



Statistical principles and quality

*TOSSD Task Force Issues Paper¹ - Agenda item 4.a
24 – 25 January 2018*

I. INTRODUCTION

1. Quality is central to statistics. In order to establish TOSSD as an international statistical standard, it will need to be associated with robust quality requirements.

2. Several tools exist in the Global Statistical System² in order to promote and ensure quality in statistics:

- Charters, codes of practice and other regulatory instruments which define the principles of statistical quality;
- Quality requirements in national statistical legislation;
- Quality frameworks developed at the national or international level, which specify a set of quality dimensions along with implementation guidelines and good practices;
- Regular monitoring and reporting on implementation of these quality frameworks; and
- Assessment questionnaires (self-assessment and / or peer reviews).

II. INTERNATIONAL PRINCIPLES FOR STATISTICAL ACTIVITIES

3. The reference set of principles governing the Global Statistical System is the United Nations Fundamental Principles of Official Statistics (FPOS)³ which were adopted by the United Nations Statistical Commission (UNSC) in 1994 and by the United Nations General Assembly in 2014. The 10 principles enshrined in the FPOS cover a range of topics including relevance, impartiality and equal access; professional standards, scientific principles and professional ethics; accountability and transparency; prevention of misuse; sources of official statistics; confidentiality; legislation; national coordination; use of international standards; and international co-operation. Other reference sets of principles that have been developed are generally inspired by the FPOS. Examples include:

- European Statistics Code of Practice, revised edition (2011), Eurostat;
- Code of Good Practice in Statistics for Latin American Countries and the Caribbean⁴ (2011), adopted by the Statistical Conference of the Americas (SCA – ECLAC) in 2011;
- ASEAN Community Statistical System (ACSS) Code of Practice (2012), ASEAN; and
- Recommendation of the OECD Council on Good Statistical Practice (2015) (first OECD legal instrument on statistics).

¹ Jointly drafted by Aussama Bejraoui (Aussama.Bejraoui@oecd.org) and Julia Benn (Julia.Benn@oecd.org).

² The Global Statistical System includes international and regional organisations that are providers of statistics, the UN Statistical Commission, the Committee for Co-ordination of Statistical Activities, international statistical standards, principles and other instruments governing international statistical activities and National Statistical Systems.

³ The United Nations Fundamental Principles of Official Statistics derive directly from the Fundamental Principles of Official Statistics which were developed by the Conference of European Statisticians in 1991 and subsequently adopted by UNECE in 1992.

⁴ The Code of Good Practice in Statistics for Latin America and the Caribbean is largely inspired by the European Statistics Code of Practice.

4. These regulatory instruments generally aim at guiding institutions in the national statistical system⁵. Significant differences exist between national and international organisations producing statistics. For example, the requirement to report to a National Statistical Organisation (NSO) is generally based on a statistical law, whereas at the international level the requirement for a country to report hinges on voluntary agreements between countries. To address the particular case of international organisations, the Committee for the Coordination of Statistical Activities (CCSA)⁶ adopted in 2005 the Principles Governing International Statistical Activities. The Principles are, to a large extent, similar to the Fundamental Principles of Official Statistics⁷. They have been endorsed by many international agencies⁸ such as the OECD, Eurostat or the World Bank.

5. In order to leave room for flexibility, the Committee noted that “the mandate and governance of individual international organisations may be such that not all these principles and practices may be applicable under current circumstances, even though they can be used for future improvements”. Several other international organisations have developed their own version of statistical principles which are generally influenced by the FPOS and the Principles Governing International Statistical Activities. The OECD issued in 2011 the Quality Framework for OECD Statistical Activities, which takes into account the Organisation’s own context.

6. The Secretariat proposes to include a statement on quality commitment in the TOSSD Reporting Instructions. In this regard, the Task Force is invited to discuss what international principles should be cited as a reference. Given that TOSSD aims at being an international statistical standard, it might be most appropriately guided by principles issued by United Nations bodies, whether the Fundamental Principles of Official Statistics or the Principles Governing International Statistical Activities.

III. QUALITY DIMENSIONS AND IMPLEMENTATION GUIDELINES

7. Charters or codes of practice are generally supported by or replaced by⁹ quality assurance frameworks, also called Data Quality Assurance Frameworks (DQAF) or Statistics Quality Assurance Frameworks (SQAF). These frameworks generally identify a set of quality dimensions through which the underlying principles are applied, and provide implementation guidelines and good practices. Although they generally share the main ideas, there is no single internationally agreed version of quality dimensions and each organisation has adapted them to its own context.

