istat
working
papers

N. 9
2012

The standardization of statistical processes: a framework for methodological standards

Giulio Barcaroli, Marco Broccoli, Nicoletta Cibella, Claudia De Vitiis, Francesca Inglese e Stefania Macchia

Istat
Istituto Nazionale di Statistica

istat working papers

N. 9
2012

The standardization of statistical processes: a framework for methodological standards

Giulio Barcaroli, Marco Broccoli, Nicoletta Cibella, Claudia De Vitiis, Francesca Inglese e Stefania Macchia

Istat

Istat Working Papers

The standardization of statistical processes:
a framework for methodological standards

N. 9/2012

ISBN 88-458-1724-5

Istituto nazionale di statistica
Servizio Editoria
Via Cesare Balbo, 16 – Roma

The standardization of statistical processes: a framework for methodological standards

Giulio Barcaroli, Marco Broccoli, Nicoletta Cibella, Claudia De Vitiis, Francesca Inglese, Stefania Macchia

Sommario

Nel definire i propri obiettivi prioritari, il Sistema Statistico Europeo si è posto anche quello di procedere alla industrializzazione dei processi di produzione dell'informazione statistica. Requisito fondamentale per il raggiungimento di tale obiettivo è quello di definire degli standard di riferimento validi per l'intera comunità della statistica ufficiale, standard che riguarderanno i vari segmenti della "statistical business architecture": quello più propriamente statistico, quello information technology (IT), quello operativo-gestionale e quello relativo a ricerca e formazione. Per garantire l'avvio del processo di definizione degli standard, l'ESSC (European Statistical System Committee) ha lanciato una serie di iniziative, tra cui un ESSnet ("Preparazione alla standardizzazione"), con il compito di esplorare le problematiche relative alla definizione, implementazione e gestione degli standard. L'ESSnet ha concluso i propri lavori producendo, tra le altre cose, uno schema concettuale degli elementi costituenti uno standard statistico. L'ISTAT ha avuto un ruolo fondamentale nella definizione di tale schema, che è stato implementato in un repository utilizzabile per l'analisi di regole e raccomandazioni derivabili dai manuali metodologici in uso negli Istituti nazionali di statistica, e per una definizione di standard condivisi.

Parole chiave: standardizzazione dei processi statistici, standard statistici, GSBPM.

Abstract

One of the highest priority of the European Statistical System is to carry out the industrialisation of production processes of statistical information. Fundamental requirement for the achievement of this objective is to define standards valid for the entire community of official statistics; these standards should cover the various aspects of "statistical business architecture": statistic production processes, IT environment, operational-management framework and research and training. To ensure the start of the standard definition process, the European Statistical System Committee has launched a series of initiatives, including an ESSnet ("Preparation for standardization"), with the task of exploring the issues relating to the standard definition, implementation and management. The ESSnet has concluded its works also producing, among other valuable outputs, a conceptual framework describing the constituent elements of a statistical standard. The role of ISTAT has been crucial in the definition of this framework, that has been implemented in a repository of rules and recommendations, used for the analysis of the methodological manuals in use by National Statistical Institutes, and potentially useful for the definition of common standards.

Keywords: statistical process standardisation, statistical standards, GSBPM.

Index

1. Introduction[1]

While the European Statistical System (ESS) has a long tradition in input and output harmonisation (definition of statistical units, characteristics to be measured, classifications to be used, definition of aggregates to be produced and disseminated), it is not possible to say the same with regard to the harmonization of production methods, processes and systems.

The "vision" for the next decade (COM 404/2009) recognizes the need for an industrialization of statistical production processes, that should be based on common and standardised processes, transforming raw data into statistical products according to generic and commonly accepted concepts.

Consequently, the Joint Strategy Paper, adopted by the ESS Committee in May 2010, states that "*The integration that is envisaged will require more harmonisation and standardisation of statistical methodologies for data collection, data validation, dissemination and communication within the ESS, access to microdata for researchers, harmonising the IT infrastructure and sharing IT tools as a way to facilitate the use of agreed statistical methods, and harmonising metadata to permit easy and efficient data and metadata exchange, leading to better quality and higher productivity of the statistical data processing.*"

The goals of standardisation are:

- efficiency: to develop methods and tools only once; to smoother interoperability; to prepare the integration of processes;
- quality: to get higher comparability over Member States; new output resulting from integration of data over statistical domains;
- flexibility: human resources can be reallocated with less costs;
- growth: investments in methods and tools will pay off more easily and resources saved can be reallocated to new domains.

The concepts and possible activities on standardisation were discussed by the Eurostat group coordinating the ESS Directors of Methodology (DIME), and also in the Workshop on Standardization (October 2010). There was a general agreement on two initiatives:

1. to launch a preparatory ESSnet in order to pave the way for further activities on standardisation. In particular, the specific targets of this ESSnet were to (i) clarify the meaning of the term "standard" in the statistical environment; (ii) review the current ESS handbooks, as potential methodological standards, proposing assessment criteria and advising on their status; (iii) investigate other kinds of standards (IT tools, data formats, metadata systems, etc); (iv) define a possible model for the definition, implementation and maintenance of standards;
2. to set up a mandate for a Sponsorship on standardisation. The Sponsorship would have to support the standardisation initiative by setting priorities, testing the standardisation procedure and deciding on further development activities.

The ESSnet, named "Preparation of standards (Stand-Prep)", was launched on December 2010, and ended its works on September 2011, on time to deliver its results to the Sponsorship, whose first meeting has been held on September 22.

The ESSnet was organised along three different work packages (WP):

- WP1 had to define the concept of standard in statistics, in particular the standard corresponding to "statistical methods", and apply these definitions, in order to find out for each of six selected methodological handbooks which parts can be considered to be mandatory rules, and which are rather to be classified as guidance;
- WP2 had to systemise standards other than statistical methods and examine issues in the adoption of standards;
- WP3 had to make suggestions to the future work on standardisation, especially to the sponsorship on standardisation.

[1] The paper is the results of the joint efforts of the authors. In any case, paragraphs 1, 2, 3 and 9 are mainly due to G. Barcaroli and S. Macchia, paragraph 4 to C. De Vitiis, paragraph 5 to M. Broccoli, paragraphs 6 and 7 to F. Inglese, paragraph 8 to N. Cibella.

Partners in the ESSnet were: INSEE (France), ONS (UK), Destatis (Germany), FSH (Hungary), GUS (Poland), CBS (Netherlands) and ISTAT (Italy).

ISTAT was responsible for WP1. In this work package the "core" activities were related to the analysis of the methodological handbooks in use inside the Statistical Institutes and the definition of a general scheme to document methodological standards.

2. Definition of statistical standard

The ESSnet decided to adopt the ISO/IEC definition of standard: "*A standard is a document, established by consensus and approved by a recognized body, that provides, for common and repeated use, rules, guidelines or characteristics for activities or their results, aimed at the achievement of the optimum degree of order in a given context. Note: Standards should be based on the consolidated results of science, technology and experience, and aimed at the promotion of optimum community benefits.*"

In defining the concept of standard, the ESSnet team decided to make use of the ISO/IEC vocabulary.[2]

Moreover, in the statistical context a relevant instrument useful for standardisation is the Generic Statistical Business Process Model (GSBPM),[3] which provides a framework to describe the statistical production process in terms of standard components (phases and sub-processes). It is intended to apply to all activities undertaken by producers of official statistics, at both national and international level, which result in data outputs. It is designed to be independent of the data source, so it can be used for the description and quality assessment of processes based on surveys, censuses, administrative archives, and other non-statistical or mixed sources.

The starting point is given by the concept of *normative document*, which can be defined as a document that provides rules, guidelines or characteristics for activities or their results.

If the normative document prescribes technical requirements to be fulfilled by a product, a process or a service, it is a *technical specification*.

If the normative document recommends practices or procedures to be adopted in the design or execution of a process, it is a *code of practice.*

When a normative document is established by consensus and approved by a recognized body, i.e. by an organisation that is in charge for standardisation, it can be considered a proper *standard*.

Technical specifications and code of practices, on the basis of the way they have been adopted, may be considered standards or parts of a standard.

Another important concept is the '*provision'* which is an expression in the content of a normative document, that can take the form of:

- *statement* (provision giving information on how to do);
- *instruction* (provision imposing actions to be performed);
- *recommendation* (provision conveying advice or guidance);
- *requirement* (provision indicating criteria to be fulfilled).

These types of provisions can be distinguished by the wording they are expressed with; e.g. instructions are expressed in the imperative mood, recommendations by the use of the auxiliary "should" and requirements by the use of the auxiliary "shall".

[2] See Annex 1 ISO/ IEC Guide 2 - Standardization and related activities - General vocabulary.

[3] See the paper "Generic Statistical Business Process Model - Version 4.0" (UNECE April 2009) www.unece.org/stats/gsbpm.

3. A framework for the analysis

The objective was to define the concept of *standard* in statistics, in particular the *standards* corresponding to "statistical methods", and to apply these definitions in order to find out, for each of the six selected methodological handbooks, which parts can be considered as *standards*.

Actually, it became soon clear that this work of analysis could yield a value added: the job was not only to excerpt the *core* of the methodological handbooks, consisting in the *provisions* that are the basic elements of a standard; but also to identify the relationships with other fundamental items and references that are contained in the same handbook, as *methodologies, methods, IT tools, classifications, definitions,* etcetera. Furthermore, it was also considered the usefulness of linking each provision (or group of provisions) to a given phase or sub-process of GSBPM, in order to have the possibility to collect and compare all provisions available for a given step of the production process of statistical information.

The relationships among all the elements cited above are rather complex: it was therefore necessary to model them by defining a conceptual scheme representing all the entities that can be found in a set of methodological handbooks, and the relations connecting these entities.

This conceptual scheme was translated in a logical one, the basis for a repository able to contain the knowledge elements to be extracted from the handbooks.

It was also necessary to develop an IT tool enabling all partners in the ESSnet to populate the repository in a coordinated and assisted way. To this aim, a procedure was defined, consisting of the following steps:

1 each single handbook was assigned to one or two partners, that conducted the analysis on it;
2 if two different analyses had been carried out on the same handbook, a reconciliation step was performed to get a unique and agreed result;
3 the results of all the analyses have been introduced in a final and unified repository, that contains all the provisions individuated in the six handbooks.

This final repository has to be considered as a proof-of-concepts: starting from the vast amount of handbooks related to the statistical production processes, each of them to be considered as a normative document, it is possible to obtain something that can be regarded as an instrument to get information at a detailed level on all rules concerning a particular subject, regardless of the source. This would not be possible if the handbooks had been described in a not homogeneous way and only at the aggregated level, without going at the elementary level of the "provision".

At the end of the work it is possible to interrogate the repository for instance in order to list all the provisions related to a given GSBPM sub-process obtaining information on the source (a given handbook), the type (indicating the associated degree of freedom), the related methodologies and methods.

In summary, all the entities which contribute or can contribute to compose a standard were listed, the attributes of each of them were defined and the relationships linking each one to the others were identified so as to design a conceptual scheme which allows to represent whatever methodological standard and helps, thanks also to its graphical layout, to identify and document it.

