WP 14
10 February 2010

ENGLISH ONLY

UNITED NATIONS
ECONOMIC COMMISSION FOR EUROPE
CONFERENCE OF EUROPEAN STATISTICIANS

EUROPEAN COMMISSION
STATISTICAL OFFICE OF THE
EUROPEAN UNION (EUROSTAT)

ORGANISATION FOR ECONOMIC CO-OPERATION
AND DEVELOPMENT (OECD)
STATISTICS DIRECTORATE

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

QUALITY REPORTING WITHIN THE EUROSTAT AND ESS METADATA SYSTEMS

SIMPLIFICATION AND IMPROVEMENT OF ESS DATA QUALITY REPORTING

August Götzfried, and Håkan Linden, Eurostat

I. INTRODUCTION

1. The purpose of this paper is to describe how the current situation when it comes to data quality reporting in the European Statistical System (ESS) can be simplified and more streamlined through better integration of the quality reporting into the Eurostat and ESS metadata systems.

2. This would, hopefully, lead to improved quality reporting and minimised reporting burden. It should also lead to more harmonised and better documentation of data quality within the ESS.

3. The proposals are in direct response to the new Commission Communication (COM 404/2009) calling for a better integration of the production methods on European statistics, also with regard to metadata and quality reporting.

II. PROBLEM STATEMENT

4. It should be recognised that detailed and extensive quality reporting exists in many statistical domains. In particular in domains where specific regulations on quality evaluation[1] exist. However:

- Quality reports do not exist for all the statistical processes in Eurostat or within the ESS.
- The structure of the national and Eurostat quality reports show certain heterogeneity across statistical domains.
- Some quality reports are not public available (e.g. only available in restricted circa groups or elsewhere).

[1] For an overview of regulations in the field of statistics containing provisions on quality: http://epp.eurostat.ec.europa.eu/portal/page/portal/quality/quality_reporting/data_quality

- No harmonised IT production system exists for supporting the production (no central repository for storage, updating and reusability of quality information) and dissemination of quality reports.

III. PROGRESS MADE IN 2008/2009

5. A number of recent developments on quality in the ESS will pave the way for the proposed solutions:

- The new versions of the ESS quality reporting documents, the ESS Standard for Quality Reports (ESQR) and the ESS Handbook for Quality Reports (EHQR)[2] were released in January 2009. The ESQR and EHQR promote harmonised quality reporting across different types of statistical processes and across Member States. The ESQR and the EHQR provide detailed quality reporting requirements for producers of statistics in accordance with the European Statistics Code of Practice[3] and the quality criteria as defined in the Article 12 in the EC/Council Regulation No 223/2009 on European statistics (see annex).
- The Euro SDMX Metadata Structure (ESMS) was introduced in Eurostat in 2008 and within the ESS through the Commission Recommendation No 2009/498/EC on reference metadata for the ESS. The ESMS structure contains the statistical concepts related to data quality. These quality concepts are consistent with Article 12 in the EC/Council Regulation No 223/2009 and as well with the ESS standards for quality reports and the Commission Recommendation on ESMS.
- The ESS Quality and Performance Indicators[4] (QPI's) from June 2009. Harmonised methodologies have been developed for a set of 21 QPI's in order to better support the monitoring of data quality across statistical processes and over time.
- Two SDMX compliant IT applications used within the Eurostat Metadata Handler are or will be available: EMIS for Eurostat and the national reference metadata storage and the National Reference Metadata Editor for the national ESMS production, storage and transmission.

IV. WHAT IS THE EUROSTAT METADATA HANDLER?

6. The Eurostat Metadata Handler is a Eurostat portal regrouping several IT applications related to Metadata and making them accessible through one single user interface.

7. The main purposes of this Eurostat Metadata Handler are:

- To assure the production and web dissemination of structural and reference metadata within Eurostat and the European Statistical System;
- To accommodate and disseminate the ESS metadata (based on standards) within the European Statistical System and beyond;
- To stimulate the use of the different types of metadata for different purposes (e.g. for improving the statistical business processes, for monitoring and improving of the data quality, etc.);
- To provide information on data and metadata flows within the European Statistical System;
- To stimulate the improvement and better integration of the Eurostat and ESS metadata systems in providing transparency regarding the pertinent or commonly used and/or harmonised metadata.

