

Statistical Data Quality in the UNECE

2010 Version

Steven Vale,

Quality Manager, Statistical Division

Contents:

- 1. The UNECE Data Quality Model – page 2**
- 2. Quality Framework – page 3**
- 3. Quality Improvement Strategy – page 7**
- 4. Quality Improvement Programme 2010 – page 8**

1. The UNECE Data Quality Model

Elements of the model

Quality Framework

Provides the basis for ↓↓

Quality Improvement Strategy

Manifests itself as ↓↓

Annual Quality Improvement Programmes

Composed of ↓↓

Quality Improvement Tasks

Governance

Owned and agreed by senior managers on behalf of all staff

Developed by the Quality Manager based on inputs from all staff

Owned and agreed by Database Coordination Group based on cost-benefit assessments of proposed tasks, and available resources

Owned by individuals or groups of staff, reporting to the Database Coordination Group via the Quality Manager

2. Quality Framework

Introduction

The main purpose of national and international statistical organisations is to provide information to inform the public and policy makers on the functioning of society, the economy and government. To fulfil this role it is important that the statistical outputs of these organisations meet certain quality standards, and that sufficient information on quality is provided to allow users to judge for themselves how reliable these data are.

Statistical organizations have responded to this requirement by producing quality frameworks to define the concept of quality, and to provide criteria against which the quality of outputs can be judged. These frameworks do not necessarily provide precise and objective measures of quality, but do encourage a standard and systematic approach to quality within the agency.

To avoid duplication of work, the UNECE data quality framework draws heavily on frameworks and experiences from other international statistical organisations, particularly Eurostat, the OECD and the IMF, adapting them to the UNECE context, whilst also taking into account work by the Committee for the Coordination of Statistical Activities (CCSA) to harmonize these frameworks. It defines quality and the various components of quality relevant to statistical data outputs, to provide a basis for practical quality improvement measures.

What is Quality?

ISO standard 9000:2005 defines quality as the "degree to which a set of inherent characteristics fulfils requirements". Therefore, under this definition, the quality of statistical data can be determined by the extent to which they meet user needs. This approach to quality is also reflected in the "Fundamental Principles of Official Statistics"¹, where the first principle refers to the need for official statistics to "meet the test of practical utility".

It is important to be able to identify the users of our data, to assess their relative importance, and determine their requirements. Each user has different needs, and therefore has a different view of quality. The requirements of key users of our data (such as other parts of the UN, national governments and international organisations) should obviously take priority. Although other users, particularly in the academic and private sectors may be seen individually as of lower importance, they are a large and growing group, so their needs are also becoming increasingly important.

This approach to quality, based on user needs, may conflict to some extent with the traditional view that the quality of statistical outputs is determined by how closely they reflect reality. For example, in certain cases, users may prefer timelier but less precise estimates, rather than data that have a high degree of accuracy, but take

¹ See <http://www.unece.org/stats/archive/docs.fp.e.htm>

longer to produce. Thus data quality should be seen as a balance of a number of different components.

Components of Quality

The seven components to be considered when assessing the quality of UNECE data are:

- **Relevance** – The degree to which statistics meet the needs of current and potential users. Relevance therefore refers to whether the statistics that are needed are produced, and whether the statistics that are produced are needed. It also covers methodological soundness, and particularly the extent to which the concepts used (definitions, classifications etc.) reflect user needs;
- **Accuracy** – The closeness of statistical estimates to true values, with the proviso that absolute accuracy can be difficult to determine when data are taken from other sources rather than directly collected. As this is often the case for the UNECE, two factors should be considered: the credibility of the source, and the plausibility of the data, i.e. the extent to which they look reasonable when compared to other periods, to similar countries, and to the values the statistician would expect;
- **Timeliness** – The length of time between data being made available and the event or phenomenon they describe
- **Punctuality** – Punctuality refers to the time lag between the release date of data and the target date when they should have been released.
- **Accessibility** – The physical conditions in which users can obtain data: where to go, how to order, delivery time, clear pricing policy, convenient marketing conditions (copyright, etc.), availability of micro or macro data, various formats (paper, files, CD-ROM, Internet...), etc.;
- **Clarity** – Clarity refers to whether data are accompanied by sufficient and appropriate metadata, whether illustrations such as graphs and maps add value to the presentation of the data, and whether information on data quality is available;
- **Comparability** – The extent to which differences between statistics are attributed to differences between the true values of the statistical characteristic, or to methodological differences. Comparability includes:
 - **Comparability over time** – The extent to which data from different points in time can be compared.
 - **Comparability through space** – The extent to which data from different countries and/or regions can be compared. The provision and application of international standards is particularly important here.
 - **Comparability between domains** – The extent to which data from different statistical domains can be compared.