8. A set of quality dimensions and implementation guidelines are proposed for the TOSSD system in Table 1 and Annex 1 respectively. The following documents have been used as a reference:

- Quality Framework and Guidelines for OECD Statistical Activities (2011), which was developed by the OECD to help manage quality within its own organisation and to demonstrate its quality commitment to users;

⁵ The European Statistics Code of Practice applies also to the Community statistical institutions, such as Eurostat.

⁶ The members of the Committee for the Coordination of Statistical Activities comprise international and supra-national organisations whose mandate includes the provision of international official statistics. The Secretariat of the CCSA is situated at the United Nations Statistics Division (UNSD).

⁷ Some elements which are not relevant to international statistics are not included (Legislation, National coordination, etc.)

⁸ Endorsements can be seen: https://unstats.un.org/unsd/methods/statorg/Principles_stat_activities/endorse.asp

⁹ In some cases, such as for the OECD, the statistical principles are included in the quality framework.

- European Statistical System Quality Assurance Framework (ESS QAF) (2011), Eurostat. The ESS QAF provides guidance for the implementation of the European Statistics Code of Practice;
- Implementation guidelines of the United Nations Fundamental Principles of Official Statistics (2015), United Nations Statistics Division;
- United Nations Statistics Quality Assurance Framework (UN-SQAF) (2016), Committee of the Chief Statisticians of the United Nations System. The UNSQAF “is designed to go a step beyond these principles for international organisations by introducing a common understanding of the quality dimensions and quality assurance for all relevant UN agencies”.

9. While there are strong similarities among these frameworks, there are also differences, which are partly explained by their purpose and scope¹⁰. Narrower approaches to statistical quality dimensions focus on the quality of the statistical output. More comprehensive versions tend to include process and institutional aspects. In line with the UN-SQAF and the regional Codes of Practice, the proposed quality dimensions for the TOSSD framework have been grouped in the three categories: output quality, process quality and the quality of the institutional environment.¹¹

a) Statistical output quality

10. It is generally agreed that the quality of a statistical product is determined by its “fitness for use”. The output quality dimensions are generally common to all frameworks. They include relevance, coherence, accuracy, reliability, timeliness, punctuality, interpretability and accessibility. Annex 2 provides definitions for the different statistical dimensions used. These definitions are derived from the UN Statistics Quality Assurance framework.

11. Ensuring the relevance of statistics requires the identification of user groups and their needs. In the context of the TOSSD cross-border pillar, users will include recipient governments¹², provider countries, international organisations, researchers, civil society and the public in general. Regular consultation processes will need to be established in order to reflect changes in user needs. In order to be coherent, the statistics need to be consistent internally and comparable with other datasets over time and across countries. This notably implies that the data are based on common concepts, definitions, classifications and methodology. The role of metadata in explaining changes in concepts or methodologies over time or across countries is essential.

12. The accuracy of a statistical product is, to a large extent, determined by the definition of the statistical concepts and methodologies and by the accuracy of the data received from reporting countries or organisations. However, data accuracy can be positively influenced at the data processing stage through the development of quality checks. One other aspect of accuracy, also referred to as reliability, is the closeness of the initially released values of a statistical product to the

¹⁰ For example the European Statistics Code of Practice is designed to apply both to the national and Community statistical authorities.

¹¹ Some quality dimensions, such as sound systems or mandate for data collection, are addressed at the organisational level, and were therefore not included in the proposed quality dimensions although they constitute important aspects which influence the quality of a statistical system.

¹² Two pilot studies in Senegal and the Philippines have been carried out to assess the needs of recipient countries. Additional pilot studies are expected to be undertaken in 2018 and 2019.

values that are subsequently released for the same reference period. In order to minimise revisions, these will need to be regularly analysed in order to adjust and improve the statistical processes.

13. Data need to be released in a sufficiently timely manner to be acted upon. Timeliness typically creates a trade-off in terms of accuracy. A release calendar will need to be established in order to improve the capacity of external users to make use of the statistics and the capacity of internal users to plan the data collection process. Punctuality is another important quality aspect related to timeliness.

14. The interpretability or clarity of the data relies, to a large extent, on the adequacy of the metadata, including a definition of the concepts and methods used as well as a description of the strengths and limitations of the statistical output. In order to be usable, the statistical output also needs to be easily accessible on an impartial basis. Interpretability and accessibility need to be supported by user support services

b) Statistical process quality

15. Other important aspects of statistical quality relate to the statistical processes. The quality of the statistical process has an impact on the quality of the statistical output. We propose two process quality dimensions: i) sound methods and ii) cost-efficiency and non-excessive burden on reporters.

