

Journal of Official Statistics, Vol. 31, No. 2, 2015, pp. 325–347, http://dx.doi.org/10.1515/JOS-2015-0020

Statistical Metadata: a Unified Approach to Management and Dissemination

Marina Signore¹, Mauro Scanu¹, and Giovanna Brancato¹

This article illustrates a unified conceptual approach to metadata, whereby metadata describing the information content and structure of data and those describing the statistical process are managed jointly with metadata arising from administrative and support activities. Many different actors may benefit from this approach: internal users who are given the option to reuse information; internal management that is supported in the decision-making process, process industrialisation and standardisation as well as performance assessment; external users who are provided with data and process-related metadata as well as quality measures to retrieve data and use them properly. In the article, a general model useful for metadata representation is illustrated and its application presented. Relationships to existing frameworks and standards are also discussed and enhancements proposed.

Key words: Generic Statistical Information Model (GSIM); Generic Statistical Business Process (GSBPM); organisational and support processes; SDMX; paradata; quality indicators.

1. Introduction

Metadata systems are extremely important management tools for all public and private institutions managing statistical data, and especially for today's National Statistical Institutes (NSIs). These systems are vital for the dissemination of statistics; they are the pillars of statistical data warehouses, and are gaining increasing importance in statistical data exchange among different organisations, especially at supranational level, where they are used in addition to facilitate communication and a common understanding of the data being exchanged. Important metadata system examples are described in the 18 statistical metadata case studies available in the United Nations - Economic Commission for Europe (UNECE) statistics wiki METIS (www1.unece.org). Many NSIs rely on sets of specific metadata systems, such as the eight metadata systems declared by Statistics Sweden, including one for classifications (KDB), one for products (the product data base) and one for governance (FMOD, applied in the management of processes related to the statistical production process). Others are mainly composed of a unique centralised system, such as the Statistics Canada Integrated Metadata Base (IMDB), obtained through the consolidation of already existing separate metadata systems, whose aim is to display data sources and methods for each statistical programme and survey, collecting reference metadata including survey methodology and data accuracy, definitions of concepts, variables and classifications for all subject-matter areas (domain-specific metadata

¹ ISTAT, Via Cesare Balbo 16, Rome 00184, Italy. Emails: signore@istat.it, scanu@istat.it, and brancato@istat.it

systems for business or social surveys exist, but they are highly integrated with IMDB, see, Greenhough et al. 2014). All these metadata systems strive to rely on conceptual models that act as current international standards: for example, European NSIs should be compliant with the European Statistics Code of Practice (CoP) that sets the standard for developing, producing and disseminating European statistics (Eurostat 2011). More generally, international institutions and NSIs worldwide agree on the Generic Statistical Business Process Model – GSBPM (UNECE 2013a) and the Generic Statistical Information Model – GSIM (UNECE 2013b). However, these standards and the related metadata systems can still be improved with regard to the relationships between different metadata typologies, such as those specifying the meaning of data and those describing the underlying processes (Signore et al. 2013a). The Generic Activity Model for Statistical Organizations – GAMSO (UNECE 2015), while only at version 0.2, seems to be a promising endeavour in connecting activities such as Strategy, Capability and Corporate Support to statistical production, thus supporting the views expressed in the present article.

In this article, we propose a unified conceptualisation of metadata that allows a thorough description of data and processes (both of statistical and support nature). Well-known metadata classes are reorganised and the role of business-related metadata, that is, metadata supporting the management of statistical organisations, is emphasised.

More specifically, three main metadata typologies (i.e., metadata related to data structure and content, process-related and business-related metadata) are discussed in this article. They are introduced in Section 2 and compared with the GSIM conceptual framework. We propose a unified metadata management system that comprises activities both at process/product level as well as at the institutional level. In this respect, this approach could be useful for assessing, for instance, the compliance of ESS Members to the European Statistics CoP that involves assessment at the abovementioned levels. One example of the applicability of this model is furnished by the unified metadata system developed by the Italian National Institute of Statistics (Istat) described by Signore et al. (2013b).

In Section 3, we introduce a metadata conceptualisation useful for connecting the three metadata typologies to the statistical process phases. After this, we present a model for describing metadata related to data structure and content as well as data transformations along the statistical process (Section 4). Existing limitations in standards such as GSIM are discussed and possible enhancements are proposed.

Process-related metadata, including quality indicators, and business-related metadata are described in more detail in Sections 5 and 6, respectively. As a result of the proposed approach, the ties between metadata and quality are given greater importance. Even though there are increasing examples of a joint use of process-related metadata and quality indicators (see, for instance Götzfried et al. 2011), a common conceptualisation is useful for both disciplines, which to a large extent developed independently of one another. For this reason, we propose some enhancements to GSBPM. One consequence of such an integrated approach is a broader perspective: quantitative information (i.e., quality indicators and paradata, Couper 1998) should be considered alongside the abovementioned typologies of metadata (qualitative information).

In Section 7, we comment on the joint use of process- and business-related metadata for improving statistical processes and products. We conclude with some final remarks in Section 8.

2. A Unified Metadata Management System

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

Hence the need to explicitly consider three metadata categories is advocated:

- 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.

This approach matches the definitions provided by Androvitsaneas et al. (2006) and SDMX (2009), as well as the information objects identified in GSIM (UNECE, 2013b), but is more suitable for a unified vision. In particular, metadata related to data structure and content correspond to respectively structural and conceptual reference metadata in Androvitsaneas et al. (2006) and SDMX (2009) terminology. Process-related metadata correspond to methodological and quality-reference metadata. Business-related metadata are not explicitly considered in Androvitsaneas et al. (2006) and SDMX (2009) terminology. Process-related metadata they are identifiable within the business top-level group of GSIM, although they are not fully defined. Figure 1 maps the metadata categories as defined in this article and the main information objects included in the four top-level GSIM groups.