The notion of “cost-efficiency” should also be mentioned. Whilst this is not considered to be a dimension of quality, it is a factor that must be taken into account in any analysis of quality. If a product can be produced more efficiently with the same

quality, then resources released can be used to improve the quality of that product or other products.

Similarly, other considerations at the level of the organisation, rather than at the level of the individual data sets, can include the concepts of professionalism, integrity, credibility, and legal and organisational environments.

Mapping of Quality Components Used by International Statistical Organisations

UNECE	OECD	Eurostat	IMF
Relevance	Relevance	Relevance	Pre-requisites of quality (part)
			Methodological soundness
Accuracy	Accuracy	Accuracy	Accuracy and reliability
Timeliness	Timeliness	Timeliness and Punctuality	Serviceability (part)
Punctuality			
Accessibility	Accessibility	Accessibility and Clarity	Accessibility
Clarity	Interpretability		Assurances of integrity (part)
Comparability	Coherence	Comparability	Serviceability (part)
		Coherence	
(Considered more relevant at the level of the organisation)	Credibility		
			Pre-requisites of quality (part)
			Assurances of integrity (part)

Other Aspects of Quality

This framework focuses on the quality of statistical data. Most data work can be characterised by one or more inputs, some kind of transformation process, and the resulting outputs (which may also be inputs to some other statistical process).

It should also be remembered that statistical data are not the only output of a statistical organisation. Other important outputs include statistical standards or norms (such as concepts, classifications, manuals and methodologies)², technical assistance (including the provision of training or software applications) and, more generally, any interaction with external people or organisations (including the organisation of meetings and contributions to the coordination of international statistical activities). For an international organisation such as the UNECE, these actions all contribute to improving the quality of data from national statistical organisations, which are used as a source for UNECE outputs.

A total quality management³ approach would require us to measure the quality of all of these outputs (along with associated inputs and processes), but if we did this, it would distract too many resources from regular tasks. This, in turn, would have an adverse effect on quality. At this point it is necessary to return to the notion of cost-efficiency; we should focus quality assessment and improvement work on areas where it will be most beneficial. Therefore, although we should think about quality (in terms of meeting user needs) when producing all types of outputs, the main focus of the quality assessment work should be on statistical data.

Communicating Quality

It is often the case that users say they want data to be of “good quality”, but they do not have the necessary evidence to judge how “good” the data are. For this reason, it is vital that the metadata accompanying statistical outputs give users sufficient information to judge the quality of those outputs for themselves. The components of accessibility and clarity are implicit in this context, but information on the other five components should be given wherever possible, focussing on the components that users consider to be most important. There should be no stigma attached to warning users about comparability (or other quality) issues, as this is part of our professional duty. If users are given an honest assessment of quality, this will help to increase credibility and trust, which will have a positive impact on their perceptions of quality in the longer term.

Assessments of the quality of data sourced from other (national or international) statistical organisations that are perceived as overly critical, could risk damaging inter-organisation relations. To avoid such problems, rather than commenting on individual data cells, assessments should be made at a higher level, such as the data cube.

² See the report by Eurostat on assessing the quality of statistical norms at: http://epp.eurostat.ec.europa.eu/pls/portal/docs/PAGE/PGP_DS_QUALITY/TAB47143266/QUALITY%20OF%20NORMS.PDF

³ For more information on the total quality management (TQM) approach see: http://en.wikipedia.org/wiki/Total_Quality_Management

3. Quality Improvement Strategy

Introduction

The purpose of the Quality Improvement Strategy is to identify the broad areas where quality improvement work should be concentrated. It acts as a bridge between the theoretical Quality Framework, and the more practical annual Quality Improvement Programmes. It therefore takes a longer-term view than is possible in a programme relating to a specific year. It should, however, be seen as a living document, and should be reviewed each year, before the next Quality Improvement Programme is prepared, to ensure it remains relevant.

