UNITED NATIONS
ECONOMIC COMMISSION FOR EUROPE
CONFERENCE OF EUROPEAN STATISTICIANS

EUROPEAN COMMISSION
STATISTICAL OFFICE OF THE
EUROPEAN UNION (EUROSTAT)

ORGANISATION FOR ECONOMIC COOPERATION
AND DEVELOPMENT (OECD)
STATISTICS DIRECTORATE

Work Session on Statistical Metadata (METIS)
(Geneva, Switzerland, 10-12 March 2010)

METADATA AND QUALITY INDICATORS REUSE FOR QUALITY REPORTING

Marina Signore and Giovanna Brancato
Istat - Division for Users' needs, Integration and Territory, Italy

I.   ISTAT DOCUMENTATION SYSTEM: METADATA AND QUALITY INDICATORS REUSE FOR NATIONAL REQUIREMENTS

1.   Istat has developed an information system, named SIDI, to document reference metadata and standard quality indicators (Brancato et al., 2004). In particular, methodological and quality metadata on survey processes are described to a great extent, whereas conceptual metadata are only partially included.

2.   The system is fully available on Istat intranet, and a reduced version, i.e. the metadata, is accessible for external users in Italian and English language on Istat website, under the name of SIQual.

3.   By developing the documentation system, a common language as well as standards for survey process documentation have been introduced at Istat, fostering the activity of documentation and the habit to measure quality. Nowadays the intranet version of SIQual documents 330 statistical processes, 77% of which are currently being carried out, and the remaining suspended or ceased.

4.   The system also represents the major source for quality evaluation, both within Istat self-assessment and auditing strategy and for specific ad hoc analyses on quality indicators. Standard quality indicators are indeed available for an increasing number of processes. The percentage of completeness of the quality indicators groups ranges from 100% for timeliness and punctuality to about 67% for Coverage and Unit Non Response groups of indicators.

5.   The metadata are annually updated and certified by about 100 quality pilots, purposely trained for this task. In addition, the metadata undergo a validation procedure from the central quality group. The standard quality indicators are annually monitored to ensure the continuation of the series and the gradual computation of new ones.

6.   It has to be mentioned that SIDI was initially planned to respond to survey manager quality control and monitoring activities, and its purposes have widened in the subsequent years, becoming an important source for survey process and survey quality documentation for the external users.

7.   Up-to-now, three main experiences have been carried out concerning the metadata and quality indicators reuse to produce standard documentation for external users:
    i) short methodological notes;
    ii) process quality reports;
    iii) product quality reports.

8.   The short methodological notes are compact reports included in the Statistical Yearbook. They are organised in few blocks of information concerning: general conceptual metadata (i.e. the observed issues, analysis units); methodological process metadata (i.e. reporting units, periodicity, European regulations, sampling design, data collection modes); quality metadata (i.e. the main activities carried out to prevent, monitor and evaluate unit non-response, item non-response and measurement error, and the validation techniques); quality indicators on timeliness of provisional and final estimates and geographical and sector level of aggregation dissemination. The metadata are slightly differently defined depending if the process is represented by a direct survey or a secondary study.

9.   All the metadata and quality indicators published in the short methodological notes are directly drawn from SIDI. Since, in general, SIDI contains more information than that required by the note and in a more detailed form, by a selection and a recoding process the metadata are properly transformed to fit the short note requirements. However, the coherence between the two is completely maintained.

10.   With respect to the timeliness quality indicators, the quality pilot in charge of the note compilation can decide if to extract from SIDI and include in the Yearbook either a punctual value of the indicator or an average over a given period of time. Indeed the Yearbook publishes tables with data that may be referred to one or more than a year, and therefore the most appropriate punctuality can be declared.

11.   The process quality report is a document describing accurately the survey production process. It can be produced starting from a pre-filled in text, that includes the information on the process drawn from SIDI, i.e. all the reference metadata regarding the survey activities. The standard items used to describe in a unique way the statistical process, are combined into simple sentences, following a logical structure. Afterwards, the document is further edited and adjusted to turn it in a fluent language.

12.   The great advantage of such an approach is to guide who is preparing the document to follow a standard structure, without forgetting any issues and leaving enough flexibility to customise the content. In addition, the meta-information available in SIDI is reused extensively.

13.   The product quality report is a short quality report, ranging from 2 to 4 pages, including a minimum set of standard quality indicators to be disseminated to the external users, jointly to press releases. The report is organised into sections following the main quality components, such as relevance, accuracy, timeliness and punctuality, transparency and accessibility and coherence.