---

[2] http://epp.eurostat.ec.europa.eu/portal/page/portal/quality/quality_reporting

[3] http://epp.eurostat.ec.europa.eu/portal/page/portal/quality/documents/code_practice.pdf

[4] ESS Quality and Performance Indicators – 2009. Working Group on Quality in Statistics, Doc. ESTAT/B-1/Quality/2009/05C.

8. Over time the Eurostat Metadata Handler should accommodate all types of metadata which are used from end-to-end of the statistical business process (i.e. from data/metadata collection to data/metadata dissemination). For more detailed information see METIS 2010 paper "The Eurostat Metadata Handler – Crucial for harmonising statistical business processes within the European Statistical System" (E. Clement and A. Götzfried).

V. THE NATIONAL REFERENCE METADATA EDITOR

9. The National Reference Metadata Editor (NRME) is an IT component of the Eurostat Metadata Handler directed to the production and the transmission of national reference metadata. This IT tool is mainly addressed to National Statistical Authorities within the ESS in offering them an on-line production IT tool for producing and transmitting their national reference metadata to Eurostat or other recipients (also in SDMX-ML).

10. National files based on the Euro SDMX Metadata Structure (ESMS) can be produced in using the NRME, but countries can also be able to produce and transmit their more detailed quality information by using the NRME.

11. In general, the national metadata will be sent from National Statistical Authorities to Eurostat via Edamis (the Eurostat Single Entry Point). The information is then inserted into EMIS where the files get accessible for the Eurostat domain managers. This IT component of the Eurostat Metadata Handler will be available in the course of 2010

VI. ESS QUALITY REPORTING – FURTHER HARMONISED

12. All producers of statistics needs full scale reports with qualitative and quantitative information dealing with all important aspects of output and process quality in detail. Thus, the quality reports require not only a description of processes and quality measurements but also quantitative quality assessments. Even though the producer oriented quality reports should contain all information needed for assessing the quality, it should be simplified as far as possible in order to minimise the reporting burden. Data quality information disseminated to the users should focus on the main quality aspects of importance for correct use of the statistics. This often requires more explanatory information on how to interpret the quality information in the specific statistical context.

13. **Following this, it is proposed to have only two stable and harmonised types of quality reports:**

(1) The **ESQR** for producers of statistics which should contain all aspects of relevance for producers to assess the data.
(2) The **ESMS** for users of statistics which should summaries the information provided according to the ESQR in order to cover only the main quality aspects having an impact on the use of the statistics.

14. The efficient production and exchange of metadata between the Member States and Eurostat requires also the use of standard format and concepts for automatic reporting and re-usability of information. Therefore, it is proposed to:

(3) Fully integrate the production and dissemination of the ESQR into the **Metadata Handler**.