16. Sound methods that rely on international standards and best practices help ensure process and product quality. Although they are not considered as quality dimensions by all organisations¹³, cost-efficiency and excessive burden on reporters can significantly impact statistical quality. For example, the resources released through cost-efficiency can be used to increase the quality of the product. Creating an excessive burden on data providers may result in reduced quality in the data reported.

c) Quality of the institutional environment

17. The perceived quality of statistics is strongly related to the trust and confidence the public places in these statistics, which in turn depends on the credibility of their producer. The quality of the institutional environment has a significant influence on the credibility of a statistical authority. One important aspect of institutional quality is related to statistical confidentiality, although this may apply only to a limited type of resource flows included in TOSSD.¹⁴

18. Institutional quality requires also that statistics are produced in an objective, transparent and independent manner. The real or perceived lack of professional independence of the statistical producer, notably from political pressure, can undermine the credibility of statistics produced. For example, one of the main criticisms that have been addressed to the ODA statistical system is the perceived political influence of members of the Development Assistance Committee.

19. The institutional quality is intrinsically linked with the governance of a statistical system. This aspect should therefore also be discussed in the context of the TOSSD governance framework.

¹³ For example the OECD does not regard cost-efficiency as an element of quality dimensions but nevertheless includes it in the analysis of quality as it can affect all other quality dimensions.

¹⁴ For example company names in the case of operations with the private sector.

Table 1: Proposed quality dimensions

UN Statistics Quality Assurance Framework	OECD Quality Framework	European Code of Practice	Proposed quality dimensions for TOSSD
Statistical output quality			
Relevance Coherence Accuracy Reliability Timeliness Punctuality Accessibility Interpretability	Relevance Accuracy Timeliness Accessibility Interpretability Coherence	Relevance Accuracy and reliability Timeliness and punctuality Coherence and Comparability Accessibility and clarity	Relevance Coherence Accuracy Reliability Timeliness and Punctuality Interpretability Accessibility
Statistical process quality			
Sound methods and systems Cost-efficiency	Cost-efficiency	Sound methodology Appropriate statistical procedures Non-excessive burden on respondents Cost-effectiveness	Sound methods Cost efficiency and non-excessive burden on reporters
Quality of the institutional environment			
Objectivity, impartiality, transparency, and credibility	Credibility ¹⁵	Professional independence Mandate for data collection Adequacy of resources Commitment to quality Statistical confidentiality Impartiality and objectivity	Objectivity, impartiality, transparency, and credibility

Issues for discussion

- *Does the Task Force agree to include a reference to statistical principles in the Reporting Instructions? What international statistical principles should TOSSD refer to?*
- *What quality dimensions should be taken into account in the TOSSD statistical framework? Should process and institutional quality aspects be included?*
- *Do members have any comments regarding the proposed implementation guidelines (Annex 1)? Should any of them be excluded? Or should any additional implementation tools be included?*

¹⁵ In the Quality Framework for OECD Statistical Activities credibility is considered as an aspect of the statistical output quality. However, given that this is mainly related to the credibility of the statistical authority we have decided to include it in the institutional quality.

ANNEX 1: IMPLEMENTATION GUIDELINES

Quality dimension	Elements	Implementation guidelines
Relevance	1. Users are identified and their needs are investigated	1.1 Regular consultation mechanisms with users are in place 1.2 Users are categorised according to their type of use and data needs 1.3 Key users are consulted on changes to statistical products 1.4 Scientific community is consulted regarding statistical methodologies and research
	2. User satisfaction is monitored and followed-up on a regular basis	2.1 Quality indicators on relevance are regularly assessed 2.2 User Satisfaction surveys are conducted on a regular basis 2.3 The results of these assessments are published
Coherence	3. Statistics are internally coherent	3.1 Procedures and guidelines to ensure internal coherence exist (accounting identities, consistency between preliminary and final data, etc.) 3.2 Regular assessments of verification and imputation procedures are carried.
	4. Statistics are comparable over of time and across countries	4.1 Data is checked with data for previous years to check for comparability 4.2 Discrepancies in statistics are identified and corrected 4.3 Harmonisation adjustments (e.g. definitions, coverage and classifications) are made in accordance to internationally accepted standards, and are clearly documented in the metadata associated with data products 4.4 Breaks in time series are communicated and clearly explained
	5. Changes and variations in methodology that affect data values are well explained	5.1 Changes in methods are clearly identified and their impact measured to facilitate reconciliation
	6. Use of common standards	6.1 Existing international statistical guidelines and recommendations are used for concepts, definitions, units, classifications and nomenclatures. Divergences from these international standards are documented and justified
	7. Statistics from different sources and of different periodicity are compared and reconciled	7.1 Appropriate procedures are in place to systematically compare and reconcile statistics from different sources
Accuracy	8. Source data, intermediate results and statistical outputs are regularly assessed and validated	8.1 Internal quality assessment procedures address accuracy issues 8.2 Comparison of results with other sources of information in order to ensure validity
	9. Accuracy and Errors are measured and systematically documented	9.1 Methods and tools for preventing and reducing errors are in implemented 9.2 Regular quality reporting on accuracy is implemented and published 9.3 Variations in accuracy and coherence over time and across countries are communicated.
Reliability	10. Revisions are regularly analysed in order to improve statistical processes	10.1 A revision policy is produced and made public 10.2 Explanations on revisions are provided 10.3 Regular analysis of revisions is used to improve the statistical process
Timeliness	11. Timeliness follows international dissemination standards	11.1 Compliance with international dissemination standards 11.2 Publication of an advance release calendar 11.3 Quality indicators of timeliness are analysed and published