This classification calls for a unified metadata management system as outlined in Figure 2, in which metadata typologies are linked to the institutional macro phases and to generic statistical process phases. Connections with the areas of the European Statistics Code of Practice (Eurostat 2011), namely institutional environment, statistical processes and statistical outputs, are also represented in Figure 2. In particular, planning and assessment are performed at an institutional level, thus reflecting strategic and overall activities, while *execution* pertains to the generic statistical process (and related outputs). The latter also includes a specific planning phase, identified by the GSBPM phases (UNECE 2013a) "Specify needs", "Design" and "Build", as well as a proper assessment phase, "Evaluate", which are all performed at a process level. The need to go beyond the generic statistical process to model metadata was already recognised by Androvitsaneas et al. (2006), who introduced the concept of the statistical system as a whole and of quality dimensions such as metadata on objectivity and credibility (pertaining to the institutional level). According to Sundgren (2004), the metadata managed in a system should be captured as early as possible in the process of developing, implementing and operating a production system; moreover, the same metadata should not be captured more than once.



GSIM top-level groups and		Proposed				
	information objects	classification				
BUSINESS	Statistical Program	Business-related				
	Statistical Support Program					
	Statistical Need	metauata				
	Business Process	Process related				
	Process Step					
EXCHANGE	Information Provider					
	Information Consumer					
	Exchange Channel metadat					
	Provision Agreement	incladata				
	Product					
STRUCTURES	Referential Metadata Set					
	Referential Metadata Structure					
	Information Resource	Metadata related to data structure and content				
	Data Set					
	Data Structure					
CONCEPTS	Variable					
	Population					
	Concept					
	Unit					
	Statistical Classification					

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



Fig. 2. The Istat metadata management system

One possible application of the proposed model is the Istat Unified Metadata System (Sistema Unitario di Metadati, SUM; described by Signore et al. 2013b). 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).

3. Modelling Metadata in the Statistical Business Process

The model for metadata identification described in the present section highlights the relationships between metadata related to data structure and content and the process- and business-related metadata that will be described in the next sections.

The proposed approach, properly adapted to building a model for statistical business process description and relative metadata, traces back to Saint Thomas Aquinas's eight circumstances determining the ethic of an action, and is commonly used in the fields of journalism and police investigation as questions for gathering basic information. This paradigm is not new and has also been used in similar contexts, for example, in metadata classification (Statistics Sweden 2008) and in other applications, for example, in the definition of a framework for Enterprise Architecture (Zachman 1987).

According to GSIM (UNECE 2013c), a business process is "[t]he set of Process Steps to perform one or more Business Functions to deliver a Statistical Program Cycle or Statistical Support Program". In other terms, the process steps can be viewed as a sequence

of logically connected elementary operations, or subprocesses in the GSBPM terminology (UNECE 2013a), aimed at producing statistical information, that is, statistical product or output. Besides these, other support processes exist that can be directly oriented towards the production goal or that are more cross-sectional, such as staff training. GSIM partly recognises the role of such processes, introducing the Statistical Support Program, whereas they are not explicitly included in GSBPM.

In order to categorise metadata within the statistical business process or relate them to it, the following fundamental questions should be answered: *cur, quid, quomodo, quibus auxiliis, quis, ubi, quando,* and *quantum* (respectively: why, what, how, by which means, who, where, when and how much). This set of questions allows us to exhaustively represent a circumstance, fact, situation and, in the field of official statistics production, a statistical business process. However, additional specifications have to be added concerning the quality dimension, that is, the level of performance attained and the fitness for use of the produced statistical data.

Cur (why) – In the philosophical interpretation, this element represents the cause of the action. Translating this concept into the metadata model, it represents the main production objective of the statistical process, responding to both users' needs and national or international regulations.

Quid (what) – The *what* element identifies data at any stage of the production process. It can refer to final products as well as to intermediate outputs. Data are formally defined by metadata related to data structure and content (Section 4) and described for users by process-related metadata (Section 5).

Quomodo (how) – The *how* component is the pivotal element, representing the way activities are carried out to meet the *why* objective. These activities can vary in nature: *i*) statistical operations; *ii*) management or administrative activities (included in GSIM as Statistical Support Program); *iii*) quality control actions.

Statistical activities are operationalised in elementary operations describing a statistical process, usually arranged in hierarchical structures, thus facilitating their organisation (UNECE 2013a; Brancato and Simeoni 2012). They represent the key elements for the description of statistical business processes; together with other components (*who, by which means*) they constitute the core of process-related metadata. Depending on the objective, elementary operations can have a more generic nature, as in GSBPM, or can be more detailed to provide hints about the methodology, as in the Istat documentation system (Brancato and Simeoni 2012). A thorough description of the methodology supports the construction of metadata related to data structure and content on the transformations the data undergo (see, Section 4). For example, in the phase "Process", process-related metadata describe that a given editing and imputation procedure has been applied, for instance "Deterministic error and outlier detection and imputation based on deterministic rules (IF-THEN)"; metadata related to data structures and content describe the variables involved and the exact parameters of the rules, thus supporting a full understanding of the data.

However, there are a number of support activities of an administrative and management nature concurrent with the statistical production. One example is the procedure needed to select a private company to carry out data collection in outsourcing. These operations are among the key elements of business-related metadata (see, Section 6). Finally, during survey execution, an extensive number of activities is devoted to quality controls, constituting the third type of the *how* elements. Similarly to statistical operations, quality control activities can be structured hierarchically (Brancato and Simeoni 2012; D'Angiolini et al. 1998).

Each *how* element can be ordered. Although, as already pointed out, operations or subprocesses follow a logical sequence, they may be performed in different orders. In addition, some operations can be performed more than once, generating cycles (UNECE 2013a).

The *how* elements are strictly connected to the *what* elements, the latter representing inputs and outputs (UNECE 2013a). The output of each operation or activity is the result of a transformation process and represents the input for the next one.

Quibus auxiliis (by which means) – In the proposed model, this element is represented by the IT tools used to perform the elementary operations, that is, a software application or a system, implementing one or more methods. For example, for the CAPI technique (*how*), the software used to implement the electronic questionnaire (*by which means*) can be specified. IT tools' characteristics and rules for their use within the administration are usually documented and stored in software and application catalogues.

Quis (who) and ubi (where) – The who element represents what GSIM identifies as the Agent. The following who elements can be defined: *i*) the person/s responsible for the operation; *ii*) the person/s executing the operation; *iii*) other actors involved in the operation.

The *where* element specifies the organisational unit where the activity is carried out and is strictly related to the *who* element.

For instance, persons in charge of and performing the activities may belong to the same entity, but not necessarily so. In addition, depending on the organisational structure, an agent such as the interviewer may belong to different units within the statistical office, or may depend on an outsourcing company in charge of the interviewing activity.