14.   The quality indicators that are included in the product quality report, are a significant subset of Istat and Eurostat wider sets of quality indicators. In particular, only indicators on unit no response, sampling error, imputation, revision policy, coherence and timeliness are included. Not all the indicators have to be computed for a given report, but only those that have a significant meaning for that product. The product quality report is produced at least annually, although it can be updated more frequently for processes repeated within the year. The indicators are referred to the previous year and, in some cases, i.e. for monthly or quarterly surveys, are summarised into yearly averages.

15.   The level of reuse of SIDI metadata for the product quality report is quite low, since as already mentioned, SIDI metadata are mainly process metadata and up-to-now, no quality dimensions descriptions have been systematically documented in the system. However, following SIDI approach, the metadata included in the product quality reports will be stored and made available for dissemination and further update. Almost all the standard quality indicators included in the product quality report are drawn from SIDI, apart from those on sampling design, that have not yet been implemented in SIDI and those on the revision policy, represented by OECD revision triangles (Di Fonzo, 2005), differently defined in SIDI.

16.   Reusing SIDI metadata and quality indicators permits to reduce documentation burden on survey managers and ensures validity and coherence of the metadata and indicators disseminated in different places.

17.   The experience gained on the reuse of the metadata and quality indicators available in SIDI to respond to national documentation requirements, can be exploited to find efficient ways to fulfill Euro SDMX Metadata Structure (ESMS) implementation at Istat and it can be of interest for other National Statistical Institutes (NSIs) facing a similar task.

II.   METADATA AND QUALITY INDICATORS REUSE FOR ESMS REQUIREMENTS

18.   The ESMS defines a series of statistical metadata to be used for documenting statistical data. It is applicable at Eurostat and at National Statistical Institute levels. The concepts have been harmonised to match with different documentation templates apart from Eurostat, i.e. the OECD Metastore and the Data Quality Assessment Framework (DQAF) of the IMF standards. In 2004, a comparison between Eurostat and IMF general quality frameworks has been carried out (Laliberté et al., 2004) showing that the two approaches were quite complementary rather than overlapping. Such differences still seem not completely overtaken in the ESMS template, that strongly relies on Eurostat framework.

19.   The ESMS concepts and the type of information required have been analysed to evaluate the better approach for its implementation within Istat.

20.   First of all, the information included in the template is prevalently under the text form. In some cases, the information assumes a different meaning depending if it is referred to European or to national Member States level. In addition, not only metadata, but also quality indicators are included in the scheme.

21.   Some concepts refer to the organization rather than to the statistical process, e.g. the institutional mandate, the confidentiality legislation, the release policy, and the quality management system. To be efficient, the description of these items should be univocally produced independently to the process being documented, and should not be updated and collected several times from different domains.

22.   Other blocks of metadata concern: concepts that are referred to the data set (classifications, variables, units, population, etc.); features on the dissemination frequency and formats; product quality; process practices (e.g. data collection, validation, adjustment, etc.).

23.   While SIDI documents extensively the metadata on the process practices and it is moving towards the documentation of some features relative to product quality, it only partially covers the concepts relative to the data set. Being the system integrated with all the publications and data bases disseminated on the web, it also documents the issues related to the dissemination, although not yet the release calendar.

24.   As an experiment, a survey fully documented in SIDI has been considered and the ESMS scheme has been filled in using the metadata available from the system. It resulted that more than 60% of the information required could be drawn from the SIDI system.

25.   With respect to the standard quality indicators, up-to-now there are 21 indicators included in the ESMS documentation template. They have different nature: some of them are to be computed by Eurostat only, others are also required from the Member States. Some are quality process indicators while others are more oriented to measure costs and burden. SIDI manages some well accepted process quality indicators and few product quality indicators. Although priority has been set on a subset of these indicators, times and computation rules have not yet been fully defined.

III.   THE HARMONISATION OF QUALITY REPORTING FOR PRODUCERS AND USERS

26.   ESMS has a great importance with respect to the harmonisation of documentation and quality reporting at European level. Indeed, it permits standard documentation of statistics provided to Eurostat by Member States, allowing for comparisons and evaluations of statistics that are input for the construction of European aggregates. At the same time, it represents the first step towards standard quality reporting to users of European statistics. In principle, ESMS could provide metadata and quality indicators needed to fulfill both requirements. Nevertheless some critical issues should be taken into account. For the rest of this paper we will draw some considerations from the Istat experience.