Punctuality	12. Preliminary results of acceptable aggregate accuracy can be released when considered useful	12.1 Review of the possibility of disseminating preliminary results and reporting on their quality, taking into account data accuracy 12.2 Policy for scheduled revisions is in place
	13. Divergence from the dissemination time schedule is explained and publicised in advance	13.1 A procedure to monitor and assess punctuality is in place 13.2 Divergences from the pre-announced time are clearly explained and communicated in advance and a new release date is announced 13.3 Quality indicators of punctuality are analysed and published
Interpretability and clarity	14. Statistical products are presented in a format that facilitates interpretation and comparability	14.1 The presentation of statistical products is clear and unambiguous and uses the most appropriate formats 14.2 Users are consulted about dissemination 14.3 Various forms of dissemination are considered to allow for better understanding and comparisons of particular results by various audiences
	15. In order to maximise their interpretability data products are accompanied by appropriate metadata	15.1 Data products are accompanied by appropriate metadata that facilitates understanding 15.2 Metadata are presented in layers of increasing detail, matching user needs 15.3 Metadata is regularly updated and procedures to ensure the updating are available
Accessibility	16. Metadata are compiled using internationally accepted standards and practices	16.1 Metadata are compiled following definitions contained in the SDMX Metadata Vocabulary 16.2 All relevant definitions used are stored in the OECD Glossary of Statistical Terms
	17. Users are kept informed about the methodology of statistical processes	17.1 Publication of user-oriented quality reports and methodological documents
	18. Data quality documentation is provided to users to ensure proper use of the data	18.1 Users are provided with information necessary to understand both the strengths and limitations of the data 18.2 Documentation on methodology allows users to assess whether the data adequately approximate what they wish to measure and whether data are produced with tolerances acceptable for their intended use 18.3 Users are provided with information necessary to avoid inconsistency between OECD statistical products and other national and international statistical products
	19. A help service supporting the data and metadata is available to users	19.1 A help service supporting the data and metadata is well identified and available during working hours
	20. Dissemination methods facilitate accessibility use modern information and communication technology and facilitate accessibility.	20.1 The website and statistical databases are the main means for disseminating statistical results and facilitate self-tabulation in the most appropriate formats. 20.2 An information service is available for answering requests and clarifications of statistical results. 20.3 all products include a contact name, e-mail address, in order to 20.4 A publication catalogue is available to users 20.5 Statistical results are disseminated using tools and formats that facilitate re-dissemination 20.6 promote products externally via wikis and blogs and social media to ensure that they reach the widest possible audience

