Results Achieved, Difficulties Encountered And Challenges Still To Be Faced In Implementing Istat Quality Strategy

Giorgia Simeoni, Gabriele Ascari, Maria Francesca Loporcaro and Mauro Scanu

Istituto Nazionale di Statistica – ISTAT, Roma, Italy

1. Introduction

In September 2020, once the modernization process started in 2016 reached a good level of maturity, Istat renovated its commitment to quality by re-constituting a Quality Committee, that had been already working from 2010 to 2016, and formalising the role of Quality Manager.

In 2021 a new quality strategy, developed by the Quality manager with the Quality committee support, was approved by the Istat top management. It includes a set of actions to be conducted in the following 3 years with the aim:

i) to renovate or improve existing quality tools like metadata systems;

ii) to develop and implement a new procedure for the quality assessment of Istat statistical processes;

iii) to improve the quality of statistics produced by the National Statistical System.

Concerning metadata systems, Istat is currently equipped with two separate metadata systems, SIDI-SIQual for reference metadata, process and quality documentation, and SUM for structural metadata. Both the systems need to be renovated and we are currently designing a new unique metadata system called METAstat. It will be integrated with the several Istat Systems to facilitate e.g.: metadata harmonisation and their reuse in statistical production, statistical process documentation and quality report production according to international standards like GSBPM [8,9] and SIMS [3], quality indicators collection and analysis on statistical processes and products.

With regard to quality assessment of statistical processes, the evaluation methods defined in the quality strategy are differentiated by statistical process types, depending on their levels of maturity, methodological soundness and standardisation:

- For traditional processes (e.g. surveys) a checklist was developed to verify the compliance of these processes to sound methodologies and good practices. The "compliant" processes obtain an internal label of conformity. Improvement actions are identified for the non-compliant and audits are going to be applied to a small subset of them.

- For complex multisource statistical processes that are at the basis of our Integrated System of Statistical Registers, a metadata model and a system of quality indicators to monitor and evaluate the different steps of such processes have been defined and tested. It is now starting to be implemented in some pilot registers. Such metadata and quality indicators will be managed in the METAstat system.

- Concerning big data and Trusted Smart Statistics it is planned to establish an ad-hoc quality framework. A review of the proposal made in international projects and in literature has been made and it will be the basis for developing the Istat proposal.

As for the National statistical system, different actions were planned for Bodies of the system that produce European Statistics (called Other National Authorities – ONAs) and the others. Specifically, for the ONAs several audits were already carried out in the period 2018-2021, while for not-ONAs, a guide for the implementation of the Italian Code for the quality of official statistics has been recently released [5,6].

The paper will describe the main actions implemented and results achieved, with particular reference to the METAstat system and to the checklist for traditional surveys assessment.

2. *Istat quality policy and checklist assessment*

In 2016-2020 Istat carried out a modernisation process that implied great changes in the statistical production. The Integrated System of Statistical Registers (ISSR) has been built. It is a new paradigm for statistical production. A system of services (Methodological, IT,…) has been set up to provide support to the production process. Furthermore, Istat is taking up the challenge of integrating new data sources, such as Big Data, and traditional sources in statistical production.

After the modernisation, a renovation of quality policy was deemed necessary to adapt the quality methods and tools the new production environment. In this new environment, in 2020, a Quality Committee has been re-constituted (a first Quality Committee was in charge from 2010 to 2016) and a Quality Manager has been appointed. The task of defining a new quality policy proposal for the Institute was assigned to them. This new proposal has been developed in line with the existing approach, methods and tools but it also takes into account the innovations due to modernisation process.

Quality management has a long tradition at Istat. It roots back to the '90s, when Istat adopted a systematic approach in order to ensure quality to statistical products, processes and services offered to the community. The Istat quality management system is based on several tools like Information Systems for the documentation of the statistical process and products and their quality, quality guidelines, quality reporting, a training programme on quality, etc...

As described in [1], the new Istat quality policy, spanning over a time horizon of 3 years, is inspired by the common quality framework of the European Statistical System and its tools. The core of the Istat approach to quality is the quality assessment program in which different quality assessment methods are implemented, according to the type (and maturity) of statistical processes.

Quality assessment is the part of quality assurance that focuses on an evaluation of how well quality requirements (the stated needs or expectations) are fulfilled. A systematic assessment of data quality allows the statistical Institute to control the various statistical processes. Problems detected during the assessment phase can be analysed, and results can be utilized to improve statistical production processes.