The *who* and the *where* elements may assume increasing importance given the tendency to abandon the classical stovepipe production process model in favour of adopting more integrated production models, the former being characterised by a concentration of responsibilities, the latter by more distributed tasks. When external bodies are involved in process operations, the identification of responsibilities may be attributed in administrative documentation, such as contracts setting out the terms and conditions of the service (e.g., when resorting to outsourcing). Finally, some elementary operations may involve actors external to the Institution, such as the reporting unit in the data collection activity.

Quando (when) – The *when* element is vital metadata to properly document validity of elementary operations, as well as changes in procedures and methodologies. Each elementary operation and its characteristics therefore need to be associated to a time reference or a validity period, as also proposed by Banca D'Italia (2007).

Periodicity is another *when* element, that is, the frequency of an operation (release, estimate, etc.). In the majority of statistical processes, all operations (e.g., data collection, processing and dissemination) are performed according to the same periodicity (monthly, yearly, etc.), although there are cases in which this is not true. In addition, the concept of estimate periodicity can be defined, that is, the time frequency of results or estimates released to users. It corresponds to the Frequency of data compilation in the Metadata Common Vocabulary terminology (SDMX Content-oriented guidelines 2009).

Quanto (how much) – This element is proposed to document resources (human, IT) and costs (financial, time) required to carry out statistical, administrative and quality control activities. It represents the typical business-related metadata useful for measuring efficiency and performing cost analysis.

4. Metadata Related to Data Structure and Content

Metadata related to data structure and content define statistical data. These data are usually structured in a table of macro data or a data set of micro data. Hence, their primary content consists of metadata defining i) the structure and ii) the structure components of the data structure by means of the appropriate concepts. These metadata are defined along the lines of well-established models, such as UNECE (1995) and GSIM (UNECE 2013c). The approach outlined in this article pinpoints, and wherever necessary reorganises, some aspects of these models in order to foster industrialisation. The main aim is to ensure data traceability along a data production process, and metadata reuse in different phases of a process and among processes. In order to be able to trace metadata along the production process, we support the GSIM idea of describing data as the input or output of a process, so that:

- any output is a function of one or more inputs;
- the function transforming the inputs into an output is an operator, that is, a method where all the parameters characterising a transformation should be carefully described, so the same result can be reproduced whenever necessary.

These operators include: population subsetting, variable transformations (variable sums, new categorisations including coding, etc.), microdata integration, imputation and editing of micro- or macrodata, transformations useful to avoid data disclosure, transformations from micro- to macrodata (averages, medians, variable totals, concentration indexes, variances, etc.), transformations from macro- to macrodata (balances, ratios, index numbers, percentage variations, etc.).

Following this approach, any output is connected to other data according to a backwardoriented "plug in" approach (from dissemination to collection). This is consistent with UNECE (1995), making explicit the way data are transformed through the operator (using notation in UNECE, 1995, a macrodata output is defined by the formula < O(t), v(t), f >where O(t) are microdata at time t, v(t) are variables observed on the units of O(t) and f is an operator of aggregation from micro- to macrodata). The use of a syntax based on statistical concepts is appropriate for describing the input and output content of an operator, enhances the possibility of reusing the same metadata throughout the whole data production process and allows metadata harmonisation between different data production processes. This syntax is mostly available in GSIM and includes the whole set of Concept metadata, some GSIM Structure Metadata (i.e., those with a specific focus on data structures both for micro- and macrodata), and specifications of other concepts related to data transformation along the process phases, such as the GSIM Process Step and Rule.

As a matter of fact, GSIM does not directly address transformation tools, instead leaving them to the GSBPM description, which is usually quite generic and not operative. These aspects are deemed to be crucial to enable a statistical system to trace data and the corresponding metadata transformations along the statistical process. For this reason, the main concepts for describing the content of a data structure are proposed in the following, pinpointing the main differences with the corresponding terms in GSIM.

Reference population – Data (either micro- or macrodata) may have a reference population, that is, the set of individuals, entities or objects on which a set of phenomena are observed (for microdata) or to which a statistical operation is applied in order to obtain macrodata. The concept of reference population includes the GSIM concepts Unit, Unit Type and Population. The distinction between Population and Unit becomes apparent once a population term is associated to either a microdatum or a macrodatum, respectively. As far as Unit Type is concerned, this concept is managed through hierarchical (subgroup) relationships as well as relationships between entities.

Variable – Our concept is in line with GSIM. The variable Value Domain specifies the statistical variable type (categorical or numerical/quantitative).

Classifications – Classifications are managed mostly, though not completely, as in GSIM. Classification versions are organised in levels containing mutually exclusive and collectively exhaustive categories. The most relevant exception compared to GSIM is the possibility of defining a classification version not only in a strictly hierarchical way (i.e., lower-level categories are subcategories of a category the next level up). In this context it is mandatory for a classification to have a lowest level of mutually exclusive and collectively exhaustive categories that are not split into subcategories (elementary events, as defined in probability theory). Any other category in use for the definition of at least one figure can be obtained by the set operations of complementation, union of and intersection of finitely many sets of events (i.e., categories in the algebra of the events as defined in probability theory). These categories should be included either in the classification (if a level of higher order than the lowest with such a category can be defined) or in a classification variant. Higher-order levels actually in use are not necessarily organised in a hierarchical way: complex graphs can describe the relationship between levels. Classification variants should not necessarily fulfil the rule of being composed of level(s) consisting of mutually exclusive and collectively exhaustive categories, as is the case for the list of classification categories actually in use in a disseminated contingency table.

Operator – This concept is useful for defining data on the basis of the way it was transformed from the previous phase in the statistical process, with all its parameters.