Priorities of the Quality Improvement Strategy

The Quality Improvement Strategy is driven by requirements expressed by users, as well as those identified by statisticians. Key inputs are the annual user survey, staff ideas, and other ad-hoc feedback. The four priority areas currently identified are:

1. **User needs** – the UNECE reforms shifted the focus from mainly serving internal customers to developing a much wider customer base. Whilst we have some knowledge of the needs of the key users of our statistical data (other parts of the UN, national governments and international bodies), we know less about the needs of our growing numbers of users in the academic and private sectors. Improving our knowledge of user needs, and gaining awareness of changing patterns is therefore a high priority.
2. **Improving accessibility and visibility** – our data are available free of charge on our web site, but that is not enough to ensure they are reaching all potential users. A number of improvements to the database web site have been implemented in response to user feedback, but more are needed. We are subject to the constraints of the UNECE web design policy, so it is important to continue to influence the development of that policy. Statistical publications, such as the biennial “UNECE Countries in Figures” help to bring our data to a wider audience, so should be reviewed periodically to ensure they remain as relevant as possible.
3. **Improving processes and quality control** – staff are encouraged to take an active role in improving processes and quality control mechanisms. To support this approach, a number of new rules and tools are needed, to improve efficiency, promote better standards for documentation and metadata, and improve data validation processes.
4. **Communicating quality** - we need to communicate the quality of data to our users more effectively, bearing in mind that user requirements are often different. We will continue to improve and standardize the metadata we provide, so that more advanced users have better information to assess the quality of our outputs themselves. We will also explore the possibility of summary quality indicators for users with limited time or domain-specific knowledge.

4. Data Quality Improvement Programme 2010

Introduction

This is the third annual Data Quality Improvement Programme for the UNECE Statistical Database. It seeks to build on the successes and experiences gained during the implementation of the two previous programmes, and to set priorities for quality improvements during 2010. The Database Coordination Group and the Director and Section Heads of the Statistical Division have approved this work plan, and agreed that it should be circulated to all statistical staff.

Actions and Priorities

The quality improvement tasks planned for 2010 are listed below. Their relationship to the broad themes of the Data Quality Strategy is indicated, and owners, deadlines and success criteria have been established.

Follow-up

The Quality Manager will report progress against this programme every three months to the Database Coordination Group. This group may agree to modify the tasks or deadlines to reflect changing priorities, and may ask the Quality Manager to bring any important issues to the attention of senior managers. At the end of the year, a final report will be produced showing the extent to which the quality improvement tasks were completed. This will be presented to senior managers, and circulated to all statistical staff.

The requirement for annual Quality Improvement Programmes should be reviewed by the Database Coordination Group and senior managers each year. New programmes should only be started if they add sufficient value to our outputs.

Quality Improvement Tasks - 2010

Task Description	Priority Area	Timing	Person Responsible	Success Criteria
Repeat the user needs survey for users of the database web site. Analyze the results and propose improvements.	User needs	Survey live during October / November 2010, Results by December 2010	Steve	At least 100 responses. Report approved by the Database Coordination Group.
Repeat the annual analysis of usefulness of data cubes.	User needs	By December 2010	Marlen	Report approved by senior staff and the Database Coordination Group.
Provide RSS feeds or similar automatic updates, to allow users to be notified when data are updated.	Improving accessibility and visibility	By August 2010	Marlen, Issoufou, ISU	High sign-up rate and positive user feedback.
Develop the use of Confluence as a repository for reference metadata that are too detailed for footnotes, including the statistical glossary.	Improving accessibility and visibility	By September 2010	Database Managers / Issoufou	All reference metadata in Confluence.
Improve the database user interface following the implementation of PX-Web 2008.	Improving accessibility and visibility	By December 2010	Marlen / Steve / ISU	Over 75% of respondents to the 2010 database user survey rate usability as good or excellent.
Review the contents of the country profiles for the 2011 edition of UNECE Countries in Figures.	Improving accessibility and visibility	By June 2010	Issoufou / Database Coordination Group	Contents agreed by the Database Coordination Group.
Implement improved data validation tools in the context of database re-engineering.	Improving processes and quality control	By December 2010	Marlen / Database Coordination Group	New tools in place and used by statistical staff.

Extend the quality checks of historical data on the Internet started by the Economic Statistics Section.	Improving processes and quality control	By September 2010	Intern (?) / Database Coordination Group	Report approved by the Database Coordination Group.
Review and update all process documentation.	Improving processes and quality control	By December 2010	Database Coordination Group	All documentation reviewed at least once during 2010.
Extend automatic data updates to more statistical domains and more data sources.	Improving processes and quality control	By December 2010	Marlen / Database Coordination Group	Automatic data updates applied to all data sourced from Eurostat, and extended to at least one other source.
Prepare a paper on definitions of summary quality indicators (traffic lights or similar) for agreement at the Senior Staff Meeting.	Communicating quality	By June 2010	Steve / Database Managers	Senior staff agree the proposed definitions.
Add summary quality indicators for each cube.	Communicating quality	By October 2010	Database Managers / Steve	Summary quality indicators implemented.
Implement the SDMX Content-oriented Guidelines in the statistical database, particularly the Cross-domain Code Lists.	Communicating quality	By September 2010	Database Managers / Steve	Content-oriented Guidelines fully implemented.