On the basis of this conceptual scheme, the logical scheme representing the data base for the repository of standards was designed; the following step was the definition of the physical scheme and the development of the IT tool to populate and process it.

4. The conceptual scheme

To design the conceptual scheme used to represent all the entities that can be found in a set of methodological handbooks, and the relations connecting these entities, the "entity/relationship" approach has been adopted.[4] This method stems from information systems design methodology,

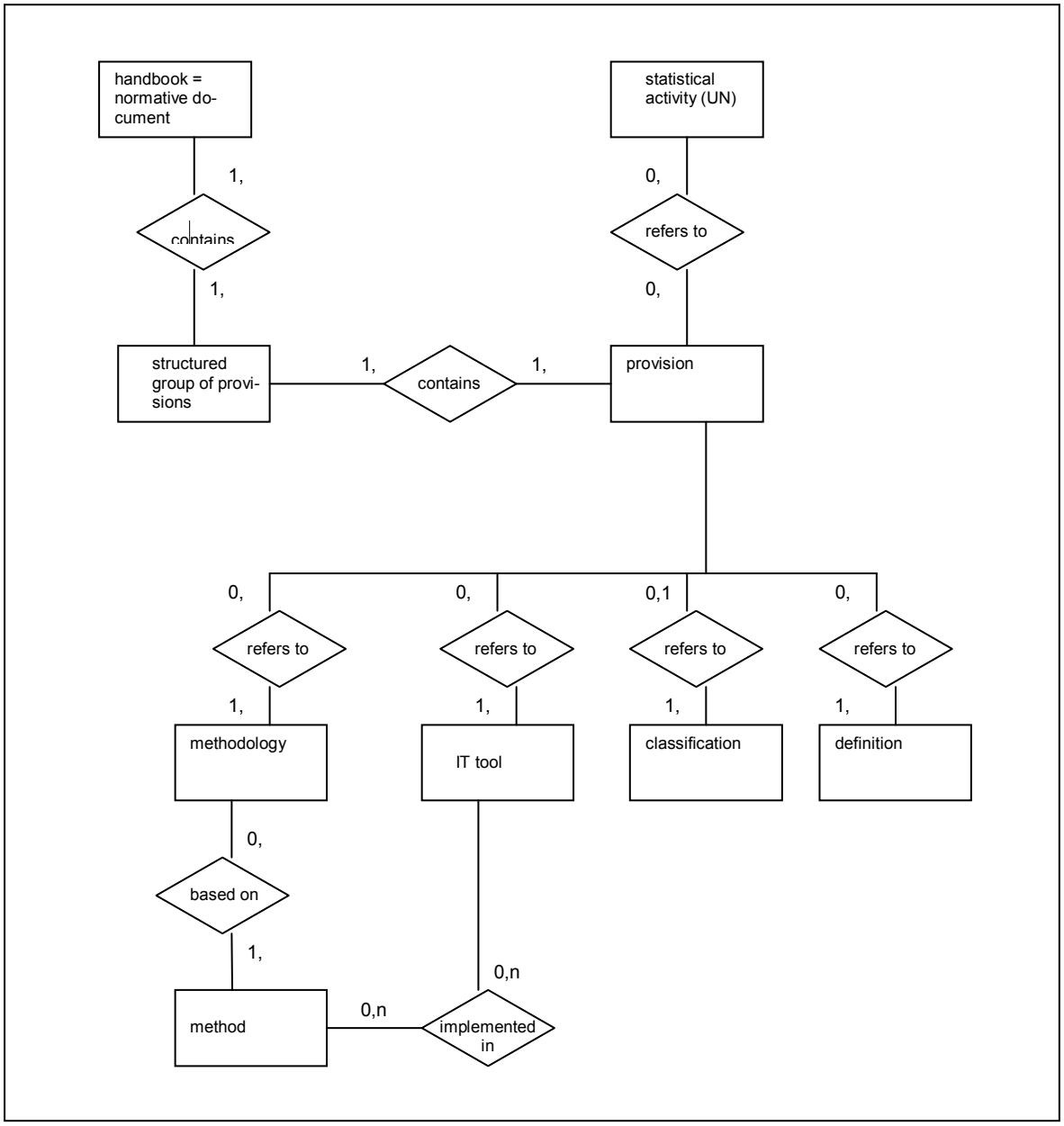
[4] Chen P.P., "The entity-relationship model-toward a unified view of data", ACM Transactions on Database Systems (TODS), September 22-24, 1975.

which can be applied to represent any kind of database. Entity/relationship schemes provide an overview on concepts, items and relationships covered in the database. The basic structure of entity/relationship schemes consists of entities, logical links between entities (relationships), as well as entities' attributes.

In figure 1, the conceptual scheme of the repository for statistical standards is reported.

In the following, we illustrate the meaning of each entity, their attributes and the relationships among entities.

Figure 1 - The conceptual scheme of statistical standards in methodological handbooks



Normative document

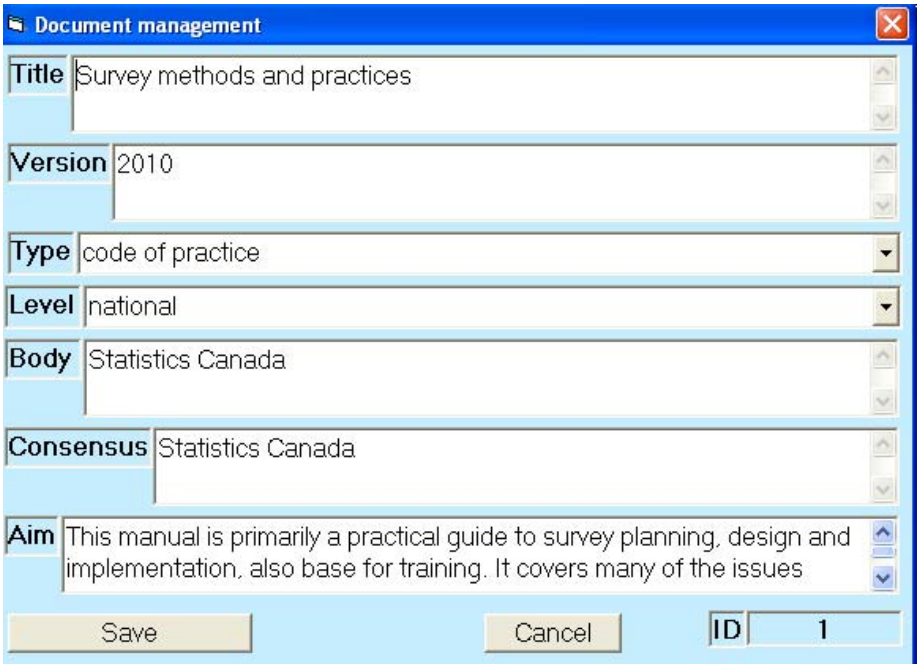
In the model described in the Entity Resolution scheme, a statistical methodological handbook may be defined as a "*normative document*" ("a document that provides rules, guidelines or characteristics for activities or their results").

The attributes related to this entity are:

1. title: the title of the document;
2. version: indication of the version, if available (otherwise, year of publishing);
3. type: the normative document can be:
   - a "standard" if the document has been established by consensus and approved by a recognized body;
   - a "pre-standard" if the document is adopted provisionally by a standardizing body and made available to the public in order to gain the necessary experience from its application;
   - a "technical specification" if it prescribes technical requirements to be fulfilled by a product, process or service;
   - a "code of practice" if it is a document that recommends practices of procedures;
   - a "regulation" if provides binding legislative rules.
4. level: the area of influence of the document. It can be:
   - international;
   - regional;
   - national;
   - provincial.
5. body: the institution(s) that is (are) responsible for the development of the normative document;
6. consensus: the procedure (if any) followed to reach a consensus, required to declare the normative document as a standard;
7. aim: the purpose of the normative document.

In general, the handbooks we are dealing with can be considered as of the kind "code of practice".

As already said, the conceptual scheme was the basis for the development of the repository of standards and of the IT tool to populate it. In order to clarify how the entities listed in this paragraph, their attributes and the relationships among entities are managed in the data base, we also show the forms dedicated to each of them that are in the software developed for the management of the repository (see paragraph 5).



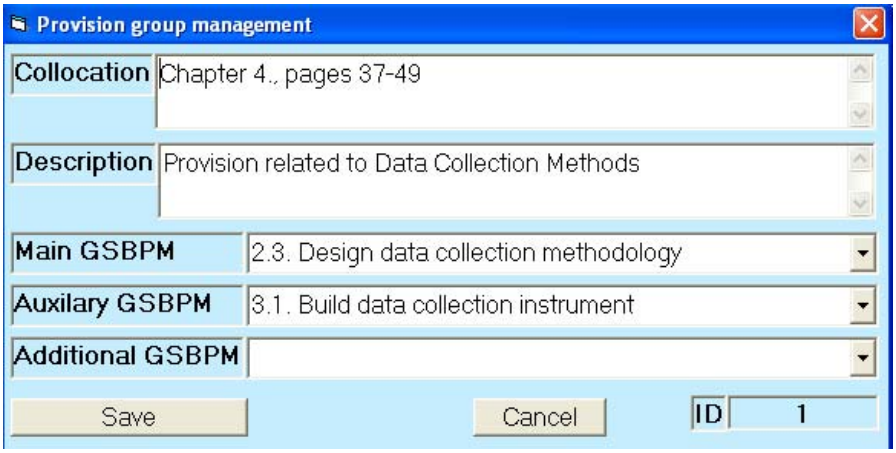
Group of provisions

A normative document must contain at least one "group of provisions", in general many of them (cardinality = 1,n).[5] One group of provisions belongs to only one normative document.

In the case of methodological handbooks, the identification of the groups of provisions should take into account the following elements: the structure of the handbook (sections and chapters) and the relations with the most detailed level of GSBPM (sub-processes). A group of provisions must refer to only one sub-process in GSBPM, and should be contained in only one chapter (maybe in more paragraphs).

Attributes of the entity group of provisions are:

1. collocation: indication of the logical (section, chapter) and physical (from page to page) collocation inside the handbook;
2. description: any important feature to be reported;
3. main GSBPM sub-process;
4. auxiliary GSBPM sub-process;
5. additional GSBMP sub-process.



Provision

Accordingly to the definitions given in Paragraph 2, a "provision" is an expression in the content of a normative document that takes the form of a statement, an instruction, a recommendation or a requirement.

A group of provisions must contain at least one provision, in general many of them (cardinality = 1,n).