Data Content (Measure) – Data Content is a key aspect to be included among metadata related to data structure and content: it is defined as the combination of input data and method of transformation (statistical operator), and feeds the GSIM concept Measure in a data structure. Based on the description of output as a function of input, this approach identifies different forms of the Data Content. In the case of macrodata, the Data Content (macrodata output) has a simple and a composite form. Simple macrodata result from the combination of reference population, a numerical variable of interest (if not present, the counting variable) and the statistical operator used to pass from the micro- to the corresponding macrodata: this is the case, for instance, for average household income (households = population; income = statistical variable; average = statistical operator). Composite macrodata are instead described by the operation between already existing macrodata, as is the case for index numbers, ratios, balances, etc.: for instance, the net occupancy rate of beds in hotels and similar accommodation is defined as the ratio (statistical operator) between the nights spent

in hotels and similar accommodation in a month (first macrodatum) and the number of beds on offer in the given month (second macrodatum). Each component of a composite macrodatum inherits its detailed description as a simple (as in the previous example) or composite macrodatum, depending on their nature. The Data Content for microdata results from the combination of reference population and the set of numerical variables associated to the population units, as well as the statistical operator used for their production. The importance of the interaction between these elements is clearly stated in metadata literature (e.g., in Section 1.2.2 in UNECE 1995, and references therein), and traces back to the foundation of statistical inference: for example, Fisher (1925) states that "Statistics may be regarded i) as the study of populations, ii) as the study of variation, iii) as the study of methods of the reduction of data", and gives definitions of and justifications for this in the introduction to the monograph. While GSIM declares that "measures correspond to Represented Variables with uncoded Value Domains (Described Value Domains)" (UNECE 2013b, item 101), the proposed approach is to code each output, relate it to its statistical description (in terms of reference population, statistical variables, statistical operator and so on) and allow its use as a tool for easing metadata traceability and reuse.

Data Structure – Data structure is the detailed description of the content of a data set of either macro- or microdata produced alongside the statistical process. There are essentially four distinct groups of data produced in a statistical process: collected data (Collect phase); validated data (Process phase); data analysed by means of a statistical procedure (Analysis phase); disseminated data (Dissemination phase). Data sets in the previous four phases can be also archived; however, this is an overarching activity (see, GSBPM) and not relevant in this framework. In line with GSIM, two mandatory components are to be included in a data structure: measure and dimensions. The measure defines the actual content of the table, and is consequently described by the Data Content (with the indicator or the list of indicators of a macrodata set, or the variables observed on a microdata set). The dimensions are aimed at clearly defining how the measure is further disaggregated according to a set of categorical statistical variables or other coded concepts. Macrodata dimensions consist mainly (but not only) of a set of categorical variables crosscutting the measure (for instance average household income per region, number of household components and household typology). Microdata also include a unit identifier among dimensions, in order to appropriately list the set of microdata records in the data set and possibly link them with those in other data sets. A fundamental dimension of both microand macrodata sets is time. In order to better specify the content of a data set, or of each data-set cell, a data structure can also have attributes such as the observation status (if a number is provisional or definitive) or the confidentiality status of a cell.

Other concepts – In addition to the previous list of metadata, our approach also considers all concepts useful for a data structure description, such as those related to time (from the most general as frequency and time period, to specific ones such as school year or the edition of national accounts data), to visualisation issues (number of decimals, unit multiplier, etc.) and comprehension issues (observation status, confidentiality status, etc.) with their corresponding code lists.

This type of metadata organisation as well as the strict definition of Data Content represents a useful tool for tackling a number of issues. First of all, the statistical content of a data set is maintained in a unique list (Data Content) and it is therefore easy to see actual

outputs (or, in other words, the set of products obtained in each phase), while there is no need to consider intersections between different metadata in order to reach this goal. Secondly, users find the content of each number in a data set in the description of a single item of the Data Content dimension, simplifying the understanding of data: in other words, our approach does not allow for more than one GSIM measure in a data structure, and this measure is organised as a Data Content. Furthermore, the organisation of Data Content helps data producers to always be exhaustive in the data definition, without taking anything for granted. Finally, the organisation of the data structure as explained above addresses one of the aspects highlighted in Gelsema (2012): describing the connection between data and metadata content in the flow of a statistical process in a functional way. This last aspect plays a fundamental role in the management of metadata related to data structure and content: these should not be used just for the static description of a data set, but should show how data and the associated metadata evolve and change along the production process steps (see, Bergamasco et al. 2013 for a description of the metadata relationship throughout different data production phases). As a real-life example, the structural metadata organisation introduced in this section has been used to model metadata in SUM and in the Istat corporate dissemination system I.Stat (http://dati.istat.it). This approach has been really useful in creating a metadata system with specialised search tools.

One key issue is using the organisation of metadata related to data structure and content depicted in this Section in accordance with the available metadata standards, such as SDMX and DDI. The Appendix shows how SUM relates metadata related to data structure and content and the SDMX artefacts.

5. Process-Related Metadata

In the last years, activity on reference metadata (Androvitsaneas et al. 2006) has been focused on conceptual harmonisation across the European Statistical System, as requested by 2009 Commission Recommendation on reference metadata (European Commission 2009). This recommendation invited NSIs to adopt the set of statistical concepts and subconcepts attached to the recommendation based on the Euro SDMX Metadata Structure, ESMS (Eurostat 2009). More recently, within the framework of quality reporting, an ultimate reporting strategy is being fostered, integrating these concepts with additional ones, as well as with quality indicators such as sampling error, response rate, timeliness and punctuality among others (Götzfried et al. 2011). With respect to process-related metadata, the current standards are thus provided on the one hand by Eurostat reference metadata templates and on the other hand by the GSBPM business process documentation.

The framework for modelling metadata described in Section 3 corresponds to a great extent to Eurostat as well as to the GSBPM standards. The area concerning quality management represents one exception. In GSBPM, this is defined as an overarching process, providing little coverage of quality standards (Eltinge et al. 2013). We believe that integrating the business process model with quality activity and quality indicators is necessary in order to correctly interpret and evaluate standard quality indicators in the light of the relevant metadata, and set improvement goals based on objective information.

As an example of this integration, Istat has developed a system named SIDI-SIQual, which currently collects, stores and disseminates most of the process-related metadata and

quality indicators included in the abovementioned standard structures (Brancato et al. 2004; D'Angiolini et al. 1998). The system is developed according to the model presented in Section 3 and represents a major component of SUM.

