How Metadata Systems Support the Modernization of Statistical Production: the Istat Experience

Marina Signore, Giovanna Brancato, Mauro Scanu, Giorgia Simeoni[1]

Abstract

The modernization process undergoing in several National Statistical Institutes (NSIs) strongly relies on the introduction of office-wide standards for the production and dissemination of statistics. NSIs are also increasing their cooperation for the development, governance and updating of common agreed standard models such as GAMSO, GSIM, GSBPM, and standard tools like SIMS - the European Integrated Single Metadata Structure for quality reporting relying on SDMX.

The paper describes how the above-mentioned supranational standards supported the design of the Istat unified metadata system - SUM, addressing the reasons why deviations from or ad hoc-tailoring of the standards were introduced in specific cases.

The SUM system manages the following metadata typologies: i) metadata related to data structure and content (roughly corresponding to "structural" metadata), ii) process-related metadata (roughly corresponding to "reference" metadata) and iii) business-related metadata (i.e. metadata supporting the management of statistical organizations). The two main components of the SUM system are: the SUM-MS managing the first typology of metadata and the SIDI-SIQual system, managing the other two metadata typologies including quality indicators.

The modernization process undertaken at Istat aims at overcoming the stove-pipe production processes and calls for a rethinking of the quality evaluation and quality control systems developed so far. A unified metadata system, such as SUM, can support such a modernization process and the paper provides some examples of possible uses.

Key Words: Metadata systems, modernization, GSIM, SIMS

1. Introduction

Office-wide standards are being developed and introduced by National Statistical Institutes (NSIs) in order to support the modernization of the production and dissemination of statistics. Given that NSIs face common problems and aim at common targets, cooperation among the statistical community is increasing remarkably also for what concerns the development and maintenance of standard models.

Particularly relevant models are those overseen by the UNECE High Level Group for the Modernisation of Official Statistics (HLG-MOS), namely the Generic Activity Model for Statistical Organizations (GAMSO), the Generic Statistical Information Model (GSIM) and the Generic Statistical Business Process Model (GSBPM). They provide generic models for statistical organizations to describe and organize i) their activities (GAMSO); ii) the statistical production processes or business processes in GSBPM terminology (GSBPM); iii) the information objects and flows (GSIM) and iv) the overarching framework of the Common Statistical Production Architecture (CSPA) which encompasses the other ones (UNECE 2015a, 2015b, 2013a, 2013b). NSIs are implementing the above standards to gain efficiency and

[1] Marina Signore, signore@istat.it; Giovanna Brancato, brancato@istat.it; Mauro Scanu, scanu@istat.it; Giorgia Simeoni, simeoni@istat.it; Italian National Institute of Statistics - Istat , via Cesare Balbo 16, 00184 Rome, Italy.

efficacy for modernized statistics, in the so-called standard-based modernization. At the same time, new requests were raised by NSIs to get support and guidelines in the standards implementation, priority setting and analysis of interrelationships among standards. Those issues are being addressed by one major collaborative project, undertaken under the HLG-MOS for 2016, named *Implementing Modernstats Standards*, targeting among others at developing a "Modernisation Maturity Model to help statistical organizations assess their current levels, and a Modernisation Roadmap to help them progress to the next levels as efficiently as possible"[2].

This paper describes how the above-mentioned standards supported the design and the development of Istat's corporate unified metadata system, named SUM (Sistema unitario di metadati, Signore et al., 2015). SUM is composed by two information systems , namely the SUM-MS system, that manages structural metadata (Scanu, 2015) and the SIDI-SIQual information system managing reference metadata and quality indicators (Brancato et al. 2006, Brancato et al. 2004). Even though the sub-systems differ for some IT solutions adopted and interfaces, they are highly integrated from a conceptual point of view, sharing whenever desirable the same concepts and definitions (e.g. populations, units,..).

As a matter of fact, with these "federate" systems we manage at corporate level the following metadata typologies (Signore et al., 2015):
i) metadata related to data structure and content (roughly corresponding to "structural" metadata),
ii) process-related metadata and standard quality indicators (roughly corresponding to "reference" metadata)
iii) business-related metadata (i.e. metadata supporting the management of statistical organizations)

The last category of metadata is still at an initial stage of definition and management in Istat corporate metadata system.