Attributes of the entity provision are:

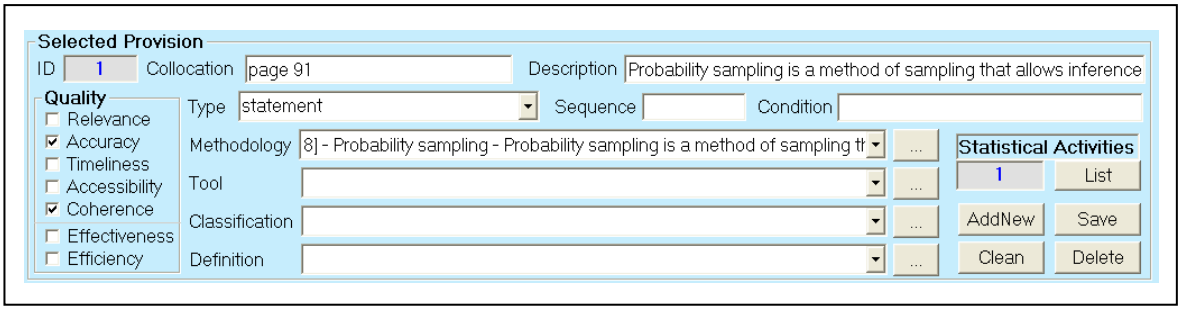
1. indication of the physical collocation (page) inside the handbook;
2. description: text, as it appears in the document;
3. type:
   - statement;
   - instruction;
   - recommendation;
   - exclusive requirement;
   - optional requirement;

[5] In a given relationship between entities A and B, the cardinality is a couple whose first element indicates the minimum number of instances of B linked to one instance of A, while the second element indicates the maximum number.

   - deemed-to-satisfy provision;
   - descriptive provision;
   - performance provision.
4. sequence: progressive number, defined by users, unique inside the same structured group; this attribute is useful if there must be an order in applying provisions;
5. condition(s) of application;
6. components of quality:[6]
   a) the ones referred to the product:
      - relevance;
      - accuracy;
      - timeliness and punctuality;
      - accessibility and clarity;
      - coherence and comparability.
   b) the ones referred to the process:
      - effectiveness;
      - efficiency.

A provision may refer to:

- statistical activities;
- a methodology;
- an IT tool;
- a classification;
- a definition.



Statistical activities

A provision may or may not concern a given "statistical activity", or many of them (cardinality: 0,n). A given statistical domain can be referred by no provision, or by many provisions (cardinality: 0,n).

The different statistical activities considered in this scheme are a subset of the complete set that has been considered for the Global Inventory of UNECE.[7] In particular, this subset contains all the subject matter areas:

- D1: demographic and social statistics;
- D2: economic statistics;
- D3: environment and multidomain statistics.

with the exclusion of the other areas as:

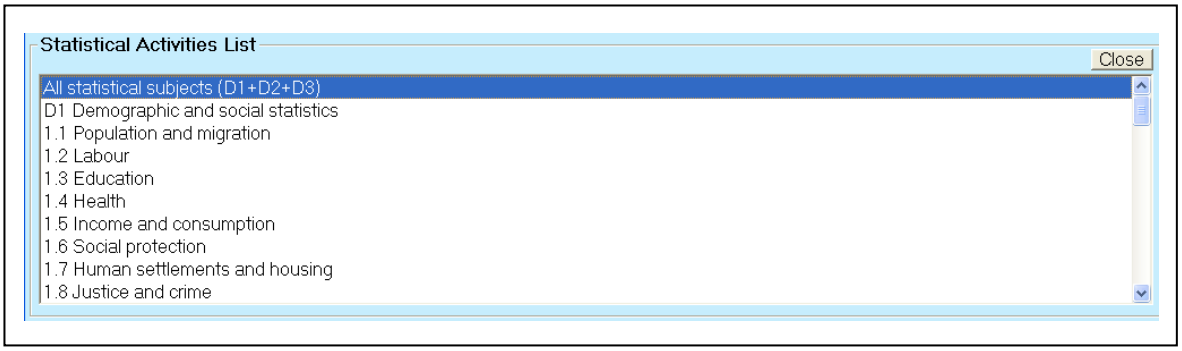
- D4: methodology of data collection, processing and dissemination;
- D5: strategic and managerial issues of official statistics.

[6] We consider the EUROSTAT quality dimensions, as they have been defined in Eurostat (2009) ESS Handbook for Quality Reports, 2009 Edition.
[7] See the Annex I to the paper "Developing a Global Inventory of Statistical Standards" (UNECE, 18 August 2010).

This is why the items contained in D4 and D5 have been judged as out-of-the-scope with respect to statistical standards, or better represented by the GSBPM.

The attributes of the entity statistical activity are:

1. identifier: see the United Nations set of codes;
2. description: text (see the United Nations set of descriptions).



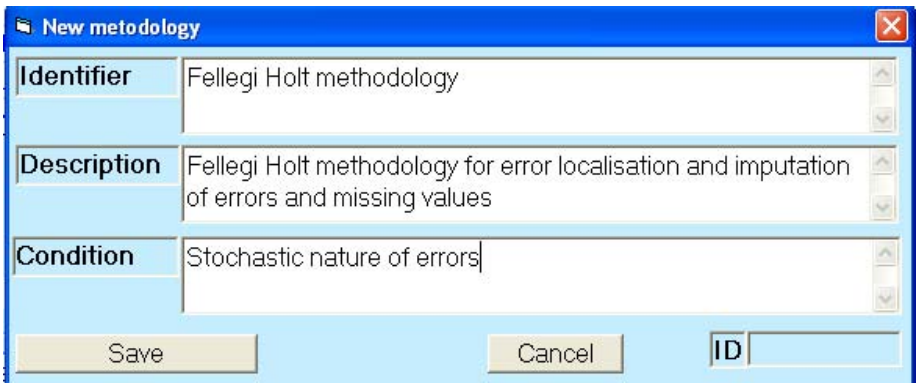
Methodology

In this context, we can define a "*methodology*"[8] as "*a structured approach to solve a problem*". More specifically, "*a set of research methods and techniques applied to a particular field of study*".[9]

A given provision may or may not indicate one methodology (cardinality: 0,1). A given methodology, if present, must be indicated by one provision, in general by many of them (cardinality: 1,n).

The attributes of the entity methodology are:

1. identifier: short description;
2. description: extended description of the methodology;
3. condition(s) of application.



Method

In this context, we define a "*method*" as "*An established, habitual, logical, or prescribed practice or systematic process of achieving certain ends with accuracy and efficiency, usually in an ordered sequence of fixed steps*".[10]

[8] OECD Glossary of Statistical Terms, available at: http://stats.oecd.org/glossary/search.asp.
[9] Statistics Canada Glossary: http://www.statcan.gc.ca/edu/power-pouvoir/glossary-glossaire/5214842-eng.htm.
[10] See the site "Business dictionary": http://www.businessdictionary.com/definition/method.html.

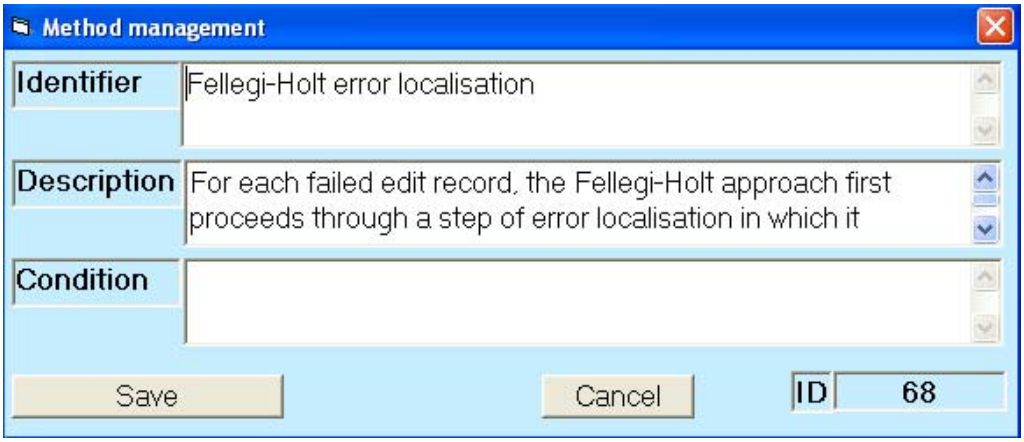
A given methodology may or may not indicate one method, in general more than one (cardinality 0,n). A given method has to be linked to at least one methodology, or to many of them (cardinality: 1,n).

The attributes of a method are:

1. identifier: short description;
2. description: extended description of the method;
3. condition(s) of application.

Relationships between methodologies and methods are difficult to define in a univocal way. Basically, a methodology should indicate a general approach, that may include a set of methods enabling to follow that approach.[11]

In addition, a given method may or may not be implemented in an IT tool (see IT tool).



IT tool

In this context, we define an "*IT tool*" as "*a software application or system, implementing one or more methods*".

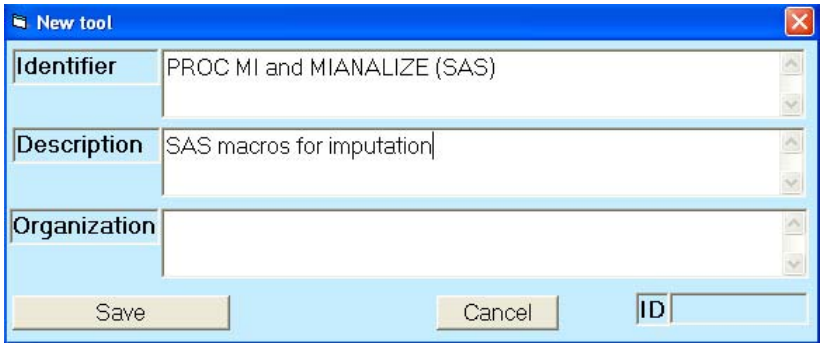
A given provision may or may not indicate one IT tool (cardinality: 0,1). A given IT tool, if present, must be indicated by one provision, in general by many of them (cardinality: 1,n).

The attributes of the entity "IT tool" are:

1. identifier: name or acronym;
2. description: text;
3. organisation.

A given IT tool may or may not implement a given method, or many of them (cardinality: 0,n). A given method may or may not be implemented in a given IT tool, or in many of them (cardinality: 1,n).

[11] For instance, in the field of editing and imputation the so-called "Fellegi-Holt methodology" is a general approach following the principle of the minimum change. Minimum change can be obtained by adopting the Fellegi-Holt error localisation method ("For each failed edit record, the Fellegi-Holt approach first proceeds through a step of error localisation in which it determines the minimal set of variables (fields) to impute, as well as the acceptable ranges(s) of values to impute.") and the Fellegi-Holt imputation method ("A single donor is selected from passed edit records by matching on the basis of other variables involved in the edits").



Classification

We define a "*classification*" as a "a *set of discrete, exhaustive and mutually exclusive observations, which can be assigned to one or more variables to be measured in the collation and/or presentation of data*".[12]

A given provision may or may not indicate one classification (cardinality: 0,1). A given classification, if present, must be indicated by one provision, in general by many of them (cardinality: 1,n).

The attributes of the entity "classification" are:

1. identifier: name or acronym;
2. description: text;
3. version.

Definition

A definition is "*a statement of the precise meaning of something*".[13]

A given provision may or may not indicate one "*definition*" (cardinality: 0,1). A given definition, if present, must be indicated by one provision, in general by many of them (cardinality: 1,n).

The attributes of the entity "definition" are:

1. identifier: name or acronym;
2. description: text;
3. version.

5. The IT tool for the management of the repository

The conceptual scheme has been implemented in a relational database. The physical scheme is reported in figure 2.