There are three assessment levels [2]: the first one is the basic assessment level and deals with measurement of processes and products. The information obtained have to be structured in order to become meaningful for data quality assessment; the second level, medium assessment level, is based on the information compiled on the first level and deals with the conformity of the statistics to internal or external standards, by means of audits or peer reviews and self-assessment questionnaires. The third step goes from evaluation to conformity to requirements set out in standards. On the third level the approaches of labelling and certification are coming into operation to facilitate this task. Both methods have in common that compliance is certified with a whole set of defined requirements.

The choice in the quality policy was to employ a tool (a checklist) that is often used at the second level in self-assessment to evaluate the conformity to Istat quality standards at the third level. In particular, compliance to the checklist is certified by a label of quality.

The checklist was originally described in [1], as a tool which follows the GSBPM model and it is used for assessment of most Istat statistical processes. The checklist has been developed by the experts on quality of the Institute with the collaboration of the methodological, data collection and dissemination experts, who were consulted for specific phases. It takes the form of an electronic questionnaire filled by the process managers of each process. More than 200 processes have been involved and the response rate was about

90%. Different categories of processes were involved in the procedure: census or sample surveys, processes based on administrative data, secondary studies (processes acquiring their data from other statistical processes), processes based on other sources (Internet, publications, etc.) (table 1). Table 1 shows that even traditional processes often have been based on different sources (for example processes based on administrative data, or processes that integrates administrative data with processes acquiring their data from other statistical processes).

Table 1. Processes, by data sources (absolute and percentage values)

	Yes		No	
	freq.	%	freq.	%
Censuses or sample surveys	113	56,2	88	43,8
Administrative data	110	54,7	91	45,3
Data from other Istat statistical processes	109	54,2	92	45,8
Other sources, e.g. available on internet or publications	68	33,8	133	66,2

The main results of the survey will be presented in this paragraph.

Since one of the purposes of the checklist was to assess the conformity of the processes to the standard quality practices of the Institute, an analysis was carried out to assign an outcome to each process. The preliminary outcomes were identified as follows: "totally compliant", "improvement action(s) proposed", "clarifications needed", "clarifications+actions". The latter two are provisional outcomes needed to be finalized after returning to some of the respondents for more details. In table 2 the frequency of the outcomes is presented. The compliant processes were the 9,5% of the total, the improvement actions needed (without clarifications) were 24,6%.

Table 2. Outcomes of processes (percentage values)

| Outcome	%
Totally compliant	9,5
Clarification(s) needed	14,6
Improvement action(s) proposed	24,6
Clarifications + Action(s)	51,3
Total	**100,0**

The following figure shows the improvement actions proposed.

Figure 1. Improvement actions
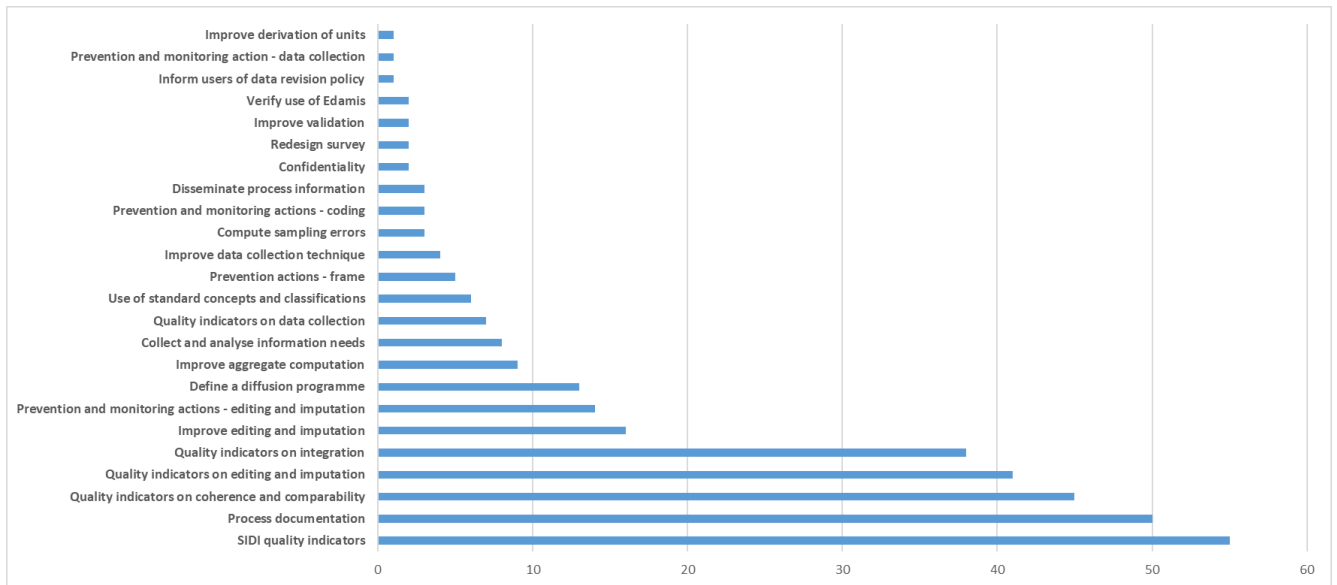
As evident, the most recurring actions are related to the computation of quality indicators. There may be multiple reasons for this. One of these reasons could be the availability of information: as shown in figure 2, indicators related to timeliness and punctuality and data collection monitoring are the ones computed by most processes. Simple information is needed in the case of timeliness and punctuality (reference date of the data, release date and planned release date), and for data collection it may be the case that most information is collected through the process and is immediately available for the computation of the indicators. On the other hand, accessibility indicators often require information that may not be easily available to the researcher of the NSI, such the number of accesses to a dataset or the number of visits to a web page.

