ISSN 1977-0375

eurostat
Methodologies and Working papers

ESS Handbook for Quality Reports

2009 edition

eurostat
EUROPEAN COMMISSION

Luxembourg: Office for Official Publications of the European Communities, 2009

ISBN 978-92-79-07855-2
ISSN 1977-0375
Cat. No. KS-RA-08-016-EN-N

Theme: General and regional statistics
Collection: Metholodogies and working papers

FOREWORD

The role of quality reporting will be strengthened in the forthcoming revision of the basic legal framework for European Statistics, reflecting the importance of quality in the world of statistics. Producers of statistics will have to demonstrate that high quality standards have been applied and achieved through all steps of the statistical production processes. Furthermore, users of statistics will be guaranteed access to appropriate metadata describing the quality of statistical outputs, so that they will be able to interpret and use the statistics correctly.

In 1998 Eurostat set up a Working Group on Quality, comprising members of the European Statistical System (ESS), which has developed and secured agreement on many aspects of the current quality framework, such as the ESS definition of quality and the ESS standard quality reporting documents. In addition, the ESS Leadership Expert Group on Quality, established in 1999, has produced a comprehensive set of recommendations for improvement of the ESS, which have been implemented by National Statistical Institutes through Eurostat-supported projects. This has resulted in the development of process and output oriented quality assessment handbooks and tools, such as the "Development of a Self-Assessment Programme (DESAP)", the "Handbook on Improving Quality by Analysis of Process Variables" and the "Data Quality Assessment Methods and Tools (DatQAM)", all of which are publicly available from the Eurostat Quality website.

The development of the European Statistics Code of Practice was a logical continuation of the focus on quality. It provides a broad conceptual framework for viewing quality and sets standards for the ESS institutional environment, statistical processes and statistical outputs.

The ***ESS Standard for Quality Reports*** will assist National Statistical Institutes and Eurostat in meeting the Code of Practice standards by providing recommendations for preparing comprehensive quality reports for the full range of statistical processes and their outputs. The ***ESS Handbook for Quality Reports*** provides much more detailed guidelines and examples of quality reporting practices. Both documents replace the 2003 versions and contain significant updates reflecting the advances in quality practices over the past five years. In particular, the documents are built around the fifteen principles articulated in the Code of Practice. The coverage of statistical processes using administrative sources or involving multiple data sources has been improved and more quality and performance indicators have been included.

The Standard and the Handbook are applicable to National Statistical Institutes and Eurostat in their roles as producers, compilers and disseminators of statistics. A key objective is to promote harmonised quality reporting across statistical processes and across Member States and hence to facilitate cross-comparisons of processes and outputs.

The new documents have been prepared by two statistical consultants - Michael Colledge and Jörgen Dalén - in cooperation with the members of the Working Group on Quality and staff in Eurostat. I would like to thank all colleagues in the ESS who have contributed to the development and finalisation of the documents.

Walter Radermacher

Director General Eurostat

Table of Contents

ABBREVIATIONS AND ACRONYMS

(See Part III for definitions of terms.)

CoP	European Statistics Code of Practice
CPI	Consumer Price Index
DatQAM	ESS Handbook on Data Quality Assessment Methods and Tools
DESAP	ESS Checklist for Survey Managers
ESQI	ESS Standard Quality Indicators
EHQR	ESS Handbook for Quality Reports
ESQR	ESS Standard for Quality Reports
ESS	European Statistical System
HICP	Harmonised Index of Consumer Prices
IQAPV	ESS Handbook on Improving Quality by Analysis of Process Variables
MCV	Metadata Common Vocabulary
NA	National Accounts
NSI	National Statistical Institute
NSO	National statistical office (NSI or other office producing official statistics)
PPI	Producer Price Index
PPP	Purchasing Power Parity
SPPI	Services Producer Price Index

PART I: OBJECTIVES, CONTENT AND CONCEPTS

1 Introduction

1.1 EHQR Objectives

The general aim of this Handbook (EHQR) is to provide guidelines for preparation of comprehensive quality reports for a full range of statistical processes and their outputs. In this context the term *statistical process* means sample survey, census, use of administrative data, production of price or other economic index, or any other statistical compilation commonly performed by a national statistical office, and the term *national statistical office (NSO)* refers to the national statistical institute (NSI) that plays the lead role in a national statistical system or to any other national agency or unit that produces official statistics of relevance to the European Statistical System (ESS).

The specific objectives of the Guidelines are:

- to promote harmonised quality reporting across statistical processes and their outputs within a Member State and hence to facilitate comparisons across processes and outputs;
- to promote harmonised quality reporting for similar statistical processes and outputs across Member States and hence to facilitate comparisons across countries; and
- to ensure that reports include all the information required to facilitate identification of statistical process and output quality problems and potential improvements.

There are numerous ESS documents relating to quality concepts, management and reporting, which can be rather confusing. Thus, a secondary aim of the Handbook is to summarise the relevant quality concepts and to explain how the various documents, including this one, relate to, and blend with, one another.

1.2 EHQR Users and Uses

Seven possible sets of users and uses of the EHQR are envisaged:

1. NSOs, for their own internal assessments of process and output quality;
2. NSOs, as the starting point for preparation of user-oriented quality reports;
3. NSOs, for submission to corresponding Eurostat Units;
4. Eurostat units, to prepare quality reports for their own statistical processes and outputs;
5. Eurostat units, to summarise process and output quality across the Member States based on NSO submissions into ESS level quality reports and to report, for example to the European Parliament or the Council;
6. Eurostat units, to report to users of European statistics; and
7. Eurostat units who are preparing statistical regulations or guidelines and wish to incorporate material on quality reporting.

The Handbook is primarily designed to assist NSOs in internal self-assessment and reporting to Eurostat (the first and third items above). However, as the Handbook puts considerable emphasis on output quality, it includes all the information necessary for user-oriented quality reporting (the second item). In addition, it provides some guidance on the preparation of European level quality reports (the fourth, fifth and sixth items) and gives very specific guidance to those who develop ESS regulations (the seventh item).

1.3 Changes from Previous Version

The previous version of the EHQR, entitled *How to Make a Quality Report*, was produced in October 2003. Since then there have been several significant developments in the ESS quality scene, including the introduction of the *Quality Declaration of the ESS* and the *European Statistics Code of Practice (CoP)*, and the preparation of the *ESS Standard Quality Indicators (ESQI),* the *ESS Checklist for Survey Managers (DESAP),* the *ESS Handbook on Improving Quality by Analysis of Process Variables (IQAPV),* and the *ESS Handbook on Data Quality Assessment Methods and Tools (DatQAM)*. Thus, a primary objective of the revision has been to take into account the new material in these documents.

- There is an updated introductory section on quality concepts, including reference to the ISO 9000 series of quality standards, the European Foundation Quality (EFQM) Excellence Model, and the ISO 20252 standard for Market, Opinion and Social Research.
- The term "statistical process" is introduced to identify the primary object of a quality report.
- A distinction is introduced between six different types of statistical process for which quality reporting (especially accuracy) needs somewhat different approaches.
- The notions of process quality and user perceptions of quality are expanded.
- The Handbook takes particular note of the 15 principles articulated in the CoP.
- The definitions of quality component, in particular comparability and coherence, have been clarified.
- More quality and performance indicators have been included, starting with, but not confined, to the indicators in the ESQI.

The companion ESS document entitled *Standard Quality Report* (SQR) has been revised in parallel with this document and in now entitled *ESS Standard for Quality Reports (ESQR)*. It is an extract from the guidelines presented in this document, summarising the essential requirements for a comprehensive quality report.

1.4 Types of Statistical Process

The methods of producing ESS statistics show a great diversity from a technical statistical perspective. A standard error structure is only well developed for surveys based on probability sampling from a frame of sampling units. Hence a single set of recommendations, especially those regarding accuracy, cannot apply to all statistics regardless of their production; it is necessary to introduce some distinctions.

A typology of statistical processes is needed. Such a typology can be drawn up in a variety of different ways. For the purpose of this Handbook six types of statistical process are distinguished. Defining these six types should be regarded simply as a pragmatic device solely for the purpose of the Handbook. It is expected that, in the future, new categories and improved distinctions will emerge.

1. **Sample survey.** This is a survey based on a, usually probabilistic, sampling procedure involving direct collection of data from respondents. For this kind of survey there is an established theory on accuracy that allows reporting on well-defined accuracy components (sampling and non-sampling errors).

2. **Census.** This can be seen as a special case of the sample survey, where all frame units are covered. There are population, economic and agricultural censuses.

3. **Statistical process using administrative source(s)**. This sort of process makes use of data collected for other purposes than direct production of statistics. An example is where statistical tabulations are produced from an administrative database maintained by the agency responsible for higher education.

   If, on the other hand, a questionnaire is sent by an NSO to a sample of (or all) educational institutions asking for information on students, teachers, courses etc., this is considered to be a survey (census) regardless of how, or from what, administrative sources the institutions retrieve the information. The key point here is that the questionnaire, including the definitions of the variables, is designed by the statistical agency.

   When discussing accuracy, three main types of process using administrative sources are distinguished: tabulations based on one register, integration of several registers, and event-reporting systems.

4. **Statistical process involving multiple data sources**. In many statistical areas, measurement problems are such that one unified approach to sampling and measurement is not possible or suitable. For example, in a structural business survey in which basic economic data - production, finance, etc - about businesses are aggregated, different units, questionnaires, sampling schemes and/or other survey procedures may be used for different segments of the survey. Furthermore, one or more segments may depend upon administrative data.

5. **Price or other economic index process**. The reasons for distinguishing economic index processes as a special type of statistical process can be described as altogether fourfold (not everyone being strong enough on its own): (i) there is a specialised economic theory to define the target concepts for economic indexes; (ii) their error structure involves specialised concepts such as quality adjustment, replacement and re-sampling; (iii) sample surveys are used in several dimensions (weights, products, outlets), mixing probability and non-probability methods in a complex way; and (iv) there is a multitude of these indexes playing a key role in the national statistical systems and the ESS.

6. **Statistical compilation**. This statistical process assembles a variety of primary sources, including all of the above, in order to obtain an aggregate, with a special conceptual significance. Mainly, but not only, these are economic aggregates such as the National Accounts and the Balance of Payments.

1.5 Content of Document

In addition to describing the objectives and users of the EHQR, Part I indicates the basis on which the guidelines in Part II were constructed. Readers who want simply to refer to the guidelines can skip the rest of Part I.

Part II provides guidelines for national quality reports. They are organised by statistical output and process quality components, with the primary section headings being:

1. Introduction to the statistical process and its outputs – an overview to provide context;
2. Relevance – an output quality component;
3. Accuracy - an output quality component;
4. Timeliness and punctuality - output quality components;
5. Accessibility and clarity - output quality components;
6. Coherence and comparability - output quality components;
7. Trade-offs between output quality components;
8. Assessment of user needs and perceptions – covering all aspects of output quality;
9. Performance, cost and respondent burden – process quality components;
10. Confidentiality, transparency and security – process quality components;
11. Conclusions – summary of principal quality problems and proposed improvements.

Part III lists all relevant references, including international quality standards, and ESS quality documents and quality regulations. It also includes a glossary of terms, as further elaborated below, and copies of some of the shorter key reference documents for ease of access.

1.6 Terminology

To the extent possible, definitions of the terms used in this document are in line with the ESS Glossary. Where a term is not in the ESS Glossary its definition is drawn from another international source where available, such as the Metadata Common Vocabulary (MCV), otherwise it is created for this document.

Using the term "statistical process" to describe the primary object of a quality report is not ideal as the same term could equally well be used to describe each of the various functions, such as questionnaire design, or editing, of which a statistical process is made up. However, it is the best choice. The alternative commonly used term "survey" is even less exact.

2 General Quality Concepts and Assessment Methods

2.1 Introductory Remarks

The aims of this chapter are to summarise general, internationally accepted quality concepts, to describe the way in which they have to be interpreted within quality frameworks in the particular context of national and European statistical systems, and to define the statistical process and output quality components that provide the basis for the quality reporting guidelines in Part II.

The chapter also outlines the spectrum of ESS quality assessment methods and indicates where the ESQR fits.

2.2 International Quality Management Concepts

Evolution of Quality Management

There is pictorial evidence that quality management was practiced thousands of years ago by the Egyptians when building the pyramids. Modern quality management has its origins *statistical quality control* as designed for production lines in factories in the 1920s and still in practice. Walter Shewhart, sometimes called the “father of statistical quality control” introduced the notion of variation due to assignable and to random causes and the elimination of the former using process quality control charts. The focus of quality control is inspection and correction. From a batch of production output, a sample is selected and each item is inspected for defects. The number of defective items is measured and, if that number exceeds a certain predetermined maximum, the whole batch is rejected, meaning that it is scrapped or sent back to the production line to be reworked. Otherwise, the defective items in the sample are corrected and the batch is passed on to the next stage of production. By this mechanism the quality of the output is controlled.

Subsequently, in the 1940s and 1950s, more emphasis was placed on preventing defects occurring rather than correcting them, for example feeding back the results of quality inspection as the basis for modifying the problems in the production lines and for employee training. This was referred to as *upstream quality control*, and the broader range of quality measures were referred to as *quality assurance*.

Over the next 30 years, as a result of papers written and advice given by quality gurus such as J Edwards Deming, renowned for his *Fourteen Principles of Management*, and Joseph Juran, who championed the human dimension, the notion of quality assurance was further extended to *total quality management (TQM)*. TQM principles are typically expressed along the following lines.

- *Customer focus:* an organisation depends upon its customers and thus must understand and strive to meet their needs; customers are central in determining what constitutes good quality; quality is what is perceived by customers rather than by the organisation.
- *Leadership and constancy of purpose:* leaders establish unity of purpose and direction of an organisation; they must create and maintain an internal environment that enables staff to be

fully involved in achieving the organisation's objectives; quality improvements require leadership and sustained direction.

- *Involvement of people:* people at all levels are the essence of an organisation; their full involvement enables their abilities to be fully used.
- *Process approach:* managing activities and resources as a process is efficient; any process can be broken down into a chain of sub processes, for which the output of one process is the input to the next.
- *Systems approach to management:* identifying, understanding and managing processes as a system contributes to efficiency and effectiveness.
- *Continual improvement:* continual improvement should be a permanent objective of an organisation.
- *Factual approach to decision making:* effective decisions are based on the analysis of information and data.
- *Mutually beneficial supplier relationships:* an organisation and its suppliers are interdependent and a mutually beneficial relationship enhances both.

ISO 9000 Quality Management System Series

The statement of TQM principles given above was extracted from the ISO 9000 series, which are the most widely used quality standards in the world. There are actually three standards:

- *ISO 9000: 2005 Quality Management Systems - Fundamentals and Vocabulary* describes the fundamentals of quality management systems, including quality principles (as reproduced above), terminology and models;
- *ISO 9001: 2000 Quality Management Systems – Requirements* specifies the required characteristics of a quality management system (as further described below); and
- *ISO 9004: 2000 Quality Management Systems – Guidelines* is designed to assist organisations having mature quality management systems to obtain performance improvements.

ISO 9001: 2000 Quality Management Systems – Requirements

This is the best known of the three standards as it is the one with respect to which organisations may seek to be certified. The requirements are presented under five headings, as follows.

- *Quality Management System (QMS)*: the organisation must establish, document, and implement a QMS and continually improve its effectiveness. This includes identifying the QMS processes and ensuring their effective operation and continual improvement. QMS documentation must include a quality policy, a quality manual, quality procedures and the information required for effective management.
- *Management Responsibility*: senior management must provide evidence of its commitment to QMS development and operation by establishing and promulgating a quality policy, stressing the importance of determining and satisfying customer requirements, ensuring that quality objectives are established and resources are available for QMS development and operation.

- *Resource Management*: the organisation must provide the resources required to implement, maintain and continually improve the QMS and to enhance customer satisfaction. Staff operating the QMS must be competent, well trained and aware of the importance the quality objectives and their role in assuring they are achieved.
- *Product Realisation*: the organisation must design and develop the processes needed for effective production. This includes determining product requirements and quality objectives, establishing effective customer communications, handling of requests and dealing with feedback and complaints, designing and developing inputs, processes and outputs and corresponding verification, validation, monitoring, inspection, and testing procedures.
- *Measurement, Analysis and Improvement*: the organisation must design and implement the monitoring, measurement, and analysis processes required to demonstrate that products conform to specifications and to continually improve the effectiveness of the QMS.

EFQM Quality Standard

The European Foundation for Quality Management has produced what is in effect a quality standard somewhat similar to the ISO 9000 series. Whilst originally designed for use by commercial organisations in the European Union, the EFQM Excellence Model is now globally recognised with thousands of organizations around the world using the EFQM principles to guide their business strategies and operations. *Excellence* is defined as *outstanding practice in managing the organisation and achieving results* and is said to require leadership commitment to principals along similar lines to the ISO 9000 principles, namely: results orientation; customer focus; leadership and constancy of purpose; management by processes and facts; people development and involvement; continuous learning, innovation and improvement; partnership development; and corporate social responsibility.

ISO 20252:2006 Market, Opinion and Social Research – Vocabulary and Service Requirements

This new international standard establishes terminology and service requirements for organizations conducting market, opinion and social research and sets a common level of quality for market research globally. It has its origins in the British Standards Institute standard BS7911, which it supersedes. It comprises sections on quality management systems, managing the research, data collection (including fieldwork, self-completion, and collection from secondary sources), data management and processing (capture, coding, editing, analysis, security), and reporting.

Six Sigma

Six Sigma is a business management strategy with wide-spread application. Whilst originally developed by Motorola, there is now no particular owner or organisation responsible for it. The basic aim of Six Sigma is to identify and remove the causes of defects and errors in manufacturing and business processes. It uses a particular set of quality management methods, including business process mapping and cause and effect diagrams, and it creates a collection of people within an organization ("black belts", etc.) who are experts in these methods. Each Six Sigma project follows a defined sequence of steps and has specific financial targets in the form of cost reduction or profit increase.

Balanced Scorecard

By focusing on human issues as well as financial outcomes, the *Balanced Scorecard* helps provide a more comprehensive view of an organisation. It is a performance planning and measurement framework, with similar principles to Management by Objectives, publicized by Robert and Norton in the early 1990s. Having realized the shortcomings of traditional management control systems, and following a one-year research project involving 12 companies, Kaplan and Norton published a book entitled *The Balanced Scorecard*. Since its introduction, the Balanced Scorecard has been awarded a prize by the American Accounting Association as the "best theoretical contribution in 1997", and its industry and academic attention has placed it alongside approaches such as Activity Based Costing and Total Quality Management.

