
Challenges in re-designing the national metadata system according to international standards

Giorgia Simeoni, Mauro Scanu (Istat, Italy)

simeoni@istat.it, scanu@istat.it

Abstract

Istat has a long tradition in metadata systems. Concerning reference metadata and quality indicators of statistical processes, the SIDI/SIQual system (Information System on Quality of Statistical Processes) was developed in the 90s'. It is in production since 2001 and is currently updated. SIDI/SIQual model for the statistical business process documentation can be considered a precursor of the Generic Statistical Business Process Model (GSBPM), and the system also includes several quality indicators that have been adopted as standard by the European Statistical System. Nevertheless, maintaining the "national" approach when international standards come out has its cost: it means mapping national concepts with international standard concepts and developing tools that translate as automatically as possible what is in the national system to be compliant with international standards. Until now it appears to be worthwhile to maintain the national approach, since it allows to comply both with different international requirements and national user needs. Nowadays, however, the SIDI/SIQual system needs to be renovated for several reasons, such as to better document the production process of multisource statistics and also for technological adaptation.

A system devoted to the description of the content and meaning of data sets (i.e. on structural metadata) has been under construction in Istat since 2010 under the acronym of SUM (Sistema Unitario dei Metadati). The system took advantage of the availability of the Generic Statistical Information Model (GSIM) in order to identify its main metadata concepts. In this case, the current challenge is not on the difference between the adopted model with respect to the international standard (as for SIDI/SIQual), but on the joint use of GSIM and GSBPM. More precisely, SUM should describe the content of data structures produced along the data production process as depicted by GSBPM: in this context, structural metadata on data structures content should be aligned and completed with process metadata, that could take advantage of the concepts in the Business Group of GSIM as well as those in GSBPM.

Considering the issues outlined above, both the SIDI/SIQual and the SUM need to be re-designed as a new comprehensive national metadata system. The new system should satisfy several demanding requirements from internal, national, and international users.

In order to satisfy Istat internal needs, the metadata system should be able to:

- preserve information already collected in existing metadata system SIDI/SIQual and SUM, reorganising it if necessary;
- document, trace (and drive) different kinds of statistical production processes, from traditional surveys to the creation of multisource statistics;

- promote the use of harmonised metadata across the institute;
- reduce the “documentation” burden on production units obtaining metadata as a by-product of production processes, or through integration with other systems and re-using metadata in the system to satisfy external requests.

In order to satisfy national user needs, a metadata system should be able to provide metadata to accompany data dissemination, tailored for different kinds of users, from less to more expert ones. Such metadata should facilitate users in finding data they are searching, and support their interpretation.

Finally, to satisfy international requirements, in particular European Statistical System ones, a metadata system should be able to produce metadata according to ESS conceptual and technical standards, e.g. with Standard Code Lists, according to the Single integrated metadata Structure SIMS 2.0, translated in the Standard for Data and Metadata eXchange SDMX.

GSBPM and GSIM are considered as standard models supporting the system re-design. They should facilitate the identification of relevant concepts and information objects to be taken into account. Indeed, GSBPM seems to be useful to reorganise statistical process documentation already available in SIDI/SIQual, in a way that could be more comparable at international level; GSIM has already proved to be helpful in the design of SUM as far as the Concepts and Structure groups concepts are concerned.

However, the joint application of GSBPM and GSIM appears to be more challenging. Focusing on the information objects belonging to the GSIM Business Group leads immediately to terminology issues with GSBPM; at the same time, the interpretation and concrete implementation of some GSIM concepts is not always straightforward. Similar difficulties are currently under discussion at international level in the Supporting Standard task team on Linking GSBPM–GSIM.

In addition, it should be assessed if GSIM information objects devoted to reference metadata and/or GSBPM Quality management and metadata management overarching process are sufficiently detailed to describe the complexity of SIDI/SIQual contents, given that it is not only a metadata system because it includes also quality metadata and standard quality indicators.

The paper will discuss the above mentioned issues, encountered in the design of the new system, as a case study of the joint application of GSBPM and GSIM standards, trying to enhance strength and weaknesses of the approach, with the aim to stimulate discussion on the issue and proposing solutions that could also bring to improvements to the standard models themselves.

Keywords

National metadata system, GSIM, GSBPM