Following the approach presented in this article, this system documents quality not only through quality indicators but also in terms of quality control actions, that is, activities aimed at: *i*) preventing errors; *ii*) monitoring process operations and *iii*) evaluating quality. In order to describe all the statistical activities, a hierarchical thesaurus of standard terms, from planning to dissemination, is available (operations thesaurus) whose mapping to the GSBPM items is ensured; for each process operation, a set of possible quality control activities is defined and organised in another hierarchical thesaurus (quality control actions thesaurus). These operations and quality control actions represent the *how* elements of the proposed conceptual model, and will be complemented with the documentation of the support activities in the future when realising the business-related metadata area. For each *how* element, also the *by which means* elements, that is, the software used to perform the activity, as well as the *when* elements, that is, the time reference of the operation/quality control action and various periodicities (periodicity of data collection, processing, dissemination and main releases), are documented in SIDI-SIQual.

The *who* element, that is, who is in charge of each operation, that was initially available within the Istat system was deleted because it was deemed irrelevant. The shift from a stovepipe production model to a more decentralised one might require the reintroduction of this item.

Together with the relevant process-related metadata, standard quality indicators are also stored in the Istat system: coverage, nonresponse, coding, editing and imputation, revision policy, timeliness and punctuality, comparability, coherence and resources (Brancato et al. 2004).

As the repository of the majority of the metadata items and quality indicators required by Eurostat for quality reporting, SIDI-SIQual is currently being enhanced in order to produce SDMX-compliant reference metadata files, (Simeoni 2013) following ESMS and ESQRS structures (Eurostat 2010).

6. Business-Related Metadata

As previously stated, the proposed approach also enables the modelling of business-related metadata, in particular those more strictly linked to the statistical process, in order to provide pieces of information useful for the more efficient management of statistical activities.

Business-related metadata have been defined as metadata arising from administrative and management activities performed to support statistical production and dissemination (i.e., support activities). As pointed out above (see, Figure 1), in GSIM they correspond to the metadata pertaining to the Business group which "is used to capture the designs and plans of Statistical Programs, and the processes undertaken to deliver those programs" (UNECE 2013b) or, in other terms, for high-level management. The GSIM conceptual model can be applied together with GSBPM at different levels of specification. At the most aggregated level, GSBPM corresponds to the statistical process as a whole, while GSIM specifies the corresponding information objects, namely Statistical Need for the input and Assessment for the output.

Business-related metadata are needed particularly by managers to assess efficiency, user satisfaction and acceptance by respondents and funders as well as to balance the needs of statistical information against other needs (Sundgren 2004). Thygesen and Nielsen (2012) further exploit GSBPM, suggesting a model for managing user needs and feedback.

GSIM provides a framework for the conceptualisation of business-related metadata: given their usefulness for the management of an NSI, business-related metadata can be described according to the macro phases *planning*, *execution* and *assessment*, already introduced (see, Figure 2). The proposed approach encompasses the GSIM input and output objects, as well as supports activities connected to statistical production (i.e., Statistical Program in the GSIM terminology) that are only partly represented in GSBPM as overarching processes (e.g., quality and metadata management).

Below, a first identification of business-related metadata for each macro phase is proposed, without claiming to be exhaustive or asserting that all business-related metadata should be managed in a unique repository. The intention is to contribute to fostering a more structured and detailed description of support activities in reference models, such as GSIM and GSBPM, particularly in connection with statistical metadata and quality.

With regard to the *planning* phase, the classes of business-related metadata generated during the different stages relate any statistical activity (data and processes) to the NSI strategic objectives (long- and mid-term goals) and to the different sectorial plans (e.g., methodological, IT, dissemination, etc.), thus setting out resources, schedules and timetables. The final aim is to ensure high quality and an efficient statistical production (regulated by Annual Statistical Planning). Annual Statistical Planning is the repository (in the form of either a document or information system) containing process-related metadata as well as metadata related to data structure and content.

In the *execution* phase, documents containing business-related metadata can be the service charters (for internal and external users), institutional policies, service-level agreements and contracts concerning, for instance, survey operations carried out in outsourcing. They give rise to classes of business-related metadata such as service standards, expected outputs, procedure targets and quality targets, time schedules of subprocesses and associated responsibilities and costs. Links to quality indicators can also be established when administrative procedures and contracts concern the execution of survey operations in outsourcing (e.g., interviewer selection, data collection, data capture, etc.). With regard to the provision of administrative source.

Finally, in the *assessment* phase done at institutional level, business-related metadata can originate from users sending comments through the website, visiting contact points, calling a toll-free number, in addition to more formalised user satisfaction surveys (or image studies, etc.) as well as from internal staff satisfaction surveys, staff evaluation, and from the assessment of the effectiveness of human resource development initiatives (e.g., staff training courses). These classes of metadata can be useful for complementing process and product quality both from a qualitative and quantitative point of view.

Table 1 provides some examples of business-related metadata (classes of information) for each macro phase, also showing where such information is generated or documented (repositories/information systems). The last column provides a link to quantitative information that originates together with business-related metadata, as well as some

PLANNING	Classes of information Users' needs Human Resources IT Resources Scheduling of activities Costs Critical events	Example of business-related metadata Repositories/information systems Long-term goals Mid-term goals Multiyear Statistical Planning Annual Statistical Planning Sectorial plans (e.g., methodological investments, IT plan, events scheduling, dissemination plan, training plan) Decisions by high-level Committees (e.g., Statistical Council, Quality Committee) Documentation from the phases "Specify needs", "Design" and	Link to other metadata Eink to other metadata Process-related metadata Metadata related to data structure and content Regulatory requirements Process-related metadata Metadata related to data structure and content Regulatory requirements Process-related metadata Metadata related to data structure and content Regulatory requirements
		"Build" for a given statistical process Risk management	and content Regulatory requirements
		Coordination of the National Statistical System	Process-related metadata Metadata related to data structure and content Regulatory requirements

Table 1. Description of business-related metadata

Table 1. Continued

	Link to other metadata	Process and product quality assessment	Performance indicators (benchmarking results to planned ones)	Users/Usages (usage rate)				
Example of business-related metadata	Repositories/information systems	Users satisfaction surveys	Web accesses to data and metadata	Feedbacks from stakeholders, journalists, public at large, etc.	Contact points	Staff satisfaction surveys	Staff evaluation	Human resources development initiatives (e.g., training courses)
	Classes of information	Users' satisfaction Users' complaints Stakeholders' requirements Staff needs (e.g., training) Available expertise Lacking expertise Lack of resources Costs (for users) Training information (e.g., attendance, topics, duration)						
		ASSESSMENT						

Table I. Continued

reference to process-related metadata and metadata related to data structure and content, which are deemed important.