2.3 Developing Quality Management Systems in the ESS Context

Need for Interpretation of International Quality Standards

With the exception of ISO 20252, the quality standards and models described in the previous section are intended to cover all organizations whatever their organisational structure, processes and products and are thus expressed in very general terms. Thus, as stated in the introduction to IS0 9001:2000:

- *the design and implementation of an organisation's QMS is influenced by varying needs, particular objectives, the products provided, the processes employed and the size and structure of the organisation. It is not the intent of the standard to imply uniformity in the structure of QMSs or uniformity of documentation.*

In other words, in developing a quality management system (QMS), each individual organisation or group of similar organisations must interpret the standards within its specific context. This is the subject of this section.

ESS Context

The European Statistical System (ESS) has two levels: the national level comprising national statistical offices (NSOs) within each Member State, of which the nominated National Statistical Institute (NSI) plays the leadership role; and the European level comprising essentially the units in Eurostat. The organisational units in this context may be characterized as follows.

- They are government not private enterprise. They are not profit based. They supply data to non-paying *users* rather than to paying *customers*. For the most part, the users cannot influence quality through purchase decisions.
- Some of the users are actually *internal users*, for example the national accounts area is a user of numerous direct and register based surveys as well as a producer.
- The primary inputs at national level are typically data from individual enterprises, households and persons, whether collected directly or through administrative processes.
- The core production processes at national level are transformations of these individual data into aggregate data, i.e., statistics, and their assembly into national statistical products.
- The core production processes at European level are further aggregations and transformations of these aggregate data and their assembly into European level statistical products.

- The primary products, referred to in this document as *outputs,* are statistics and accompanying services.

This is the context within which international quality standards have to be interpreted. ISO 20256 Market, Opinion and Social Research goes part way to providing an appropriate quality standard for NSOs, in particular placing considerable focus on the need for a quality management system. However, the standard is not entirely appropriate, being aimed at commercial organisations, and thus not in accordance with the first element of the context noted above.

Elements of Quality Management System in ESS Context

In the ESS context, a quality management system (QMS) is typically referred to as a quality management framework (QMF). The elements of which a QMF is comprised vary from one organisation to another but typically include the following:

- Quality policy – a short statement by senior management indicating the extent of its commitment to quality management;
- Quality model – a definition of what is meant by quality, usually elaborated in terms of quality components;
- Quality objectives, standards and guidelines – target quality objectives together with international or local standards and guidelines adopted by the organisation;
- Quality assurance procedures - part of, or embedded in, the production processes to the extent possible;
- Quality assessment procedures – sometimes incorporated in the quality assurance procedures, more frequently conducted on a periodic basis, for example based on a self-assessment checklist such as the DESAP checklist;
- Quality measurement procedures - specifically including a set of quality and performance indicators, with procedures for collecting the corresponding data values being embedded in the production processes to the extent possible;
- Quality improvement procedures – continual improvement and re-engineering initiatives.

Following sections describe some particular examples of these elements in the ESS context, with the emphasis on European rather than individual national instances.

2.4 Quality Policies, Models, Standards and Guidelines in the ESS

Quality Policies

The UN Statistical Commission provides some guidance to NSOs and European level units on the sort of environment within which quality management can flourish through two documents.

- The UN Fundamental Principles of Official Statistics for national statistical systems were promulgated by the UN Statistical Commission in 1994. There are ten principles (copy annexed), and whilst none of them explicitly relates to quality, they are all fundamental to establishing a quality management system.

- The UN Principles Governing International Statistical Activities were adopted by the UN Committee for the Coordination of Statistical Activities in 2005. The first principle is that "High quality international statistics, accessible for all, are a fundamental element of global information systems".

Within the ESS, the ESS Quality Declaration was adopted by the Statistical Programme Committee in 2001 as a formal step towards total quality management in the ESS, in line with the EFQM Excellence Model. It is just one and a half pages long, comprising the ESS mission statement, the ESS vision statement and ten principles which borrow from the UN Fundamental Principles and tailor them in the ESS context.

The logical continuation of this work was the European Statistics Code of Practice (CoP) promulgated by the European Commission in 2005, which commits Eurostat and the Member States NSOs to fifteen principles covering the *institutional environment, statistical processes* and *outputs*. For each principle, the CoP defines a set of indicators reflecting good practice and providing a basis for assessment.

The quality requirements are integrated into the forthcoming basic legal framework on European Statistics (ref. proposal for a Regulation of the European Parliament and of the Council on European Statistics, 14280/08 of 16 Oct. 2008) where Article 2 provides the statistical principles that shall govern the development, production and dissemination of European Statistics. It is also stated that these statistical principles are further elaborated in the CoP (as referred in Article 11). Statistical quality is addressed under Article 12 where the quality criteria that are to be applied are described in line with the ESS Quality Definition. The Article 12 states:

"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 time lag 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. Such 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."*

Quality is also a consideration in other regulations adopted by the Council and the Parliament creating the legal basis for the provision of European statistics in various domains. Council Regulations are themselves quality assurance mechanisms, setting specific timeliness targets establishing methodological standards leading to enhanced accuracy and comparability, and covering relevance in the form of the needs of European institutions for national statistics.

Several NSIs have formulated their own individual quality policies. For European level statistics, the European level policies are paramount.

Quality Models

The development of a quality management system depends upon a precise definition of quality. The starting point is the definition of output quality from a customer/user perspective. The most general and succinct definition of product quality is *fitness for use*. In slightly more detail, but rather obscurely, ISO 9000 defines product quality, which in the case of statistics is synonymous with output quality, as:

- *the degree to which a set of inherent characteristics fulfils requirements*.

A somewhat more comprehensible definition provided by ISO 8402:1986 is:

- *the totality of features and characteristics of a product or service that bear on its ability to satisfy stated or implied needs*.

These broad definitions provide a basic notion of output quality, but need to be accompanied by a more precise interpretation of the *components* of output quality. This is provided in the ESS Quality Definition, which was presented at the October 2003 meeting of the ESS Working Group *Assessment of Quality in Statistics*. Although never given a more formal status, this document has provided the basis for defining output quality components in all subsequent quality related documents, including the CoP and the forthcoming basic legal framework on European Statistics.

Output Quality Components

In line with the ESS Quality Definition and the last five CoP Principles, output quality in the ESS is assessed in terms of the following components.

- *Relevance:* outputs meet current and potential users' needs.
- *Accuracy and Reliability:* outputs accurately and reliably portray reality.

- *Timeliness and Punctuality:* outputs are disseminated in a timely and punctual manner.
- *Accessibility and Clarity:* outputs are presented in a clear and understandable form, disseminated in a suitable and convenient manner, made available and accessible on an impartial basis, and accompanied by supporting metadata and guidance.
- *Coherence and Comparability:* coherence means that outputs are mutually consistent and can be used in combination; comparability is an aspect of coherence means that outputs referring to the same data items are mutually consistent and can be used for comparisons across time, region, or any other relevant domain**.**

Process Quality Components

There are other aspects of quality. *Output quality* is achieved through *process quality*. Process quality has two broad aspects:

- *Effectiveness:* which leads to the outputs of good quality; and
- *Efficiency:* which leads to their production at minimum cost to the NSO and to the respondents that provided the original data.

Guidance on the definition of more detailed process quality components is provided by the first ten CoP principles. The quality components thus formulated are in two groups: components of the *institutional environment* within which a programme of statistical processes is conducted, and components of *individual statistical processes*.

Institutional Environment

- *Professional independence:* the professional independence of the staff (responsible for the process and output dissemination) from other policy, regulatory or administrative departments and bodies, as well as from private sector operators, is required to support the credibility of outputs.
- *Mandate for data collection:* the organisation has a clear legal mandate to collect the particular information required. Where a survey is conducted under the statistics act providers can be compelled by law to allow access to or to deliver data.
- *Adequacy of resources:* the resources available are sufficient to meet systems and processing requirements.
- *Quality commitment:* staff commit themselves to work and cooperate according to the principles stated in the 'Quality declaration of the European statistical system.
- *Statistical confidentiality:* the privacy of data providers (households, enterprises, administrations and other respondents), the confidentiality of the information they provide and its use only for statistical purposes is absolutely guaranteed.
- *Impartiality and objectivity:* production and dissemination of statistics respect scientific independence and are conducted in an objective, professional and transparent manner in which all users are treated equitably.

Individual Statistical Process

- *Sound methodology:* sound methodology, including adequate tools, procedures and expertise underpins quality statistics.
- *Appropriate statistical procedures:* appropriate statistical procedures are implemented from design through data collection to data validation and evaluation.
- *Non-excessive burden on respondents:* the reporting burden is in proportion to the needs of the users and not be excessive. It is monitored over time and targets are set for its reduction.
- *Cost effectiveness:* resources are effectively used.

Several NSIs have formulated their own individual quality models, mostly in line with the ESS output quality components. For reporting quality at ESS level and for cross country comparisons, the ESS model is appropriate

Quality Standards and Guidelines

As previously mentioned the ESS Quality Definition provides the basis for defining output quality components. However, it would benefit from a minor revision to clarify the concepts in the light of recent experience.

How to make a Quality Report was presented at the ESS Working Group meeting in October 2003. Its purpose was to serve as a handbook for reporting on the quality of statistical data in the ESS. In addition to presenting reporting guidelines, it discusses quality concepts, costs, and EU legislation relating to the quality of statistics. It makes extensive use of examples drawn from Member States and includes a substantial bibliography. It is superseded by the *ESS Guidelines for Quality Reporting*, i.e., this document.

The Standard Quality Report, which was also presented at the ESS Working Group meeting in October 2003, provided the ESS standard for quality reporting. It was aligned with and an extract from How to Make a Quality Report. It is superseded by *ESS Standards for Quality Reporting* now in preparation.

The ESS Quality Glossary was presented at the same ESS Working Group meeting in October 2003 as a companion document to How to Make a Quality Report. It provides a short definition of a range of technical terms, indicating the sources of the definitions. The document would benefit from revision, first, as a number of key terms are not included and, second, as several other glossaries that have subsequently appeared.

In addition, there are a number of other ESS guidelines described in the following section on quality assessment, and most NSIs have their individual quality standards and guidelines.

2.5 ESS Quality Assessment Methods

In view of the relevance of quality assessment to quality reporting, the various guidelines currently available relating to quality assessment are described in some detail in this section.

ESS Handbook on Data Quality Assessment Methods and Tools

The Handbook on Data Quality - Assessment Methods and Tools (DatQAM) details the full range of methods for assessing process and output quality and the tools that support them. It makes recommendations on how these methods and tools can be implemented. It is primarily targeted at ESS quality managers, enabling them to introduce, systematise and improve the work carried out in the field of data quality management.

DatQAM defines *assessment method* as being an approach to evaluation, for example documenting/reporting, calculating (indicators), auditing, self-assessing, or questioning the user. It defines *assessment tool* as referring to the actual form by which the method is implemented, for example, producing a quality report, calculating key indicators, using an auditing procedure, a checklist, or a user survey. In this sense a quality report is a quality assessment tool.

DatQAM notes that there are two types of prerequisite for quality assessment:

- documentation describing the key components of the process and outputs together with some basic systematic quality measurements (which it refers to as *preconditions*); and
- internal or external references with respect to which the assessment can be made, for example international standards, general quality guidelines, process specific best practices, and user requirements (which it refers to as the *external environment*).

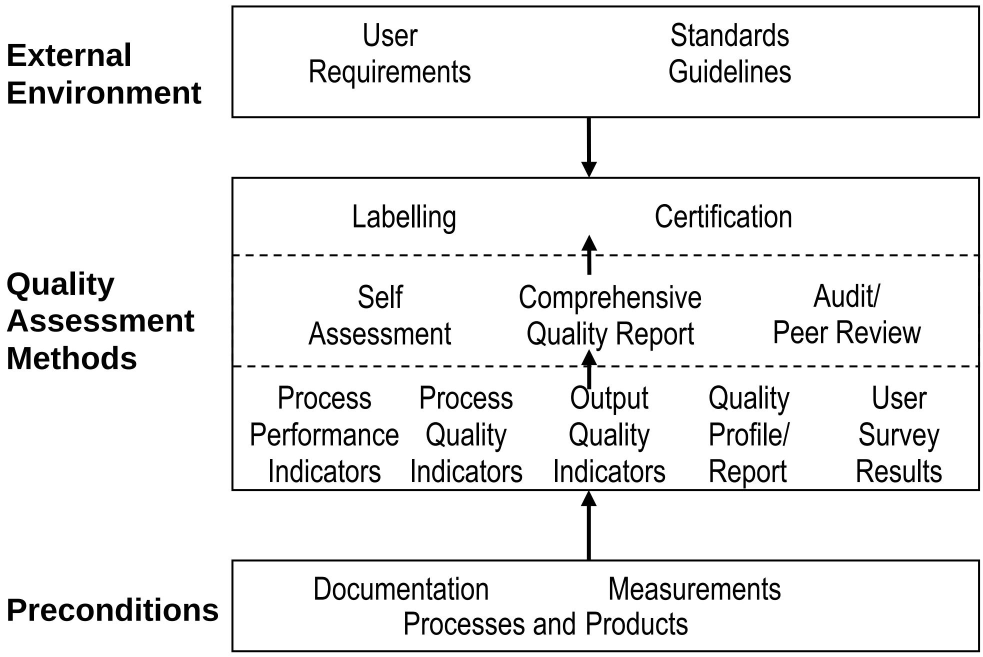

Figure 1. Quality Assessment Methods in Context

Figure 1 indicates a spectrum of assessment methods on three levels. It is based on a corresponding figure in DatQAM. The main reason for the changes that are made here is to show a comprehensive quality report in the middle, on the second level. The figure here is also somewhat more detailed than the original one. These changes do not oppose DatQAM messages. The DatQAM figure shows the level where an assessment method is introduced. Moreover, DatQAM provides a strategy for introducing and combining different quality assessment methods into packages with successively increasing ambition level.

- *Level 1: Output quality indicators, process quality and performance indicators, quality profiles and short reports, and user feedback/survey results*. At this level, detailed data obtained from measurements of processes and outputs (sometimes referred to as paradata) are selected and structured into quality and performance indicators, which may themselves be grouped in quality reports. For example, data gathered from interviewer control are processed in order to identify possible problems in data collection. User surveys provide user perceptions of product quality. Quality reports are prepared to meet legislated or other specific information requirements. Quality profiles are prepared to provide succinct summaries for user purposes.
- *Level 2: self-assessments; comprehensive reports and peer reviews/audits*. At this level, using information from the Level 1, processes and products are assessed against comprehensive internal or external standards or checklists. Assessment may be self assessment, for example based on the DESAP checklist, or audit/ peer review involving staff from other areas or external experts. Such assessments give a more comprehensive picture than Level 1. They may cover a range of precision and detail, from answers to a mostly coded checklist to a very comprehensive report.
- *Level 3: labelling and certification*. At this level processes and products are assessed against a quality standard. Labelling is for the benefit of users and involves assigning a label to the statistical products of those processes that meet specific NSI defined quality requirements. Certification is similar to labelling but is against an external standard such as ISO 9001 or ISO 20252 and is applied to statistical processes rather than specific products. Labels and certificates thus provide highly condensed quality information regarding the statistics released.

ESS Handbook on Improving Quality by Analysis of Process Variables

The Handbook on improving quality by analysis of process variables describes a general approach and useful tools for identifying, measuring and analysing key process variables. It includes practical examples of the application of the approach to various statistical processes.

ESS Self Assessment Checklist for Survey Managers

The European Self Assessment Checklist for Survey Managers (DESAP) is a tool for ESS survey managers that enables the conduct of quick but systematic and comprehensive quality assessments of statistical processes and outputs and identification of potential improvements. It is compliant with the ESS Quality Definition and is generic in the sense that it applies to all statistical processes irrespective of the specific subject matter area or the survey methodology. It is totally compatible with, and, in essence, a checklist version of this Handbook.

There is a Condensed Version of Checklist for survey managers (DESAP) that contains only selected key questions and there is an Electronic Version of Self Assessment Checklist with a Electronic Version User Guide

ESS Standard Quality Indicators

Standard Quality Indicators was presented to the May 2005 meeting of the ESS Working Group *Assessment of Quality in Statistics.* Its purpose is to present a standard set of indicators for use by producers in summarizing the quality of their statistical outputs. It is aligned with and elaborated the indicators in *How to Make a Quality Report*. It is essentially superseded by the EHQR.

2.6 Quality Reporting Structure

Quality reporting underpins quality assessment, which in turn is the starting point for quality improvements. Thus, standards and guidelines for effective quality reporting are an essential aspect of the quality management framework. The reporting structure, i.e., set of headings and subheadings that is envisaged for a comprehensive quality report is outlined in the following paragraphs. It is the structure on which the guidelines in Part II are based.

The output and process quality components are the starting point in choosing an appropriate structure for a quality report. However, given that process quality leads to product quality, if the structure required an explicit assessment of quality in terms of each of process and output quality component there would be considerable duplication between the sections. Thus, the proposed quality reporting structure is based, in essence, on the output quality components and supplemented by headings covering those aspects of process quality that are not readily reported under any of the output components. The primary section headings, with their justification, are as follows.

1. *Introduction to the statistical process and its outputs*. An overview is required to provide context for the whole report.
2. *Relevance.* Relevance is one of the output quality components and merits its own section.
3. *Accuracy.* Likewise, accuracy is an output quality component and merits its own section.
4. *Timeliness and punctuality.* Timeliness and punctuality are output quality components that merit a section.
5. *Accessibility and clarity.* Accessibility and clarity are output quality components that merit a section.
6. *Coherence and comparability.* Coherence and comparability are output quality components that merit a section.
7. *Trade-offs between output quality components.* The output quality components are not mutually exclusive in the sense that there are relationships between the factors that contribute to them. There are cases where the factors contributing to improvements with respect to one component lead to deterioration with respect to another. The most commonly quoted example involves the factors contributing to accuracy and timeliness. Thus, there is the need to a section that deals with the trade-offs that have to be made in such circumstances.
8. *Assessment of user needs and perceptions.* Users are the starting point for quality considerations and, typically, information regarding their needs and perceptions is obtained for all output components at the same time, not just each one individually. Hence there is a need for a separate section on this topic.
9. *Cost, performance and respondent burden.* Cost, performance and respondent burden are important process quality components that cannot be readily covered under any of the headings based on output quality. There are invariably trade-offs to be considered between all the output quality components and cost, performance and response burden.

10. *Confidentiality, transparency and security.* Likewise, confidentiality, transparency and security are process quality components that cannot be readily covered under any of the headings based on output quality.
11. *Conclusion.* The report should conclude with a summary of principal quality problems and improvements proposed to deal with them.

2.7 Types of Quality Report

There is a wide range of different possible quality reports according to the scope of the report, the level of detail, the producer or user orientation, and the perspective of process or output. The various types and how they are covered in the guidelines are described in the following paragraphs.