Figure 2. Quality indicators by section of the questionnaire (percentage values)
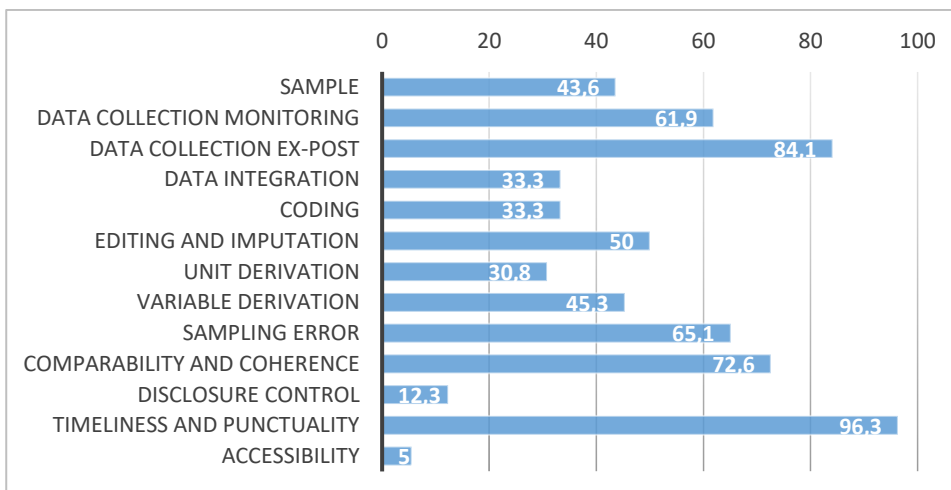
Furthermore, other actions that were frequently recommended are linked to the documentation of the process. This is confirmed in figure 3, where the actions are identified by their related phase or subprocess (which, in turn, represent distinct sections of the questionnaire): documentation and evaluation, followed by actions related to non-sampling errors and results validation, are again among the most recommended

actions. Less frequent are the actions on unit derivations and confidentiality, which means that most processes resulted compliant with regard to these aspects.