As stated in Section 5, process-related metadata usually include quantitative measures used for quality analyses and reporting. More specifically, these measures are represented by paradata and quality indicators whose commonly accepted definitions are:

- *i) paradata* are data about survey operations, for example, times of day on which interviews were conducted, interview duration, frequency of contacts with each interviewee or attempts to contact the interviewee (Couper 1998; Groves and Heeringa 2006);
- *ii) quality indicators* are measures of the quality of statistical products or processes. Some quality indicators are obtained as a by-product of the production process, thus overlapping with paradata. In this case they are also called process variables (Ehling and Körner 2007).

However, support processes and management activities also produce quantitative information, that is, specific paradata (for example, number of available resources by professional expertise, punctuality and timeliness of administrative activities, evaluation score of training courses, compliance with quality targets/service goals, etc.) that can be exploited for a more comprehensive and detailed analysis of statistical activities (see, next section).

Indeed, business-related metadata and their associated paradata can complement information on quality components such as: *i*) accuracy, for instance whenever administrative procedures, service level agreements or contracts establish quality targets; *ii*) timeliness and punctuality, whenever support activities set time schedules that might affect statistical operations; *iii*) relevance and accessibility of data and metadata, by providing information collected through contact points, from the website, and by user satisfaction surveys; *iv*) completeness and coherence of the national statistical system as a whole, through coordination and planning activity.

In addition, business-related metadata provide information on costs for producers, respondents and users. Even though costs are not a quality dimension, they are undoubtedly a constraint for an NSI and need to be assessed to increase efficiency.

7. Complementing Process-Related Metadata With Business-Related Metadata

National Statistical Institutes are currently coping with challenging demands for new information needs and for more accurate and timely data while at the same time facing resource and financial restrictions. Thus it is vital to increase efficiency, reorganise outputoriented production, and rationalise, standardise and industrialise statistical processes. Business-related metadata can support such activities as they are useful for assessment performed at an institutional level (as referred to in Figure 2) if combined with process-related metadata, particularly with quality indicators. A similar assumption can be found in Eltinge et al. (2013), who propose a framework for improving statistical production systems by modelling, assessing and balancing multiple performance criteria including data quality, cost, risk and stakeholder utility. Furthermore, GSIM (UNECE 2013b; UNECE 2013c) introduces the concept of assessment for the Business group, defining it as "the result of the analysis of the quality and effectiveness of any activity undertaken by a statistical organisation and recommendations on how these can be improved", even though the related metadata are not further specified among the conceptual objects.

Nevertheless, the role of business-related metadata in supporting the decision-making process and quality assessment needs to be more thoroughly investigated.

In fact, business-related metadata and paradata are important not only for efficiently managing statistical processes but also for ensuring process and product quality. For instance, the administrative procedure for selecting an external CATI company will affect the timeliness of survey data if not completed in a given period of time. Since support processes may affect data quality, measurements of such impact (i.e., business-related paradata) are required.

The range and number of potentially useful business-related paradata to be collected and monitored could be extremely wide; this speaks for a stepwise approach, as also proposed by Sundgren (2004), starting from those typologies more linked to quality indicators. Similarly, Thygesen and Nielsen (2012) encourage a more interdisciplinary approach in statistical quality and metadata definition and implementation, focusing on fulfilling user needs.

As possible examples of this, the efforts being made at Istat to try and establish stricter links between quality assessment and management activities can be mentioned. Above all, the improvement actions resulting from quality auditing and self-assessment are currently reported in the Annual Planning, thus allowing for monitoring and reporting on quality improvements, their associated costs, allocated resources and time schedule (Signore et al. 2012).

Following the needs expressed by Istat managers, Annual Planning is being exploited for business-related metadata to complement quality indicators already available in SIDI-SIQual. Information about the staff employed in each business process, as well as information about the costs of carrying out a survey (e.g., printing paper questionnaires and instructions, interviews, data capture) will be taken directly from Annual Planning. In future, the duration of each operation will also be taken from Annual Planning, thus providing a more meaningful understanding of timeliness, and allowing for the office-wide analysis of process operations to screen for bottlenecks.

The exploitation of business repositories and the integration of some business-related metadata in the SUM system will support overall assessment by enabling middle and top management to accomplish a more comprehensive analysis of statistical production processes, widening the range of improvement actions beyond statistical aspects. This activity will also lead to a greater harmonisation of concepts used in different sectors of the Institute (administrative and support processes on the one side and statistical production processes on the other) to satisfy internal users' needs.

8. Final Remarks

This article proposes a unified approach for metadata conceptualisation and management. By jointly modelling metadata related to data structure and content and process- and business-related metadata, it is possible to fully describe data and the underlying statistical production processes. The proposed model also enables the exploitation of businessrelated metadata for high-level planning and assessment, even though this is quite a challenging task that requires step-by-step implementation.

This approach also supports harmonisation. As a matter of fact, similar pieces of information are often collected by different sectors of the organisation according to a "silo" approach, resulting in multiple collections of the same items, in information that cannot be reused (by other sectors or in other process steps), or in less accurate information. The adoption of a unique metadata conceptual model facilitates overcoming these problems that many NSIs face by ensuring terminological coherence across different sectors within the organisation and consistency of targets, actions and results.

For this reason, the proposed approach is useful for NSIs in streamlining metadata and quality-related initiatives and in supporting ongoing standardisation and industrialisation processes. The implementation of a unified metadata management system can also lead to increased process efficiency and efficacy as well as to quality enhancement, as already highlighted. This approach is widely applicable in different contexts even though implementation requires adapting it to office-wide infrastructures and technical standards.

The present article also debates some current limitations and suggests possible enhancements of current standards. GSIM could be enhanced with a better representation of the data transformation over the course of the production process (e.g., the use of data content within a data structure). GSBPM could be improved by further detailing the quality management overarching process and proposing generic quality indicators. Both standards could better address business-related metadata and their relationships with other metadata.

In developing SUM, Istat has demonstrated the practical applicability of the conceptual model proposed in the present article, ensuring compliance with standards at the same time.

Of all the lessons learned in implementing SUM, it is worth mentioning the importance of addressing different experts in statistical organisations (e.g., statisticians, quality managers, survey managers, metadata-modelling and IT experts) who often work on specific aspects according to their expertise, thus lacking a view of the issue as a whole. In Istat's experience, a common perspective was found by assigning a prominent role to statistics (i.e., the statistician's point of view) in modelling and using metadata.