Scope/level of Report

A quality report can have narrow or wide scope, from dealing with a specific indicator and the process that produced it, to the whole ESS, as illustrated in Figure 2. The guidelines in this document are primarily aimed at describing all quality aspects of a ***statistical process*** (direct or register based survey, price index or other major statistical compilation as previously defined) ***at national or at European level***, in other words the row in bold italic in Figure 2. The guidelines can also be used for lower level domains (in the bottom two rows of Figure 2) but not for higher level domains (the upper level rows in Figure 2).

Figure 2: Scope/Levels of Quality Reporting

Scope	**National level**	**European level**
Institution	NSI and all other NSOs	Whole ESS
Broad statistical domain (e.g. health, agriculture)	All statistical processes within broad statistical domain	All statistical processes in all Member States within same broad statistical domain
Statistical process	***Process with full set of outputs, as determined by NSO***	***Same process and outputs, as determined by ESS for all Member States***
Subdomain within statistical process	Subgroups or specific data items for which outputs are produced	European aggregates* for same subgroups or specific indicators
Specific indicator(s)	Outputs in the form of single numbers or time series of such numbers	European aggregates* of single numbers or time series of such numbers

* European aggregates are functions (averages, totals, etc.) of national estimates for EU-27, EEA, Euro Zone, etc.

ESS level report

Based on quality reports from Member States, quality reports may be produced for European (ESS) level statistics. Such reports may not only to bring together in one place information about the quality of all the national outputs and the processes that produced them but also present the quality of aggregated estimates at the European level, comparisons between countries, and specific uses of European level data.

Two aspects of ESS level statistics stand out as distinct from national statistics and hence of special importance.

- European level statistics may include aggregations (averages, sums etc.) of national estimates applicable to a European entity (EU-27, EEA, Euro area etc.). If so, the quality report will refer to these aggregations.
- European level statistics may include comparisons and contrasts of national estimates. If so, the quality report will refer to the comparability of outputs across Member States.

Thus, two possible objectives of an ESS quality report are to provide information, first, on the quality of aggregate statistics and, second, on the quality of comparisons of national statistics. In addition, there is a third possible objective, namely to give a condensed overview of the quality of national outputs.

Producer/User Orientation of Report

A quality report may be user-oriented, producer-oriented or both. There are producers and users at various levels. A producer of statistics may at the same time be a user of other statistics. Reports may be required to communicate quality information between the producers. Users of final outputs can be advanced analysts and researchers, or the public at large, often represented by media.

These guidelines are ***producer-oriented*** with a special focus on what is needed for ensuring the quality of the ESS system. User-oriented quality reporting requires its own set of guidelines. However, a quality report produced according to these guidelines will include all the information required for the production of user-oriented reports as well.

Process/Output Orientation of Report

A quality report may have a process or an output orientation. As noted earlier in this section the guidelines in Part II have an output orientation even though the primary target users are producers.

Level of Detail in Report

A quality report can range from very short and concise to very detailed. For example, a quality profile may cover only a few specific attributes and indicators; a completed DESAP checklist covers all aspects of the statistical process and its outputs, but not in great detail.

The guidelines in this document are aimed at the most comprehensive form of report commonly prepared, i.e., a full scale report with qualitative and quantitative information, dealing with all important aspects of output and process quality in detail. Thus, the guidelines require not only a description of processes and quality measurements but also quantitative quality measures or assessments and discussions of how to deal with deficiencies.

Related Documentation

The quality report is one type of documentation for statistical processes. Many others are used as well and in this regard national practices differ widely. Some countries produce technical reports and the like where the statistical methodology is described in detail, for example with estimation formulas, etc. When such documentation exists, the quality report can refer to it and need not repeat all the same information in the body of the report. However, when no such documentation is available, information on methodology must be included in the quality report itself.

Reporting Frequency

Quality reports may be prepared for every cycle of the statistical process, annually, or periodically. Typically the more frequent the report the less detail. The guidelines in this document are aimed at the sort of comprehensive document that will be produced periodically, say every five years, or after major changes. In between such comprehensive quality reports it is envisaged that less detailed reports may be prepared, for example quality and performance indicators for every survey occasion, and a checklist completed annually.

Whilst it is not the role of this document to dictate the quality policy of NSOs or the ESS units, it is felt in some quarters that annually updated quality reports should be the norm. This would not actually place an undue burden on the report writers, since, if no major changes have taken place, material could be cut-and-pasted from one year to the next and the only new material would be in the form of updated quality and performance indicators.

2.8 Role of Quality Reporting

Within a quality management framework, a quality report is a means to an end, not an end in itself. Thus, it should not only provide a factual account of quality according to the reporting structure, it should include recommendations for quality improvements and justification for their implementation.

PART II: GUIDELINES FOR QUALITY REPORTS

1 Introduction to the Statistical Process and Its Outputs

To facilitate an understanding of its technical parts, a quality report should include some background information on the statistical process and outputs that are the subject of the report. This is the purpose of the Introduction.

It is natural to start by providing a brief history. When was the process in question initiated and what were its initial objectives? What major changes have subsequently been made and why? This should be followed by a general description of the process and its outputs, and their evolution over time.

The broad statistical domain (or domains) to which the statistical outputs belong should be stated and related outputs in the same domain listed. The boundary between the process and outputs described in the quality report at hand and those described in other reports should be made clear. This boundary is sometimes not obvious since outputs with different names and conceptual targets can have one or more subprocesses in common and even share the same micro-data base. Where this is an issue, the reasons for the chosen boundary should be explained

An overview of all outputs associated with the process should be given, including:

- all media (Internet, paper reports, reports to general statistical compilations like yearbooks etc.);
- national outputs as well as outputs reported to international organisations;
- outputs to the ESS system should be separately listed.

The format and structure chosen for the quality report should be motivated by the general type and characteristics of the process. The most important quality problems should be indicated.

References, preferably by hyperlinks, should be given to other documentation on the methodology of the process and quality of the outputs. References concerning specific quality aspects should also be given in other places in the quality report. (This statement applies for each quality component but is not repeated.)

ESS level

In an ESS level quality report, an overview of the European regulations (if any) governing the statistical outputs and the processes by which they are to be produced should be given, together with a list of the Member States that have produced quality reports and the general coverage of these reports.

Quality and Performance Indicators

None specifically identified.

Summary

What should be included in the Introduction

- A brief history of the statistical process and outputs in question.
- The broad statistical domain to which the outputs belong; related statistical outputs.
- The boundary of the quality report at hand and references to related quality reports.
- An overview of all output produced by the statistical process.
- References to other documentation, especially on methodology.

2 Relevance

ESS Quality Definition.

Relevance is the degree to which statistical outputs meet current and potential user needs. It depends on whether all the statistics that are needed are produced and the extent to which concepts used (definitions, classifications etc.,) reflect user needs.

2.1 For All Statistical Processes

A ***content-oriented description*** of all statistical output should be given, typically including:

- key indicators (especially those emphasised in press releases, e.g., national unemployment rate, 12-month inflation, GDP growth);
- variables, e.g. turnover, consumption, employment, salaries;
- subdomains, (for which indicators are shown separately);
- estimates of level versus change (time series); and
- reference period (month, quarter, year, etc.) and frequency of release.

An assessment regarding the ***key outputs/ indicators*** desired by different categories of users should be given and any shortcomings in outputs for important users should be mentioned. This could, for example, involve insufficient breakdown of data into sub-domains, time series that are too short, or outputs that are too infrequent, for example quarterly instead of monthly. Not all user needs can be met, reasons being either budgetary or technical. The quality report should include information on unmet user needs, the reasons why certain needs cannot be fully satisfied, and any plans to satisfy needs more completely in the future.

Definitions of ***statistical concepts*** and their relation to the definitions that would be ideal from a user perspective should be given. Concepts defined during the design and planning of the statistical process include target population, target definition of units, and aggregation formula. It is often the case that what is ideal differs between users and, if so, this should be noted. Sometimes it is possible to apply different definitions to the same set of micro-data and present all the results. More usually this is not possible and a single definition has to be selected, in which case the motivation for the chosen definition should then be given. Any discrepancies between the definitions used and accepted ESS or international definitions should always be clearly pointed out.

Numerical illustrations of the likely sensitivities of the results to the chosen definitions can be very informative and should be provided whenever possible. The basis for these illustrations could be sensitivity analyses or simulations. Such illustrations inform users of the risks of a ***relevance problem*** for their particular application, i.e., of a discrepancy between the definitions used and what the user wants.

Definitions also affect coherence and comparability and thus, instead or as well, can be discussed under that heading (see Chapter 6).

There is a grey zone between certain relevance problems and accuracy, as further discussed in connection with sampling errors (using a cut-off threshold) and coverage errors in Chapter 3.

If certain indicators, variables and/or domains foreseen by the ESS or other international regulations/ guidelines are not covered, the statistics are ***incomplete***. An explicit statement of the degree of completeness in terms of ESS regulations should be given where relevant, including plans for improvements in this respect in the future. Completeness can also be measured relative to a national target.

ESS level

In an ESS level quality report, national compliance with agreed ESS or other international definitions should be described in detail. Other important differences in definitions between Member States should be noted.

The completeness of the national statistical outputs should be analysed. In this respect, two dimensions are important:

- Are any Member States not producing the statistics in question?
- Are important variables missing from the outputs of some Member States?

2.2 For Statistical Processes Using Administrative Source(s)

When administrative data are used for statistical purposes, the registered population and definitions of the included variables are already fixed based on the primary purpose of the administrative register or transaction database. These definitions are often not ideal for statistical purposes and may give rise to constraints when defining the target population and target variables. The quality report should include definitions of important variables including population definition in the register/database and discuss their relation to the definitions desired by key users of the statistics.

ESS level

An overview over national definitions and sources should be given.

Example 2.2.A: European Statistics on Accidents at Work – Methodology, 2001

A methodology handbook exists for **Serious accidents at work** that serves some of the functions of a quality report. Pages 24-27 present a detailed definition of an *accident at work*, including different variants in the Member States.

2.3 Price Index Processes

In price indexes, although defined in general terms by economic theory, the target of estimation is usually impossible to specify exactly and is even open to some controversy. A quality report should discuss important issues concerning the target of estimation and its relation to approaches and methods chosen, also relating these to recommendations in international manuals and legal documents in the ESS system.

Example 2.3.A: CPI Target

There is a debate on whether the CPI should try to estimate a cost-of living index in a sense defined by economic theory or whether some other concept (fixed basket, cost of goods) should be declared as the target. A quality report should include report on the stance taken on this issue and how it influences the choice of detailed methodology approaches.

Example 2.3.B: PPI Target

There is a similar issue for the theoretical target for Producer Price Indexes, which, however, is less often discussed and may have a smaller practical influence on actual methodology. In any case, the general approach including definitions and coverage should be discussed as a relevance issue.

Example 2.3.C: Construction Input and Output Concepts

For Construction Price Indexes, there is a sharp distinction between input and output concepts, each with its own conceptual framework.

Example 2.3.D: Index Aggregation Formulae

The choice of index formula at different levels of aggregation is important. Any price index involves a number of calculation stages and choice of formula at each stage. This choice involves issues involving both target definition and accuracy. It could also be referred to a special methodology section.

2.4 For Statistical Compilations

The quality report needs to relate to the definitions and conceptual choices made in line with recommended international manuals or other forms of general agreement.

For the National Accounts there are two relevant manuals, the System of National Accounts 1993, at international level and ESA95 at the EU level.

For the Balance of Payments there are, for example, the Eurostat BoP Vademecum, the OECD benchmark definition of Foreign Direct Investment (FDI) and the Task Force FDI concerning the compilation of flows and stocks.

2.5 Quality and Performance Indicators and Summary

Quality and Performance Indicators

R1. Rate of available statistics.

General definition: the number of output data elements provided in accordance with a relevant ESS regulation as a ratio of those required by the regulation.

To be further defined for subject-matter domain: (i) the set of relevant data elements; (ii) possible weighting, distinguishing key and non-key data elements.

Remark 1: This indicator is applicable only if there is an ESS regulation or guideline.

Remark 2: Not all output data elements are of equal importance. Thus, an appropriate weighting system will often improve the usefulness of this indicator.

ESS level

(i) Presentation of R1 over all Member States;

(ii) Presentation of an overall (weighted or un-weighted) R1 over all Member States.

Summary

What should be included on Relevance

- A content-oriented description of all statistical outputs.
- Definitions of statistical target concepts (population, definition of units and aggregation formula) including discrepancies from ESS/international concepts. (May also be discussed under Coherence and Comparability.)
- Information on completeness compared with relevant regulations/guidelines.
- Unmet user needs, including reasons for not meeting them.
- Available quality indicators.

3 Accuracy

ESS Quality Definition

The accuracy of statistical outputs in the general statistical sense is the degree of closeness of estimates to the true values.

3.1 Overall Accuracy for All Statistical Processes

A purpose of statistics is to produce estimates of unknown values of quantifiable characteristics of a target population. Estimates are not equal to the true values because of ***variability*** (the statistics change from implementation to implementation of the statistical process due to random effects) and ***bias*** (the average of the possible values of the statistics from implementation to implementation is not equal to the true value due to systematic effects; the bias of an estimator equals the difference between its expected value and the true value).

There are several types of error originating from all the various production processes and a typology of errors has been developed. ***Sampling errors***, which apply only to sample surveys; are due to the fact that only a subset of the population is selected, usually randomly. ***Non-sampling errors***, which apply to all statistical processes, may be categorised as:

- coverage errors:
- measurement errors;
- nonresponse errors; and
- processing errors.

Model assumption errors are not considered an independent type of error. Usually models are used precisely in order to reduce other errors. If so, they are second-order error types and do not merit a separate heading. However, it is important to distinguish different cases with regard to the use of models in official statistics (see Section 3.8.1 below.)

The above forms of non-sampling errors have clear definitions in probability sample surveys, but for other statistical processes their meanings are not so well established and need more elaboration as the error typology above may not be the one best suited for reporting accuracy. Therefore the error profiles for each type of statistical process are discussed in separate sections.

As previously noted, there is a grey zone between certain relevance problems and accuracy. This occurs when the definitions most appropriate for users are modified so as to fit the practical measurement circumstances, with the consequence that the statistical outputs become less relevant to the users. To avoid ambiguity, accuracy as here defined refers to the difference between the estimates and the true values as defined in the practical situation.

The section on accuracy in a quality report has to be accompanied by a presentation of the methodology used. A methodology description serves two purposes. First, it serves to ensure the reader that the methodology in use is sound and in accordance with best practice and internationally accepted standards, as required by the ESS Code of Practice Principles 7 and 8.

Second, it enables the reader to understand better the particular quality assessments regarding sources of error.

There are several ways to present the methodology. It can be in an annex, or a separate chapter, or it can be integrated into the discussion of types and sources of error. A mixed approach can also be used. It could also be presented as a separate document to which reference is made, preferably by hyperlink. In the latter case there should still be a summary presentation of methodology in the quality report.

A section on overall accuracy is required in any quality report. The section should begin with identification of the main sources of random and systematic error in the statistical outputs. Distinctions should be made between the main variables, domains of study and estimates. The section should contain a summary discussion of all errors, random and systematic. For key indicators an assessment of the aggregate risk of random and systematic errors should be made.

Although random variation can be associated with all types of error, its major source is normally sampling. The detailed presentation of random errors is, therefore, usually best made in a sub-section on sampling. However, an assessment of the risks of bias for important estimates is often best made in the overall accuracy section, except where bias is associated mainly with a particular source of error, in which case the assessment can be included in the relevant sub-section on that type of error.

According to the state of knowledge of the producer, the assessment of bias can be in quantitative or qualitative terms, or both. It should reflect the producer's best current understanding including actions taken to reduce bias. A qualitative assessment should refer to the likely sign of the net bias and include a statement referring to its order of size, for example using general terms like *negligible, small,* or *large* or by stating its likely maximum value. The basis for this statement should be included as well.

Specific sources of error should be described in separate sub-sections under accuracy. Different types of statistics are affected by different types of errors and the relative importance of each type varies. Therefore, the detailed organisation of the section on accuracy in a quality report needs to be unique for each statistical process and outputs. Domain-specific regulations may give more guidance. This document provides advice for each specific type of statistical process.

A useful general reference on reporting accuracy is Measuring and Reporting Sources of Error in Surveys produced by the US Office of Management and Budget (2001). An example of national quality guidelines is Statistics Finland's Quality Guidelines for Official Statistics (2007).

ESS level

In an ESS level quality report, the key methodological divergences by Member States from ESS and/or international norms should be described under Accuracy or in the Introduction. There should also be information on other important differences affecting accuracy between Member States.

An assessment of the most critical issues concerning accuracy should be included. Separate sections should deal with each of these issues. Where European aggregates are calculated their computation should be explained and their specific error profiles, based on national estimates, should be analysed based on currently best available knowledge.

The detailed structure of the accuracy section depends on the key issues for each type of statistical process and its outputs. The ultimate objective is to provide the best overview assessment possible of the possible margins of error associated with the estimates in the national and European level outputs. Special emphasis should be on how these margins of error could affect comparisons between Member States.

Evaluation

To report on accuracy it is necessary to ***evaluate*** accuracy, i.e., to acquire the relevant information about accuracy. However, apart for the vast literature on variance estimation for sampling errors in probability surveys, not much is written about the evaluation of accuracy. Thus evaluation methods are indicated in several of the subsections below. Note there is a distinction between *quality control* (meaning ensuring quality of output) and *evaluation* (meaning acquiring information about the quality of output).

Two classical works on the evaluation of accuracy in official statistics are Morgenstern (1965) and Zarkovich (1966). There are also two documents, both in the Swedish language, that treat this topic: Dalén (1981) and Andersson, Lindström and Polfeldt (1999). These, especially the latter, are extensively used in this Chapter.

The methods and approaches for evaluation described below are less well defined than variance estimation used for evaluating sampling error, for which there is a solid statistical theory. Often they have a more common sense character and the results they provide have to be used with judgement and accompanied by a discussion of the possible risks of error.

At this point it is appropriate to note that there is no method to estimate the total error of an estimate. However, various approaches for acquiring some indication of total error exist.

The first approach is to make a *comparison with another source*. For example, employment is often estimated by labour force surveys as well as by business surveys. In practice the differences observed in comparisons between such sources are combinations of errors and differences in definitions (as further discussed in the chapter on Coherence.) An analysis aiming at decomposing the differences can shed light on total error.

Consistency studies can be used when there are known relations between different parameters, for example:

- *number of married men equals number of married women (according to traditional marriage laws);*
- *number of dwellings in year 1 = number of dwellings in year 0 + new construction – demolition (+ net change in use);*
- *income = expenditure + saving - new loans.*

Relations need not be exactly satisfied by the data. However, significant discrepancies require further exploration of possible errors or mistakes. Consistency studies should normally be done before statistical results are published but for reasons of time this is not always possible. If different parameters are estimated independently, inconsistencies between estimates for them could be a starting point for analysing errors in each one of them.