Figure 3. Improvement actions by section of the questionnaire
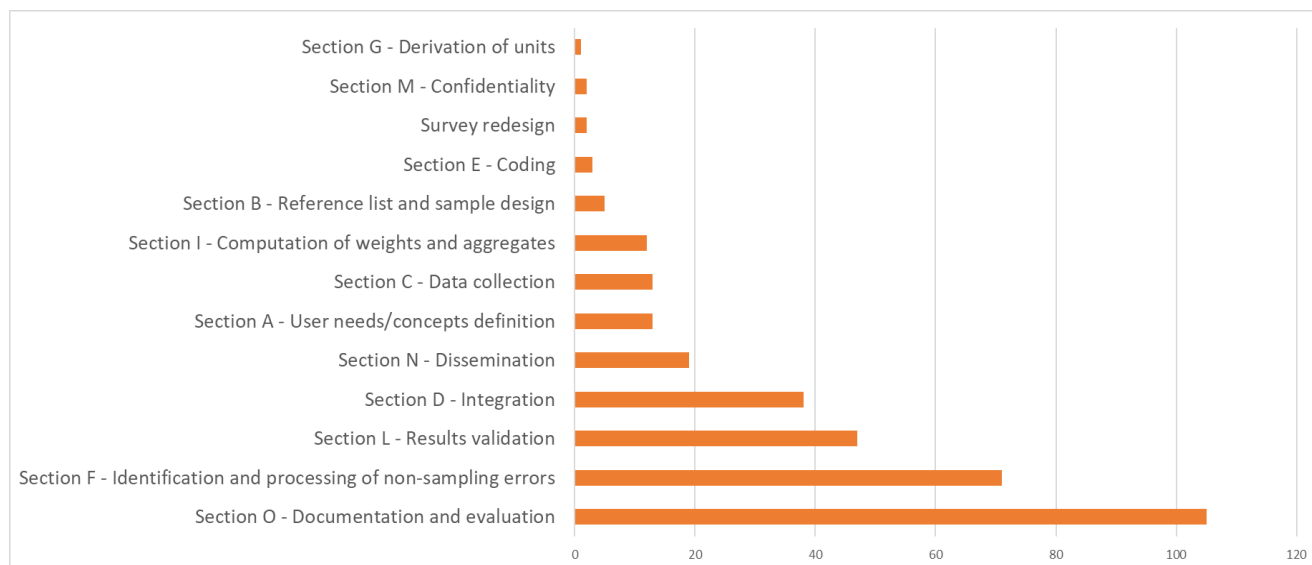
Lastly, it may be interesting to present the results regarding the methodological support. As it has been noted, the quality policy follows the implementation of the modernization programme, which rejected the traditional "silos" structure in favour of an interconnected approach to statistical productions. Under this approach, distinct sectors such as Data collection, Information Technology and Methodology provide services and support to the production domain sectors. Aspects regarding the aforementioned supporting sectors were investigated in the questionnaire, but special attention was focused on Methodology. The most frequent activities for which the production sectors of the Instituted have asked the Methodology Directorate for support are presented in figure 4.

Figure 4. Methodological support requests by method (percentage values)
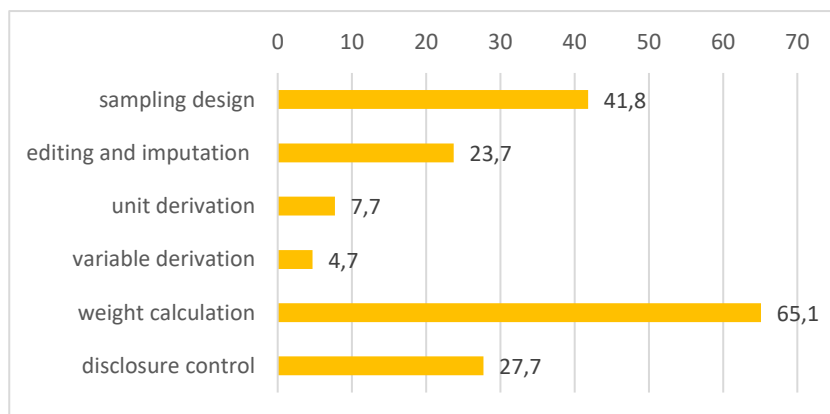
3. *Other activities on quality*

The checklist shown in the previous paragraph was developed to verify the compliance of traditional processes to sound methodologies and good practices. Indeed, while the internal quality assessment procedure is general enough to encompass statistical processes that incorporate traditional survey data and administrative data, specific attention has to be paid to more complex processes such as the ones based on statistical registers and on big data sources. Furthermore, a part of the quality policy is dedicated to processes

outside the Institute, that is the one carried out by the Italian Other National Authorities and the other Italian Statistical System organization. In these directions actions that can be considered as parallel to the main action of the internal quality assessment procedure are ongoing.

For the processes involving the ISSR, internal task forces have been working on defining metadata and quality indicators according to the GSBPM and GSIM definitions that could be used to describe and assess the quality of processes that are carried out through the ISSR.