From the implementation point of view, SIDI-SIQual was conceived in the late '90s and the first release was in 2001, while SUM-MS has been first released in 2015. Therefore, they could benefit in a different way from the above-mentioned standards.

Particularly, SIDI-SIQual was developed much earlier before GSBPM and the mapping to GSBPM has been extensively described in Brancato and Simeoni (2012). Nevertheless SIDI-SIQual can still benefit from reference models such as GAMSO since business-related metadata could complement reference metadata and quality indicators. SIDI-SIQual is also the main repository for implementing standard quality reporting to Eurostat following the European Single Integrated Metadata Structure (SIMS).

For what concerns SUM-MS, structural metadata were modelled according to GSIM which represented the main standard reference.

The role of international standards in designing Istat's corporate information systems is described in more detail in section 2.

Section 3 describes how Istat's corporate metadata systems support the modernization process undergoing in our office. It also provides some considerations on how reference models such as GSBPM could be enhanced in order to better support modernization.

2. The role of standards in designing, implementing and maintaining information systems at Istat

[2] For details on the project *Implementing Modernstats Standards please refer to* http://www1.unece.org/stat/platform/display/hlgbas/Implementing+Modernstats+Standards

The standards mentioned in section 1 proved to be useful also in supporting the design and the maintenance of the corporate unified metadata system, developed at Istat.

Apart from GAMSO and GSIM, this section considers the European Single Integrated Metadata Structure (SIMS) for quality reporting developed by Eurostat and adopted by European NSIs to produce user- and producer-oriented quality reports for statistical products, according to the Article 12 of Regulation 223/2009 on European statistics.

In the following, some examples on how the standards were introduced are presented together with comments on any adaptions (as for GSIM) that were deemed necessary to enhance the applicability of the standard. Furthermore, it is suggested that enhancement of SIMS can derive from incorporating GSBPM.

2.1 Structural metadata in SUM and GSIM

The part of the SUM on structural metadata, that is SUM-MS, contains:

1. The definition of micro and macro data-structures produced along any statistical program undertaken by Istat.
2. The necessary concepts for micro and macro-data structures definitions.

Traditionally, at Istat these two kinds of metadata were not centrally managed. On the one hand, as in many other situations, there was a stove-pipe based management of metadata, where each pipe was a specific statistical program. On the other hand, metadata related to data produced in different process phases were also managed according to the storing and management facilities available for each specific phase, leading to a patchwork-like structural metadata management, actually unsuitable for any kind of industrialization process.

The SUM-MS was asked to modify this situation, centrally storing and managing definitions of data structures and the necessary concepts. Two main objectives were assigned to this new system: metadata harmonization between and within (along) statistical programs and traceability of each data production process.

These two objectives were perfectly in line with the GSIM part on Concepts and Structures groups, whose definitions and components were immediately adopted in the SUM-MS. Although the nature of GSIM is to describe all the information objects available in a statistical program, including also those of the business part, the use of GSIM was restricted only to those concepts that refer to data (hence, Concept and Structure groups). The Business and Exchange parts were not used, yet.

- The fact that GSIM organizes information objects as input and output of process phases facilitates traceability. Traceability should be also complemented by an appropriate language (e.g. the Validation and Transformation Language, VTL) that describes operatively what kind of transformations has been performed.
- The use of the same Concepts and Structures components of GSIM for the definition of the input and output data structures promotes comparison and consequently harmonization between and along data production programs.

Table 1 describes which concepts characterize micro and macro data sets as well as indicators, pinpointing their specific contribution to data comprehension.

Table 1.

Concepts roles	GSIM concepts for micro datasets	GSIM concepts for macro datasets
Terms that position a data	Statistical program	Statistical program

set in a process	Phase	Phase
	Reference time	Reference time
Definition of the data structure	Reference population	Statistical indicator (code list) with a full description of input and method of transformation
	Represented variables -categorical: add a code list -identifying: link to a population -numerical: describe the domain, unit measure and unit multipliers Other concepts: -weights -paradata	Represented variables -categorical: add a code list Other dimensions -time -operative
	Attributes	Attributes
Relationship with other data sets	Data input	(included in each statistical indicator in the dataset)
	Method of transformation	(included in each statistical indicator in the data set)
	Logical record (links to other populations observed in the same statistical program)	

Note that both data input and method of transformation are described at the indicator level in a macro data set. According to the method of transformation, the data input can be either a micro dataset (e.g. means, frequencies, totals,…) or other indicators (e.g. ratios, index numbers,…). Hence, it is the specification of the statistical indicator in its data content (as described in Signore et al. 2015) that allows the use of the same set of concepts (e.g. reference population, numerical variables,…) for both micro and macro datasets.