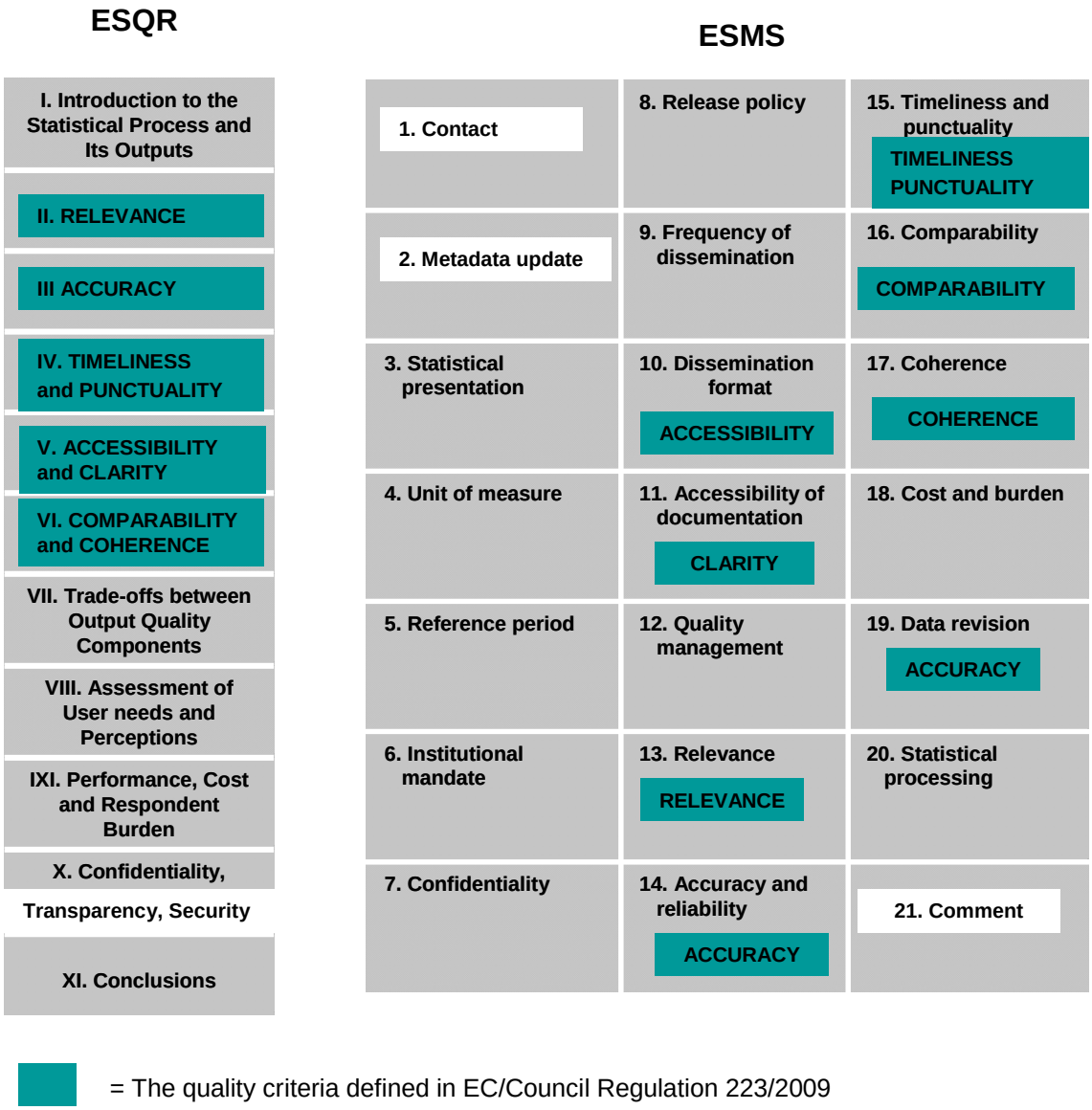
15. We propose to fully integrate the production and dissemination of the ESQR into the Metadata Handler. This recommendation is based on the fact that identical work processes can be used for the ESMS and the ESQR production, dissemination and exchange. This also means that the national ESQR files are transmitted via Edamis to Eurostat.

16. This integration will also permit the putting in place of the automatic transfer of certain parts of the quality information from the domain specific ESQR files to the domain specific ESMS files – this is further explained below.

VII. MORE INTEGRATED QUALITY REPORTING

17. In order to ensure consistency in concepts and descriptions with the SDMX standards and as well to maximise the re-usability of information in order to minimise the reporting burden a detailed mapping of the ESQR and ESMS has been done[5]. For an overview of the ESQR structure and the ESMS concepts see figure 1 below.

Figure 1. Overview of the ESQR chapters and the ESMS cross domain concepts.



[5] Ref. Eurostat working document "Mapping of ESS Standard for Quality Reports (ESQR) and Euro SDMX Metadata Structure (ESMS)". 14 December 2009.

18. The main results of the mapping are:

- There is consistency between ESMS and ESQR when it comes to the coverage of the different quality aspects at a general level (e.g. the main chapters and corresponding reporting requirements in ESQR are covered by the ESMS cross concepts and sub-concepts). Only with a few exceptions: the ESQR quality aspects "Transparency" and "Security" are not included in ESMS; and the ESMS concepts: “1 Contact”, “2 Metadata update” and “21 Comment” are not included in ESQR.
- The main quality criteria in 223/2009 are defined/ described in the same way in both ESMS and ESQR. However, it should be noted that the concept “Relevance” in ESQR contains a more content-oriented description of the statistical process and its outputs then the “Relevance” concept in ESMS which covers more the assessment of user needs and user satisfaction. These assessment aspects are in ESQR covered by chapter VIII “Assessment of User Needs and Perceptions”. The concept "Clarity" does not exist as such in ESMS (covered indirectly by other concepts).