Although the ISSR is fed by great volumes of data, they mainly come from administrative sources and on-field survey. The term "big data" is usually adopted to point out the most recent and innovative sources of raw information and (potential) statistical data, such as smart meters, satellite images, mobile network data, and so on. The exploration of the usability of these sources for official statistics is currently an ongoing research, not only at Istat but in the overall international statistical community. Of course, the quality of these data and the related methodologies should be assessed, but the traditional measures for quality evaluation may fall short for some categories of big data. Concerning big data and Trusted Smart Statistics it is planned to establish an ad-hoc quality framework. A review of the proposal made in international projects has been made and it will be the basis for developing the Istat proposal. For the aspects related specifically to Trusted Smart Statistics, an internal technical note [7] has been produced, which will help the research that will be carried out by an internal working group that has recently been implemented.

Lastly, the quality action towards the external National Statistical System continue in the direction of mirroring the structure and quality framework at the European level: the recent addition to the already existing set of tools available to the organizations of the Sistan is a guide that recommends specific methods for each principle of the Italian version of the European Code of Practice; in other words, a guide that mirrors the purposes of the Quality Assurance Framework of the European Statistical System. The volume, called *Guida per l'implementazione del Codice Italiano per la qualità delle statistiche ufficiali (Guide for the implementation of the Italian Code for the quality of official statistics)* [5], was released in October 2023 and is aimed at the not-ONAs organizations of the system, since the ONAs can use the European Code of Practice and the QAF as a reference [4].

4. The development of a new metadata system

In this section the proposal for the development of the METAstat metadata system is sketched, with an overview of the expected benefits that would arise from the implementation of a centralized metadata system for modernization purposes, as well as a summary of the results obtained so far.

Characteristics of the METAstat project

METAstat is the proposal for a new metadata system, which includes already existing but separated systems:

- the glossary of terms,
- the specialization of terms in data documentation (structural metadata) throughout the life cycle of the data,
- the description of the processes (referential metadata)
- quality indicators.

This new system has a particularly innovative character as METAstat is not just a documentary system, but also a proactive system providing various Istat system using metadata with validated and standardised ones Thus, METAstat has the ambition of adding functionalities to the ones already present in the existing systems.

In the first year of activity of this project, it was possible to complete the analysis of the stakeholders' requirements, the definition of the Institute's governance on metadata (which impacts relationships with and

between data systems that use metadata) and the completion of an overall draft project, through its organization into modules.

The design and development activities of the various METAstat modules is starting. It is expected that they will focus primarily on the core METAstat processes, i.e.:

- the terms management system (glossary);
- the management of the registry of statistical processes;
- the management system for classifications.

As a final result, we expect that the system will allow:

- Reuse of terms and definitions from the glossary in all systems, including those exposed to external users (data browser, Armida, publications,...);
- Accurate and consistent documentation with the measures useful for assessing the quality of all statistical processes, including those of the ISSR, as well as quality reporting as required by European standards;
- Dissemination to all systems that contain data (from acquisition to dissemination systems) of updated classification versions or the changes that need to be defined on the same classification version (suggestion of the ATECO Committee), also as Linked Open Data;
- Structuring of a transcoding system aimed at promoting interrelationship with the acquisition of data sources, facilitating correspondences between what happens in the administrative field and what is needed in the Institute (required functionality of the work register).

The system, at the end of its development, would also allow data to be accessed directly from the metadata, for example from the glossary or the classifications displayed on the institutional website, projecting the institute towards metadata driven and therefore favoring a friendlier query of the data. A further aspect to highlight is that METAstat is designed to promote the reuse of metadata not only between different production processes but also between different phases of the same process, making greater efficiency possible, avoiding duplication of similar activities.

4.1. METAstat modules and development priorities

As mentioned, we have organized METAstat into modules. Furthermore, taking into account the availability of some functions in active although obsolete systems (for example SIDI-SIQual), the Metadata Committee that supervises the project proposes the following order of priority in the development of the various METAstat modules.

1. *Glossary, DB and management*. The maximum priority is linked to the fact that this part is currently managed in Access or Excel. This part is preparatory to the development of the structural and referential components of the metadata and is therefore useful for the other modules.
2. *Statistical process registry management*. As with the glossary, the process registry is useful for the development of the other modules.
3. *Basic structural metadata management*. This module allows to manage essential structural metadata (variables, units and classifications) without organizing them into data structures (see point 6).
4. *Management of referential metadata - "sub-process" level of detail*. Management of the documentation of the phases and sub-processes of the statistical processes according to the metadata model developed for the processes of the ISSR. Process quality indicators could also be developed in this module.
5. *Management of product quality indicators*.