Quality and Performance Indicators

None

Summary

What should be included on <u>Overall Accuracy</u>

- A presentation of the methodology sufficient for (i) judging whether it lives up to internationally accepted standards and best practice and (ii) enabling the reader to understand specific error assessments.
- Identification of the main sources of error for the main variables.
- If micro-data are accessible for research purposes, it may be motivated to make additional comments to assist such uses.
- A summary assessment of all sources of error with special focus on the key estimates.
- An assessment of the *potential for bias* (sign and order of magnitude) for each key indicator in quantitative or qualitative terms.

3.2 For Sample Surveys

3.2.1 Sampling Errors – Probability Sampling

Sampling can be of two types: probability sampling, meaning that each unit of the frame population has a known, non-zero probability of being selected in the sample, and non-probability sampling.

For probability sampling, sampling theory provides techniques for the estimation of the expected value and variance of specific indicators over all possible samples, Therefore, the random variation due to sampling can be calculated. Furthermore, sampling biases are normally zero or negligible so that the variance can be taken to represent total sampling error (subject to full response - see nonresponse errors).

The variability of an estimator around its expected value may be expressed by its variance, standard error, coefficient of variation (CV), or confidence interval. As regards non-sampling errors, computation of the bias requires knowledge of the true population value and detailed knowledge of the survey processes. In practice it is often possible to get an idea about whether the bias is positive or negative but rarely possible to estimate its size well. The total error of an estimate relative to the unknown true population value is expressed as the *root mean square error (RMSE)*, defined as the square root of the sum of variance and the square of the bias. Although being the most relevant direct measurement of accuracy from a user point of view, the RMSE can rarely be estimated. Therefore a report on accuracy needs to take a more indirect approach based on separate assessments of the various types of non-sampling errors as previously listed. The types of errors that occur and their likely magnitudes vary according the survey and outputs in question.

Sampling errors should be reported for all estimates resulting from a statistical process where sampling is involved. Where they are significant, and there is a scientific basis for their calculation, they should be given in quantitative terms along with the estimation and variance formulas. In this context, there are several presentational devices that can be used.

The *standard error* is the square root of the variance of an estimator. Usually the standard error is not suitable for use by itself since its interpretation is not obvious to the average user.

The coefficient of variation (CV) is defined as the standard error divided by the expected value of the estimator. It is the standard error in relative (percentage) terms. It is the most suitable sampling error statistic for quantitative variables with large positive values, which are common in economic statistics. It is not recommended for proportions, for estimates that are expressed in percentage terms or for changes, where it could easily be misunderstood. It is also not usable for estimates that can take on negative values such as profits, the net export/import value etc.

The *confidence interval* is defined as an interval that covers the true value with a certain probability. In most cases where it is reasonable to assume the estimator follows a normal distribution, the interval that results from taking ± 2* estimated standard error from the point estimate results in a 95 % confidence interval. Taking instead ± 2* estimated CV expresses the interval in percentage terms.

For key indicators the sampling error should be expressed as a confidence interval, since this is the most rigorous and clear way of demonstrating sampling variability.

For large sets of estimates in tables, confidence intervals often lead to a rather clumsy presentation and CVs or CV intervals are more natural to use. A CV interval could, for example, state that the CV is in a certain range (5-10 %, say) of the estimate. Different ranges can be denoted with letters (e.g., A= <2 %, B= 2-5 %, C= 5-10 %, D= >10 %). Use of ranges is also appropriate because estimates of sampling variability are not exact.

Especially in economic surveys, *outliers* can greatly influence the estimates and lead to major sampling errors. The quality report should state clearly, whether, how and why outlying sample units have received special treatment in the estimation process.

In household surveys, results are often presented as proportions or percentages and it is not usually appropriate to present random sampling errors in the form of CVs. Confidence intervals are a better choice. It is sometimes possible to present simplified indicators of sampling errors, where a certain range of estimated proportions are associated with a certain level of sampling error according to the well-known formula Variance = p(1-p)/n, where p is the proportion and n the sample size.[1]

For business surveys, especially where large positive numbers (of production, turnover, export, etc.,) are targeted, estimated CVs are normally the best way to express sampling error. The size of the sampling error relates to the sample size for the domain to which the estimates relate, so, for a large table with many cells that would be overburdened with an estimated CV in every cell, they are instead best presented in a separate table.

Where CV thresholds are included in regulations, a comparison between estimated CVs and the relevant thresholds should be included.

[1] For example if n=10,000 and p is between 0.2 and 0.8 the standard error will be between √0.000016 and √0.000025 or 0.004-0.005. The confidence interval for these proportions can thus be approximated as ±0.01.

Some further technical points concerning the presentation of sampling errors are:

- Non-response should be taken into account, i.e., the sample size should be the effective sample, after deduction of non-response.
- The original stratification should be applied, i.e., the sampling error should not be artificially reduced by first moving outliers to a special stratum. Also note that variance estimation should be in accordance with the actual sampling and estimation method applied.
- Sampling errors for estimates of change are of great importance, although sometimes more difficult to calculate due to non-independence between samples in adjacent periods. Nordberg (2001) and Wood (2008) discuss this problem at a fairly general and technical level. It should be remembered that an assumption of independence normally leads to an overestimate of the sampling error for a change (since the covariance term is actually negative). If this is the case a statement like "*The sampling error for the change between Q3 2007 and Q3 2008 is at most X"* is valid, where X is calculated under an assumption of independence.

Example 3.2.1.A: Presentation of CVs, (UK Index of Production, ONS (1999b, page 101).

Estimated standard error of the Index of Production (IoP) (Manufacturing) at section and sub-section

All manufacturing	DA	DB	DC	DD	DE	DG	DH	DI	DJ	DK	DL	DM	DN
0.8	2.5	2.1	5.3	3.7	2.6	1.2	3.5	1.9	2	2.3	3.4	1.8	2.8

The estimates were made using both the Taylor linearization methods and parametric bootstrapping. Assumptions had to be made about the likely variance of the PPIs (which were not based on a probability sample) and the effect of neither stock adjustment nor seasonal adjustment was included.

The results in the above table refer to 12-month changes in the IoP, that is, the IoP data were treated as if they were from a period of twelve months after the base year, which was 1990 at the time of the study. The estimate of the standard error at the Manufacturing level using the bootstrap simulation was 0.82, close to the estimate shown above. The confidence interval for the overall manufacturing index is approximately 1.6 percentage points.

Example 3.2.1.B: CVs for UK Annual Production and Construction Inquiries (ONS, 1999b, page 41).

The coefficients of variation (%) of the main variables at NACE 2-digit level in 1996 are shown in the following table:

Industry	IoP Weight (per 1000)	Total employment	Total sales and work done	Total purchases	GVA at basic prices	Total net capex
10	7.05	2.98	5.30	5.26	5.64	8.96
14	8.45	10.71	3.61	3.61	4.48	11.91
15	99.31	0.57	1.19	1.49	1.02	1.08
17	23.59	0.92	1.35	1.53	1.44	8.57
18	14.62	1.32	2.58	3.60	1.76	9.03
19	5.42	1.89	3.12	3.57	3.95	14.36
20	10.96	3.27	4.61	5.42	4.36	11.81
21	26.31	0.87	1.44	1.76	1.31	3.30
22	72.63	0.85	1.26	1.48	1.43	4.06
23	17.86	0.64	0.34	0.57	0.86	0.22
24	90.73	0.59	0.84	1.05	0.59	5.83

25	39.66	0.85	1.15	1.50	1.03	3.07
26	29.77	0.98	2.21	2.85	1.62	3.66
27	33.58	1.01	3.27	4.15	2.23	3.53
28	59.02	1.01	1.40	1.83	1.40	3.86

The lowest coefficients of variation for gross value added are for divisions 24 (chemicals) and 34 (motor vehicles), both of which are dominated by large (and hence completely enumerated) businesses. Some of the highest coefficients of variation are in divisions 37 (recycling) and the textiles industries (19 and 20) that are dominated by small firms.

Example 3.2.1.C: Reporting of CVs at ESS level for Structural Business Statistics (Eurostat, 2001a)

Coefficients of variation for EU totals - "aggregated" CVs

The coefficient of variation must take into account non responses, the misclassification errors and the sampling error. Yet it is very often reported that the misclassification error is not taken into account. The non response rate is not always taken into account either. For instance, some Member States report, for a given characteristic, a CV=0 while the non response is different from zero.

The Regulation foresees that CVs at national level cannot be disseminated. Therefore the following numbers refer only to the distribution of these CVs for the main Industries and variables:

Coefficients of variation for EU-totals at Nace 3-digit level (in %)

Manufacturing, Gas and Water supply, Construction (NACE Rev.1 sections D, E and F)

	Mean	Max	P75	Median	P25
Turnover	1.1	8.3	1.2	0.8	0.4
Value added	1.2	6.0	1.6	1.0	0.5
Personnel costs	0.8	5.4	0.9	0.5	0.3
Tangible investment	3.5	41.5	3.8	1.9	0.9
Employment	0.9	5.4	1.3	0.6	0.4

The "aggregated" CVs (at EU level for Manufacturing) seem to be relatively low. There is hardly any CV higher than 1.5%. A possible explanation could be that many countries use administrative sources, at least for a significant part of their data collection. Yet misclassification problems and even non response may often occur in that case without being always taken into account.

The variability of the estimators varies from one characteristic to the other. Gross investment in tangible goods is to be distinguished from other characteristics, as the related CV is much higher than for the other characteristics. For manufacturing the median CV (at 3-digit) is 1.9 %, almost four times higher than for personnel cost (0.5%). The precision of employment characteristic (0.6%) is almost as good as the one of personnel costs (0.5%), while the CV is a bit higher for value added and turnover (0.8% and 1%).

At NACE Rev.1, 2-digit level (in %)

Coefficient of variations for EU totals at 2-digit level for Manufacturing, Gas and water supply, Construction (NACE Rev.1 sections D, E and F)

	Mean	Max	P75	Median	P25
Turnover	0.5	1.6	0.6	0.4	0.2
Value added	0.6	1.8	0.7	0.5	0.3
Personnel costs	0.3	1	0.5	0.2	0.1
Tangible investment	1.4	4.9	1.9	1.1	0.6

Employment	0.5	2.2	0.6	0.4	0.2

At 2-digit level, the variability is lower. For manufacturing and construction (NACE Rev.1 sections D, E and F), the coefficients of variation at 2-digit level vary from 0.2% to 0.4% from one characteristic to the other, while it varies from 0.5% to 1% at 3-digit level (except investment)."

Example 3.2.1.D: Confidence intervals in Labour Force Surveys

Table 4.1.1-4.1.2 (page 11-12) Labour Force Survey,

This EES level quality report provides confidence intervals for key indicators of employment and unemployment for Member States and European aggregates in a coherent and lucid way.

3.2.2 Sampling Errors – Non-Probability Sampling

When *non-probability sampling* is applied, random error can not be estimated without reference to a model of some kind. Furthermore, sampling biases may well be significant and need to be assessed as well. There are many types of non-probability sampling, each of which require their own evaluation depending on the situation at hand.

One type of non-probability sampling that is frequently applied in economic surveys and therefore needs special attention is the use of a *cut-off* threshold. Units (businesses, enterprises, establishments) below a certain size threshold, although belonging to the target population, are not sampled at all; there is a term cut-off sampling for such a procedure. Technically this situation is similar to undercoverage (further discussed below under *coverage errors*) but with the distinctive feature that the cut-off is intentional and there is register information for the excluded units, which gives a better opportunity for model-dependent estimation. Two of the reasons for a cut-off threshold are reduction of the response burden for small units and considerable contributions to the errors (sampling and non-sampling) of the design-based estimator.

The introduction of a cut-off threshold results in a different situation than probability sampling, including a bias (according to the design-based survey sampling paradigm) due to the sampling probability being zero. On the other hand, if, by definition, the target population refers only to the sampled portion of the population, then instead of an accuracy problem there is a relevance problem for those users who are interested in properties of all units and not just of those above the threshold. When the population below the threshold is included in the target, a model-based estimator is natural. From this perspective, the quality reporting rather belongs to Section 3.8.1 below, but it is put here.

A cut-off threshold is often combined with probability sampling above the threshold and in this case can be called ***sampling with cut-off*** as opposed to ***census with cut-off*** where all units above the threshold are included. For an example of census with cut-off see Example 3.8.1.C below.

For reporting on sampling with cut-off the most suitable approach is two-fold. For the sampled portion of the population, random sampling error may be presented as above. For the non-sampled portion a discussion about the (explicit or implicit) model used in the estimation process should be included. Often this model simply assumes that the units cut off behave similarly to those in the sampled portion. This assumption should be analysed as far as possible. Such an analysis is useful also where the cut-off is considered as a relevance problem rather than contributing to sampling error. If the accuracy has been evaluated on an intermittent basis by sampling in the cut-off portion this should be reported.

For other forms of non-probability sampling, for example those applied for price indexes, it may be reasonable to apply standard error estimators as if the sample is effectively random, using an assumption for the design or some model based approach. This approach has, however, to be complemented with a discussion of possible sampling bias and of possible limitations in the sampling model used. For example it can often be determined whether (and why) the estimates of sampling error thus derived are "conservative" (i.e., upper limits) relative to the real errors.

It is not enough just to declare that a sample is "purposive" or "subjective" without providing more information. Technical details on how the sample was selected should always be reported. The rationale for not using probability sampling should be stated as well as an assessment of how the sampling procedures can affect the estimates.

Quality and Performance Indicators

A1. Coefficient of variation (CV).

General definition: as given above.

To be further defined for subject-matter domain: list of variables and domains for which CVs are to be provided as well as devices for summarising the information.

ESS level:

(i) CVs for variables and Member States;

(ii) CVs for European aggregates (if any).

Remark. CVs are useful primarily for variables taking on large values. They are not appropriate for proportions or for indicators that can take on negative values.

Summary

What should be included on <u>Sampling Errors</u>

<u>Always applicable</u>

- Where sampling is used there should be a section on sampling errors.
- As far as possible sampling error should be presented for estimates of change in addition to estimates of level. If necessary, reasonable assumptions can be used.

<u>If probability sampling is used:</u>

- There should be a presentation of sampling errors calculated according to formulas that should also be made available. If the estimators include adjustments for non-sampling errors, for example nonresponse, this should be explained and included also in the accuracy assessment.
- The most appropriate presentational device should be chosen, normally CVs, ranges of CVs, or confidence intervals.
- If outliers have received special treatment in estimation, this must be clearly described.

<u>If non-probability sampling is used:</u>

- For sampling with cut-off an assessment of the accuracy due to the cut-off procedure should be included in addition to the presentation of sampling error for the sampled portion of the population (see also Section 3.8.1 below).
- For other forms of non-probability a sampling model can be invoked for the estimation of sampling error. A motivation for the chosen model and a discussion of sampling bias should be included.

3.2.3 Coverage and Other Frame Errors

The ***target population*** is the population for which inferences are made. The frame (or frames, as sometimes several frames are used) is a device that permits access to population units. The ***frame population*** is the set of population units which can be accessed through the frame and the survey data really refer to this population. The frame also contains sufficient information about the units for their stratification, sampling and contact.

The concept of a frame is traditionally used for sample surveys, but applies equally to censuses. For some other types of statistical process the concept may also be useful but has to be defined in each case.

Coverage errors (or frame errors) are due to divergences between the frame population and the target population.

Three types of coverage error are distinguished:

- ***Undercoverage:*** there are target population units that are not accessible via the frame (e.g., persons without a phone will not be listed in a telephone catalogue);
- ***Overcoverage:*** there are units accessible via the frame which do not belong to the target population (e.g., deceased persons still listed in a telephone catalogue);

- ***Multiple listings (duplication)***: target population units are present more than once in the frame (e.g., persons with two or more telephone connections).

Other sorts of frame deficiency that can cause errors involve incorrect classification, contact and auxiliary information about the units included in the frame. Such deficiencies can also cause errors other than coverage errors. For example, wrong contact information (address, phone number) may result in non-response error, or if the size of a unit as recorded in the frame is smaller than its actual size, the sampling error may increase (sometimes dramatically where an outlier is created).

Overcoverage can be detected during the measurement process, is straight forward to handle in the estimation procedure, and results in increases in sampling error and survey costs.

Multiple listings, if recognised, can be handled by statistical methods and also result in an increase of sampling error and cost but no significant biases. However, multiple listings of smaller units for which sampling rates are low are difficult to detect. If there is a significant risk of such error this should be reported.

As a matter of good practice, in annual or less frequent survey, the frame information for every contacted unit should be checked to see whether it is accurate. For subannual surveys, frame information should be checked for all new units and periodically, say annually, for continuing units. In this way overcoverage, inaccurate classification, contact and auxiliary information and multiple listings can be detected. The extent of these problems among the selected units can give an idea about their extent over the whole frame. Response burden has to be taken into account when deciding how to check the accuracy.

Quantitative information on overcoverage and multiple listings is normally easy to obtain in sample surveys and censuses. This information should be included in the quality report in sufficient detail with respect to important sub-domains. For other statistical outputs, frame and coverage errors should be included where relevant.

Undercoverage cannot be detected in the measurement process and is the most serious type of coverage error. The resulting bias depends on the units outside the frame population but in the target population and the differences between the characteristics of these units and those in the frame population. Thus, a qualitative description of these units is a first step in assessing the undercoverage bias. Methods to detect undercoverage and assess its effects include, for example (i) when there is a time lag in registering frame units, a later frame version can provide information, and (ii) comparisons with another frame or other external information. Where undercoverage is suspected to be significant, an assessment is always needed. As far as possible estimates of undercoverage (extent and effect) should be included in the quality report.

Undercoverage can be “defined away” by limiting the target population to what is covered in the frame. If so, the coverage error is transformed into a relevance problem and should be treated under that heading instead.

Whilst the quality of the survey frame is important, the main objective of a quality report is to indicate the effects of frame deficiencies on the statistical outputs. To this end, information on the frequency and timing of frame updates is useful to include in the quality report as well as their likely consequences for the survey estimates at hand.

References to any documents describing frame quality should be made. Sometimes a summary description of how the frame is derived and its general properties (reference period, updating actions) is useful. In particular, frames for economic surveys are typically derived from a central business register that serves a number of surveys. Frames for household surveys are often drawn from a household register or from a general purpose area frame constructed by listing households in areas selected by probably sampling. Thus the quality report should include a description of the register or other frame source in so far as this assists in understanding coverage errors and their effects.