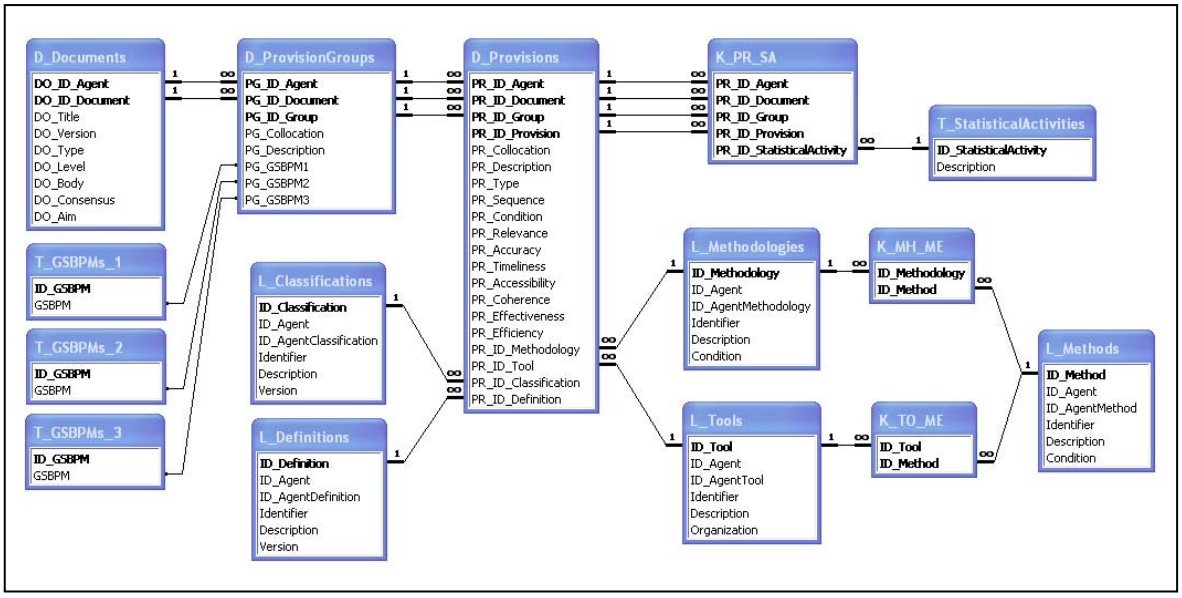
The standard of the table names gives the context of the information contained in the system.

The records stored in the "D_" tables are the instances concerning the information of the handbooks.

The "T_" tables are the fixed list of items of the conceptual schema as the GSBPM and the statistical activities. The "L_" tables contain the information stored by any user of the system and containing the information concerning any collection of items as classifications, definitions, methodologies, tools and methods. The "K_" tables are the link tables for the many to many relations present on the conceptual schema.

[12] OECD Glossary of Statistical Terms.
[13] Ibidem.

Figure 2 - Structure of the repository

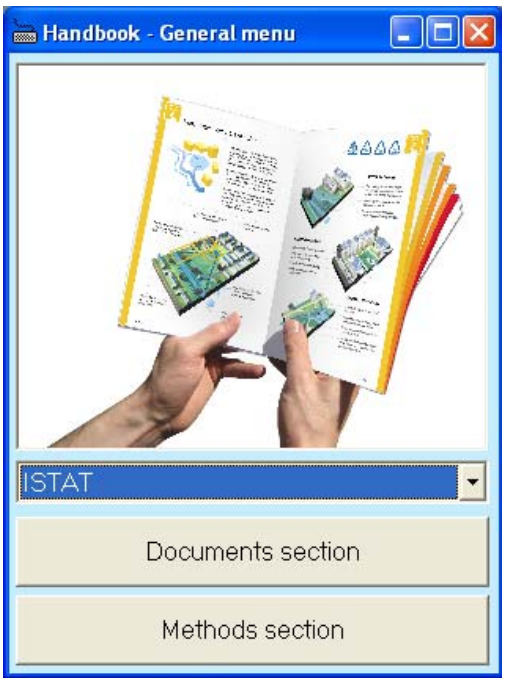


In order to allow each partner in the ESSnet project to populate the repository in an assisted way, a common identifier, including the identification of each partner (ID_Agent), has been assigned to each handbook repository table. Each attribute starts with a prefix as DO for documents, PG for provision groups and PR for provisions.

In the following we describe the most important characteristics of the software.

General flow of the operations

Running the application, the first form that will appear is the "general menu" form. The dropdown list located just under the handbook image permits to select the institute (or the specific agent) identifier. The system will enable the two section buttons when the institute has been selected.



After the identification of the Institute, the system permits to move in one of the two management sections: "Documents" (or handbooks) and "Methods".

It was suggested to proceed in this way:

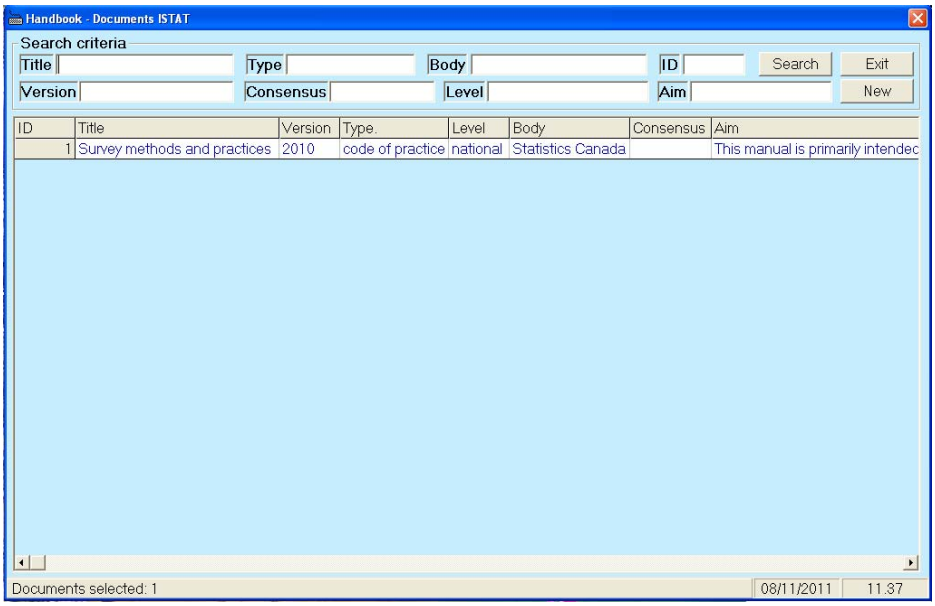
1. when analysing a given methodological handbook, start to select the "Documents" section, and proceed to introduce all necessary information related to the entities in the scheme up to the entity "methods" excluded;
2. once a new method has to be introduced, leave the "Documents" section and, from the "Main menu" select "Methods" section;
3. once the methods have been introduced, then get back to the "Documents" section and proceed to link methodologies to methods.

Documents List Form

The documents list form, as the methods one, is divided in two areas: the search frame and list grid.

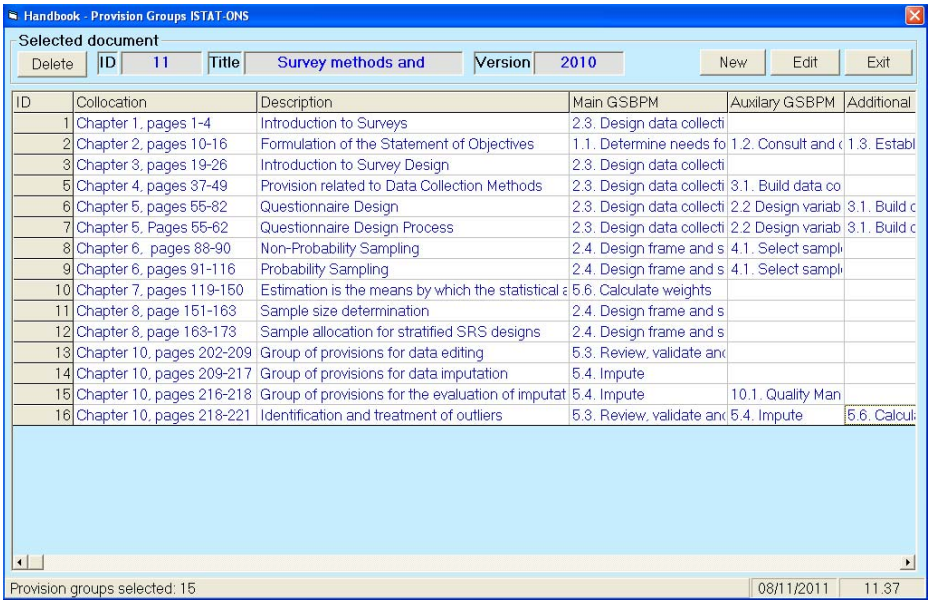
The search frame permits to display the documents by title, type, body, ID, version, consensus, level and aim. All these criteria are in "and" search condition. By pressing the "New" button, it is possible to insert a new instance for the "Document" entity.

To select a document, the user has just to click any cell of the document row, the application will show the provision group form concerning the information related at the selected document.



Provisions Groups List Form

Once a "Document" has been introduced, the Provisions Groups List Form appears. Also in this case, by clicking on the "New" button, it is possible to insert a new instance.



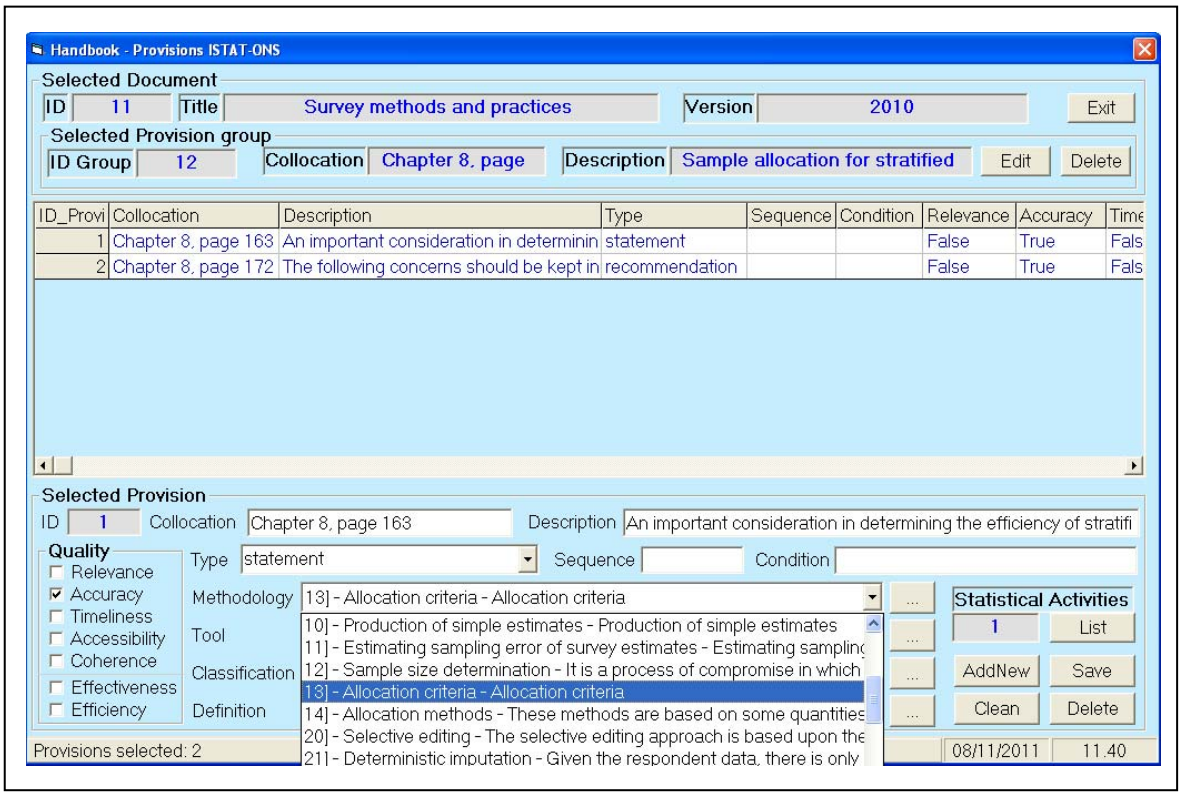
Provisions Form

For each selected Provision Group, it is possible to define new instances of "provisions". Once introduced, for each of them the values of the attributes are reported in the lower part of the form.