6. *Management of data structures.* This module includes the connection of data structures and data sets to the process phases.
7. *Reuse functions*. This module derives methodological notes, quality reports, and so on from the available information in the system.
8. *Development of services on reporting and consultation*. This module include the dissemination of metadata in various formats.

5. Conclusions

A quality strategy, especially in official statistics, needs to be flexible enough to keep track of all the innovations that are constantly being introduced in this field, especially regarding new data collection techniques and available sources. Istat quality policy, developed after a profound renovation programme in the organisational structure of the Institute, aims to encompass all these changes. The most recent focus of the policy was a new assessment procedure to explore the quality aspects of most statistical processes within Istat. This has been an ambitious action not only for the number of processes (and process managers) involved, but also because each specific assessment aimed to explore all the phases of the process, from the beginning to the end. Therefore, experts from many methodological fields and not only on quality were involved in the preparatory work of the checklist, the main tool through which the assessment was conducted.

Furthermore, other activities have been implemented to investigate the quality of other categories of processes that could not be included in the checklist assessment, mainly the ones based on the ISSR and on Trusted Smart Statistics. Considering also the activities dedicated to support the quality improvements of the National Statistical System, a comprehensive picture emerges of a strategy that has to face the issues brought by the many aspects of the complex and multifaceted official statistics production.

The new metadata system METAstat has the ambition to sustain the whole set of activities as well as the new ones (e.g. those based on registers and the TSS) in our NSI by making use of an integrated and harmonised set of terms to be used for data produced along the whole data production process described at a very high level of detail. In this context, methods, terms, processes and quality interact in the same system in order to sustain statistical processes and get information for enhancing quality.

References

[1] Ascari G., Boggia A., Loporcaro M.F., Scanu M., Simeoni G. (2022). *The new Istat quality policy for statistical production: challenges for a new data ecosystem*. *Proceedings of Q2022 European Conference on Quality in Official Statistics*, June 8 - 10, 2022, Vilnius, Lithuania.

[2] Bergdahl M., Ehling M., Elvers E., Földesi E., Körner T., Kron A., Lohauß P., Mag K., Morais V., Nimmergut A., Sæbø H.V., Timm U., Zilhão M.J. (2007). *Handbook on Data Quality Assessment Methods and Tools*. EUROSTAT, Luxemburg.

[3] Eurostat (2015) Single Integrated Metadata Structure V 2.0 (Sims V2.0) and Its Underlying Reporting Structures the Ess Quality And Reference Metadata Reporting Standards Esms 2.0 And Esqrs 2.0. Available at: https://ec.europa.eu/eurostat/documents/64157/4373903/SIMS-2-0-Revised-standards-November-2015-ESSC-final.pdf

[4] Eurostat (2017). *European Statistics Code of Practice*. Available at: https://ec.europa.eu/eurostat/documents/4031688/8971242/KS-02-18-142-EN-N.pdf/e7f85f07-91db-4312-8118-f729c75878c7?t=1528447068000

[5] Istat (2023). *Guida per l'implementazione del Codice Italiano per la qualità delle statistiche ufficiali*. Letture statistiche – Metodi. Available at: https://www.istat.it/it/files/2023/10/Guida-per-limplementazione-del-Codice-Qualita_Ebook.pdf (in Italian only).

[6] Istat, Sistan (2022). *Codice italiano per la qualità delle statistiche ufficiali.* Available at https://www.sistan.it/fileadmin/redazioni/IMMAGINI/pdf/Codice-Italiano-Qualita__-Statistiche-Ufficiali.pdf | https://www.sistan.it/fileadmin/redazioni/IMMAGINI/pdf/Codice-Italiano-Qualita__-Statistiche-Ufficiali.pdf

[7] Lima R. (2023). *Dai Big Data alle Trusted Smart Statistics Nota tecnica sulla Qualità.* Technical internal note, Istat, Roma (in Italian only).

[8] United Nations Economic Commission for Europe - UNECE (2019*). Generic Statistical Business Process Model (GSBPM). High Level Group – Modernization in Official Statistics*. Geneva, Switzerland. Available at: https://statswiki.unece.org/display/GSBPM/GSBPM+v5.1

[9] United Nations Economic Commission for Europe - UNECE (2020). *Generic Statistical Information Model (GSIM): Specification. High Level Group – Modernization in Official Statistics*. Geneva, Switzerland. Available at: https://statswiki.unece.org/display/gsim/GSIM+Specification