In populating the system it appeared clear that this set of concepts was sufficient for a complete description of each data, thus no other concepts were included (as stated in Signore et al. 2015 just one specific concept, the measure in a data structure of macrodata, was reinterpreted through the data content that formalizes each indicator produced as an output of a statistical process in terms of GSIM concepts). This set of concepts resulted also necessary: dropping anyone of them make it impossible to describe data as a product of a process step in a statistical program (a classic example is SDMX as described in Signore et al. (2015) whose concepts have been described as a grouping of GSIM concepts, useful for data and metadata exchange in the dissemination phase but unable to represent data as a product along a statistical program).

In order to help the description of the data production process, SUM-MS has been very clear in assessing if an output data set of a GSBPM phase was a macro (or in GSIM terms *dimensional*) or micro (in GSIM terms *unit*) data set. In fact, the first ones are characterized by aggregate indicators, usually playing the role of measure dimensions in a data structure. The description of their data content, as outlined in Signore et al. (2015) and described in detail in Scanu (2015), is essential for understanding their meaning in terms of the foundational concepts of GSIM (population, numerical variable,..). On a unit data set, indicators should not be present. The indicator (data content) is the key connecting an input data set into the output data set.

2.2 Business metadata in SIDI-SIQual and GAMSO

Business-related metadata are those metadata useful for the management of an NSI in planning, executing and assessing both statistical and support activities (Signore et al., 2015). They are partly accounted for in GSIM in the *Business* area that deals with the planning of statistical programs and the support processes undertaken to deliver those programs in order to satisfy statistical needs.

Given that business-related metadata arise from support activities such as administrative and management activities undertaken to support statistical production and dissemination, GAMSO seems to be the most suitable reference model to handle those metadata. Actually, GAMSO can support the identification and classification of relevant categories of business-related metadata since it complements GSBPM with those additional activities needed to support, manage and enhance statistical production in a statistical organization.

When handling business-related metadata, the amount of information potentially useful is huge. This requires an identification of those classes of business-related metadata that are more relevant to start with.

Following GAMSO, at Istat we started with those business-related metadata more closely related to the statistical production process and that were already managed in the internal planning system, namely metadata related to human resources, costs and duration of phases of the business process (see Signore et al. 2015).

Table 2 shows the areas of GAMSO for which we are going to analyse the business-related metadata. The GAMSO area *Capability management* is not taken into account because it deals with the development cycle of capabilities, and "When a capability is fully integrated in *Production*, its support is transferred to *Corporate Support*" (UNECE 2015b). It can be therefore considered that capabilities will start to produce business-related metadata on a regular basis once they are fully integrated in Production and managed in Corporate support.

As a matter of fact, we are not going to duplicate information or collection processes. The plan is to integrate different information systems using the more suitable IT solutions, e.g. via web-services or other ways to share information. The idea is to exploit some metadata that are already available in internal repositories and manage them in the corporate metadata systems.

The main objective is to complement the information on quality already available in the SIDI-SIQual system (i.e. the corporate information system that manages quality indicators and reference metadata) with some structured information from support activities (i.e. business-related metadata) related to resource planning and measurement of costs of the production of statistical products.

Table 2. GAMSO areas to be investigated for business-metadata

Area	**Activity**	**Sub-activity**
Strategy & leadership	Govern&lead	Define annual statistical programme Allocate project and programme portfolio budgets
Corporate support	Manage business and performance	Manage business performance Manage change and risk
	Manage finances	Manage procurement and contracts
	Manage humane resources	Manage employee performance Manage and develop skills
	Manage statistical methodology	Manage cross-cutting statistical methodologies
	Manage information and knowledge	Manage information standards and rights Manage metadata and data
	Manage consumers	Manage communications and media relations Manage stakeholder consultations Manage cross-product user support
	Manage data suppliers	Manage data sharing agreements
Production - GSBPM	Overarching quality management	Seeking and analysing user feedback Setting of global quality criteria

As known, GSBPM is part of GAMSO dealing with the area *Production*. The contribution of GSBPM in identifying business-related metadata is twofold: i) with regard to a single statistical process, business-related metadata come into play whenever a sub-process (or a process step) involves support activities or concerns support processes such as hiring or training interviewers (e.g. sub-process 4.2 Set up collection); and ii) as an overarching process when dealing with quality management. For instance, overarching quality management involves activities such as *Seeking and analysing user feedback* that generates business-related metadata or *Setting of global quality criteria* that can be better accomplished with the availability of information on constraints such as resources and costs or duration of external procedures (e.g. hiring interviewers).