In this lower part, it is possible also to associate "Methodologies", "Tools", "Classifications" and "Definitions".

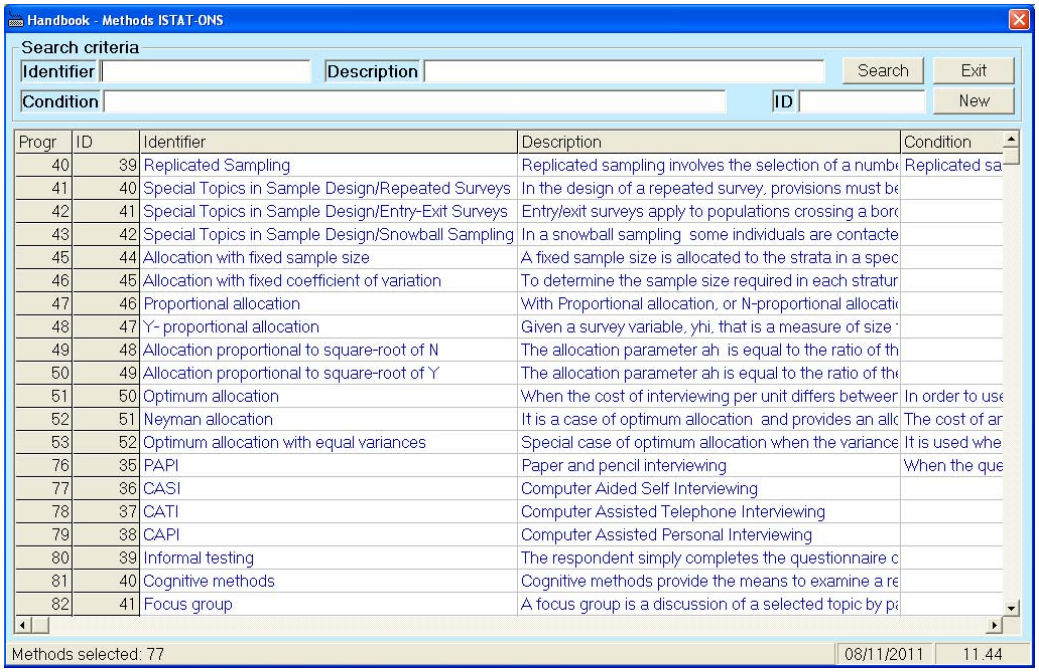
If one of these objects does not yet exist, by clicking on the corresponding "…" button it will be possible to introduce a new one that from that moment can be selected for any further provision.

It is also possible to associate one or more "Statistical activities" by clicking on "List" button: the list of statistical activities appears on the upper part of the form, hiding the provision grid and it is possible to link the ones required in multiple selection mode.



Methods List Form

From the initial form, it is possible to select the "Methods" form. In this form all the methods introduced so far will appear. The "New" button permits to add a new method. To select the displayed method the user has to click in any cell of the selected method. When a method has been selected, it will be possible to associate the methodologies to which it belongs. This procedure will link permanently the method to the methodology for each provision.



6. Procedure followed for the construction of the unified repository

For the review of the current international methodological handbooks and guidelines, it was established that six handbooks should be selected from the list of the ESS methodological handbooks, appended to Annex I, taking into account the results of the "Survey about the use of methodological standards".[14]

The choice was also guided by the criterion of guaranteeing coverage of most GSBPM phases.

In order to analyse the chosen methodological handbooks, each of them was assigned to different partners:

HANDBOOK	Organisation(s) carrying out the analysis	Organisation carrying out the reconciliation
Handbook of recommended practices for questionnaire development and testing in the European Statistical System	ONS & Insee	ONS
Eurostat sampling reference guidelines: Introduction to sample design and estimation techniques	Destatis	-
Survey methods and practice (Stat.Can.)	Istat & ONS	ISTAT
Guidelines for statistical metadata on the Internet	Insee	-
ESS handbook for Quality reports	GUS &KSH	KSH
Edimbus: editing and imputation of cross sectional business statistics	KSH & Destatis	Destatis

In some cases, the analysis (i.e. the introduction into the repository of the information required for each entity in the scheme) has been carried out by a single Institute.

In some other cases, the analysis has been performed by two Institutes. The aim of this choice was to compare the results, analyse the differences and assess the degree of arbitrariness in the process.

[14] Survey about the use of methodological standards, EUROSTAT, 2010 (Eurostat internal document).

In order to build a unique repository for all the handbooks, for the three cases in which the analysis has been done by two partners, it was decided to assign to one of them the responsibility for a reconciliation step. The partner responsible for the compared analysis had to detect:

1. inconsistencies between the two repositories, that could regard
   - provisions: differences in definition and/or attributes;
   - other objects (methodologies, methods, IT tools, etc.): differences in definitions and/or in links;
2. non-correspondences (missing items in one repository with respect to another).

The responsible partner had to list all the identified problems, and discuss them with the other partner, in order to agree on the final result, that is a unique repository of validated provisions and other objects, for the given handbook.

Once for each handbook a unique repository has been defined, then it has been possible to build the final, unified repository containing all the information derived by the analysis of the six methodological handbooks.

This final repository has been analysed in order to describe its content and also to illustrate with some examples how it is possible to use it in order to obtain useful information regarding provisions coming from different handbooks and dealing with the same field of application.

Construction of the repository for the Handbook Survey Methods and Practices

In order to provide an idea of the complexity of the process of constructing of the repository, in this paragraph we summarise the analysis conducted by ISTAT, together with ONS, on the handbook Survey Methods and Practices. This example also shows the influence of subjectivity in this task, that can be overcome by defining strict criteria and rigorous rules for the definition of standards.

The only chapters in common between ISTAT and ONS were Chapter 4, Data Collection Methods, and Chapter 5, Questionnaire Design.

About the first of the two chapters, a general comment is that ISTAT and ONS followed a similar approach: both defined a generic provision regarding data collection as a statement, but while ISTAT identified three methodologies (one for data collection in general, one for self-interviewing and one for interviewer assisted techniques) each of them linked to the related provisions, ONS identified the general methodology linked to all the provisions. As a matter of fact ONS linked all the identified methods to the general methodology, while ISTAT distinguished the methods to be linked to each methodology.

With regard to Chapter 5, Questionnaire Design, the main difference regards the number of identified Provisions Groups: ONS defined only one group (Questionnaire design), while ISTAT split the different aspects concerning the questionnaire design treated in the chapter in separated groups. Furthermore, ONS identified two methodologies, Questionnaire design and Questionnaire testing, while ISTAT did not consider Questionnaire design as a methodology itself.

As a general remark, it is possible to conclude that the two approaches were quite similar and that the framework of the repository allowed not only to highlight the differences but also to reconcile them. Finally, an observation valid for both the chapters analysed in common is that ISTAT assigned the attribute “statement” in a smaller number of cases than ONS did, often preferring the attributes “instruction” or “recommendation”.

The analysis of the common chapters led to the final version, combining the two analyses. Regarding Chapter 4, the final version contains the provisions identified by both NSIs, together with all the other provisions defined separately by each. Three methodologies were chosen as this allows a clearer link to the methods. Relatively to Chapter 5, the integrated version defines two provision groups; the first for Questionnaire design in general (in which the majority of provisions were grouped) and a second group regarding the Questionnaire design process, which corresponds to 7 provisions to be performed in the specified sequence. Provisions referring to concepts identified by both NSIs were maintained, but when pertaining to the same issue they were defined in a single provision. In general, a number of attributes concerning the type of provision were discussed and sometimes consequently modified.

7. Description of the unified repository for methodological standards

The analysis of the unified repository has been carried out through the construction of a set of descriptive tables with the aim to highlight differences and peculiarities of the six handbooks (HB in the following) examined. In these tables the handbooks have been analysed through some synthetic information on the number of identified entities, such as provision groups, provisions, methodologies and listing the statistical activities and the quality dimensions. A further step has been to list all the combination of GSBPM sub-processes indicated for all HB all together, considering the repository as a unique DB.

It is important to consider, when analysing the results obtained, that the unified repository cannot be thought at this step as the definitive and exhaustive set of the methodological standards treated in the HBs for two different reasons:

- these HBs were not written with the precise purpose of defining standards, so their identification by each partner was often affected by a certain degree of discretion due also to the fact that the conceptual scheme of statistical standards has been designed in this project for the first time and the agreement on the precise meaning of each entity and attribute was reached 'on the job' while using it to analyse HBs;
- in addition, not all the HBs have been examined thoroughly; for some of them, due to the scarce time at disposal, only some chapters have been analysed.

On the other hand, the quantitative results shown in the following tables, together with the analysis of the content defined by partners in this repository (see par. 8), are extremely useful to clarify concepts, to reduce the level of subjectivity and to improve the future work.

The first two tables focus on the analysis of provisions and methodologies for each HB, considered in terms of number and type.

Table 1 - Counts of entities by handbook

HANDBOOK	# PGs	# PRs	# MHs	# MEs	# TOs	# CLs	# DEs
Sampling guidelines	2	5	4	20	3	0	0
EDIMBUS	9	35	9	29	0	2	0
Guidelines for Statistical Metadata	3	20	0	0	0	0	2
Quality Report	10	197	15	36	0	6	12
Questionnaire Design and Testing	6	17	14	13	0	0	0
Survey Methods and Practices	15	161	19	76	0	0	0
TOTAL	**46**	**436**	**61**	**175**	**3**	**8**	**14**

Table 2 - Distribution of provisions by type and handbook

HANDBOOK	Type of provision				Total
	# Statements	# Instructions	# Recommendations	# Other types	
Sampling guidelines	0	0	5	0	5
EDIMBUS	11	11	14	0	36
Guidelines for Statistical Metadata	5	1	11	3	20
Quality Report	105	0	91	1	197
Questionnaire Design and Testing	0	0	17	0	17
Survey Methods and Practices	64	48	47	2	161
TOTAL	**185**	**60**	**185**	**6**	**436**

A first consideration regards the total number of provision groups (PG) for each HB (Table 1): generally this number is higher for the multi-topic HB, such as Survey Methods and Practices (SMP) and Quality Report (QR), with respect to the HB devoted to specific methodological topic. The same consideration can be done for the number of provisions and methods. It is necessary to underline that for Quality Report (QR) a high number of entities has been indicated, even though only one chapter has been analysed. Another observation is about the fact that a relation exists be-

tween the number of PGs and Methodologies on one side and the number of provisions and links to methods on the other side.