For household surveys, the frame is often based on a census from a number of years back. If not updated, undercoverage and classification errors will result from changes that have occurred since then. If it is updated, the updating procedures and resulting lags will be important for determining the remaining undercoverage, so this is an aspect that needs to be dealt with in a quality report. Some persons are often left out of population registers, such as recent immigrants, people without a permanent registered dwelling or institutional households. In some surveys people without telephone are left out. The quality report should try to assess and preferably quantify the errors resulting from all these sources of undercoverage.

Business surveys normally use a business register. The business register updating frequency and procedures determine the coverage properties of a survey frame drawn from the business register as of a certain date, and the quality report should try to assess this. In addition, classification issues may influence the effective coverage of business surveys, more so than the case of household surveys. In particular, the economic activity codes of economic unit determine whether or not they are in scope for the survey, and thus wrong codes may cause undercoverage (which cannot be detected) or overcoverage (which can).

There may also be a coverage issue in terms of the particular type of unit that should be the target unit for a survey. NACE refers to four possible standard types of unit for use in business statistics – *enterprise, kind of activity unit, local unit* and *local kind of activity unit*. If the largest of these unit types (enterprise) is chosen as the survey target then there may be some enterprises that are not in scope for a particular survey even though at a more detailed level (say kind of activity unit) there would have been one or more units in scope for that survey.

Evaluation of Coverage Errors

Possible methods include the following.

Matching with a different register. The sampling frame is matched with a control register that wholly or partly covers the same population as the frame. If the sampling frame and control register are not both electronically stored then matching can be done on a sample basis. If the control register is of superior quality, then errors in the frame can be directly assessed. Otherwise a reconciliation process, involving checking (a sample of) the non-matches is needed to determine the extent of errors in the survey frame.

Analysis of lag structure. Every frame is updated with a certain lag: the birth, death or change of a unit is registered with a delay. Due to this the frame will always, to a smaller or larger degree, have a less than perfect coverage at the time of use. The lag effect can be studied for example by matching two consecutive register versions and establishing which of the units in the latter version should, by definition, have been included in the former. Other approaches are also possible. Register errors can be studied in several consecutive versions. It may be possible to observe certain stability in error levels that can be assumed to continue into the future. The degree of under- or overcoverage as well as changes in contact data etc., can thereby be estimated. (It is also possible to use this kind of information for a *model-based adjustment* of the estimates themselves.)

Quality and Performance Indicators

A2. Rate of overcoverage.

General definition: proportion of units accessible via the frame that do not belong to the target population.

Remark 1. Overcoverage is best reported together with nonresponse in a coherent manner so that, for example, the treatment of units with unknown status is made clear.

Remark 2. It is also possible to define rates of misclassification, incorrect contact details and multiple listings in straight-forward ways. However, in most cases these indicators are not as important as A2.

Remark 3. Although the rate of undercoverage is the most important indicator it is not usually directly observable and thus not included in the list.

ESS level

Individual values and aggregates of A2 over Member States.

Summary

What should be included on Coverage Errors

- Quantitative information on overcoverage and multiple listings.
- An assessment, preferably quantitative, on the extent of undercoverage and the bias risks associated with it.
- Actions taken for reduction of undercoverage and associated bias risks,
- Information on the frame: reference period, updating actions, and references to other documents on frame quality.

3.2.4 Measurement Errors

Measurement errors are errors that occur during data collection and cause the recorded values of variables to be different from the true ones. Their causes are commonly categorized as:

- Survey instrument: the form, questionnaire or measuring device used for data collection may lead to the recording of wrong values;
- Respondent: respondents may, consciously or unconsciously, give erroneous data;
- Interviewer: interviewers may influence the answers given by respondents.

Measurement here implies measurement *at the unit level*, for example the monthly income of a person or the annual turnover of a company. The result of a measurement may be viewed as comprising the true value plus an error term that is zero if the measurement is correct. This implies that a true value exists, which is sometimes subject to debate.

Measurement errors can be systematic or random. Random errors are often associated with the idea of replication, i.e., if the measurement process is repeated many times for the same unit under fixed conditions the registered measurement values will vary randomly whereas the systematic error will stay constant. The following simple model can be used to represent this fact for the registered value y_k:

- $y_k=Y_k+B_k+e_k$, where Y_k is the true value, B_k the systematic error and e_k the random error for unit k.
- e_k has an average of 0 over repeated measurements whereas B_k is constant for a given unit.

More complex and realistic models can be obtained by splitting B and e according to the causes of error, e.g., questionnaire, respondent, collection method, or interviewer. Biemer and Stokes (1991) give an overview over many possible measurement models.

Measurement errors may cause both bias and extra variability of statistical outputs. Bias is usually the main problem. The evaluation of measurement errors depends on the type of data at hand. The quality report should identify the main risks in terms of measurement error for the statistical process under consideration.

Respondent errors are often caused by the desire to appear socially acceptable, sensitive questions and the like. Where such factors are at play in the survey data, a specific discussion of possible resulting measurement errors is necessary

Data editing identifies inconsistencies. They are usually the consequence of errors in the original data, but may also be the result of processing errors due to coding or data entry. Information from the data editing process should be included in the quality report, since it is indicative of the risk of measurement error. The failure rate of each edit rule can be calculated over the records to which the edit is applied. Clerical correction and/or automatic imputation are usually applied in order to remove inconsistencies in the data. The failure rates, therefore, are an indication of the quality of data collection and processing and not of the quality of the final data. The amount of detail on data editing in a quality report should be related to the importance of measurement errors in the survey in general and for the key indicators.

Questionnaires used in the survey should be attached to the quality report as annexes (or as hyperlinks if they are large). The efforts made in design and testing the questionnaires should be briefly described.

Important measurement errors are unique for each survey and thus need to be accompanied by any available analyses, or, in the absence of such analyses, the producer’s best knowledge.

Evaluation

When the risk of substantial bias is considered high, evaluation studies are needed. Respondent error can be assessed by a re-interview study in which the respondent is asked to provide the same data on a second occasion. If there is no memory effect, the two interviews may be considered independent and the difference between the responses is an indication of the size of the measurement error.

In order to assess instrument or interviewer effects, repeated measurements can be made with different instruments (e.g., alternative phrasing of questions) or different interviewers. Alternatively an experiment can be carried out with subsamples being randomly allocated to different instruments and /or interviewers. This approach is mostly appropriate for surveys on attitudes/opinions or where memory effects are involved. Information on relevant aspects of interviewer training could also be included.

For data of a factual nature, especially economic data, the potential for finding other databases with similar data is often good. Such databases may contain similar data with a time lag and can be used for evaluating earlier versions of the present statistical output. However, when comparing two sets of data, it is necessary to distinguish measurement errors from comparability issues, such as differences in definitions, with which they may be confounded.

Another method for finding errors is to subject economic data to accounting rules and reasonableness checks. These approaches are usually used in the editing stage in order to correct the data before final estimation.

Four groups of methods are applicable for evaluating errors at unit level. Such errors could have been generated in the measurement phase, the processing phase or they could have existed already in the sampling frame.

Comparisons with other information at the unit level. This is of course the best way to obtain a quality check provided there is a common unit identification scheme for both sources. Matching of registers, as mentioned under coverage errors above, can be used also for this purpose, provided the control register can be assumed to have good information about the units for certain variables. Care must be taken to distinguish actual errors from differences in definition or measurement points in time.

Control at source /re-interview with superior method. Control at source means that the evaluator gets access to source data (company accounts or records kept at an agency etc.) A re-interview with a superior method may use an expert interviewer or face-to-face instead of mail interview. Another approach is to use the same interview method once again (but with a different interviewer) and use a reconciliation procedure (for example an expert panel) where different responses are obtained. Such methods capture all types of errors that have occurred during measurement and processing, whether due to respondent, questionnaire, interviewer or data entry. They are best done for a random sample of units resulting in unbiased estimates of error.

Replication. Replication means that there are two ore more observed values for a sampled survey unit. Such values can obtained by different interviewers, from different respondents (answering for the same sampled unit) or simply by repeating the measurements after sufficient time for the respondents not to remember their initial responses. The differences between the measurement values can be used for learning how stable the measurement process is. Formal analyses of replication often assume that errors are independent between replications. This assumption is rarely fully met in practice. The method is used for estimating the random variation due to measurement. Under some circumstances (for example if an expert interviewer or respondent is used) it can also provide some information on the systematic error (bias).

Effects of data editing. By comparing results from original and edited data the extent of initial measurement error can be deduced. Of course, this gives a minimum estimate of the error levels, since not all errors will be detected in the editing process. Such analyses provide ideas for improving the measurement methods, but no information on the undetected measurement errors nor how they affect the statistical outputs.

Example 3.2.4.A Report on interviewer effect

Berthier and Néros (1998) applied a method for measuring interviewer effect on the French results of the European Household Panel Survey. Their basic conclusions were as follows.

- The interviewers were asked to give details of the type of non response (no contact, long absence, inability to answer and refusal). Analysis of non response types showed that interviewers and interview duration both had a high effect on non response.
- The interviewer effect was non-existent for evaluation of the standard of living; it was small for amount of earnings; and it was slightly higher for the correlation between these two variables.
- Respondents were given two options for declaring earnings: either to state their exact earnings or to choose one among predefined earnings classes. The interviewer had an effect on respondents' choices.

Example 3.2.4.B: Report on response consistency

Särndal *et al* (1992, p. 604) report part of the results of an evaluation study of the 1980 US Census of Population and Housing carried out by the Census Bureau. A sample of households was re-interviewed and their tendency to give different answers to the same question was assessed. The following table concerns the answers of a sample of 8596 households to the question: "How many automobiles are kept at home for use by the members of the household".

Census	Re-interview				
	None	One	Two	Three or more	Total
None	1050	230	49	6	1335
One	119	3308	618	81	4126
Two	13	339	1895	248	2495
Three or more	2	32	171	435	640
Total	1184	3909	2733	770	8596

Quality and Performance Indicators

A3. Edit failure rate.

General definition: proportion of responding units for which an error signal is triggered by a specified checking algorithm.

To be further defined for subject-matter domain: the relevant checking algorithms.

Remark. This is an indirect indicator only, since it does not tell how much measurement error remains after editing. However, it provides information about the quality of the initial responses.

***ESS level*:**

Individual values and aggregates of A3 over Member States.

Summary

What should be included on <u>Measurement Errors</u>

- Identification and general assessment of the main risks in terms of measurement error.
- If available, assessments based on comparisons with external data, re-interviews or experiments.
- Information on failure rates during data editing.
- The efforts made in questionnaire design and testing, information on interviewer training and other work on error reduction.
- Questionnaires used should be annexed (if very long by hyperlink)

3.2.5 Nonresponse Errors

Nonresponse is the failure of a sample survey (or a census) to collect data for all data items in the survey questionnaire from all the population units designated for data collection. The difference between the statistics computed from the collected data and those that would be computed if there were no missing values is the ***nonresponse error***.

There are two types of nonresponse:

- *unit nonresponse* which occurs when no data are collected about a population unit designated for data collection, and
- *item nonresponse* which occurs when data only on some but not all the survey variables are collected about a designated population unit.

The extent of response (and accordingly of nonresponse) is measured in terms of response rates of two kinds:

- *unit response rate*: the ratio of the number of units for which data for at least some variables have been collected to the total number of units designated for data collection;
- *item response rate*: the ratio of the number of units which have provided data for a given variable to the total number of designated units or to the number of units that have provided data at least for some data items.

Other ratios are sometimes used instead of, or as well as, these ratios of counts. They are:

- *design-weighted response rates,* which sum the weights of the responding units according to the sample design;
- *size-weighted response rates,* which sum the values of auxiliary variables multiplied with the design weights, instead of the design weights alone.

Mathematical definitions of non-response rates

The American Association for Public Opinion Research (AAPOR) (2008) provides exact definitions of unit and item response rates. However, the weighted variants are not included there. Here slightly more simplified definitions are provided, which also cover the weighted cases. The definitions are, with small modifications, translated from a document in Swedish - Statistikersamfundet (2005).

The sample can be divided into the following categories:

- R: Responding units belonging to the target population; of which
  - F: Responding units (in R) for which full responses were obtained;
  - P: Responding units (in R) for which only partial responses were obtained;
- N: Non-responding units which belong to the target population;
- U: Units with unknown target population status (either nonresponse or overcoverage);
- O: Units not belonging to the target population (overcoverage).

The number of sample units in each category is denoted n_X, with X equal to one of the categorisation letters in the above list.

The total sample size $n = n_R + n_N + n_U + n_O$ and $n_R = n_F + n_P$.

The design weight d_j of unit j in the sample is its inverse inclusion probability. For the size-weighted case value measure of unit j is x_j.

For the units with unknown status, it is assumed that proportion *α* is nonresponse. Unless there are strong reasons to the contrary, it is recommended to set *α=1* which gives a conservative (upper bound to) the nonresponse rate.

The definitions in Table 1 apply for the unit response rates.

Where nonresponse exists, unit response rates thus defined should always be included in the quality report using the most relevant variants (unweighted, design-weighted or size-weighted) in each case. The rates should also be presented for important sub-domains. A breakdown of the nonrespondents into refusals, no contact and other causes is also informative.

Table 1: Definitions of unit response and nonresponse rates

	Response rate	*Nonresponse rate*
Unweighted	$Rr_{uw} = \dfrac{n_R}{n_R + n_N + \alpha n_U}$	$NRr_{uw} = 1 - Rr_{uw}$
Design-weighted	$Rr_{dw} = \dfrac{\sum_R d_j}{\sum_R d_j + \sum_N d_j + \alpha \sum_U d_j}$	$NRr_{dw} = 1 - Rr_{dw}$
Size-weighted	$Rr_{sw} = \dfrac{\sum_R d_j x_j}{\sum_R d_j x_j + \sum_N d_j x_j + \alpha \sum_U d_j x_j}$	$NRr_{sw} = 1 - Rr_{sw}$

For business surveys, size-weighted nonresponse rates are normally the most relevant but it may also be informative to include several measures side by side.

In all definitions of response rates, sampling units identified as overcoverage should neither be included among the respondents nor among the non-respondents. However, it is often informative when presenting the non-response rates to also include overcoverage as a separate category.

The exact definition of response rates (formulas etc.) should normally be included in the quality report along with the numerical information on the rates.

The impact of nonresponse on the statistical outputs is likely an introduction of bias and an increase in sampling error. Sampling error increases simply because the available number of responses is reduced. Bias, which is the main problem with nonresponse, is introduced since nonrespondents are not similar to respondents for all variables in all strata whilst standard methods for handling nonresponse assume they are.

Item nonresponse reporting

For item nonresponse rates there is basically a choice between two reporting approaches, which can also be used in parallel. If the focus is on a particular variable Y, response rates with regard to that variable can be defined as in Table 1 above but with R defined as "responding to variable Y". These rates are the most relevant ones for judging the accuracy of an estimate for variable Y and should be used for all key variables in a survey. They are referred to as *item Y response rates.*

In cases of item nonresponse, there is a choice of explicitly imputing, or not, the values of missing data. Practices regarding imputation should be included in the quality report together with an assessment of their impact on estimates and sampling errors for all data items. (Imputation is further discussed in Section 3.8.3.)

Overall rates of *full response* with regard to all variables are also of interest and their definitions are given in Table 2 below:

Table 2: Definitions on full response and nonresponse rates

	Full response rate	*Full nonresponse rate*
Unweighted	$IRr_{uw} = \dfrac{n_F}{n_R + n_N + \alpha n_U}$	$INRr_{uw} = 1 - IRr_{uw}$
Design-weighted	$IRr_{dw} = \dfrac{\sum_F d_j}{\sum_R d_j + \sum_N d_j + \alpha \sum_U d_j}$	$INRr_{dw} = 1 - IRr_{dw}$
Size-weighted	$IRr_{sw} = \dfrac{\sum_F d_j x_j}{\sum_R d_j x_j + \sum_N d_j x_j + \alpha \sum_U d_j x_j}$	$INRr_{sw} = 1 - IRr_{sw}$

The following provides a numerical example of the above definitions with two variables Y1 and Y2.

Example 3.2.5.A: Numerical example of unweighted response rates

n	Sample size	1000		***Percentage rates***	***Response rate***	***Nonresponse rate***
n_R	Unit response	800		***Unit …***	***83,3%***	***16,7%***
n_{RY1}	Response to variable Y1	700		***Item Y1 …***	***72,9%***	***27,1%***
n_{RY2}	Response to variable Y2	750		***Item Y2 …***	***78,1%***	***21,9%***
n_F	Full response	650		***Full …***	***67,7%***	***32,3%***
n_O	Overcoverage	40				
n_U	Unknown status	30				
α		*1*				
n_N	Unit nonresponse	130				
n_{NY1}	Variable Y1 nonresponse	230				
n_{NY2}	Variable Y2 nonresponse	180				
n_N+ (n_R-n_P)	Full nonresponse	280				

Effects of Nonresponse

Response rates provide an indication of the risk of bias risk but the actual bias depends also (and mainly) on the average differences between the respondents and nonrespondents with respect to survey variables. Normally there is some evidence, although rarely firm, on this matter, which should be included in the quality report in the form of a qualitative assessment.

As previously noted in Section 3.2.1, the increased sampling errors due to nonresponse can and should be taken into account when computing CVs or confidence intervals.

Efforts and measures, including response modelling, to reduce nonresponse in the primary data collection and follow-ups should be described. The technical treatment of nonresponse at the estimation stage (by imputation or by exclusion) should also be clearly stated. Efficient use of auxiliary information can sometimes improve precision considerably in the presence of nonresponse - see Hörngren (1992) for a Labour Force Survey example.

Evaluation

The increase in sampling error due to nonresponse is monitored through the sampling error, as described in Section 3.2.1. The remaining and more difficult issue is how to obtain information on nonresponse bias. The basic approach is to compare the response and nonresponse strata with respect to any variables that are available for both these strata.

Complementing with register data. The method assumes that there is a strong enough correlation between a survey variable for which there is nonresponse and another variable in the sampling frame or another register. This information can be utilised in various ways. For evaluation, one way is to compare the "estimate" of this other variable derived from the whole sample with that derived from the sample excluding non respondents. A small difference provides some indication of a small nonresponse bias for the survey variable as well. The better the correlation is between the two variables, the better, of course, is the judgement that can be made in this way.

Special data collections. These methods aim to show how the nonresponse error would change if it were possible to increase the response rate. The studies are done so that a higher response level is reached than the one achieved with normal effort. For example, more effort can be set aside for tracing, more effort by other staff for persuading refusers to respond, increased time for field work, allowing other collection forms, reducing response burden by concentration on fewer variables or by offering incentives to the respondent. The differences in estimates thus obtained will reflect not only nonresponse error but also measurement and random sampling errors.