27.   Eurostat and other international organisations reusing data collected by NSIs require detailed information on metadata and quality indicators associated with input data. ESMS is quite comprehensive regarding metadata and work is ongoing for including standard quality indicators into it. Once implemented, the transmission of data to Eurostat will be accompanied by a set of standard quality indicators that will permit data quality assessment. In addition, the metadata provided will enable to relate the quality indicators to the underlying production process for a better understanding of both product and process quality indicators. In this way Eurostat will have at disposal an information basis that can be reused for communicating quality to users. Indeed it would be recommended to tailor quality reporting to the needs of different users typologies and to the different dissemination media. In the previous paragraphs, Istat approach in tailoring quality reporting to users needs has been described. In a similar way, ESMS items can be used to extract quality information targeted to users of European statistics. In particular for statistics disseminating on the website, summary reports are most useful and effective. Thus, ESMS content should be analyzed with the purpose of identifying a subset of relevant metadata and quality indicators to be disseminated through Eurostat web site. The experience of Member States can support the identification of such a subset and possibly differentiating it for particular domains or processes (e.g. secondary studies). To this aim, also Istat experience on consultation of different users' profiles with respect to metadata and quality documentation supply, project supported by Eurostat that will yield its results in autumn 2010, could be of interest.

28.   The identification of a subset of relevant indicators is a first step, however some requisites should be fulfilled for their dissemination to users of European statistics. One of them concerns the harmonisation of quality indicators across Member States and statistical domains. Quality reporting with not fully harmonised indicators could indeed result in not consistent and potentially misleading information. Awareness of the importance of this issue is increasing both at European and National level and several initiates have been undertaken in order to reach such a goal. Eurostat has recently revised and updated the set of standard quality indicators (Eurostat, 2009) and it is envisaged that ESMS will include all or part of them. Nevertheless, it is vital that actual discrepancies in EU regulations with regard to quality reporting will be overcome.

29.   The need to set up new procedures or to revise those actually in place in order to produce quality indicators according to European standards may slow down the implementation of ESMS in Member States. The impact on Member States of European standard quality indicators will depend on the extent of changes and on the flexibility of their current quality assurance procedures. Thus, the identification of a subset of quality indicators and their gradual implementation over domains will help in reducing the additional workload for Member States and in warranting harmonisation.

30.   The identification of a minimum set of quality indicators might be a demanding and controversial task given that their dissemination should be aimed at helping users in properly use the data. Indeed, expert users that are going to perform statistical analyses will require detailed information on quality and on any limitations affecting the analyses. On the opposite, the general users and the public at large might not be able to understand and use information provided by quality indicators. A compromise can be reached considering the data dissemination via website. In this case, the set of quality indicators can be limited to a well selected minimum set of relevant and easy to understand indicators. Additional quality indicators could be released to expert users on request or included in ad hoc publications. As a consequence, ESMS should include the minimum set of quality indicators for external users plus additional indicators useful for Eurostat for assessing the quality of input data of EU statistics. It is not recommended, for the time being, to include all standard quality indicators defined by Eurostat (Eurostat, 2009) in ESMS even though they should serve as guidelines for quality assessment by Members States in their current activities.

31.   The frequency of collection of quality indicators is another item to consider. While reference metadata are quite stable in time, unless changes happen in the production process, quality indicators are related to each survey occasion. In particular for short term statistics it should be decided whether it would be better to collect average quality indicators or not (e.g. annual average response rate for monthly statistics or monthly response rates).

32.   Finally, ways of presentation of quality indicators to users should be examined. Major issues are how to summarize the information collected from Member States and to investigate which are the most suitable presentations for quality indicators. For example, it could be examined in which occasions it would be better to present numerical values of the indicators, or the range of variation in EU countries, or an average value over countries. In some cases, numerical values could be integrated or even substituted by graphical representations that, even though less precise, might be easier to understand.

REFERENCES

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

Brancato G, Pellegrini C, Simeoni G. e Signore M. (2004) "Standardising, Evaluating and Documenting quality: the implementation of Istat information system for survey documentation – SIDI". Proceedings of the European Conference on Quality and Methodology in Official Statistics (Q2004). Mainz, Germany.

Di Fonzo, T. (2005) "The OECD Project on revision analysis, first elements for discussion". Technical Report, OECD Short-term Economic Statistics Expert Group (STESEG). http://www.oecd.org/dataoecd/55/17/35010765.pdf

Euro SDMX Metadata Structure (release 2, February 2009).

Eurostat (2009). Eurostat Methodologies and Working papers "ESS Handbook for Quality Reports". http://epp.eurostat.ec.europa.eu/cache/ITY_OFFPUB/KS-RA-08-016/EN/KS-RA-08-016-EN.PDF

Eurostat (2009). Eurostat Methodologies and Working papers "ESS Standards for Quality Reports".http://epp.eurostat.ec.europa.eu/portal/page/portal/ver-1/quality/documents/ESQR_FINAL.pdf

International Monetary Fund (2003). "Data Quality Assessment Framework" http://dsbb.imf.org/Applications/web/dqrs/dqrsdqaf/

Laliberté L., Grünewald W., Probst L. (2004). „Data Quality: A comparison of IMF's Data Quality Assessment Framework (DQAF) and Eurostat's quality dimensions". http://dsbb.imf.org/vgn/images/pdfs/dqaf_eurostat.pdf