Regarding the provision types (table 2), there is in general a prevalence of statements and recommendations. Moreover we see a consistent number of instructions in particular in Survey Methods and Practices and none of them in Quality Report. This result perfectly reflects the different purposes of the two multi-purpose HBs: the first is aimed at 'driving' the researcher in all the survey process, while the second one, relating to quality aspects of the survey, provides guidelines to assess quality. The coherence of this result with the handbooks characteristics confirms the ability of the unified repository to provide a true representation handbooks content.

As far as the *quality dimensions* are concerned, when analysing Table 3 the dimensions prevalently indicated are accuracy and efficiency: in general accuracy is indicated for the 89% of the provisions, with a small variability among HBs, with the exception of Metadata, while efficiency is indicated for the 31% of the provisions but with a high variability among HBs. It is worthwhile to note that for QR only accuracy has been analysed.

Table 3 - Percentage of provisions by quality dimensions and handbook (product - process)

HANDBOOK	Percentage of provisions for quality dimensions for each handbook						
	Product					Process	
	Relevance	Accuracy	Timeliness	Accessibility	Coherence	Effectiveness	Efficiency
Sampling guidelines	0.0	80.0	0.0	0.0	0.0	20.0	60.0
EDIMBUS	0.0	100.0	36.1	0.0	72.2	63.9	77.8
Guidelines for Statistical Metadata	0.0	0.0	5.0	45.0	10.0	20.0	20.0
Quality Report	0.5	99.0	0.0	0.0	0.0	0.0	0.0
Questionnaire Design and Testing	41.2	88.2	5.9	0.0	17.6	0.0	35.3
Survey Methods and Practices	19.9	87.6	16.1	10.6	14.3	30.4	58.4
TOTAL	**9.2**	**89.7**	**9.4**	**6.0**	**12.4**	**17.7**	**31.0**

From the analysis of Table 4, where the indicated *statistical activities* are listed, it is possible to underline that the multi-topic HBs refer in general to all statistical activities, while the specific ones, as Edimbus, refer to specific activities both for the survey context and for the survey phase. Moreover, QR indicated all statistical activities but also specific contexts, highlighting that peculiar quality indicators exist for specific survey contexts.

Table 4 - List of statistical activities by handbook

HANDBOOK	LIST of distinct statistical activities
EDIMBUS	All statistical subjects (D1+D2+D3) D2 Economic statistics 2.3 Business statistics 4.4 Data editing and data linkage 4.7 Data analysis
Guidelines for Statistical Metadata	4.4 Data editing and data linkage 4.5 Dissemination, data warehousing 5.5 Management and development of technological resources (including standards for electronic data exchange and data sharing)
Quality report	All statistical subjects (D1+D2+D3) D1 Demographic and social statistics D2 Economic statistics 2.1 Macroeconomic statistics 2.2 Economic accounts 2.3 Business statistics 2.4 Sectoral statistics

Table 4 - continued

HANDBOOK	LIST of distinct statistical activities
Quality report	2.4.1 Agriculture, forestry, fisheries 2.4.2 Energy 2.4.3 Mining, manufacturing, construction 2.4.4 Transport 2.4.5 Tourism 2.4.6 Banking, insurance, financial statistics 2.5 Government finance, fiscal and public sector statistics 2.6 International trade and balance of payments Intranet 2.8 Labour cost 2.9 Science, technology and innovation D3 Environment and multi-domain statistics 3.1 Environment
Survey Methods and Practices	All statistical subjects (D1+D2+D3) D4 Methodology of data collection, processing, dissemination and analysis 4.4 Data editing and data linkage

With regard to the GSBPM model, in Table 5 are listed the GSBPM sub-processes indicated as the main ones. Even though the funding of this list depends on the degree of completeness of the analysis conducted for each HB, as a general consideration, the multi-topic HBs cover quite a high number of sub-processes, while the other ones a more limited set.

Table 5 - List of GSBPM sub-processes by handbook

HANDBOOK	LIST of distinct GSBPM sub-process
Sampling guidelines	2.4. Design frame and sample methodology 4.1. Select sample 5.4. Impute 5.6. Calculate weights
EDIMBUS	5.3. Review, validate and edit 5.4. Impute 6.3. Scrutinize and explain
Guidelines for Statistical Metadata	7.2. Produce dissemination products 7.3. Manage release of dissemination products 7.4. Promote dissemination products
Quality report	2.3. Design data collection methodology 2.4. Design frame and sample methodology 2.5. Design statistical processing methodology 4.1. Select sample 10.1. Quality Management
Questionnaire Design and Testing	1.1. Determine needs for information 1.5. Check data availability 2.3. Design data collection methodology 3.1. Build data collection instrument
Survey Methods and Practices	1.1. Determine needs for information 1.2. Consult and confirm needs 1.3. Establish output objectives 2.2. Design variable descriptions 2.3. Design data collection methodology 2.4. Design frame and sample methodology 3.1. Build data collection instrument 4.1. Select sample 5.3. Review, validate and edit 5.4. Impute 5.6. Calculate weights 10.1. Quality Management

From the unified repository it is also possible to analyse the combinations of sub-processes with respect to the number of GPs, provisions, methodologies and methods (see the table 6).

At a more aggregated level, as it can be seen in the following graphic, the higher number of Provisions relates to GSBPM processes:

- 2.3. Design data collection methodology (109 provisions);
- 5.3. Review, validate and edit (52 provisions);
- 10.1. Quality Management (108 provisions).

Analysing the contents of these provisions:

- those relating to process 2.3. are mostly extracted from the HB Survey Methods and Practices, as three of the analysed chapters regarded the data collection phase, and from Questionnaire Design and Testing HB;
- those relating to process 5.3. are extracted from the HB Survey Methods and Practices, as one of the analysed chapters regarded the edit and imputation phase, and from EDIMBUS HB;
- those relating to process 10.1. are mostly extracted from the HB Quality report.

Figure 3 - Number of provisions for GSBPM sub-process in the integrated repository

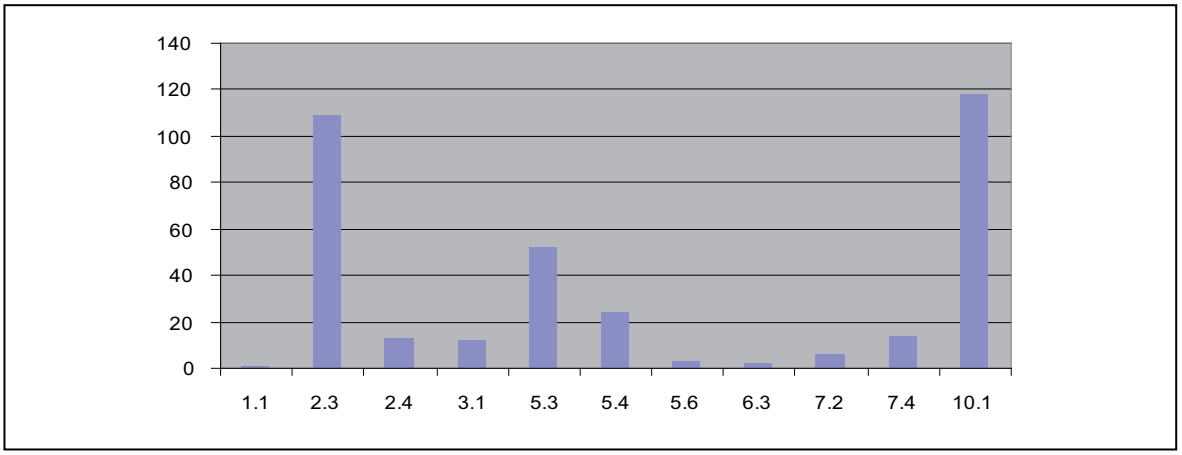


Table 6 - Number of entities for combinations of GSBPM processes in the integrated repository

GSBPM sub-process combination			# PGs	# PRs	# MHs	# MEs	# TOs	# CLs	# DEs
Main GSBPM	Auxiliary GSBPM	Additional GSBPM							
1.1. Determine needs for information	1.2. Consult and confirm needs	1.3. Establish output objectives	1	3					
1.1. Determine needs for information	1.5. Check data availability		1	1	1	1			
2.3. Design data collection methodology			2	34					
2.3. Design data collection methodology	2.2 Design variable descriptions	3.1. Build data collection instrument	2	63	1	7			
2.3. Design data collection methodology	3.1. Build data collection instrument		3	12	6	6			
2.4. Design frame and sample methodology			2	7	3	9			
2.4. Design frame and sample methodology	4.1. Select sample		2	3	2	16			
2.4. Design frame and sample methodology	4.1. Select sample	5.6. Calculate weights	1	3	3	14	2		
3.1. Build data collection instrument			2	6	5	5			
3.1. Build data collection instrument	2.3. Design data collection methodology		1	6	5	5			
5.3. Review, validate and edit			7	48	6	18		2	
5.3. Review, validate and edit	5.4. Impute	5.6. Calculate weights	1	4	2	5			

Table 6 - continued

GSBPM sub-process combination			# PGs	# PRs	# MHs	# MEs	# TOs	# CLs	# DEs
Main GSBPM	Auxiliary GSBPM	Additional GSBPM							
5.4. Impute			3	19	5	19			
5.4. Impute	10.1. Quality Management		1	3	1	1			
5.4. Impute	5.6. Calculate weights	4.1. Select sample	1	2	1	6	5		
5.6. Calculate weights			1	3	3	21			
6.3. Scrutinize and explain			2	2	2	4			
7.2. Produce dissemination products	7.3. Manage release of dissemination products	7.4. Promote dissemination products	1	6					
7.4. Promote dissemination products			2	14					2
10.1. Quality Management			7	96	8	19		1	11
10.1. Quality Management	2.4. Design frame and sample methodology	2.3. Design data collection methodology	1	14	2	2		2	
10.1. Quality Management	2.4. Design frame and sample methodology	2.5. Design statistical pro cessing methodology	1	8				1	1
10.1. Quality Management	2.4. Design frame and sample methodology	4.1. Select sample	1	79	10	30		2	
TOTAL			**46**	**436**	**66**	**188**	**7**	**8**	**14**

8. Examples of compared analyses of standards derived from different handbooks for the same GSBPM sub-processes

With a unique repository containing the results of the analysis carried out on different methodological handbooks, as the one obtained by following the procedure described in par. 3, it is possible to compare, among other things, all the objects pertaining to a given GSBPM sub-process.