VIII. THE ESS STANDARD FOR QUALITY REPORTS STRUCTURE (ESQRS)

19. Given that there is good consistency between the ESMS and the ESQR it is proposed to go for a pragmatic quality reporting approach where **The ESQR Structure to be used for the reporting of metadata on quality (ESS Standard for Quality Reports Structure = ESQRS) should only detail further the output quality criteria as defined in EC/Council Regulation 223/2009** – see figure 2 below. This is supported by the fact that many existing domain specific implementing Commission Regulations on quality are structured according to these output quality criteria.

Figure 2. Overview of the ESS Standard for Quality Reports Structure (ESQRS)

ESQRS	Concept name
I	Introduction
II	Relevance (user needs and perceptions)
II.1	User needs
II.2	User satisfaction
II.2.1	User satisfaction index (US1)
II.2.2	User satisfaction survey – date
II.3	Completeness
II.3.1	Rate of available statistics (R1)
III	Accuracy
III.1	Overall accuracy
III.2	Sampling error
III.2.1	Coefficient of variation (A1)
III.3	Non-sampling error
III.3.1	Coverage and other frame errors
III.3.1.1	Rate of overcoverage (A2)
III.3.2	Measurement errors
III.3.2.1	Edit failure rate (A3)
III.3.3	Non response errors
III.3.3.1	Unit response rate (A4)
III.3.3.3	Item response rate (A5)
III.3.4	Processing errors
III.3.4.1	Imputation rate (A6)
III.3.5	Model assumptions
III.3.6	Mistakes (A7)
III.3.7	Data revision
III.3.7.1	Data revision – policy
III.3.7.2	Data revision - practice
III.3.7.3	Average size of revisions (A8)
III.3.8	Seasonal adjustment

ESQRS	Concept name
IV	Timeliness and Punctuality
IV.1	Timeliness
IV.1.1	Timelag – first results (T1)
IV.1.2	Timelag – final results (T2)
IV.2	Punctuality
IV.2.1	Punctuality – publication (T3)
V	Accessibility and Clarity
V.1	News release
V.2	Publications
V.2.1	Publications – number (AC1)
V.3	On-line database
V.3.1	On-line database – accesses (AC2)
V.4	Micro-data access
V.5	Other
V.6	Documentation on methodology
V.6.1	Metadata –completeness (AC3)
V.7	Quality documentation
VI	Comparability
VI.1	Comparability – geographical
VI.1.1	Assymetries for statistics mirror flows (CC2)
VI.2	Comparability - over time
VI.2.1	Length of comparable time series (CC1)
VI.3	Comparability – domains
VII	Coherence
VII.1	Coherence – cross domain
VII.1.1	Coherence –subannual and annual statistics
VII.1.2	Coherence – National Accounts
VII.1.3	Coherence with other statistics
VII.2	Coherence – internal