The main challenge ahead for SUM is to better conceptualise and integrate businessrelated metadata into the system, as well as to test their usefulness for institutional global assessment. Based on this experience, we deem it likely that other NSIs implementing unified metadata systems will encounter the same main challenges.

Appendix

The UNECE Frameworks and Standards for Statistical Modernisation group has already investigated the relationship between GSIM concepts with two standards available, SDMX (2009) and DDI (2014), which are usually used in NSIs. This mapping is extremely useful for metadata related to *data structure and content* (Section 4) because their aim is to define a standard to describe data content (DDI) or to allow data and metadata exchange (SDMX). Restricting attention to the "Concept" and "Structure" parts of GSIM used in Section 4, the corresponding concepts are more easily mapped in DDI (with some exceptions, such as classification) than in SDMX. As a matter of fact, SDMX lacks the

typical statistical organisation of metadata in terms of reference population, statistical variables, statistical operations, and so on. Given the mandatory application of SDMX in the European Statistical System, attention was focused on how to map SDMX concepts with those defined in Section 4. As a matter of fact, the use of the statistical role of metadata in conjunction with the SDMX machinery can be a tool to enhance integration within a National Statistical Institute and across different organisations through a unique representation of data structures. According to the SDMX standard, metadata related to data structure and content are mainly categorized in three large groups: *concepts*, that is, any element playing a role in a table description; *code lists*, specifying those concepts with a finite number of instances; *data structure definitions* (DSD), consisting of concepts, their associated (when necessary) code lists and the assignment of concept roles as dimensions, attributes and measures. This is the relationship between these SDMX concepts and the corresponding metadata related to data structure and content (Section 4).

Code list: define specific code lists for populations, numerical variables, statistical operators and Data Content. The Data Content code list can be defined in terms of its components as simple or composite macrodata by considering a number of code annotations: reference population, statistical operator, numerical variable, and so on.

Concept schemes: in order to allow reuse for micro data also, it is advisable to collect the names of the statistical categorical variables and the names of the other concepts used in a DSD (temporal, operational and explanatory concept) in two distinct concept schemes. In a given process, categorical variable names will be reused more than other concepts; among processes, the other concepts will be reused more than categorical variable names (most of the other concepts are available in the Cross Domain concept scheme, although some categorical variable names are also included as the Reference Area). Furthermore, it is appropriate to separate the two groups of concepts because there are operations (as marginalisation of contingency tables) that can be performed only on those dimensions consisting of statistical variables, and not on other kinds of dimension.

DSD: The DSD is organised in the following way. If the primary measure is associated to the concept Observed value (as it usually is in the SDMX context) one (and only one) DSD dimension should be devoted to the Data Content code list. All the categorical variables will be defined as dimensions, as well as time-related concepts (mainly time dimension and frequency). If useful, other concepts can be used as dimensions, as might happen for the time series adjustment concept. In any case, it would be better to organise the noncategorical variable concepts as attributes if possible and appropriate.

One feature that SDMX does not fully cover yet is the metadata relationship over the course of the statistical process. This aspect should be well modelled in order to ensure integration, traceability and metadata phase-based search. As a matter of fact, SDMX already has a module on "Expression and Calculations", although it still lacks an expression language. The issue of an expression language is under discussion in the SDMX Technical Working Group.

9. References

Androvitsaneas, C., B. Sundgren, and L. Thygesen. 2006. "Towards an SDMX User Guide: Exchange of Statistical Data and Metadata Between Different Systems, National and International." OECD Expert Group on Statistical Data and Metadata Exchange, Geneva, 6–7 April 2006. Available at: https://sites.google.com/site/bosundgren/my-life (accessed November 2013).

- Banca d'Italia. 2007. *The Matrix Model. Unified Model for Statistical Data Representation and Processing*. Available at: https://www.bancaditalia.it/statistiche/raccolta-dati/sistema-informativo-statistico/modellazione/matrixmod.pdf (accessed May 2015).
- Bergamasco, S., A. Cardacino, F. Rizzo, M. Scanu, and L. Vignola. 2013. "A Strategy on Structural Metadata Management Based on SDMX and the GSIM Models." Work Session on Statistical Metadata (METIS), Geneva, 6–8 May 2013. Available at: http://www.unece.org/stats/documents/2013.05.metis.html#/ (accessed April 2015).
- Brancato, G., C. Pellegrini, M. Signore, and G. Simeoni. 2004. "Standardising, Evaluating and Documenting Quality: the Implementation of Istat Information System for Survey Documentation – SIDI." In Proceedings of European Conference on Quality and Methodology in Official Statistics, Q2004, CD-ROM ISBN: 3-8246-0733-6. Mainz, 24–26 May 2004.
- Brancato, G. and G. Simeoni. 2012. "Istat Statistical Process Modelling and the Generic Statistical Business Process Model: a Comparison." European Conference on Quality in Official Statistics, Q2012, Athens, 29 May 1 June, 2012. Available at: http://www.q2012.gr (accessed November 2013).
- Couper, M.P. 1998. "Measuring Survey Quality in a CASIC Environment." In Proceedings of the Survey Research Methods Section, 41–49. Available at: http://www.amstat.org/sections/srms/Proceedings/ (accessed March 2014).
- D'Angiolini, G., M. Paolucci, and M. Signore. 1998. "Developing Tools for Managing, Exploiting and Disseminating Metainformation: the Istat's Experience." New Techniques and Technologies for Statistics (NTTS) Conference, 4–6 November 1998, Sorrento, vol.1, Specialised Sessions Papers, 119–206.
- DDI Alliance. 2014. Getting Started with DDI. Available at: http://www.ddialliance.org (accessed October 2014).
- Ehling, M. and T. Körner. 2007. *Handbook on Data Quality Assessment Methods and Tools*. Luxembourg: Eurostat.
- Eltinge, J.L., P.P. Biemer, and A. Holmberg. 2013. "A Potential Framework for Integrating of Architecture and Methodology to Improve Statistical Production Systems." *Journal of Official Statistics* 29: 125–145. Doi: http://dx.doi.org/10.2478/jos-2013-0007.
- European Commission. 2009. Commission Recommendations of 23 June 2009 on Reference Metadata for the European Statistical System. Official Journal of the European Union 2009/498/EC. Available at: http://eur-ex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32009H0498 (accessed May 2015).
- Eurostat. 2009. *Euro SDMX Metadata Structure (ESMS)*. Available at: http://ec.europa.eu/ eurostat/data/metadata-structure (accessed April 2015).
- Eurostat. 2010. *ESS Standard for Quality Reporting Structure*, ESQRS release 1, October 2010. Available at: http://ec.europa.eu/eurostat/data/metadata/metadata_structure (Accessed May 2014).
- Eurostat. 2011. *European Statistics Code of Practice*. Available at: http://ec.europa.eu/ eurostat/web/quality/european-statistics-code-of-practice (accessed May 2015).