In particular, it is possible to compare the provisions, that are the elementary items that compose a standard, and all related objects (methodologies, methods, IT tools, definitions and classifications).

What is the utility of such a comparison?

Accordingly to the definition of standard given in par. 2, it is clear that the methodological handbooks that have been taken into consideration cannot be considered as "standards", at least for the whole European Statistical System. They are rather "*normative documents*" of the kind "*code of practice*", usually valid locally inside the body that produced them.

In case an initiative to produce statistical standards valid for the whole European Statistical System should take place, a *recognized body* could be instituted, with the aim of defining these standard for each important step of a statistical process, i.e. for each GSBPM sub-process. This body could take into account the most important methodological handbooks currently used inside the National Statistical Institutes and the other international statistical organisations, as a good start for the process of definition of the standards.

Each handbook could be analysed following the same procedure adopted here, thus populating a repository with the structure here proposed. Then, for a given GSBPM sub-process, the provisions originated by the different handbooks could be evaluated in order to decide which of them could be promoted to become part of the final "standard" for that sub-process.

In the following, we report an example of comparison, related to GSBPM sub-process "*5.3. Review, validate and edit*".

In this comparison, we analyse all objects in the repository that are related, directly or indirectly, to the main GSBPM sub-process "5.3. Review, validate and edit". This sub-process has been indicated in two methodological handbooks: "Survey Methods and Practices" and "Recommended practices for Editing and Imputation in Cross-Sectional Business Surveys (EDIMBUS)".

First, we compare the different groups of provisions that have been defined from the two handbooks.

Table 7 - Groups of provisions defined from the two handbooks

SURVEY METHODS AND PRACTICE	EDIMBUS
Data editing	Detection of errors
	Missing values
	Systematic errors
	Random errors
Identification and treatment of outliers	Influential errors
	Outlier

Clearly, there is a more compact way of defining provision groups for "Survey Methods and Practices" than for "EDIMBUS", but it is possible to establish a clear correspondence between the two sets of groups.

Now, we consider the different provisions, distinctly for the two handbooks.

Table 8 - Provisions defined from "Survey Methods and practice"

GROUP OF PROVISIONS	Provisions
Data editing	1.Editing should be performed at several stages of the survey.
	2. The purpose of editing is to: better understand the survey processes and the survey data; detect erroneous or missing data; follow-up with the respondent; send a record to imputation; delete a record.
	3. Editing should be used to provide information about the survey process, either in the form of quality measures for the current survey or to suggest improvements for future surveys.
	4. When starting a survey, some assumptions are made about the data. During editing, it is possible to test the validity of these assumptions. For example, it may become obvious that some range edits were too strict or that some sequencing edits failed too.
	5. Editing is the application of checks to identify missing, invalid or inconsistent entries that point to data records that are potentially in error.
	6. Edits applied throughout collection and processing should be consistent with each other.
	7. Editing can be automated by means of a computer program.
	8. Edits should be developed by staff who have expertise in the subject matter, questionnaire design, data analsis and with other similar surveys.
	9. There are three main categories of edits: validity, consistency and distribution edits.
	10. Validity edits verify the syntax of responses and check that the coded data lie within an allowed range of values.
	11. Consistency edits verify that relationships between questions are respected.
	12. Distribution edits attempt to identify records that are outliers with respect to the distribution of the data.
	13. Information on the types of edits performed and the impact of editing on the survey data should be communicated to users.
	14. Quality assurance and quality control procedures should be applied to minimise and correct errors introduced during editing.
	15. Selective editing practices are recommended, particularly for business surveys (i.e., where the population is skewed and a few businesses dominate the estimates).
	16. Follow-up is generally limited to edits failures identified during collection or arising from selective editing.
Identification and treatment of outliers	17. An outlier is an observation or subset of observations that appears to be inconsistent with the remainder of the dataset.
	18. Outliers are detected by measuring their distance from the centre of data.
	19. Outliers detected at the editing stage of the survey process can be treated in various ways.
	20. The goal of outlier treatment is to decrease the impact that the outlier has on the sampling variance of the estimate without introducing too much bias.

Table 9 - Provisions defined from "EDIMBUS"

GROUP OF PROVISIONS	Provisions
Detection of errors	1. Detection of errors consists of identifying values that are not acceptable with respect to some pre-defined logical, mathematical or statistical criteria. Since different error types may contaminate the observed data in a sample of units, usually the error detection process consists of a set of error detection methods dealing each with a specific type of error.
Missing values	2. Missing values stem from questions the respondent did not answer. Non-response can be due to several reasons; the respondent may not know the answer, may not be willing to respond or may have simply missed a question.
	3. Replacing missing values by zero is neither an acceptable imputation procedure nor an acceptable alternative to flagging missing values.
	4. Variables and cases with many missing values should be studied and a decision on the variables and observations to be imputed should be taken in view of the amount of missingness.
	5. Appropriate indicators on missing values should be calculated.
	6. The indicators should be analyzed to gain information on non-response mechanisms.
Systematic errors	7. A systematic error is an error that is reported consistently over time by responding units. It is a phenomenon caused either by the consistent misunderstanding of a question during the collection of data, or by consistent misinterpretation of certain answers in the course of coding.
	8. Systematic errors should be detected and treated before dealing with random errors, in particular when the Fellegi-Holt method is used and before selective editing.
	9. The analysis of indicators on edits helps to find systematic errors mechanisms. If systematic error mechanisms are found by examining edits, then appropriate deterministic checking rules to detect errors due to the systematic error mechanism should be added.
	10. If systematic error mechanisms are found, then improvements to the survey process (Questionnaire, interviewer training, coding, processing) should be made to prevent similar errors.
Influential errors	11. Influential errors are errors in values of variables that have a significant influence on publication target statistics for those variables.
	12. Selective editing is an appropriate method to focus attention on critical observations without generating a detrimental impact on data quality.
	13. Priorities should be reflected in the score functions of selective editing. Score functions should include a risk and an influence component.
	14. The thresholds of score functions should be chosen carefully during the tuning and testing of the E&I process. The thresholds and parameters of score functions should be revised whenever the survey process is changed (Questionnaire, data entry, removal of systematic errors).
	15. The quality of the anticipated value should be assessed at least for a sub-sample of the survey. E&I flags have to be taken into account if data of earlier periods are used as anticipated values.
	15. Units, for which the score function cannot be calculated have to belong to the critical stream to prevent biasing the estimates.
	17. Macroediting should be properly documented.
	18. Important publication aggregates, publications cells and publication sub-populations should be considered in macroediting.
	19. The quality of reference data and problems like inflation and structural differences (e.g. definitions) should be taken into account.
	20. Macroediting should be performed before releasing data for final estimation.
Outliers	21. An outlier is an observation which is not fitted well by a model. The model can be a parametric distribution or the model can be a more loosely defined concept like "close to the center of the data". In the latter case an outlier is an observation which is not close to the center of the data.
	22. Outliers and influential observations should be detected. Such observations may be errors or correct but sometimes their correctness is unclear. Errors should be treated and also correct or unclear influential or outlying observations may have to be treated and also correct or unclear influential or outlying observations may have to be treated to prevent potential bias and high variability.
	23. The influence on important results should be controlled even after selective editing and outlier detection.
	24. Outlier detection methods should be robust against outliers. Therefore methods based on means or weighted means and standard deviations, which are not robust, should be avoided.
	25. Simple univariate methods, graphical displays or more complex multivariate techniques should be used for outlier detection depending on possible models among variables. Different models may have to be applied in different sub-populations.
	26. When choosing tuning constants for detection or treatment several tuning constants should be tested and the corresponding impact on estimates and their variances observed. Often the treatment of a few outliers is acceptable.

Table 9 - continued

GROUP OF PROVISIONS	Provisions
Random errors	27. Random errors are errors that are not caused by a systematic reason, but by accident. They primarily arise due to inattention by respondents, interviewers and other processing staff during the various phases of the survey cycle.
	28. The data in each record should be made to satisfy all edits by changing the fewest possible items of data (fields).
	29. As far as possible the frequency structure of the data file should be maintained.
	30. Imputation rules should be derived from the corresponding edit rules without explicit specification.
	31. If appropriate software is available, random errors should be detected and treated by applying the Fellegi-Holt paradigm.
	32. If appropriate software for applying the Fellegi-Holt paradigm is not available, random errors may be detected by means of deterministic checking rules. However, the error localization becomes more arbitrary with deterministic checking rules.

Survey Methods and Practices (SMP) counts 20 provisions, while EDIMBUS contains 32 provisions dealing with sub-process 5.3. It is not surprising, as EDIMBUS is dedicated to the edit and imputation steps of surveys.

By comparing the two sets of provisions, we can classify them in:

1. those dealing with the same arguments, and that could be merged (having verified they are not contradictory): for instance, "selective editing" (provision # 15 in SMP, and provision # 12 in EDIMBUS): these provisions may be potentially merged ("*Selective editing practices are recommended, particularly for business surveys (i.e., where the population is skewed and a few businesses dominate the estimates), as they are appropriate methods to focus attention on critical observations without generating a detrimental impact on data quality*");
2. those dealing with the same argument, but not compatible: for instance, provision # 17 in SMP and # 21 in EDIMBUS. They are both statements defining outliers: even they are both correct, the first may be considered too generic with respect to the second, that should be adopted;
3. non overlapping provisions: for instance, provisions from # 1 to # 14 in SMP are more general than those in EDIMBUS; or provisions from #7 to # 10 do not have correspondence in SMP. In this case, once verified their validity, they could be all promoted to be part of a new standard for editing.

Let us now consider methodologies and methods:

Table 10 - Methodologies and Methods defined from the two handbooks

SURVEY METHODS AND PRACTICES		EDIMBUS	
Methodologies	Methods	Methodologies	Methods
Selective editing	Questionnaire Score Method (Berthelot-Latouche)	Detection of missing values	
	Graphical method	Detection of systematic errors	Analysis of fatal edits
	Top-down		Analysis of ratio Edits
	Aggregate method		Finite mixture Models
Methods for outliers identification	Relative distance from the centre of data	Detection of influential errors	Macroediting
	Quartile method		Selective editing
		Detection of outliers	Hidiroglou-Berthelot Method
			Median absolute deviation Rule
			Regression-based models
			Graphical editing
			One-step robustified ratio estimator
			Weighted quantiles
			Winsorised mean and trimmed mean
		Localising random errors	Deterministic checking Rules
			Fellegi-Holt paradigm

Once again, there is a higher number of methodologies and methods defined in EDIMBUS than in SMP.

Despite the different naming adopted in the two handbooks, it is possible to establish some equivalences between the defined objects:

- "*Selective editing*" methodology in SMP corresponds to "*Detection of influential errors*" methodology, but also to "*Selective editing*" method, in EDIMBUS;
- "*Methods for outliers identification*" in SMP corresponds to "*Detection of outliers*" in EDIMBUS;
- the couple "*Top-down*" and "*Aggregate method*" methods in SMP correspond to "*Macroediting*" in EDIMBUS;
- "*Graphical method*" in SMP corresponds to "*Graphical editing*" in EDIMBUS (but the former is linked to the "*Selective editing*" methodology, while the latter to the "*Detection of outliers*" methodology).