Variations over response waves. The purpose of studying responses over response waves is to show how estimates change as a larger share of data collection is accomplished. Results are of interest when intending to publish flash estimates based on data obtained before a certain date. Another use arises in the context of a need, for budgetary or timeliness purposes, to reduce the target response rates and to be able to judge in advance the consequences of such a reduction.

A more controversial use of such studies is to draw conclusions about the remaining nonrespondents based on those that responded in the last wave. Although such an approach can shed some light, further evidence is needed before drawing strong conclusions on bias.

Example 3.2.5.B: Swedish Labour Force Survey (LFS) – Hörngren (1992) and Mirza and Hörngren (2002) .

Hörngren (1992) demonstrates how to use auxiliary information for improving estimates of unemployment in Sweden. The information is taken from an external register, where the employment status of the total population in Sweden is included. The register is annually updated, therefore lagged in relation to the monthly LFS and the employment definition differs from that of the LFS. Nevertheless, the use of this register information for poststratification enables an increased precision in estimates by industry of up to 32 % and, even more remarkably, a large reduction of nonresponse bias.

In Mirza and Hörngren (2002) the estimation system is further refined, using a generalised regression estimator. Djerf (1997) applies a similar method to the Finnish LFS.

This example shows the potential of information, external to a survey for either improving estimates or estimating nonsampling bias.

Example 3.2.5.C:

Särndal *et al* (1992, p. 566) report on the outcome of a mail survey among 3116 fruit growers in North Carolina. Three mailings were carried out (each successive one among those who did not respond to the previous ones) in order to boost response. From independent sources the number of trees per farm was established. In order to assess the similarities between respondents and non respondents and the bias caused by non response the following table was created.

	Number of the mailing				
	1	2	3	Nonresponse	Total
Percentage (%) of returns	10	17	14	59	100
Average number of fruit trees per farm	456	386	340	290	329

It is obvious that respondents and non respondents are not similar; the more trees one has the more likely one is to respond early. Moreover one can see that basing estimation on respondents the average number of trees per farm would be overestimated. If only one mailing was used the response rate would be 10% and the bias 456-329=127; with two mailings the response rate would be 27% and the bias 412-329=83. Even after three mailings, the response rate is 41% and the bias is 388-329=59.

Example 3.2.5.D: Non-response reporting (EU-LFS Quality Report 2005)

Rates of nonresponse by wave. Annual average 2005 (first 9 columns of table)

%	**Waves**							
Country	***1***	***2***	***3***	***4***	***5***	***6***	***7***	***8***
BE	20.6	-						
BG	21.6	16.0	16.2	12.7				
CZ	21.5	19.2	19.3	19.3	19.7			
DK	36.1	35.2	38.5					
DE	-	-	-	-				
EE	31.2	25.9	23.7	20.5				
IE	-	-	-	-	-			
EL	-	-	-	-	-	-		
ES	24.5	18.7	18.7	18.7	18.7	18.7.............		

Division of non-response into categories. 2nd quarter 2005 (first 9 rows of table)

	Non-response (%)			
Country	***Total***	***Refusals***	***Non-contacts***	***Other reasons***
BE	20.8	2.0	9.3	9.8
BG	16.7	3.9	11.9	0.9
CZ	19.8	14.3	4.9	0.6
DK	36.6	-	-	-
DE	-	-	-	-
EE	25.1	7.7	16.5	0.9
IE	9.1	-	-	-
EL	-	-	-	-
ES	19.6	9.0	8.8	1.8

Quality and Performance Indicators

A4. Unit response rate.

General definition: as given above.

To be further defined for subject-matter domain: (i) the choice between weighted and unweighted versions; (ii) rule regarding the proportion α of units of unknown status that is to be considered nonresponse.

A5. Item response rates.

General definition: As given above.

To be further defined for subject-matter domain: (i) the choice of variables for which item response rates are to be provided, (ii) the choice between weighted and unweighted versions; (iii) rule regarding the proportion α of units of unknown status that is to be considered nonresponse.

Remark. It is straight-forward to define an indicator for full response rates as well. This indicator is normally of less interest, however.

ESS level

Individual values and aggregates of A4 and A5 over Member States.

Summary

What should be included on <u>Nonresponse Errors</u>

- Nonresponse rates according to the most relevant definitions for the whole survey and for important sub-domains.
- Item nonresponse rates for key variables.
- A breakdown of nonrespondents according to cause for nonresponse.
- A qualitative statement on the bias risks associated with nonresponse.
- Measures to reduce nonresponse.
- Technical treatment of nonresponse at the estimation stage.

3.2.6 Processing Errors in micro-data

Between data collection and the beginning of statistical analysis, data must undergo processing comprising data entry, data editing (checks and corrections), sometimes coding and imputation. Errors introduced in these stages are called processing errors. Like measurement errors they affect micro-data and evaluations of either type of error tend to involve the other type. Another type of processing error concerns macro-data, as described in Section 3.8.4.

A case where processing error is especially important to evaluate and report is where there is manual coding of response data provided in free text format. This typically occurs when information on occupation or education are requested, for example in a population census. The coding of business activity for inclusion in a business register is another example. The quality of a coding operation depends in a complex way on the coding rules, how they are interpreted in practice and on the downright mistakes committed by the coders.

Processing errors affecting individual observations cause bias and variation in the resulting statistics, just as measurement errors do. The importance of micro-data processing errors varies greatly between different statistical processes and their treatment in a quality report needs to be proportional to their importance. When they are significant, their extent and impact on the results should be evaluated. If such an evaluation has been made it should be included in the quality report.

Evaluation

Studies of effects of editing. The effects of editing are obtained by comparing edited and unedited data. By calculating the final estimates based on both data sets, the total net effect of editing can be measured. These effects can be broken down by unit in a so called top-down list, where the effects by unit are sorted in descending sequence and the most influential units can be seen. Such a list can serve several purposes. One is to check once more that the influential units have their correct values; another is to generate ideas for optimising the editing procedures. For more information on editing procedures including quality aspects the reader is referred to the UN handbook in three volumes: *Statistical Data Editing (UN), Vol 1*, *Statistical Data Editing (UN), Vol 2* and *Statistical Data Editing (UN), Vol 3*

Studies of coding variation. In an independent coding control study the coding is done twice without the coders being allowed to see each other's results. In dependent coding the second coder has access to the first coder's proposals. Dependent coding gives, as expected, smaller variation between the coders. Lyberg (1981) gives an extensive treatment of the topic of coding. High coding variation is of course an indicator of a large potential processing error.

Example 3.2.6.A: Editing in Sweden's panel census of agriculture (Medin and Wilson, 1994)

"After data entry, an editing procedure begins in which data for individual farms are subject to a series of tests for logical or probable errors. In a few cases, fully automatic corrections are made but, most often, indicated errors are handled manually by the staff of Statistics Sweden. In this phase, telephone calls are often made to farmers in order to clarify the situation. The impact of the micro editing on the final statistics diverges greatly between different variables but also between years for the same variable.

A consequence of the panel approach is that structural changes in individual farms have to be carefully investigated and recorded. Hence micro editing is supplemented by a special procedure for checking transfers of agricultural land between farms.

Macro editing is different to micro editing and performed as a kind of final check of all preceding steps in data collection and data processing. Macro editing compares aggregated data of various kinds. Special attention is paid to comparisons with corresponding figures for the preceding year. Major deviations observed are traced back to the individual farms concerned for confirmation."

Example 3.2.6.B: Comparisons of coding processes (Chenu and Guglielmetti, 2000)

"The thoughts presented here rely on two comparisons:

The first is the comparison of codification of the PCS (Profession et categorie socio-professionnelle – occupation and socio-professional category), for persons interviewed twice, once in the Census (semi-automatic coding according to the Colibri process: Codage en Ligne des Bulletins du Recensement Individuel – On line Coding of the population census forms) and once in the 1982 Employment survey ("manual" coding with the help of documentation principally consisting of one guide and one alphabetical index in print).

Then there is the comparison of the coding of the PCS and the variables closely related to it during the 1997 census test from the same forms, processed in a semi-automatic way on one hand and according to a principally manual monitoring process on the other hand.

In the 1982 comparison, the differences between the two coding processes of the PCS result from the differences in the nature of statements that the respondents supplied at the census and the Employment survey (« declaration error »), from the coding carried out by the INSEE (« classification error »), and from the combination of these two. These can be particularly frequent, because 1982 was the year when a whole new socio-professional nomenclature was used in practice for the first time. Furthermore the two procedures are not of the same type: the questionnaire of the Employment enquiry was administered by an interviewer, while the census form was left to respondents and then recollected from them by census agents: the population census is a process that requires the full altruism of the respondents, whilst the context of a survey focused on the employment is called to describe labour or unemployment in detail.

In the 1997 test, the differences result uniquely from the variations in the coding processes used on the same form.

Rate of divergence between the two socio-professional codings

Census and the 1982 Employment survey

Socio-professional group in the 1982 census	Classification in the 1982 Employment survey (%)				Total
	(a) different group	(b) same group different category	(c) same category different profession	(d) same profession	
1. farmers	5	24	28	43	100
2. self-employed	13	7	17	63	100
3. executives	24	10	19	47	100
4. intermediate professions	19	7	21	53	100
5. clerks	13	5	19	63	100
6. workers	9	28	16	47	100
Total	13	15	19	53	100

Semi-automatic process and the manual coding of the 1997 census test

Socio-professional group in the semi-automatic coding	Classification according to the manual coding (%)				Total
	(a) different group	(b) same group different category	(c) same category different profession	(d) same profession	
1. farmers	3	24	12	61	100
2. self-employed	8	6	12	74	100
3. executives	11	10	18	61	100
4. intermediate professions	12	7	12	69	100
5. clerks	8	4	11	77	100
6. workers	5	12	12	71	100
Total	8	8	13	71	100

Quality and Performance Indicators

None explicitly defined.

Remark 1. Indicators of coding errors require some form of repeated coding.

Summary

What should be included on <u>Processing Errors for micro-data</u>

- Identification of the main issues regarding processing errors for the statistical process and its outputs.
- Where relevant and available, an analysis of processing errors affecting individual observations should be presented; else a qualitative assessment should be included.

3.3 For Censuses

The objective of a census is to collect data from all units according to an agreed definition. Three important categories of census are:

- population census - the units are households and individuals;
- economic census - the units are enterprises and local units (a producing unit of an enterprise with a physical address) or other intermediate units (kind-of-activity units, local kind of activity units.)
- agricultural census - the units can be of two kinds – agricultural businesses (farms) and/or land based units.

By definition there is no sampling error in a census but what is said on non-sampling errors in Section 3.2 is relevant also for a census. The error profile of a census may be very different from a sample survey, however, and may also vary greatly depending on type of census and type of approach used. This affects the relative emphasis that should be put in the quality report.

In general the following aspects are known to be of special importance for censuses based on extensive field work.

- ***Undercoverage and overcoverage*** (also referred to as ***undercount*** and ***over-*** or ***double count*** in the census context). The quality report should assess this potential source of error, i.e., that field procedures do not reach all target units or that they reach them twice. A special, deliberate, case of not covering all units arises in the context of a ***cut-off threshold*** as previously described in Section 3.2.2 and below in Section 3.8.1.
- ***Measurement and nonresponse errors*** may well be important. The same assessment and reporting principles apply to censuses as to sample surveys - see Sections 3.2.4 and 3.2.5.
- ***Processing errors*** in the form of data entry or coding errors can be of great importance in a census. Data entry errors may occur when the information is provided by respondents on paper and is data captured either manually or through an optical reading device. Coding is a further source of error, occurring when variables like occupation, education, or economic activity are provided by a respondent in free text format and have to be interpreted by a coder in terms of a pre-determined code structure, as described in Section 3.2.6.

For US censuses there is a huge literature on the associated errors. Bureau of the Census (2002), describing the testing and evaluation program for the 2000 Population Census provides an entry into this literature as well as the Census Evaluation home page itself.

Example 3.3.A: Coverage assessment in the US population Census (Prewitt, 2000)

Prior to April 1, 2001, the Census Bureau will have completed an enumeration of the American population, including a coverage measurement survey, that is designed to improve the accuracy of the initial counts. The coverage measurement survey, called the Accuracy and Coverage Evaluation (A.C.E.), is based on the established statistical method known as Dual System Estimation (DSE) and is designed to correct for missed individuals or erroneous enumerations in the traditional enumeration. The method of Demographic Analysis will also be used to evaluate the completeness of population coverage in Census 2000 at the national level, and to assess changes from previous censuses.

The operations used to produce the apportionment counts are designed with the goal of counting and correctly locating every individual residing in the United States on April 1, 2000, and also to count federal employees and their dependants living overseas as of that date. This goal cannot be completely and accurately realised. Every decennial census, from 1790 to 1990, has included in the census counts some who should have been excluded, and has missed some who should have been included. The first source of error leads to an overcount; the second source to an undercount. Every census for which the effect of these errors has been systematically measured has shown a net undercount – that is, the number of residents who were missed was greater than the number of erroneous enumerations.

The Census Bureau has documented and measured a substantial differential undercount since the 1940 census. After the 1940 census, Census Bureau statisticians and academic researchers refined a statistical technique known as Demographic Analysis, a technique that measures coverage trends as well as differences in coverage by age, sex, and race.

Demographic Analysis uses records and estimates of births, deaths, immigration, emigration, and Medicare enrolments to develop estimates of the population at the national level, independently from the census.

Demographic Analysis, though not without its errors, reveals the persistence of the differential undercount that exists between the Black and the non-Black populations.

The following table illustrates this differential:

Demographic Analysis Estimates of percentage Net Undercount, by Race: 1940 - 1990						
	1940	1950	1960	1970	1980	1990
Total	5.4	4.1	3.1	2.7	1.2	1.8
Black	8.4	7.5	6.6	6.5	4.5	5.7
Non-Black	5.0	3.8	2.7	2.2	0.8	1.3
Difference Black/Non-black	3.4	3.6	3.9	4.3	3.7	4.4

Quality and Performance Indicators

Indicators A2-A5 apply to censuses as well.

Summary

What should be included on <u>Accuracy for a Census</u>

- An evaluation/assessment of undercoverage and overcoverage.
- A description of methods used to correct for undercoverage and overcoverage.
- A description of methods and an assessment of the accuracy if a cut-off threshold is used (see also Section 3.8.1 below).
- An evaluation/assessment of measurement and classification errors.
- An evaluation/ assessment of nonresponse errors.
- An evaluation/assessment of processing errors, especially where manual coding of data in free text format is used.

3.4 For Statistical Processes Using Administrative Source(s)

This is an area where an established theory and concepts are still missing although the recent publication by Wallgren and Wallgren (2007) entitled *Register-based Statistics – Administrative Data for Statistical Purposes* goes a long way towards filling the gap. Although the book's title implies that all administrative data come in the form of registers, there are also other cases.

For the purpose of these guidelines, three types of register-based statistics, without direct data collection, are defined.

- *Estimates produced from one register.* The target population and the variables need to be defined. The tabulations are to be made from this perspective, also the estimation of error properties: a possibly difficult task where the register is updated over time. In particular, lags in updating register units may cause errors in the results, depending on the time when data is extracted from the register for statistical purposes. (See also Business Registers, Section 3.5 below.)

- *Integration of several registers in order to obtain and describe new populations and variables*. This more complex case is treated in some length by Wallgren & Wallgren. For example, population censuses in some Scandinavian countries are currently made in this way. The interested reader is left to consult Wallgren & Wallgren and other, mainly but not only Nordic, sources directly.

- *Event-reporting systems*. Three examples are crime statistics, statistics of road accidents and statistics on the causes of death. Extra-EU external trade, as reported to Customs authorities when goods pass the EU borders, is another example. In these cases the responsible administrative authority (police, hospitals, customs, etc) reports an event, including a number of variables characterising the event. The report is triggered by the event itself rather than by a questionnaire sent by the statistical agency. The events may or may not be entered into a register before being reported to the statistical agency.

Registers, whether for administrative or for statistical purposes, cover all units according to a certain definition. Thus, as for censuses, sampling errors do not exist. Pertinent errors are:

- ***Coverage.*** Over- and under- coverage of eligible units according to the target definition, using also the register definition, should be assessed and reported. Lags in entering information into registers are crucial for understanding the coverage properties of a register. Evaluation approaches regarding these errors have much in common with those mentioned in Section 3.2.3.

- **Nonresponse.** Unit and item nonresponse (missing data) should be assessed and reported.

- ***Differences in concepts***. If target concepts differ from register concepts the effect on outputs of differences should be assessed quantitatively to the extent possible for key indicators.

- ***Errors in register variables.*** For various measurement or processing reasons a register unit may have erroneous value for one or more variables. The cause of the error may be that the value was erroneously provided or miscaptured in the first place or that a later change in the variable has not yet been recorded in the register. The lag structure associated with register updating (see Section 3.2.3) can be analysed in order to throw light on the latter aspect.

The quality report should describe the actions taken with respect to the units and variables originally included in the register - whether they are kept as they are or whether new units and/or values are derived. Models and estimation procedures used should be presented.

A more complex situation occurs for so called multi-valued variables in registers, for example, businesses active in several industries, persons with more than one job, etc. Wallgren and Wallgren give examples of how multi-valued variables can be treated when producing statistics from registers. The approach used should be described and motivated in the quality report.

The procedures for assigning economic activity codes to businesses is of crucial importance register-based business statistics and should be dealt with at length.

Event-reporting Systems

The quality of data from an event-reporting system depends first and foremost on the completeness of the reporting system. The *rate of unreported events* is a key quality factor, although sometimes difficult to estimate. It is a special type of undercoverage error (or relevance problem, if an operational definition excluding non-reported events is used).

Errors in the classifying variables (type of crime, type of accident, type of goods) can best be regarded as a processing error. Approaches to monitor these errors are normally domain related. For example in crime statistics there is an intricate system for coding main crimes and related crimes, counting crimes, etc., that depends upon principles and practices in the area of criminology. A similar situation occurs for cause of death statistics, where the measurement procedures depend upon medical principles and practices.

Quality and Performance Indicators

None explicitly defined.

Remark 1. It is often possible to define quality indicators that are specific to the particular administrative sources used.

Summary

What should be included on <u>Accuracy for a Statistical Process using Administrative Source(s)</u>

- An evaluation/assessment of overcoverage, undercoverage and nonresponse (missing data).
- An evaluation/assessment of errors in classification and other variables.
- For event-reporting systems, an estimate/assessment of the rate of unreported events.

3.5 For Statistical Processes Involving Multiple Data Sources

In many statistical areas, measurement problems are such that a single approach to sampling and measurement is not possible or suitable. For example, for Structural Business Surveys, in which the basic economic data about businesses (production, finance etc.,) are aggregated, different units, questionnaires and sampling schemes may be used for different segments of the survey.