Sound methods	21. Metadata is easily accessible	21.1 Definitional, procedural and operational metadata describing the statistical activity are readily available to all users 21.2 Final metadata are stored, as much as possible, in the common repository of metadata
	22. Access to microdata is allowed for research purposes and is subject to specific rules	22.1 Publication of the rules or protocols to access microdata.
	23. The methodological framework used follows international standards and good practices	23.1 The methodological framework and the procedures for implementing statistical processes are integrated into a standard methodological document 23.2 Divergence from international methodological recommendations are explained and justified
	24. quality self-assessment	24.1 An annual quality self-assessment of the production process and products is conducted
	25. Co-operation with the scientific community in order to improve the methodology and the effectiveness of the methods implemented	25.1 Regular contacts, e.g. through conferences, workshops, task forces, are established with the scientific community to discuss methodological and innovation developments. 25.2 External evaluation of the methods used is made from external experts where appropriate
	26. Revisions follow standard, well-established and transparent procedures	26.1 Guidelines and principles regarding the revision of published statistics are established and made known to users 26.2 Revisions are explained to users
Cost-efficiency and non-excessive burden on reporters	23. The productivity potential of information and communications technology is being optimized for data collection, processing and dissemination.	23.1 Automated processing techniques are used and regularly reviewed
	24. Maximum use is made of already existing data and databases managed by the OECD or by other international organisations.	24.1 Other relevant data currently available is used 24.2 The data managed is integrated with other OECD databases if relevant
	25. Minimise the reporting burden on providers	24.1 Needs of statistical information are regularly analysed 24.2 Each collected variable is justified 24.3 Response burden is measured and monitored 24.4 Data and metadata transmission procedures are designed to be as automated and efficient as possible.
Objectivity, impartiality, transparency and credibility	25. Statistics are produced and disseminated on an objective basis, determined only by statistical considerations	25.1 Statistical methods and outputs are chosen on the basis of statistical considerations and not by pressure from providers, users or other stakeholders. 25.2 Data release is not timed in response to political pressure 25.3 The Secretariat decides if the publication of poor quality data received from countries affects the overall credibility of the statistics and is allowed to refuse to publish poor quality data 25.4 statistical releases are clearly distinguished from political/policy statements and issued separately from them 25.5 Statistical releases and statements made in press conferences are objective and non-partisan
	26. Information on the sources, methods and procedures used is publicly available.	26.1 Methodological notes and metadata on sources, methods and procedures used are available in databases and are published on the website 26.2 Users are informed about any changes in sources and methods that might impact the outputs

27. Decisions about dissemination of statistics are only informed by statistical considerations	27.1	Non-disclosure of data is only permitted for reasons of statistical confidentiality.
28. Statistical release dates and times are pre-announced	28.1	A release calendar is publicly available
	28.2	Changes to the dissemination schedule, when deemed absolutely necessary, are publicly and promptly announced in advance and duly accounted for (stability of the release calendar)
29. Advance notice and explanation are given on major revisions or changes in methodologies	29.1	A calendar of the regular major revisions is published
	29.2	Information on major revisions or changes in statistical methodologies is communicated in advance
30. All users have equal access to statistical releases at the same time.	30.1	All users have equal access to statistical releases at the same time. Any privileged pre-release access to any outside user is limited, monitored and publicised
31. Errors discovered in data outputs are corrected as soon as possible after they are detected, and users are informed	32.1	Errors in data outputs are corrected as soon as possible
	32.2	Users are informed about errors detected and corrected
	32.1	External entities such as statistical organisations (e.g UNSD), scientific bodies or external experts monitor the scientific practices and statistical quality issues
32. External entities reinforce and monitor the scientific practices, the statistical quality aspects and the ethical conduct.	32.2	External entities such as statistical or organisations or watchdog organisations monitor the ethical conduct

ANNEX 2: DEFINITIONS OF QUALITY DIMENSIONS

The following definitions are taken from the UN Statistics Quality Assurance Framework¹⁶:

Relevance

“The relevance of a statistical output is the degree to which the data serve to address the purposes for which they are sought by users.”

Accuracy

“The accuracy of a statistical output is the degree to which the data correctly estimate or describe the quantities or characteristics they are designed to measure.”

Reliability

“Reliability is the closeness of the initially released values of a statistical output to the values that are subsequently released for the same reference period.”

Coherence

“The coherence of a statistical output reflects the degree to which it is logically connected and mutually consistent with other statistical outputs”.

Coherence includes four sub-dimensions: coherence within a dataset, coherence across datasets, coherence over time and coherence across countries.

Timeliness

“The timeliness of a statistical output is the length of time between its availability and the event or phenomenon it describes.”

Punctuality

“An output is punctual if it is disseminated in accordance with the schedule.”

Accessibility

“The accessibility of a statistical output reflects how readily the data can be discovered, located and accessed from within the Agency data holdings.”

Interpretability

“The interpretability (sometimes called clarity) of a statistical output reflects the ease with which users can understand and properly use the data.”

Sound methods

“Sound methods refers to the use of international standards and best practices through all stages of a statistical process from identification of requirements, through design, data collection, processing, analysis, dissemination, and evaluation.”

Cost-efficiency

“The cost-efficiency with which statistical outputs are produced as a measure of the costs incurred and resources expended relative to the benefits of the products.”

¹⁶ <https://unstats.un.org/unsd/statcom/48th-session/documents/BG-4j-QAF-E.pdf>