As an example of integration, we propose to consider the union of all methodologies and methods, with the necessary re-classification in order to avoid inconsistencies (for instance, "*Selective editing*" should be considered as a methodology, making use of different methods).

In the following table a possible integrated solution is reported. This unified list of methodologies and methods could be the starting point for the integration also of the provisions and provision groups, when constructing the final "standard" for this particular sub-process.

Table 11 - Integrated solution for Methodologies and Methods defined from the two handbooks

METHODOLOGIES	Methods
Detection of missing values	
Detection of systematic errors	Analysis of fatal edits
	Analysis of ratio edits
	Finite mixture models
Detection of influential errors	Top-down
	Aggregate method
Selective editing	Questionnaire Score Method (Berthelot-Latouche)
	Graphical method
Detection of outliers	Hidiroglou-Berthelot method
	Median absolute deviation rule
	Regression-based models
	Graphical editing
	One-step robustified ratio estimator
	Weighted quantiles
	Winsorised mean and trimmed mean
Localising random errors	Deterministic checking rules
	Fellegi-Holt paradigm

9. Concluding remarks

As shown in this paper, the work carried out in this ESSnet project has been very productive because it succeeded in building a framework for standardisation which is fundamental to define and clarify concepts related to standards but also to manage them with the support of a software package.

As a matter of fact, the conceptual scheme that has been designed to represent methodological statistical standards, allows to define and describe the constituent elements of a statistical standard, in terms of concepts, relationships between them and attributes which characterize them.

In addition, the repository designed to store standards according to this scheme - on the basis of contents of methodological handbooks - is able to represent a coherent, not redundant and exhaustive way to document standards, while the procedure settled to populate the repository helped to define a structured method to identify and verify consistence among standards.

In synthesis, what produced in this ESSnet project represents not only suggestions to the future work on standardisation, especially to the sponsorship on standardisation, but provides also a theoretical model and a tool, even if at a prototypal level, to define, document and manage statistical standards.

References

Breitenberg M., 2009. *The ABC of standards activities* - US National Institute of Standards and Technology (NIST).

Chen P.P., 1975. *The entity-relationship model—toward a unified view of data*, ACM Transactions on Database Systems (TODS), September 22–24.

Delaney H., Van de Zande R., 2000. *A Guide to EU Standards and Conformity Assessment* - (NIST).

Eurostat (2009), *ESS Handbook for Quality Reports*, 2009 Edition.

EUROSTAT, 2010. *Survey about the use of methodological standards.*

ISO/ IEC, 2004. *Guide 2 - Standardization and related activities* - Annex 1: General vocabulary.

ISO*,* 2005. *ISO and conformity assessment.*

OECD - *Glossary of Statistical Terms*, http://stats.oecd.org/glossary/.

STATCAN, 2007. *Canadian advisory panel on opinion surveys.*

STATCAN, Statistics Canada Glossary, http://www.statcan.gc.ca/edu/power-pouvoir/glossary-glossaire/5214842-eng.htm.

UNECE, 2009. *Generic Statistical Business Process Model - Version 4.0* - www.unece.org/stats/gsbpm.

UNECE, 2009. *Generic Statistical Business Process Model - Version 4.0* www.unece.org/stats/gsbpm.

UNECE, 2010. *Developing a Global Inventory of Statistical Standards.*

Annex 1: methodological handbooks to be reviewed

Table A1 - List of general ESS methodological handbooks

Handbook of Recommended Practices for Questionnaire Development and Testing in the European Statistical System (2004)	http://epp.eurostat.ec.europa.eu/portal/page/portal/research_methodology/documents/RPSQDET27062006.pdf
Eurostat sampling reference guidelines: Introduction to sample design and estimation techniques (2008)	http://epp.eurostat.ec.europa.eu/cache/ITY_OFFPUB/KS-RA-08-003/EN/KS-RA-08-003-EN.PDF
Benchmarking through calibration of weights for microdata	http://epp.eurostat.ec.europa.eu/cache/ITY_PUBLIC/KS-DT-05-007/EN/KS-DT-05-007-EN.PDF
Edimbus: editing and imputation of cross sectional business statistics	http://epp.eurostat.ec.europa.eu/portal/page/portal/quality/documents/RPM_EDIMBUS.pdf
Handbook on Constructing Composite Indicators: Methodology and User Guide (2008)	http://www.oecdbookshop.org/oecd/display.asp?CID=&LANG=en&SF1=DI&ST1=5KZN79PVDJ5J
ESS guidelines on Seasonal Adjustments (2009)	http://epp.eurostat.ec.europa.eu/cache/ITY_OFFPUB/KS-RA-09-006/EN/KS-RA-09-006-EN.PDF
The seasonal adjustment of short time series (2005 Edition):	http://ec.europa.eu/eurostat/ramon/statmanuals/files/KS-DT-05-002-EN.pdf
Guidelines for Seasonal Adjustment on Quartlerly National Accounts	http://epp.eurostat.ec.europa.eu/portal/page/portal/national_accounts/documents/quarterly_accounts/SAWD_RECOMMENDATIONS.PDF
The seasonal adjustment of qualitative business and consumer surveys (2005 Edition):	http://ec.europa.eu/eurostat/ramon/statmanuals/files/KS-DT-05-001-EN.pdf
Variance estimation methods in the European Union (2002)	http://epp.eurostat.ec.europa.eu/portal/page/portal/research_methodology/documents/MOS_20VARIANCE_ESTIMATION_202002.pdf
Handbook on statistical disclosure control	
Revisions Policy for Official Statistics: A Matter of Governance	http://ideas.repec.org/p/imf/imfwpa/04-87.html
Handbook for Monitoring and Evaluating Business Survey Response Burdens	http://epp.eurostat.ec.europa.eu/portal/page/portal/quality/documents/HANDBOOK%20FOR%20MONITORING%20AND%20EVALUATING%20BUSINESS%20SURVEY%20R.pdf
Benchmarking through calibration of weights for microdata	http://epp.eurostat.ec.europa.eu/cache/ITY_PUBLIC/KS-DT-05-007/EN/KS-DT-05-007-EN.PDF
International Standard Cost Model Manual – Measuring and reducing administrative burdens for businesses	http://www.oecd.org/dataoecd/32/54/34227698.pdf
ESS Handbook for Quality Reports	http://epp.eurostat.ec.europa.eu/portal/page/portal/product_details/publication?p_product_code=KS-RA-08-016
Handbook on improving quality by analysis of process variables	http://epp.eurostat.ec.europa.eu/portal/page/portal/quality/documents/HANDBOOK%20ON%20IMPROVING%20QUALITY.pdf
Quality Guidelines	http://www.statcan.ca/bsolc/english/bsolc?catno=12-539-X&CHROPG=1
Guidelines for finding a balance between accuracy and delays in the statistical surveys	http://epp.eurostat.ec.europa.eu/portal/page/portal/quality/documents/GUIDELINES_FOR_BALANCE_BETWEEN_ACCURACY_AND_DELAYS.pdf
Handbook on the design and implementation of business surveys (1997)	http://epp.eurostat.ec.europa.eu/portal/page/portal/research_methodology/documents/Handbook_on_surveys.pdf

Table A2 - The top 15 known standards (from "Survey on methodological handbooks" - October 2010)

PROCESS PHASE	International standard	N.*
6. Analyse	ESS guidelines on Seasonal Adjustments (2009)	19
10. Overarching	ESS Handbook for Quality Reports (2009)	18
1. Specify Needs	Handbook of Recommended Practices for Questionnaire Development and Testing in the European Statistical System (2004)	18
4. Collect	Handbook of Recommended Practices for Questionnaire Development and Testing in the European Statistical System (2004)	18
2. Design	UNECE: Guidelines for the modelling of statistical data and metadata (1995)	17
6. Analyse	Guidelines for Seasonal Adjustment on Quarterly National Accounts	17
7. Disseminate	UNECE: Guidelines for statistical metadata on the internet (2000)	17
10. Overarching	Handbook on data quality assessment methods and tools (2005)	17
10. Overarching	Handbook on improving quality by analysis of process variables (2002)	17
10. Overarching	Quality Guidelines, Statistics Canada (1985, …, 2009)	17
6. Analyse	Handbook on statistical disclosure control (2010)	16
6. Analyse	Variance estimation methods in the European Union (2002)	16
5. Process	Edimbus: editing and imputation of cross sectional business statistics (2007)	16
4. Collect	Eurostat sampling reference guidelines: Introduction to sample design and estimation techniques (2008)	16
5. Process	Eurostat sampling reference guidelines: Introduction to sample design and estimation techniques (2008)	16

* Number of countries reporting the international standard as "known"

Table A3 - The selected 6 methodological handbooks to be analysed in ESSnet "StandPrep"

HANDBOOK	GSBPM phases
Survey Methods and Practices (2010)	All
Handbook of Recommended Practices for Questionnaire Development and Testing in the European Statistical System (2004)	1.Specify Needs; 4.Collect
Eurostat sampling reference guidelines: Introduction to sample design and estimation techniques (2008)	4.Collect; 5.Process
Edimbus: editing and imputation of cross sectional business statistics (2007)	5.Process
Handbook on statistical disclosure control (2010)	6. Analyse
ESS Handbook for Quality Reports (2009)	10. Overarching

Informazioni per gli autori

La collana è aperta ad autori dell'Istat e del Sistema statistico nazionale, e ad altri studiosi che abbiano partecipato ad attività promosse dal Sistan (convegni, seminari, gruppi di lavoro, ecc.). Da gennaio 2011 essa sostituirà Documenti Istat e Contributi Istat.

Coloro che desiderano pubblicare sulla nuova collana dovranno sottoporre il proprio contributo alla redazione degli Istat Working Papers inviandolo per posta elettronica all'indirizzo iwp@istat.it. Il saggio deve essere redatto seguendo gli standard editoriali previsti, corredato di un sommario in italiano e in inglese; deve, altresì, essere accompagnato da una dichiarazione di paternità dell'opera. Per la stesura del testo occorre seguire le indicazioni presenti nel foglio di stile, con le citazioni e i riferimenti bibliografici redatti secondo il protocollo internazionale 'Autore-Data' del *Chicago Manual of Style*.

Per gli autori Istat, la sottomissione dei lavori deve essere accompagnata da una mail del proprio dirigente di Servizio/Struttura, che ne assicura la presa visione. Per gli autori degli altri enti del Sistan la trasmissione avviene attraverso il responsabile dell'ufficio di statistica, che ne prende visione. Per tutti gli altri autori, esterni all'Istat e al Sistan, non è necessaria alcuna presa visione. Tutti i lavori saranno sottoposti al Comitato di redazione, che valuterà la significatività del lavoro per il progresso dell'attività statistica istituzionale. La pubblicazione sarà disponibile su formato digitale e sarà consultabile on line.

Gli articoli pubblicati impegnano esclusivamente gli autori, le opinioni espresse non implicano alcuna responsabilità da parte dell'Istat. Si autorizza la riproduzione a fini non commerciali e con citazione della fonte.