When presenting a quality report for statistical process involving multiple sources, there is a need to focus on the whole picture as well as the segments. An early section of the report should contain an overall description of the organisation of the survey, the various segments, and a summary of the quality aspects. Then, for each segment the critical quality/accuracy aspects should be pointed out. There is a choice of organising the accuracy section of the quality report either by segment or by type of error (all segments together). Normally, at least where segmentation is by different types of units, organising the presentation by segment is suggested.

A segment-wise presentation of accuracy can be combined with a component-wise presentation of the other quality aspects. The choice is ultimately up to the report writer.

Maintenance of a business register can be seen as a statistical process of this kind. In comparison with, for instance, a sample survey, the character of the task is much more of a continuous output. Source data for business registers may come from tax registers, other administrative sources or from special statistical surveys. When statistical outputs, such as numbers and sizes of businesses according to NACE code, are directly drawn from a business register, the quality report must focus on the properties of the business register itself. The error profile of a business register depends on the error profiles of each of the sources and segments. A quality report concerned with direct statistical outputs from a business register should be along the lines described in the previous paragraph. On the other hand, a report on register quality with other uses in mind, especially provision of frames for sample surveys, should embody a more comprehensive and detailed view of quality as, for example, provided in by the UNECE in its Annual Questionnaire on Business Registers (also referred to as the BR Identity Card) comprising a spreadsheet with guidelines for reporting BR characteristics. For further details the reader is referred to this and other documents presented at the International Roundtable on Business Survey Frames.

Example 3.5.A: Waste report Sweden

The statistical process includes 13 surveys and the error profile of each is presented in an Annex. The description and basic approach is also presented by survey in a slightly more abbreviated form. In a final Annex the overall accuracy is assessed.

Quality and Performance Indicators

Indicators A1-A5 above can be used.

Remark 1. CVs (when applicable) are often straight-forward to calculate based on the composite sampling design.

Remark 2. With supplementary specifications other indicators, like overcoverage and nonresponse rates can also be defined.

Summary

What should be included on Accuracy for a Statistical Process Involving Multiple Data Sources

- An overall description of the organisation of the survey, the various segments and a summary of the quality aspects.
- For each segment, the items as specified in the appropriate sections in these guidelines. (These items may be grouped by segment and/or by error type/quality component.)

3.6 For Price and Other Economic Index Processes

Price indexes (CPI, HICP, PPI, SPPI, PPP, Construction and Real Estate price indexes) play an important role in the European Statistical System as well as in all national statistical systems. They are based on statistical surveys and their objective is to monitor price differences in space (PPP) or over time (all others) for all products (goods or services) within their scope, and to provide an overall estimate of price change/difference. In addition there is the Industrial Production Index and other volume indexes that can be seen as a part of a system of economic indexes.

Price, volume (and the less common productivity) indexes are economic indexes for which economic theory and index theory provide a conceptual framework, for which the target concepts are complex.

Price indexes often involve data from multiple sources and are thus a special case of a process in Section 3.5 but their importance merits special treatment. Different measurement approaches are typically used for different types of products. For example, in a CPI/HICP there is a long list of products requiring special approaches (such as cars, PCs, insurance, telecom services, and electricity). A quality report needs to include and assess the approaches for each one of these special products.

Approaches for estimating ***sampling error*** in price indexes have been pioneered by some countries but no generally agreed approach exists. The ILO CPI manual (chapter 5) includes discussions and further references on this topic. It is important to note that there are several sampling dimensions in a price index. In a CPI there is sampling of households (for weights), of products, and of outlets (by type and region). In a PPI there is usually sampling of companies, and of products within companies. In practice probability sampling is not used in all these dimensions and the nature of the purposive procedures are therefore important to describe. A quality report should include a discussion of all relevant sampling dimensions. (See also Section 3.2.2. above on non-probability sampling in general.)

Coverage errors likewise have to be considered in all sampling dimensions. Normally, there are limitations in coverage in all these dimensions that need to be reported and their consequences analysed. Where non-probability sampling is used, the distinction between sampling and coverage error becomes to some extent blurred. In any case a quality report needs to clearly state the limitations in coverage which are the consequence of the frame and sampling procedures used.

A particular and very important source of error in price indexes is ***quality adjustment including replacements and re-sampling*** – the treatment of changes over time in the product universe - between different varieties/models of a product with sometimes different values to the user. Since it is not possible to define the target of estimation exactly, the issue is often stated as a comparability problem, especially in the ESS context (HICP). But strictly speaking, quality adjustment should be regarded as a type of measurement problem. The quality report needs to describe and assess the replacement, re-sampling and quality adjustment methods used for all products, which together determine the way in which price index estimation works. ***Model assumptions*** (e.g. hedonic models), implicit or explicit, are often used in the treatment of quality change and what is said below in Section 3.8.1 is therefore also relevant here.

Non-response and other errors are normally regarded as secondary problems in price indexes. The reason for this can be expressed as an "efficient market hypothesis". No outlet or company could, in the medium or long term, deviate from the rest of the market in its price policy. If this is true then non-response bias is small. In a PPI, a distinction needs to be made between non-response in the first phase (when a company is recruited) and in the ongoing phase (when the company is asked monthly for its current prices). In the latter case there will be a temporarily or permanently missing price. In any case there should be a discussion of the non-response problem for a PPI in a quality report.

Documentation on price indexes has not yet been in the form of quality reports. As far as known, no documents labelled quality report have yet been produced, although documentation with a similar purpose exists.

Ultimately, it will be necessary to issue special standards for quality reports for each of the major price indexes. The above remarks may serve as a starting point for this work.

Example 3.6.A: Handbook of the Swedish CPI

This is basically a methodology description covering all the around 60 different surveys within the CPI system. However, there is only limited direct information on accuracy.

Quality and Performance Indicators

None explicitly defined.

Remark 1. An indicator that has been tried for the HICP is the Implicit Quality Index. See example 3.8.6.A for a detailed presentation.

Remark 2. For the Purchasing Power Parities, the quality report is primarily an issue for the ESS level, since the objective is about comparisons between countries.

What should be included on Accuracy for Price or Other Economic Index Process

- Information on all sampling dimensions (for weights, products, outlets/companies, etc).
- Any attempt at estimating or assessing the sampling error in all or some of these dimensions.
- Quality adjustment methods (including replacement and re-sampling rules) for at least major product groups.
- Assessment of other types of error, where they could have a significant influence.

3.7 For Statistical Compilations

At the top level of national and European statistical systems are economic and other aggregates that are compiled from basic statistics from a variety of different sources and that concern various aspects of the economy, society and the environment. This section discusses quality reporting for such statistical compilations.

The most well known compilations are economic aggregates, of which the best known are the National Accounts and the Balance of Payments. (A longer list is provided by Statistics Canada.) Analysing and reporting the accuracy of these economic aggregates is extremely difficult since they involve many diverse sources. It is necessary to take a very different approach than for sample surveys. Sections 3.7.1 and 3.7.2 present examples for the National Accounts and the Balance of Payments. Other statistical compilations are briefly discussed in Section 3.7.3.

3.7.1 National Accounts

There are a variety of approaches to the assessment and reporting of the accuracy of the National Accounts as illustrated in the following paragraphs.

The International Monetary Fund (IMF) has developed its Data Quality Assessment Framework (DQAF). The first three levels of the framework are generic and the same for any statistical process, the final two levels are process specific. There is a version devoted to measurement of the quality of the National Accounts. The aspects of quality covered by the framework are: (i) integrity; (ii) methodological soundness; (iii) periodicity and timeliness; (iv) accuracy; (v) revision practice and policy; (vi) stability; (vii) consistency; and (viii) accessibility. This approach shares many aspects with an approach based on the principles in the European Statistics Code of Practice, where the same criteria can be found.

In the UK, a more direct way to address accuracy has been pioneered. Illustrative extracts from the Blue Book on the UK National Accounts are provided below.

Example 3.7.1.A: Blue Book on the UK National Accounts, 2007 Edition, page 27

Accuracy and reliability

One key aspect of quality for many users is accuracy. National Statistics strives to publish timely, consistent, and coherent estimates of GDP that accurately represent productive activity in the economy. The basis of these estimates is strengthened by the inter-relationships within the system, and the subsequent requirement that the many (and often independent) data sources are internally consistent. However, it remains very difficult to comment on the accuracy of GDP.

Estimates of GDP are built from numerous sources of information, including business surveys, household and other social surveys, administrative information and survey data from HM Revenue & Customs. Data is collected monthly, quarterly, annually and in some cases from ad hoc surveys. Some of the resulting estimates that feed into GDP will be firmly based whilst others may be weaker.

Assessing the accuracy of an estimate involves assessing the errors associated with that estimate. Sampling errors can be calculated for estimates derived from random samples. At present, sampling errors are calculated for several surveys that feed into GDP, but for other surveys there remain technical problems to be solved before reliable estimates of error can be formed. A program of work is currently underway which will lead to the publication of sampling errors for all major ONS business surveys.

In addition to sampling errors, accuracy is also affected by nonsampling errors such as limitations in coverage and measurement problems. Though there is limited information about non-sampling errors it is likely that for some surveys non-sampling errors are the more important source of error. Data validation by survey statisticians, additional consistency checks and the inclusion of coverage adjustments where survey sources are known to have shortcomings reduce non-sampling error and improve the quality of the accounts.

Even if the sampling and non-sampling errors of all individual data sources were known, the complexity of the process by which GDP is estimated is such that it would be difficult to build up an overall estimate of accuracy from the component series. The process of bringing together the three approaches to GDP into one measure, which uses detailed supply and demand balances, brings in extra information about the accuracy of the raw data and its consistency with other sources. This adds significantly to the accuracy of the overall estimate of GDP, but this cannot be measured scientifically.

One alternative approach to measuring the quality of GDP estimates is to use evidence from analyses of revisions to growth rates, outlined below. The purpose is to assess the reliability of GDP estimates, referring to the closeness of early estimates to subsequently estimated values.

In summary a direct approach for measuring accuracy is not considered possible. There are two main instruments for analysing accuracy. The first one is an *analysis of revisions*. Revisions show the degree of closeness of initial estimates to subsequent or final estimates. Since all estimates are affected by error, this type of analysis can not definitively demonstrate the accuracy of initial estimates. But clearly the amount of revision is still an indicator of accuracy, since it is reasonable to assume that estimates are converging towards the true value as estimates are based on more and more reliable sources.

The UK blue book provides such analyses on pages 28-29. In the UK case there are six stages of estimation and publication. The presentation focuses on average revisions between two successive stages and between a certain stage and the last stage.

A second source of information on National Accounts accuracy is the so called *statistical discrepancy*. The following extract from the UK ONS Blue Book gives the idea.

Example 3.7.1.B: Blue Book on the UK National Accounts, 2007 Edition, pages 29-30

Further assessment of the reliability of the consolidated economic and sector accounts can be gained by examination of the capital and financial accounts – which should, in theory, show a balance between the net lending/borrowing in the capital account and financial account for each sector. However, because of errors and omissions in the accounts, such a balance is rarely achieved. The resulting statistical discrepancy items required to equate these accounts are shown in this publication.

These discrepancies provide a measure of reliability as they reflect errors and omissions in the accounts. Some components of the accounts (for example, estimates for general government) provide excellent coverage and are very reliable whilst others (for example life insurance and pension funds) are less fully covered. A detailed table, which looks specifically at the reliability of components of the sector financial accounts, is published in *Financial Statistics Explanatory Handbook*. However, because of the many sources of information that feed into the economic accounts it is not possible to generalise these 'reliability measures' to the aggregate estimates.

Meader and Tily (2007) further discuss UK National Accounts quality. They regard the most important tools for monitoring the accuracy and coherence of quarterly GDP growth estimates to be:

- internal coherence – the analysis of published adjustments (alignment adjustments and statistical discrepancies) as well as unpublished adjustments; these three measures together contribute to understanding coherence within the GDP data set;
- wider coherence – measures that indicate the degree of coherence between GDP and other ONS and external sources;
- sources – the monitoring of the quality of source data that feed into GDP. While the above three measures concentrate on GDP output, this one looks at the accuracy of ONS surveys and administrative information.

Fixler and Grimm (2007) make a similar analysis for the US National Accounts.

A key problem for the National Accounts is the non-observed economy, i.e., that part of the economy that is not covered by the usual administrative and survey sources. Measurement of the Non-Observed Economy: A Handbook (OECD et al) provides guidance on measurement of the non-observed economy as included in, and as missing from, the National Accounts.

In summary, reporting accuracy for National Accounts needs quite a different approach to that used for other statistical processes.

3.7.2 Balance of Payments

The Balance of Payments, like the National Accounts, is compiled from a wide range of administrative and statistical sources providing data on trade in goods, trade in services, capital flows, etc., and involves the same difficulty in evaluating accuracy.

The legal requirement for quality reporting is included in the Balance of Payments Regulation but there is no technical guidance on the content of a quality report. No national quality reports are published in the public sphere.

At the ESS level, however, there is the ECB Balance of Payments Quality Report, which follows the basic principles of the International Monetary Fund (IMF) Data Quality Assessment Framework. The section entitled accuracy is very brief and, as in the case of the National Accounts, the main approach is through revisions analysis.

3.7.3 Other Compilations

Environmental accounts provide an example of statistical compilation outside economic statistics. However, there are no established standards for such accounts, thus it is premature to formulate quality reporting guidelines for them.

A special case of great importance is that of statistics on *Greenhouse Gas Emissions (GGE)*, for which detailed international guidelines are provided by a UN institution. GGE statistics are compiled from a large number of national and international reports of anthropogenic emissions and removals of greenhouse gases. Instructions on managing uncertainties are included in an annex to the guidelines. There is nothing that resembles a quality report of the kind discussed in this document but there is a special chapter on *Gaps in Knowledge* that has some aspects in common with a quality report.

Quality and Performance Indicators

See relevant manuals.

Remark 1. The main indicator of accuracy for economic aggregates is revisions. See Section 3.8.5 below.

Summary

What should be included on Accuracy for a Statistical Compilation

- Information and indicators relating to accuracy for example as defined in the IMF's Data Quality Assessment Framework (DQAF) or other relevant, well accepted standard.

- Analysis of revisions between successively published estimates.

For National Accounts

- Analysis of the causes for the statistical discrepancy.
- Assessment of non-observed economy.

3.8 Some Special Issues Concerning Accuracy

There are several issues in the reporting of accuracy that are not specific to the type of statistical process. They are discussed in the following paragraphs.

3.8.1 Model Assumptions and Associated Errors

Models are often applied in statistics. Sometimes the target of estimation relies on an abstract model defined by a subject matter discipline. In other cases, such as seasonal adjustment, which is treated in the next section, the model is of a purely mathematical-statistical nature. Sometimes a model is applied in estimation in order to improve precision.

Model-assisted estimation (in the sense of Särndal et al, 1992) is the first case. Here models are only used for the purpose of reducing sampling error as defined by the design-based paradigm. Sampling error calculated according to the relevant variance estimation formulas is sufficient and no separate discussion of model assumptions is needed in the quality report. If the basic design-based estimation is extended to adjust for non-sampling errors, such as nonresponse, a description should be provided.

Model-dependent estimation is a different matter. In this case there are no design-based estimators to use and the inference depends on the model, whose assumptions need to be critically checked. When model-dependent estimation is used as a remedy for a particular non-sampling error (like nonresponse or measurement errors) the discussion of the model should be in the relevant error section rather than in a separate section. A similar case is where models are used for a sample or a census with cut-off (discussed in Section 3.2.2).

In yet other cases even the *target of estimation is model-based*. The model is then usually developed by a domain related science. Natural science models are used in environmental statistics, medical models for some parts of health statistics and economic models for concepts in economic statistics such as productivity and inflation. (The National Accounts system is an economic model.) In such cases, the model should be described in the quality report and its validity for the data at hand assessed. Whether to do this in a general section on methodology or in a section on model assumptions is a matter of choice in each case.

Example 3.8.1.A:* *Healthy Life Years Expectancy

This is a case where the target of estimation is model-based (approved by the World Health Organisation). This document describes the method for calculation, but is not a full quality report, since, for example, no discussion of accuracy is included.

Example 3.8.1.B: [Greenhouse Gas Emissions]

This is another example, where the target of estimation is model-based (and very complex). It was developed by the UN Intergovernmental Panel on Climate Change.

Example 3.8.1.C: [Foreign Trade Statistics]

This is a case where model-dependent estimation is used for a relatively simple target concept. A model is used for estimating the part of trade that is below a threshold (a census with cut-off). The actual method used is different in different Member States. In the document the effects of the estimation are shown in Table 3 and the thresholds used are given in Tables 1 and 2. Strictly, the example belongs to Section 3.3.

Quality and Performance Indicators

Model dependent.

Summary

What should be included on Model Assumptions and Associated Errors

- Models related to a specific source of error should be presented in the section concerned. This is recommended also in the case of a cut-off threshold and model-based estimation.
- Domain specific models, for example, as needed to define the target of estimation itself, should be thoroughly described and their validity for the data at hand assessed.

3.8.2 Seasonal Adjustment

ESS guidelines on seasonal adjustment (EGSA) have been adopted. Their implementation will enhance quality of seasonally adjusted figures as well as enforce the robustness and reliability of European aggregates.

For statistical processes involving seasonal adjustment, a quality report should include a section on this topic. Where full documentation exists in other places and/or the metadata template provided in the EGSA has been compiled, a reference to the relevant documents can be made and a brief summary given. The following points then constitute a minimum:

- A short description of the method used, including pre-treatment (calendar effects corrected for, calendar used, type of outliers detected and corrected, model selection and revision and decomposition scheme adopted) and specification of the seasonal adjustment tool chosen (software and version);
- Validation: specification of the quality measures and diagnostics used to evaluate the appropriateness of the identified model and the results of the seasonal adjustment process.
- Revisions: approach chosen for handling revision of seasonally adjusted data in combination or not with revision of raw data (specification of the horizon of revision seasonal factors).

In case no other documentation is available a full presentation of the process applied and of the methodological choices made with respect to each item of the EGSA (in particular for pre-treatment, seasonal adjustment, revision policies, quality of the seasonal adjustment process) should be included in the quality report.

Quality and Performance Indicators

None explicitly defined.

Summary

What should be included in the Quality Report on <u>Seasonal Adjustment</u>

- A short description of the method used.
- A report on quality aspects in line with the ESS guidelines on seasonal adjustment.

3.8.3 Imputation

Imputation is a response to deficiencies in the data received. In a sample survey or census the reasons for imputation could be nonresponse (usually item nonresponse) or to correct values affected by measurement or processing errors. In price index processes, imputation may occur due to temporarily missing prices.

The extent to which imputation is used, the reasons for it, and the imputation procedures should be described in the quality report. Where imputation is associated with a particular source of error, it is best to include its description under the relevant heading (for example nonresponse or measurement error).