Fisher, R.A. 1925. Statistical Methods for Research Workers. London: Oliver and Boyd.

- Gelsema, T. 2012. "The Organisation of Information in a Statistical Office." *Journal of Official Statistics* 28: 413–440.
- Götzfried, A., H. Linden, and E. Clement. 2011. "Standards and Processes for Integrating Metadata in the European Statistical System." Workshop on Statistical Metadata (METIS), Geneva, 5–7 October 2011. Available at: http://www.unece.org/stats/ documents/2011.10.metis.html (accessed March 2014).
- Greenough, C., K. Mechanda, and F. Rizzolo. 2014. "Metadata in the Modernization of Statistical Production at Statistics Canada." European Conference on Quality in Official Statistics, Q2014, Vienna, 2–5 June 2014. Available at: http://www.q2014.at/ papers-presentations.html (accessed May 2015).
- Groves, R.M. and S.G. Heeringa. 2006. "Responsive Designs for Household Surveys: Tools for Actively Controlling Survey Errors and Costs." *Journal of the Royal Statistical Society, Series A* 169: 439–457. Available at: http://onlinelibrary.wiley.com/ doi/10.1111/rssa.2006.169.issue-3/issuetoc (accessed March 2014).
- SDMX 2009. *Content Oriented Guidelines*. Available at: http://www.sdmx.org/ (accessed November 2013).
- Signore, M., R. Carbini, and M. D'Orazio. 2012. "Quality Assessment in Istat: the Combined Use of Standard Quality Indicator Analysis and Audit Procedures." European Conference on Quality in Official Statistics, Q2012, Athens, 29 May-1 June 2012. Available at: http://www.q2012.gr/articlefiles/sessions/6.2_Signore %20et%20al%20_Quality%20assessment%20at%20Istat%20paper.pdf (accessed November 2013).
- Signore, M., M. Scanu, and G. Brancato. 2013a. "Statistical Metadata: a Unified Approach to Management and Dissemination." NTTS – Conferences on New Techniques and Technologies for Statistics, Brussels, 5–7 March 2013. Available at: http:// www.cros-portal.eu/sites/default/files//NTTS2013%20Proceedings_0.pdf (accessed January 2015).
- Signore, M., M. Scanu, A. De Santis, A. Ambrosetti, and V. Olivieri. 2013b. La Governance del Sistema Unitario dei Metadati Metadati Strutturali. Principi e struttura organizzativa per la gestione e armonizzazione dei metadati strutturali dell'Istituto. Internal document, Istat, Rome Italy.
- Simeoni, G. 2013. "Implementing ESS Standards for Reference Metadata and Quality Reporting at ISTAT." Joint UNECE/Eurostat/ OECD Work Session on Statistical Metadata. Geneva, 6–8 May 2013.
- Statistics Sweden. 2008. "Classifications of Statistical Metadata." Joint UNECE/Eurostat/OECD work session on statistical metadata (METIS). Luxembourg, 9–11 April 2008. Paper prepared by Bo Sundgren. Available at: http://www.unece.org/stats/ documents/2008.04.metis.html# (accessed May 2015).
- Sundgren, B. 2004. "Metadata Systems in Statistical Production Processes for Which Purposes are They Needed, and How Can They Best Be Organized?" Joint UNECE/Eurostat/OECD work session on statistical metadata (METIS), Geneva, 9–11 February 2004. Available at: https://sites.google.com/site/bosundgren/my-life (accessed November 2013).

- Thygesen, L. and M.G. Nielsen. 2012. "How to Fulfill User Needs Metadata, Administrative Data and Processes." European Conference on Quality in Official Statistics, Q2012, Athens, 29 May – 1 June 2012. Available at: http://www.q2012.gr/ articlefiles/sessions/21.3_Thygesen-Nielsen_How%20to%20fulfill%20user%20 needs%20-%20metadata,%20administrative%20data%20and%20processes%20 version%20final.pdf (accessed March 2014).
- United Nations Economic Commission for Europe (UNECE). 1995. *Guidelines for the Modeling of Statistical Data and Metadata*. Geneva: United Nations. Available at: http://www1.unece.org/stat/platform/display/metis/UNECE + Guidelines + for + the + Modelling + of + Statistical + Data + and + Metadata (accessed March 2014).
- United Nations Economic Commission for Europe (UNECE). 2013a. GSBPM v5.0. Available at: http://www1.unece.org/stat/platform/display/GSBPM/GSBPM + v5.0 (accessed March 2014).
- United Nations Economic Commission for Europe (UNECE). 2013b. Generic Statistical Information Model (GSIM): Communication Paper for a General Statistical Audience (Version 1.1, December 2013). Available at: http://www1.unece.org/stat/platform/ display/gsim/GSIM + Communication + Paper (accessed April 2015).
- United Nations Economic Commission for Europe (UNECE). 2013c. Generic Statistical Information Model (GSIM): Specification (Version 1.1, December 2013). Available at: http://www1.unece.org/stat/platform/display/gsim/GSIM + Specification
- United Nations Economic Commission for Europe (UNECE). 2015. Generic Activity Model for Statistical Organizations (GAMSO) v0.2. Available at: http://www1.unece.org/ stat/platform/display/GAMSO/GAMSO+Home (accessed April 2015).
- Zachman, J.A. 1987. "A Framework for Information Systems Architecture." *IBM Systems Journal* 26, no. 3. Available at: http://www.zachmanframework.com/images/ZI_PIcs/ ibmsj2603e.pdf (accessed March 2014).

Received November 2013 Revised February 2015 Accepted February 2015