In the present approach, business-related metadata are addressed to senior management for analyzing the performance and effectiveness of (groups of) statistical processes and for recommending improvements. Additionally business-related metadata associated to quality information can support strategic planning and performance assessment for the whole organization as discussed in Signore et al., 2015.

2.3 SIMS and its implementation in SIDI-SIQual

The Single Integrated Metadata Structure – SIMS (Eurostat, 2014 and Eurostat, 2015) is a dynamic and unique inventory of the statistical concepts that represents the ESS standard for reference metadata (see Fig. 1) with definitions and reporting guidelines.

SIMS includes all types of reference metadata (see SDMX Glossary, SDMX, 2016): Conceptual metadata (e.g. the Statistical Presentation section), Methodological metadata (mainly the Statistical Processing section) and Quality metadata (the sections devoted to the quality dimensions like Relevance, Accuracy, etc.).

Figure 1: SIMS 2.0 - Source: Eurostat (2015)

Item No	Concept name
S.1	**Contact**
S.1.1	Contact organisation
S.1.2	Contact organisation unit
S.1.3	Contact name
S.1.4	Contact person function
S.1.5	Contact mail address
S.1.6	Contact email address
S.1.7	Contact phone number
S.1.8	Contact fax number
S.2	**Metadata update**
S.2.1	Metadata last certified
S.2.2	Metadata last posted
S.2.3	Metadata last update
S.3	**Statistical presentation**
S.3.1	Data description
S.3.2	Classification system
S.3.3	Sector coverage
S.3.4	Statistical concepts and definitions
S.3.5	Statistical unit
S.3.6	Statistical population
S.3.7	Reference area
S.3.8	Time coverage
S.3.9	Base period
S.4	**Unit of measure**
S.5	**Reference period**
S.6	**Institutional mandate**
S.6.1	Legal acts and other agreements
S.6.2	Data sharing
S.7	**Confidentiality**
S.7.1	Confidentiality - policy
S.7.2	Confidentiality - data treatment
S.8	**Release policy**
S.8.1	Release calendar
S.8.2	Release calendar access
S.8.3	User access
S.9	**Frequency of dissemination**
S.10	**Accessibility and clarity**
S.10.1	News release
S.10.2	Publications
S.10.3	On-line database
S.10.3.1	AC1. Data tables - consultations
S.10.4	Micro-data access
S.10.5	Other
S.10.5.1	AC 2. Metadata - consultations
S.10.6	Documentation on methodology
S.10.6.1	AC 3. Metadata completeness - rate
S.10.7	Quality documentation
S.11	**Quality management**
S.11.1	Quality assurance
S.11.2	Quality assessment
S.12	**Relevance**
S.12.1	User needs
S.12.2	User satisfaction
S.12.3	Completeness and R1. Data completeness - rate for U
S.12.3.1	R1. Data completeness - rate for P
S.13	**Accuracy and reliability**
S.13.1	Overall accuracy
S.13.2	Sampling error and A1. Sampling errors - indicators for U
S.13.2.1	A1. Sampling errors - indicators for P
S.13.3	Non-sampling error and A4. Unit non-response - rate for U and A5. Item non-response - rate for U
S.13.3.1	Coverage error
S.13.3.1.1	A2. Over-coverage - rate
S.13.3.1.2	A3. Common units - proportion
S.13.3.2	Measurement error
S.13.3.3	Non response error
S.13.3.3.1	A4. Unit non-response - rate for P
S.13.3.3.2	A5. Item non-response - rate for P
S.13.3.4	Processing error
S.13.3.5	Model assumption error
S.14	**Timeliness and punctuality**
S.14.1	Timeliness and TP2. Time lag - final results for U
S.14.1.1	TP1. Time lag - first results for P
S.14.1.2	TP2. Time lag - final results for P
S.14.2	Punctuality and TP3. Punctuality - delivery and publication for U
S.14.2.1	TP3. Punctuality - delivery and publication for P
S.15	**Coherence and comparability**
S.15.1	Comparability - geographical
S.15.1.1	CC1. Asymmetry for mirror flows statistics - coefficient
S.15.2	Comparability - over time and CC2. Length of comparable time series for U
S.15.2.1	CC2. Length of comparable time series for P
S.15.3	Coherence- cross domain
S.15.3.1	Coherence - sub annual and annual statistics
S.15.3.2	Coherence- National Accounts
S.15.4	Coherence - internal
S.16	**Cost and burden**
S.17	**Data revision**
S.17.1	Data revision - policy
S.17.2	Data revision - practice and A6. Data revision - average size for U
S.17.2.1	A6. Data revision - average size for P
S.18	**Statistical processing**
S.18.1	Source data
S.18.2	Frequency of data collection
S.18.3	Data collection
S.18.4	Data validation
S.18.5	Data compilation
S.18.5.1	A7. Imputation - rate
S.18.6	Adjustment
S.18.6.1	Seasonal adjustment
S.19	**Comment**

All ESS standard templates for reporting metadata to Eurostat are being derived from SIMS. SIMS is a quite different kind of standard with respect to the UNECE models that are analyzed in the other sections of the present paper. GAMSO, GSBPM and GSIM are reference model that can be adopted by an NSI to facilitate some activities, e.g. designing the metadata system or supporting the modernization process, therefore are highly recommended standards. On the other hand, all NSIs are asked to provide Eurostat with quality reports following SIMS for the different statistical products. In these terms, the implementation of SIMS can be considered becoming gradually mandatory.

SIMS has been obtained by integrating and harmonizing the concepts of the two already existing ESS standard templates ESMS - Euro SDMX Metadata Structure (recommended for user-oriented quality reporting) and ESQRS – ESS Standard for Quality Reports Structure (recommended for producer-oriented quality reporting). Afterwards, in version 2.0, the alignment of SIMS (and consequently of ESMS and ESQRS) with ES Code of Practice has been improved by appropriately merging and re-ordering the statistical concepts, and also the coherence between ESMS and ESQRS has been increased mainly by adding to ESQRS the conceptual metadata from ESMS that were missing in ESQRS.

Thanks to the experience gained over the last decades in documenting reference metadata and quality indicators in SIDI-SIQual, Istat cooperated actively with Eurostat and other NSIs, in each phase of the definition of the SIMS standard, promoting: i) a continuously increasing horizontal (among domains) harmonization of contents; ii) the alignment with the ES Code of Practice; iii) the establishment of rules for vertical interoperability among national metadata systems and Eurostat system (Nielsen et al., 2015).

In the next, SIMS is expected to be improved with respect to the statistical business process documentation, that is currently organized in three main subsections: data collection, data validation and data compilation. Indeed, this section was born to document Eurostat internal statistical processes that are fairly simpler than NSIs statistical processes. This part would certainly improve by adopting a standard documentation model such as GSBPM. For example, for some sub-processes, like *5.1. Integrate data* or 5.*5. Derive new variables and units*, it would be also useful to have more detailed and structured descriptions in order to better represent new and more complex statistical processes that integrate administrative data and survey results.

As mentioned, reference metadata are documented in the SIDI-SIQual system in a standardized way. Thus, the statistical concepts included in the SIMS inventory have been mapped to SIDI-SIQual contents. As a result, the majority of the required items were found to be already documented in SIDI-SIQual and consequently an IT application to compile quality reports following ESS standard templates has been developed as a subsystem of SIDI-SIQual. It allowed i) to exploit the already available information; ii) to reduce the burden on statistical domain manager; and iii) to improve the Istat metadata asset (Simeoni 2013).

Indeed, a set of procedures has been developed to dynamically extract such information and re-use it for automatically filling-in the quality reports following ESS standard templates. In order to fill-in also the conceptual metadata required by ESMS (and in the next future by ESQRS, too), the integration with SUM-MS has also been exploited. In addition, the application allows to increase the level of harmonization and coherence among reports from different domains (horizontal): indeed, the contents of the SIMS concepts that require information at Institute level, such as the Confidentiality or the Dissemination or the Quality policy, are automatically filled in with the same statements formulated at centralized level by the corresponding manager.