Imputation is a part of data processing and thus may itself cause processing error. Normally this can be assumed to be a minor problem compared with the error sources that created the need for imputation in the first place and, if so, need not be dealt with explicitly.

Imputation can also affect the calculation of sampling error. In particular if imputation based on replacement by stratum mean is used the result will be to introduce some underestimation of the real sampling error. This should be noted where sampling errors are presented unless special methods have been applied to deal with this.

Quality and Performance Indicators

A6. Imputation rates.

General definition: "Imputation is the process used to determine and assign replacement values for missing, invalid or inconsistent data that have failed edits. This is done by changing some of the responses or assigning values when they are missing on the record being edited to ensure that estimates are of high quality and that plausible, internally consistent record is created." Statistics Canada (2003).

To be further defined for subject-matter domain: the relevant imputation algorithms.

Remark. This is an indirect indicator only since it only shows to what extent imputation is done but it does not tell the effects of imputation on the estimates.

ESS level

Individual values and aggregates of A6 over Member States.

Summary

What should be included in the Quality Report on <u>Imputation</u>
▪ Information on the extent to which imputation is used. ▪ A short description of the methods used and their effects on the estimates. (Typically this information will be reported in the section(s) dealing with the errors that imputation is helping to correct rather than in a separate section.)

3.8.4 Mistakes

There are two very different kinds of processing errors. The first type, which has already been discussed in Section 3.2.6 concerns *micro-data*. The second type concerns *macro-data* and involves *serious mistakes in calculation or presentation of aggregates that are not found until after publication*. Mistakes affect all types of statistical process in essentially the same way. They are the errors most visible to the public, typically receiving a lot of negative attention. Examples are when the methodology is not applied correctly, when the wrong number is inadvertently included in a press release, and when analytical presentations or diagrams give wrong impressions. They may occur in any stage in the production of statistics: programming, calculation, report writing, editing of manuscripts, etc. The type and number of mistakes that have been officially recognised and have resulted in unplanned revisions should be presented for several years back.

Procedures to minimise the risk of gross mistakes in calculation or presentation should be described in the quality report. Policies for handling the situation when mistakes are discovered should be presented as well.

Quality and Performance Indicators

A7. Number of mistakes made, by type.

General definition: as given above.

To be further defined for subject-matter domain: (i) a classification of mistakes into the most common/crucial types; (ii) time span over which mistakes are to be summarised and presented.

Summary

What should be included on <u>Mistakes</u>
• The number and nature of mistakes over the past few years should be described. • Measures taken to avoid mistakes in the future should be described.

3.8.5 Revisions

Revisions can be planned or unplanned. Unplanned revisions are usually caused by the discovery of a mistake in a published result as discussed in Section 3.8.4. This section deals with planned revisions.

The ESS Code of Practice requires that a revision follows standard, well-established and transparent procedures. This means for example that pre-announcements are desirable and that the reasons for and nature of the revision (new source data available, new methods, etc.,) should be made clear. Whether this is the case should be stated in the quality report.

Revision practices vary greatly between countries and, especially, between statistical processes. The quality report should first state the relevant revision policy, if there is one, and then present the actual practice. The statement should detail the variables and domains that are subject to revision and the pattern of successive releases.

The quality report should also include information on the size and direction of revisions, and their spread (standard deviation) based on historical data. This information should cover all key indicators.

Measuring the size and direction of revisions can be done in a number of ways. Consider the case where a key indicator, X, referring to a specific time period is published k times at pre-defined time points. Often k is 3. The first publication may be referred to as the *initial estimates,* the subsequent publication as the *revised* estimates and the last publication as the *final* estimates. In the general case, for a given reference period, k different values $\{X_1, X_2, \ldots, X_i, \ldots, X_k\}$ of X will be available, X_k being the final estimate.

The i^{th} estimate may be compared with the final estimate (which is likely to be most accurate) taking the difference $(X_i - X_k)$. The ***revision size*** from the i^{th} estimate relative to the final, k^{th}, estimate is better defined in the converse order (so that, for instance, the revision size has a negative sign when the revision is downwards):

$$R(i,k) = X_k - X_i,$$

and the corresponding relative revision size (normally expressed as a percentage) as

$$r(i,k) = \frac{R(i,k)}{X_i}.$$

Also of interest are revisions between successive estimates, namely:

$$R(i,i+1) = X_i - X_{i+1} \text{ and } r(i,i+1) = \frac{R(i,i+1)}{X_i}$$

Functions of these statistics can serve as informative quality indicators for the i^{th} estimate and also as predictors of the subsequent estimates (i+1, i+2,…, k). Such functions can, for example, be averages or standard deviations over a time series of published estimates of *R(i,k), r(i,k), R(i,i+1), r(i,i+1)*, or the corresponding absolute quantities for different successive estimates *i*. The most suitable indicators in each case should be chosen, but it is difficult to give general advice on this point.

For comparability over time, the *k* time lags and the length of the time-vector should be the same in each year. However, this is not a precondition for the derivation of the indicator.

The indicator described above is a simple descriptive statistic aiming at assessing the size of revision. More sophisticated approaches (detection of biases and correlations, i.e., the detection of persistent patterns in the revisions) can be found in the report of the joint ECB and Eurostat task force report on the quality of balance and payments and international investment position statistics [ECB and Eurostat, 2003].

Example 3.8.5.A: ECB Balance of Payments Quality Report, Annex 1 and chart 7, page 15

Chart 7 in the documentation gives a diagrammatic presentation of revisions. Annex 1 provides a large set of possible indicators for revisions including their detailed mathematical definitions.

Example 3.8.5.B: Blue Book on the UK National Accounts, 2007 Edition, pages 28-29

Mean revisions according to definitions above are presented in tables on page 29 of this document. Documentation on source data etc. for revisions is provided, integrated with the methodological description by sector.

Quality and Performance Indicators

A8. Average size of revisions.

General definition: average or average absolute revisions as defined above.

To be further defined for subject-matter domain: (i) the data elements for which the indicators should be published; (ii) the choice of successive estimates *i* for which the quality indicators should be calculated.

Remark 1. Percentage errors facilitate cross-domain and cross-Member State comparisons.

Remark 2. Percentage errors are not suitable for data elements that are themselves percentages. Firstly, there is a risk of misunderstanding in these cases. Secondly, when these percentages refer to changes that could be zero or negative the definition breaks down.

ESS level

The above applies equally to revisions of European level data.

Revision policies and patterns in all Member States should be summarised with the main focus on how they affect published data at the European level.

Summary

What should be included on Revisions

- The revision policy.
- The average number of revisions (planned and unplanned).
- The average size of revisions (one or more measures).
- The main reasons for revisions, and the extent to which the revisions improved accuracy.

3.8.6 Subject-dependent Techniques for Evaluation

For any particular type of statistical process there are unique opportunities for error checking and evaluation. This section gives a few examples, mostly to inspire producers to invent other methods, similar or not, that are suitable for their particular statistical processes. Creativity is certainly a virtue in this field.

Mirror statistics. The classical example of mirror statistics is for foreign trade. In principle, Country A's exports to country B over a certain period must equal country B's imports from country A. In practice, the comparison is blurred by a factors such as valuation (cif/fob), timing (arrival at B may be later than departure from A), and classification differences. However, adjustments for these factors can usually be made so that the extent of the actual errors can be more or less accurately determined.

Another case where mirror statistics can be of use is for statistics on migration.

Unexplained variation over time in event-reporting. In event-reporting statistics, there is normally some stability in reporting patterns from the relevant authorities (police, hospitals, customs, etc). Lags in reporting or failure to report by a particular local institution cause undercoverage. It is simple to keep track of the reports from each institution subject to the reporting obligation. If this is done irregularities in the number of reports give rise to suspicion that something is wrong and corrective action can be taken.

Reasonability arguments. In all statistics, subject-matter knowledge on what is possible and reasonable is a useful tool. Often, all that is needed is a creative use of common sense. A more intricate example of such an argument is used in price statistics as described in the following example.

Example 3.8.6.A: A control statistic based on a reasonability argument.

For a certain product in a Consumer Price Index, one could compute a raw average price for all observations in any given month. The ratio between such average prices between two months could be called the raw price index, which will differ from the actual price index due to implicit or explicit quality adjustments. Now the following statistic can be calculated

IQI=Implicit quality index =(raw price index)/(actual price index).

If the quality adjustments are correct and the IQI shows an increase of 10 %, then this implies that there has been a 10 % quality improvement in the product concerned. This could then be tested against the general consumer experience, which may for example be that quality improvements have occurred for high-tech goods (PCs, cars, TVs, stereos etc.) but not for non-technical goods such as clothing and household utensils.

4 Timeliness and Punctuality

ESS Quality Definition

The ***timeliness*** *of statistical outputs is the length of time between the event or phenomenon they describe and their availability.*

Punctuality *is the time lag between the release date of data and the target date on which they were scheduled for release as announced in an official release calendar, laid down by Regulations or previously agreed among partners.*

Timeliness is relatively easy and straightforward to measure. A common measure is the production time measured from the end of the reference period or point to which the data refer to the day of release, averaged over a number of process implementations. The maximum production time is also useful by providing the worst recorded case. Average timeliness is meaningful for releases that are annual or more frequent.

Presentation of punctuality is likewise simple. The most relevant measure is the percentage of releases delivered on time, according to scheduled release dates laid out in Regulations, official timetables or other agreements. Such percentages are meaningful for releases that are annual or more frequent.

Some statistics are released in several versions, for example preliminary, revised and final). In this case each release then has its own timeliness and punctuality profile. The releases should be distinguished and separately presented in the quality report.

Where quality standards have been set up in domain specific regulations and the like, they can be used for benchmarking, for example by taking the ratio of, or difference between, the actual production time and the specified standard production time.

The reasons for possible long production times and non-punctual releases should be explained and efforts to improve the situation described.

ESS level

Two aspects of timeliness and punctuality should be dealt with:

- National data deliveries to the ESS. The agreed time frame for deliveries should be included as well as the achieved dates for deliveries during a past period. Where there are several stages of publication (e.g., preliminary and final results) all should be included.
- Publications from the ESS to the public. This should follow the same pattern as for national reporting, so the guidelines above are applicable here as well.

Example 4.A: The Quality Indicators Panel set up at INE Portugal (Zilhao, 2001).

"As stated in the INE-P's Quality Charter, a Quarterly Quality Indicators Panel (QIP) was designed and started to be implemented in 1996, comprising indicators concerning suggestions/complaints received, services rendered (offers, demands and supplies of statistical data), and **timeliness of the dissemination of statistical data**, for which quality standards are defined.

These indicators measure the lag between the performance and the target.

The aim of the QIP is to assess the performance in order to create a systematic basis for monitoring timeliness of statistical data - it is a management tool integrated in the Quality Management System and it is disseminated on INE-PT's Intranet.

Quality Indicators Panel results.

The QIP provides information on the performance of 31 publications and 122 statistical projects. Examples of QIP results for some publications and data in 2000 are given below.:

Publications Availability 2000

Publication	**Reference period**	**Publication availability term**	**Quality standard**	**Deviation (days)**
Business Statistics	1998	01.01.1999 to 19.07.2000	18 months	19
North Region Yearbook	1999	01.01.2000 to 31.07.2000	7 months	**0**
International Trade	1999	01.01.2000 to 28.06.2000	6 months	–2

Data Availability 2000

Statistical Project	**Reference period**	**Data availability term**	**Quality standard**	**Deviation (days)**
Animal husbandry Statistics	1999	01.01.2000 to **14.06.2000**	5 months	14
Consumer Price Index (monthly)	Month n	10th Working day month n+1	10th Working day month n+1	**0**
Labour Cost Index (Trimestrial)	4th trimester of 1999	01.01.2000 to **17.01.2000**	Trimester n + 1 month	-14

"

Example 4.B: Timeliness in Structural Business Statistics (Eurostat, 2001b).

Article 9 of the Regulation foresees that *"the results shall be transmitted in an appropriate technical format and within a period of time from the end of the reference period which shall be no longer than 18 months."*

Besides, Article 11 foresees that *"during the transitional period, derogations from the provisions of the Annexes may be accepted in so far as the national statistical systems require major adaptations."*

During the transitional period, from the reference year 1995 to the reference year 1998 (included), derogations have been granted to Member States. Eurostat accepted that, during that transitional period, some data be transmitted up to 24 months after the end of the reference year, that is 6 months after the deadline foreseen by the derogation.

However, the reference year 1999 corresponds to the first reference year without any derogation with a full entry into force of the Regulation. The main SBS series are expected before July 2001, even if some data might not be sent on time to Eurostat despite the end of the transitional period.

Main series of the SBS Regulation

Delays from regulation deadline for main series of the SBS Regulation

(Expressed in months after the regulation deadline (T+18)).

	Reference year 1996			Reference year 1997			Reference year 1998		
series	Mean	Median	Max	Mean	Median	Max	Mean	Median	Max
1A	6	7	10	4	3	12	2	1	12
2A	5	5	10	4	3	12	2	2	12
3B	6	6	10	5	3	12	3	2	12
4A	5	6	10	5	3	12	3	2	12

1A = annual enterprise statistics (services)

2A = annual enterprise statistics, detailed module for Manufacturing (NACE Rev.1 section D)

3B = annual enterprise statistics, detailed module for Trade (NACE Rev.1 section G)

4A = annual enterprise statistics, detailed module for Construction (NACE Rev.1 section F)

There is evidence that the delays for data transmission have dramatically decreased during the transitional period. However, one big country has not provided yet definitive data for the main series for reference year 1998.

Preliminary results

The Regulation foresees the transmission to Eurostat of preliminary results 10 months after the end of the reference year for some characteristics. These data have been actually sent to Eurostat for reference year 1999, and were disseminated at the very beginning of February 2001.

* First assessment of SBS data three months after the end of the reference year

Eurostat estimates SBS data three months after the end of the reference year (T+3) by using short-term indicators. The population covered is Manufacturing (NACE Rev.1 section D) and the level of detail is the NACE Rev.1 2-digit level. For instance, it is intended to disseminate in the next few weeks first SBS estimates for reference year 2000.

Of course, it is possible to assess *ex post* the quality of these assessments by using final SBS results. It appears that the quality of these first assessments (that we call "nowcasts") is satisfactory at 2-digit level for EU totals. Indeed, the gap between the assessment at T+3 and the final results are very often lower than the range of the confidence interval for the final results, as it is calculated from the EU coefficients of variations."

Quality and Performance Indicators

T1. Time lag between the end of reference period and the date of the first/provisional results.

General definition: the number of days from the last day of the reference period to the day of publication of first results.

T2. Time lag between the end of reference period and the date of the final results.

General definition: the number of days from the last day of the reference period to the day of publication of final results.

ESS level

(i) functions (average, maximum) of national T1 or T2 data, (ii) T1 or T2 for ESS level publications.

T3. Punctuality of publication.

General definition: number of days separating a previously announced date of publication and the actual date.

To be further defined for subject-matter domain: (i) the unit of time to use; (ii) the most appropriate function for a number of publication instances.

Remark. Requirements for punctuality vary widely for different types of statistics. Market-sensitive economic indicators are often published at an exact pre-announced date and time, and any delay or early disclosure is a severe drawback. A possible definition of the indicator in such a situation could be *rate of instances where the effective publication was more than one minute early or late.* In other less sensitive cases, the *rate of instances where the preannounced day of publication was missed* along with the *average delay in number of days* may be the most appropriate indicators.

Summary

What should be included on Timeliness and Punctuality

- For annual or more frequent releases: the average production time for each release of data.
- For annual or more frequent releases: the percentage of releases delivered on time, based on scheduled release dates.
- The reasons for non-punctual releases explained.

5 Accessibility and Clarity

ESS Quality Definition

The ***accessibility*** *of statistical outputs is the measure of the ease with which users can obtain the data. It is determined by the physical conditions by means of which users obtain data: where to go, how to order, delivery time, pricing policy, marketing conditions (copyright, etc.), availability of micro or macro data, various formats (paper, files, CD-ROM, Internet, etc.)*

The ***clarity*** of statistical outputs *is the measure of the ease with which users can understand the data. It is determined by the information environment within which the data are presented, whether the data are accompanied with appropriate metadata, whether use is made of illustrations such as graphs and maps, whether information on data accuracy are available (including any limitations on use) and the extent to which additional assistance is provided by the producer.*

In summary, accessibility and clarity refer to the simplicity and ease with which users can access statistics, with the appropriate supporting information and assistance.

Evaluation of accessibility can take a range of forms as accessibility is affected by the many aspects of dissemination practice, including:

- the dissemination channels;
- the form of the outputs - microdata or aggregates; and
- the pricing policies.

The quality report should include a description of the various ways the statistical outputs can be accessed - in paper publications, through the Internet, etc. Pricing policies and their likely effect on access should be described together with the limits on access set by confidentiality provisions and any other restrictions.

Clarity depends upon the quality of statistical metadata that are disseminated alongside the statistical outputs. A summary description of these metadata (documentation, explanation, quality limitations, etc.,) should be included in the quality report.

Vale (2008) makes a number of useful points regarding both accessibility and clarity based on a division of users into occasional users ("tourists") and more experienced, professional users ("harvesters" and "miners"). This division appears to be helpful especially for web-based publishing. "Tourists" typically prefer data in static formats so that they are easy to find and interpret. Therefore quality assessments for this group of users should focus on ease of access and search and on simple and clear presentation of data and accompanying metadata. "Harvesters" and "miners", however, have rather different needs. Typically they prefer a database approach to statistical dissemination where they can select and download just those data that are of interest to them, sometimes for further data manipulation and analysis.

The quality report should normally refer to the needs of each of these kinds of users and how well they have been addressed.

User feedback appears to be the best way to assess the clarity of published data from the user's perspective. Questions on user experiences regarding ease of access to the data and their exact meaning and interpretation should be included when user satisfaction surveys are designed, and this and any other user feedback should be reported.

Recent and planned improvements to accessibility and clarity should also be described

ESS level

The above applies equally to the ESS level.

Example 5.A: Accessibility and clarity for the European Community Household Panel (ECHP), from Eurostat Internal Quality Report on ECHP (version of December 1998)

"**Accessibility of the ECHP (= forms of dissemination)**

Data can be found in the following sets:

Users' Data Base

For confidentiality reasons, a Users' Data Base, which meets various "objective anonymisation criteria", is produced from the Production Data Base (PDB).

Provided that all these criteria are met, the ECHP data can be considered "non-confidential" in the sense of the Statistical Law and therefore made more widely available, although *direct access to such anonymised micro-data needs to be restricted by means of contracts stipulating the strict conditions of use*.

In December 1998, Eurostat produced a CD-ROM containing the anonymised Users' Data Base, including a very detailed documentation. