The designed governance for a central metadata management system: Istat experience

Claudia Brunini – brunini@istat.it

The METADATA MANAGEMENT SYSTEM

- The metadata management system represents a key element for the modernization of NSIs.
- It should be defined as a COMMON STATISTICAL INFRASTRUCTURE (UN, Handbook on Management and Organization of National Statistical Systems, Version 2022/A., p. 520) that must be independent from each production process and capable to support them all.
- A metadata management system evolved as a common infrastructure has many key benefits. Each element of the infrastructure can support all statistical processes and these can use the resources in order to gain efficiency and reduce the costs. An additional advantage of a common statistical infrastructure is to promote harmonization across statistical processes, thanks to the opportunity of using common methods and standards.

The core modules of the system

- The core information of the Istat central metadata management system METAstat are referential, structural and terminological metadata. These information is captured and reorganized having standard GSBPM, GSIM and ISO as a reference.
- In an active system the statistical information is also linked to other types of information such as administrative, legal, technological etc. The standard GAMSO is used to model the connection between the core modules and the cross-cutting activities.

The statistical process: the cornerstone of the system

- In the system the referential, structural and terminological metadata are connected between each other through the statistical process.
- In such a way the process becomes the cornerstone of the central system.
- Since the metadata should be reused in each phase and by all statistical processes, the standard GSBPM is used to identify all the life cycle phases of the metadata in the processes. The roles and relative tasks are accurately identified for every GSBPM's phase and sub-process involved in the metadata management. A detailed task description accompanies each specific role.

The main principle of the governance (1)

- Looking at metadata, PROCESSES ARE SPLIT IN MANAGERS AND USERS. The first ones generate new metadata and become responsible of them, the second one use already existing metadata and can't modify them.
- All metadata activities are performed by the process manager and his team. He is responsible for entering the process into the system and completing all the information. He enters referential, structural and terminological metadata. He also takes charge of the changes and the definition of the states throughout the metadata life cycle.

The main principle of the governance (2)

- EACH METADATA OPERATION IS CENTRALLY SUPERVISED by the structure that is responsible for the control and validation of the metadata management system contents. This structure guarantees the availability of all metadata produced for managers and users inside and outside the Agency. It also ensures their correct update, standardization, harmonization, consistency and integration.

The structural metadata

- Each structural metadata is loaded into the system by a process, which becomes responsible for the metadata itself. This process is in charge of the management, i.e. its initial drafting and maintenance during all phases of the life cycle.
- The process responsibility role on the metadata is the connecting element between the registry of processes and the referential metadata with the structural metadata. Therefore, between the GSBPM standard, used to model the process phases, and the GSIM standard used to model the structural metadata.

The terminological metadata

- The central metadata management system is structured to support the semantically interoperability. For this reason, it is equipped with a terminological component where every term has a proper life cycle and is connected to the structural metadata, hence to the referential one.
- The main references for the terminological component are ISO 1087-2019 and ISO 25964-2013. The ISO 1087-2019 supplies instructions on how to correctly manage the terminology, the ISO 25964-2013 helps in documenting the semantic connections.

The cross-cutting activities

- The central metadata management system has the core mission to manage metadata from the statistical production. This kind of activity is one of the three overarching processes (as GSBPM model calls)with the goal to support the statistical production (the other two are quality management and data management). Other activities that also support the statistical production but are carried out at the level of the organization, are modelled by the standard GAMSO.
- A central metadata management system, in order to properly manage the metadata from the statistical production activity, needs to capture and contain also data and metadata from the cross-cutting activities. This establishes the connection between the two standards GSBPM and GAMSO.

Conclusions

- In a central metadata management system, the governance is a key element. A well-oriented governance is a crucial component necessary to achieve and support interoperability between the standards and, on a future note, with other components connected to metadata. The governance includes many key aspects such as legal and business policies, the active adoption of standards and roles which tasks should be well identified, recognized and institutionalized.

Thank you

brunini@istat.it

Istat | Istituto Nazionale di Statistica