Finally, the SIDI-SIQual subsystem for quality reporting can be considered a good example of "modernized" process from two different points of view: first of all, because it offers a centralized service

to all the statistical production processes substituting a process that was previously managed in a strictly stove-pipe approach, since each statistical domain manager provided the quality report to Eurostat independently; secondly, because the application has been designed to interact "vertically" with the ESS-Metadata Handler, the IT application provided by Eurostat to fill in the national quality reports.

3. How corporate metadata systems can support the modernization process

Istat has undertaken a modernization process in order to move from a stove-pipe approach to a new model that supports the industrialization and standardization of production processes. As described in Barcaroli et al. (2015), the modernization process implements Istat Business Architecture and focuses on a production system centered around the exploitation of statistical registers based on administrative data sources and integrated with surveys. Furthermore, a set of specialized functions and services such as methodology, IT, data collection and dissemination are being centralized and will be provided by the "corporate support service area" to the production units. These service exchanges will be regulated by Service Level Agreements and by Operational Level Agreements in order to rule the services and clearly assign responsibilities.

The modernization process can benefit from the availability of a corporate unified metadata system such as SUM that allows for traceability of data, transparency of production processes and quality assessment. Particularly, SUM can support the use of administrative data sources to produce statistical information and set up statistical registers as well as the integration process between different data sources such as different surveys, different administrative data sources or surveys and administrative data sources. The subsection 3.1 provides some ideas on how the adoption of GSIM in the SUM-MS can support the integration process.

The implementation of SUM, while supporting the modernization process, can generate feedbacks useful to the improvement of the same standard models from which both the modernization process and the metadata system took inspiration from, thus creating a sort of virtuous cycle. Subsection 3.2 presents some suggestions on how to enhance GSBPM to better document reference metadata and quality indicators of "modernized" statistical processes.

3.1 Relationship between modernization and the use of GSIM

As introduced in Section 2.1, SUM-MS adopted the Concept and Structure terms and definitions for structuring the system, so that any data is described according to the same concepts along all process phases of a statistical program.

The use of a coherent set of concepts for data description in all the GSBPM phases (or in other words without taking into account their nature of either micro data whose inputs are other microdata, macro data whose inputs are microdata, or macrodata whose inputs are other macrodata) allowed SUM-MS to be enriched with search functionalities whose aim is to find out any data sets that:

1. are produced by a statistical program
2. are produced in a specific time period
3. refer to a certain population
4. investigate a specific variable, whatever its nature (categorical, numerical,…)
5. are obtained from a data input
6. are obtained by means of a method of transformation
7. contain data where a variable assumes a specific category (e.g. all the indicators disseminated for the municipality of Rome in the NUTS)

As far as item 7. is concerned, it is not possible to search for all those data sets that include a pair of categories of two categorical variables. This is because the system registers for a data set what are the categories of each categorical variable actually in use, hence the system does not have information on how missing data are dispersed inside the data sets.

Search functionalities are not only appropriate for data retrieval, comparisons and metadata harmonization. These tools are the ones necessary for a fundamental aspect in official statistics modernization: data integration. Given that SUM-MS reports microdata reference population in Unit Data Structures, as well as reference populations of the macrodata indicators through their data content, SUM is appropriate as a tool for listing all the data sets, from data collection up to dissemination that refers to specific populations. In this way the researcher that needs to investigate a population has the complete overview of the data sets containing information on those populations and the statistical variables reported on them, which is the first step of a data integration project. Anyway other possibilities are also available, as queries focused on gathering all the data sets that study specific variables, and then looking at the reference populations attached to each detected data set in order to complete target populations as much as possible. Under this perspective the SUM-MS is a solution to the problem posed by researchers in the data integration area in cutting the usually large amount of time spent for data retrieval and harmonization.

3.2 Relationships between the modernization process and the reference metadata and quality documentation

The documentation of reference metadata and the computation of standard quality indicators are a precondition for transparency and are the backbone of quality assessment. They continue to play a fundamental role in the modernization process. In the Common Statistics Production Architecture, GSBPM is the reference model for the Business Architecture that defines what the organization does (statistics in the specific case) and how it is done (the statistical business process).

