes with NSI's strategic objectives (e.g. long-term goals) and with the different management plans of each Institute (e.g. methodological investments); to assign responsibilities and resources to the objectives; to plan schedules and timetables of different actions; to evaluate the achievement of objectives and to assess performance and efficiency.

At present SUM is composed of two main subsystems:

- SIDI-SIQual which manages process-related metadata and quality indicators and is described in GSBPM case-study
- SUM-MS which manages structural metadata and is described in GSIM case-study

## Metadata strategy

Istat has adopted a strategy for managing statistical metadata in a harmonized way across the institute. A single centralized system, the Unified Metadata System- SUM, will store and validate all metadata related to statistical data and processes. Furthermore, it integrates quality and metadata management.

The Istat Unified Metadata System is based on the following key concepts and objectives:

- data retrieval and usability (by associating proper meaning to data as well as, methodological and quality information);
- metadata reuse (in order to harmonise concepts and reduce documentation burden);
- traceability (with the objective of statistical process transparency and process automation);
- integration (with the aim of making the different sectors within a statistical institute speak with one voice and support standardisation)

The SUM system allows for a full description of statistical data and processes and related quality aspects. It is compliant with the international standards such as GSIM and GSBPM as described in the specific case-studies.

It supports the modernization process underway at Istat aimed at overcoming stove-pipe production processes.

## Current situation

The SUM system is composed by two main subsystems:

- the SIDI-SIQual system, managing metadata describing the statistical production processes and their quality features and quality indicators together with relevant documentation (e.g. methodological manuals, regulations, procedures,...);
- the SUM-MS system, managing metadata describing all that is necessary to give a definition and a meaning to statistical data (e.g. data content and data structures)

The SIDI-SIQual system was designed back in the late 90s and it is fully implemented. It documents all Istat statistical processes (e.g. surveys, administrative data sources, statistical compilations,...) and manages time series of quality indicators for output quality dimensions for a decade (around 313.000 indicators stored so far).

The SUM-MS system was designed, in its current features, in 2013 and it already contains metadata that define disseminated macrodata for most of the current surveys. In 2015 the system will begin to contain also metadata relative to collected and validated microdata.

The SIDI/SIQual metadata management system is being enhanced in order to be compliant with the ESS conceptual and technical reference metadata standards (SDMX dissemination of metadata according to ESMS and ESQRS).

## Metadata Classification

In Istat approach, metadata are viewed as part of a wider system where different metadata typologies are related to one another and connected to quality.

To this aim, three metadata categories are considered:

- Metadata related to data structure and content include all that is necessary to give a definition and a meaning to statistical data;
- Process-related metadata describe the statistical business process in terms of methods (e.g. sampling, collection methods, editing processes) and quality (e.g. timeliness, accuracy).
- Business-related metadata are useful for the management of an NSI, in planning, executing and assessing both statistical and support activities. These metadata allow stakeholders to connect data and processes with NSI's strategic objectives (e.g. long-term goals) and with the different management plans of each Institute (e.g. methodological investments); to assign responsibilities and resources to the objectives; to plan schedules and timetables of different actions; to evaluate the achievement of objectives and to assess performance and efficiency.

With regard to SDMX terminology: i) metadata related to data structure and content correspond to respectively structural; ii) conceptual reference metadata and process-related metadata correspond to methodological and quality reference metadata, while iii) business-related metadata are not explicitly considered.

The above mentioned metadata categories find a correspondence to GSIM information objects, as shown in Fig. 1.

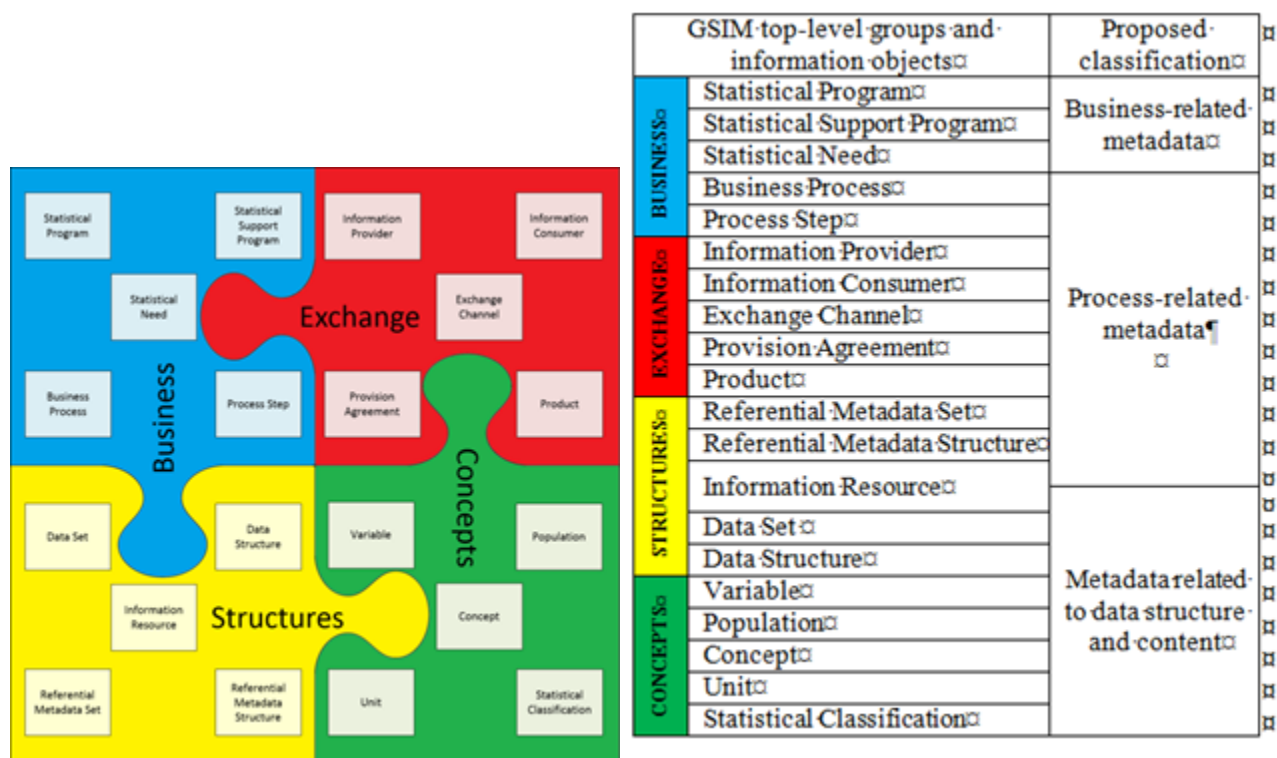


Figure 1. Mapping between the GSIM information objects and the proposed classification

## Metadata system(s)

As shown in the fig. 2, the SUM system manages the three metadata typologies allowing for:

- a description of statistical data and its transformation along the statistical production process (from the collection/acquisition to the disseminate and evaluate phase), following GSIM conceptual model;
- a description of the production process, its phases and sub-processes, the applied methodologies and the quality activities implemented and related indicators, following GSBPM
- supporting strategic planning and performance assessment by providing top and middle managers with pieces of information (i.e. business-related metadata) useful for a more efficient management of statistical activities.

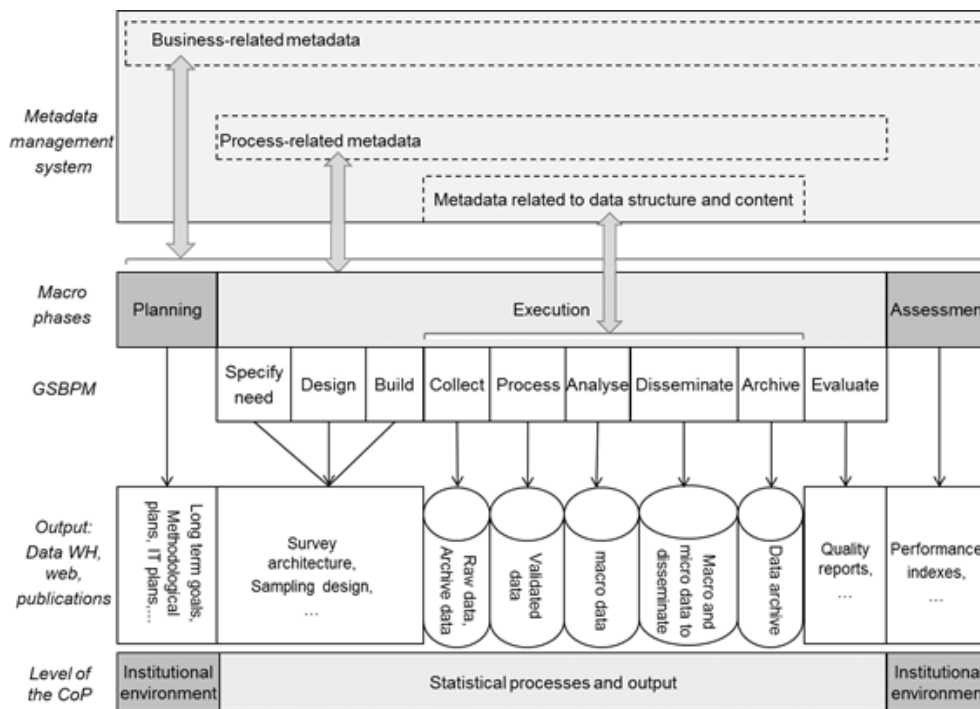


Figure 2: The Istat metadata management system

As said it is composed by two main subsystems that are interrelated:

- the SIDI-SIQual system, managing process-related metadata describing the statistical production processes and their quality features and quality indicators together with relevant documentation (e.g. methodological manuals, regulations, procedures,...);
- the SUM-MS system, managing metadata related to data structure and content thus describing all that is necessary to give a definition and a meaning to statistical data.

For more details on the two systems, see the GSBPM and GSIM case-studies.

## Costs and Benefits

Monetary costs are not available.

As a by-product, the organization of metadata as implemented in SUM-MS allows for the implementation of easy research tools of data and metadata through all the data collections in Istat.

The clear definition of metadata that define data, organized according to GSIM, enhances the possibilities to identify data sets that can be integrated, thus facilitating the use of administrative data in the statistical production processes.

Major benefits also derive from the joint management of metadata and quality, namely:

- increased quality culture;
- efficiency gains in the documentation activity;
- support to quality evaluation based on objective information
- increased transparency

## Implementation strategy

The SIDI-SIQual system is fully implemented. It is now being enhanced by adding a component for generating Quality Reports compliant with the European standards (ESS SIMS- Single Integrated Metadata Structure) in SDMX format.

Documentation in the SIDI-SIQual system (metadata and quality indicators) is regularly updated and validated. Quality pilots who received specific training, are in charge of maintaining information up-to-date in the system and calculating quality indicators. The central unit in charge of the system validates new terms suggested by quality pilots and checks that quality indicators have been calculated for each survey occasion, once a year. A working group is in charge of coordinating the activities in the production directorates, of collecting emerging needs from internal users and suggesting improvements to the system.

The SUM-MS is being developed and will be completed in 2015. The SUM-MS system has been documented centrally, starting with metadata describing macro-data disseminated on Istat corporate dissemination system Istat and the classification systems. The documentation of metadata describing micro-data will start in 2015 by collecting metadata as soon as they are generated from centralized data collection system (e.g the business portal). A working group is in charge of starting the harmonization process of metadata related to data structure and content.

As far as business-related metadata are concerned, it is foreseen to adopt a stepwise strategy starting from those types of metadata that are more strictly related to quality (e.g. costs and resources) and that could be drawn from internal repositories (e.g. annual planning information system).

## **IT Architecture**

### **Metadata Management Tools**

### **Standards and formats**

### **Version control and revisions**

### **Outsourcing versus in-house development**

### **Sharing software components of tools**

### **Overview of roles and responsibilities**

The SUM system and its sub-systems are in charge of the Division "Metadata, Quality and coordination of EU R&D projects" under the Integration, Quality, Research and Production Networks Development Department.

More specifically, the unit "Integrated metadata system" is in charge of SUM-MS and its governance. A working group is in charge of the harmonisation process of metadata being used in Istat. The working group is composed by representatives from the various information systems that manage metadata, at present: the dissemination system based on the dot Stat technology (containing all the disseminated data), the data collection system based on GX (containing some data collection processes, although it seems it will be the Istat standard for data collection), the data validation system (already containing all the validated data), the system of microdata (containing the most important archives used for the production of some data). The unit has the task to define a metadata model to be used by the different systems, to collect metadata from the systems, to propose standard metadata for the whole Institute, to propose metadata changes to the different systems and a timetable for its application.

The unit "Quality, auditing and harmonization" is in charge of SIDI-SIQual. Since 2002, for each statistical process a quality pilot is in charge of keeping updated the metainformation in the system. So far more than 226 quality pilots have been trained. In addition, a working group is in charge of coordinating the activities in the production directorates, of collecting emerging needs from internal users and suggesting improvements to the system. The "Quality, auditing and harmonization" unit is in charge of validating the metadata and quality indicators loaded in the system, monitoring the quality and completeness of the stored metadata and quality indicators and producing Institute-level quality reporting using the information stored in the system.

Other centralised tasks concern the coordination of quality reporting to Eurostat, thus ensuring coherence across the Institute and the reuse of metadata and quality indicators available in the SUM system components.

### **Metadata management team**

See above

### **Training and knowledge management**

Different training courses are regularly offered to Istat personnel:

- Training courses on structural metadata, as a part of the course devoted on how to organize data and metadata for the dissemination system
- Training courses on SDMX
- Training course for Quality Pilots: it is a specific designed training course to explain the functionalities of the SIDI-SIQual system, how to document statistical processes, how to update the information and how to calculate quality indicators for each survey occasion
- Training courses on quality

Ad hoc software was developed, interfacing the SIDI system, in order to facilitate the quality indicators upload, preventing errors and assuring homogeneity and coherence.

### **Partnerships and cooperation**

Istat is very active in international and national cooperation on metadata and quality related aspects. Istat participates in the HLG for the Modernisation of Statistical products and services and in the Modernisation Committee on Standards.

### **Other issues**

### **Lessons learned**

Links:

**Attachments**