= Concepts in common with ESMS

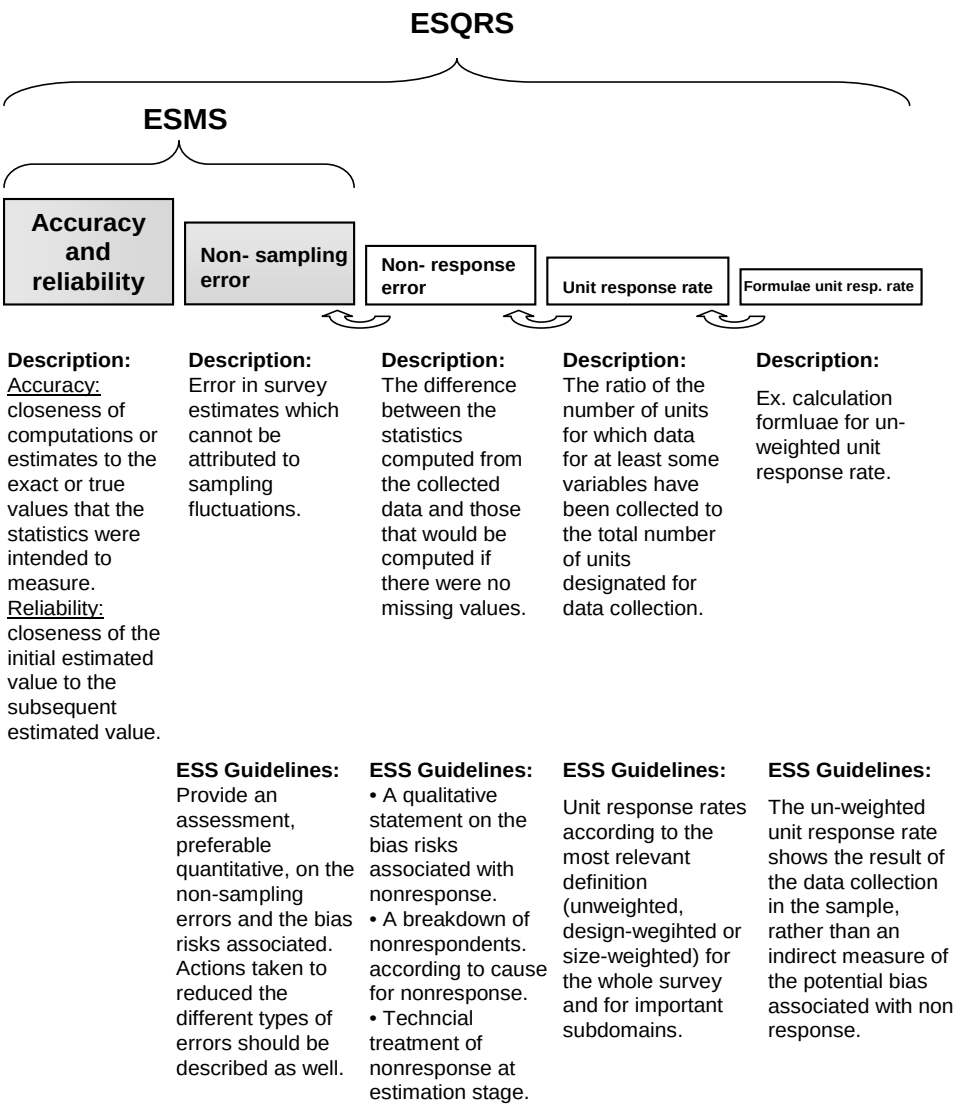
20. The remaining information from the ESQR is, de facto, covered by the ESMS (except for security and transparency) and will thus, in order to avoid duplication, be available trough the ESMS files.

21. The ESS quality reporting structure (ESQRS) is generic and should be used for any types of statistical process. The guidelines for quality reporting provided in EHQR are already used for the "ESS Guidelines" in ESMS and will be fully used also for detailing the quality reporting requirements in ESQRS for different types of statistical processes.

22. The following six processes will be singled out: Sample survey; Census; Statistical Process Using Administrative Source(s); Statistical Process Involving Multiple Data Sources; Price or other economic index process; and Statistical Compilation.

23. Figure 3 below illustrates how the information in ESQRS will be further detailed compared to ESMS. It also shows that the ESMS cross domain concepts and sub-concepts will correspond to the two first levels in the ESQRS.

Figure 3. Example of integrated quality reporting ESMS and ESQRS



IX. IMPROVED PRODUCTION AND DISSEMINATION OF REFERENCE METADATA ON QUALITY

24. Following the above illustrated process flow of reference metadata and metadata on quality, the actors for the production and dissemination of metadata on quality in the ESS are further defined in table 1 below.