Reference metadata permit to track changes in the production processes, to document the responsibilities of production processes' activities, to provide evidence of the tools used to perform any operation in the statistical production process, to attest the quality control activities performed in order to prevent, monitor and evaluate errors. Thus, reference metadata support not only quality evaluation but also the assessment of efficiency in the production processes.

The standard quality indicators permit to monitor quality and performance thus allowing for the identification of benchmarks and an evaluation of the impact of the modernization on some quality aspects.

However the model and the approach used so far in GSBPM as well as in SIDI-SIQual, may result not completely appropriate and require some tailoring and adjustments, also because there are some organizational factors that become relevant when moving from a stove-pipe production processes to a more integrated one.

GSBPM captures the main activities performed within a statistical process using administrative data, e.g. the development and maintenance of statistical register. For such processes the most peculiar sub-processes included in GSBPM are: *1.5. Check of data availability* and *5.5. Derivation of new variables and units.*

When the modernization process and the increased use of administrative data imply a different organizational model, i.e. the centralization of some services such as the data acquisition or the data validation, in order to properly document the process, the model has to be reviewed.

For example, in the centralized acquisition of administrative data to serve more statistical purposes, a new process can be defined, concerning: an internal consultation on the administrative data needs and requirements; the setting of a dialogue and agreements with the administrative data owners; the technical checks and quality controls on the acquired data sets; the integration of the acquired data sets with additional variables and/or with different classifications; the monitoring of subsequent supply of the same data sets; the release of administrative data checked and pre-treated; the monitoring on the use of the data provided for internal statistical production. Such a process, preliminary to any other internal use of the administrative data, can be documented by forcing the interpretation of some of the GSBPM cells, and exploiting the cyclical nature of the model, or it can be more easily documented by considering new sub-processes. The relationship between this process and other statistical processes using administrative data can then be ruled by means of the already mentioned internal formalized agreement (e.g. Service Level Agreements), and the quality and performance measured throughout proper indicators (Barcaroli, 2015).

Not to be neglected that, when using administrative data, some of the above listed activities can be carried out outside the statistical organization, thus becoming more difficult to be documented and assessed for quality.

With respect to the standard quality indicators that support the use of administrative data, there is a wide availability of measures, useful for different purposes (Daas et al. 2009, Daas and Ossen 2011, Admin Data ESSnet 2013). The survey-specific quality indicators can then be replaced by more meaningful quality measures such as: technical checks, input data quality indicators, indicators on the source of errors when integrating the administrative data in the statistical processes.

The comparison of quality in terms of the efficiency, costs, accuracy under the modernization approach with respect to more traditional survey-centered approach, may pose some difficulties. Indeed, whether it is quite straightforward to compare time, cost and burden under the two models, measures to compare statistical outputs' quality under the two approaches are not yet available in all situations.

Finally, reference metadata together with business-related metadata can support the design and implementation of the modernization program. For example, being able to jointly analyze the following classes of metadata for each statistical process would support a better design of a common service for the data collection phase, namely *i)* number of units to be interviewed, expected number of respondents, data collection periodicity, data collection technique and software, concerning reference metadata and quality indicators, and *ii)* contracts to be signed for outsourcing service of data collection, time of the year for each survey data collection, constraints for the use of certified mail, for business-related metadata.

4. Concluding remarks

As described, in the development of Istat metadata corporate system SUM, composed by SUM-MS and SIDI-SIQual, international standards played a guiding role. Particularly, the standards were useful at different stages: e.g. in the design phase (as for GSIM in supporting the metadata modeling in SUM-MS) or in the alignment phase of existing systems to emerging needs (as for GAMSO in identifying relevant classes of business-related metadata to be managed in SIDI-SIQual in addition to reference metadata and standard quality indicators).

In Istat experience, metadata information systems needed to evolve in order to keep satisfying users' needs in a changing landscape. Thus, international standards and reference models were useful also in maintenance of information systems.

Finally, the international community is working on defining frameworks and models to support modernization of official statistics. In this respect, corporate metadata systems developed alongside the standards can be exploited to facilitate the modernization process itself. Some examples with regard to Istat experience were provided given that Istat recently launched a modernization program and is currently reorganizing both statistical and support activity according to its Business Architecture.

We are aware that there are a lot of challenges ahead in implementing the modernization program and priorities should be carefully set. Nevertheless, a big contribution can still derive from the cooperation of the international scientific community.