Table 1. Production and dissemination of reference metadata and metadata on quality in the ESS

	Member States	**Eurostat**
ESQRS	- Production of national ESQRS - Transmission of national ESQRS to Eurostat - Dissemination of national ESQRS (if decided so)	- Production of the Eurostat ESQRS based on the national ESQRS - Dissemination of the Eurostat ESQRS (if decided so) - Checking and dissemination of the national ESQRS (if decided so)
ESMS	- Production of national ESMS- files using information in the the ESQRS - Transmission of of national ESMS files to Eurostat - Dissemination of national ESMS files (if decided so)	- Production of the Eurostat ESMS files using the Eurostat ESQRS - Dissemination of the Eurostat ESMS files - Checking and dissemination of the national ESMS files (if decided so)

25. As proposed in table 1, and similar to now, it is a priori not requested to disseminate national ESQRS and synthesised EU-ESQRS; it should be left to the decision of the Eurostat and national statistical domain managers if the ESQRS are disseminated or not. This, since domain specific ESQRS files might contain confidential information or detailed quality information that might be difficult for the general users to interpret correctly.

X. SUMMARY

26. To integrate the ESQR into the Eurostat Metadata Handler will allow for a more automatic production and dissemination of reference metadata on quality in the ESS. It should improve the situation for:

27. The **producers** of statistics by minimising the reporting burden since information can be pre-filled and reused for different reporting purposes (not only for domain specific reporting but also for cross domain monitoring of the data quality).

28. For the **users** of statistics since it would, hopefully, lead to improved availability and more harmonised information on data quality which would ease cross-national or cross-domain comparisons in future dissemination environments.
Once being available within the Eurostat metadata system the ESQRS information can also be promoted for the international exchange of quality related information if deemed necessary and appropriate.

29. For **Eurostat** as compiler of EU aggregates, it will create a large economy of scale, since:

- The common reporting structure, ESQRS, once integrated in the National Reference Metadata Editor can replace the existing domain specific templates/ questionnaires used by production units for collecting metadata on quality from countries.

- National quality information can to a large extent be used to automatically generate the EU-ESQRS. This is foreseen to be possible both for quantitative and descriptive information.

30. We estimate that already around 30 collections (and more are to come) of national reference metadata on quality exist in the various statistical domains which will be successively adapted to the ESQRS. This will not be very difficult due to the proximity of the statistical concepts used for quality.

31. For the statistical domains, which currently do have limited collections of metadata on quality, the ESQRS and the National Reference Metadata Editor, can contribute to reinforce the reporting in order to comply with EC/Council Regulation 223/2009.

ANNEX

REGULATION (EC) No 223/2009 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 11 March 2009 - extract Art. 12 on Statistical quality

Article 12

Statistical quality

1. To guarantee the quality of results, European statistics shall be developed, produced and disseminated on the basis of uniform standards and of harmonised methods. In this respect, the following quality criteria shall apply:

(a) 'relevance', which refers to the degree to which statistics meet current and potential needs of the users;

(b) 'accuracy', which refers to the closeness of estimates to the unknown true values;

(c) 'timeliness', which refers to the period between the availability of the information and the event or phenomenon it describes;

(d) 'punctuality', which refers to the delay between the date of the release of the data and the target date (the date by which the data should have been delivered);

(e) 'accessibility' and 'clarity', which refer to the conditions and modalities by which users can obtain, use and interpret data;

(f) 'comparability', which refers to the measurement of the impact of differences in applied statistical concepts, measurement tools and procedures where statistics are compared between geographical areas, sectoral domains or over time;

(g) 'coherence', which refers to the adequacy of the data to be reliably combined in different ways and for various uses.

2. In applying the quality criteria laid down in paragraph 1 of this Article to the data covered by sectoral legislation in specific statistical domains, the modalities, structure and periodicity of quality reports provided for in sectoral legislation shall be defined by the Commission in accordance with the regulatory procedure referred to in Article 27(2).

Specific quality requirements, such as target values and minimum standards for the statistical production, may be laid down in sectoral legislation. Where sectoral legislation does not so provide, measures may be adopted by the Commission. Those measures, designed to amend non-essential elements of this Regulation by supplementing it, shall be adopted in accordance with the regulatory procedure with scrutiny referred to in Article 27(3).

3. Member States shall provide the Commission (Eurostat) with reports on the quality of the data transmitted. The Commission (Eurostat) shall assess the quality of data transmitted and shall prepare and publish reports on the quality of European statistics.