References

Admin Data ESSnet (2013). *"WP6 Quality Indicators when using Administrative Data in Statistical Outputs. Deliverable 6.3/2011: Guidance on the accuracy of mixed-sources statistics"* (at http://essnet.admindata.eu/WorkPackage/ShowAllDocuments?objectId=4257)

Barcaroli G., Falorsi P.D., Fasano A., Mignolli N. (2015). "A Business Architecture Model to support the Modernisation Project within the Italian National Institute of Statistics" – Istat 60th World Statistics Congress – ISI2015, Rio de Janeiro, 26-31 July 2015

Brancato G., Simeoni G. (2012). "Istat statistical process modelling and the Generic Statistical Business Process Model: a comparison." European Conference on Quality in Official Statistics, Q2012, Athens, Greece, 29 May - 1 June, 2012

Brancato G., Carbini R., Pellegrini C., Signore M., Simeoni G. (2006). "Assessing quality through the collection and analysis of standard quality indicators: the Istat experience", Proceedings of the European Conference on Quality in Survey Statistics (Q2006), Cardiff, UK, 24-26 April 2006

Brancato G., Pellegrini C., Signore M., Simeoni G. (2004). "Standardising, Evaluating and Documenting Quality: the implementation of Istat Information System for Survey Documentation – SIDI", European Conference on Quality and Methodology in Official Statistics Q2004 Mainz, Germany 24-26 May 2004

Daas P., Ossen S. (2011). "Report on methods preferred for the quality indicators of administrative data sources", Blue – ETS Project, Deliverable 4.2.

Daas P., Ossen S., Vis-Visschers R., and Arends-Tóth J. (2009). "Checklist for the Quality evaluation of Administrative Data Sources", Statistics Netherlands, The Hague /Heerlen, 2009

Eurostat (2015). "Single Integrated Metadata Structure v 2.0 – ESMS 2.0 and ESQRS 2.0- 2015" http://ec.europa.eu/eurostat/documents/64157/4373903/SIMS-2-0-Revised-standards-November-2015-ESSC-final.pdf/47c0b80d-0e19-4777-8f9e-28f89f82ce18

Eurostat (2014). "Single Integrated Metadata Structure and its Technical Manual"– 2014
http://ec.europa.eu/eurostat/documents/64157/4373903/03-Single-Integrated-Metadata-Structure-and-its-Technical-Manual.pdf/6013a162-e8e2-4a8a-8219-83e3318cbb39

Nielsen M.G., Simeoni G., Linden H., Zdrentu B. (2015). "ESS reference metadata: implementation of standards and interoperability of national and European metadata systems." Workshop of the Modernisation Committee on Standards: International Collaboration for Standards-Based Modernisation, Geneva, 5 – 7 May 2015

Scanu M. (2015). "GSIM implementation in the Istat Metadata System: focus on structural metadata." UNECE Workshop of the Modernisation Committee on Standards: International Collaboration for Standards-Based Modernisation, Geneva, 5 - 7 May 2015

SDMX (2016). "SDMX Glossary" Version 1.0 February 2016 https://sdmx.org/wp-content/uploads/SDMX_Glossary_Version_1_0_February_2016.pdf

Signore M., Brancato G., Scanu M. (2015). "Statistical Metadata: a Unified Approach to Management and Dissemination" Journal of Official Statistics, Vol. 31, No. 2, 2015, pp. 325–347

Simeoni G. (2013). "Implementing ESS standards for reference metadata and quality reporting at Istat" UNECE Work Session on Statistical Metadata, Geneva, 6-8 May 2013

UNECE (2013a). “GSBPM v5.0”. http://www1.unece.org/stat/platform/display/GSBPM/GSBPM+v5.0
UNECE (2013b). “Generic Statistical Information Model GSIM v1.1” http://www1.unece.org/stat/platform/display/metis/Generic+Statistical+Information+Model
UNECE (2015a). “Common Statistical Production Architecture “ http://www1.unece.org/stat/platform/display/CSPA/Common+Statistical+Production+Architecture+Home
UNECE (2015b). “GAMSO Generic Activity Model for Statistical Organisations” Version 1.0: 1 March 2015 http://www1.unece.org/stat/platform/display/GAMSO/GAMSO+Home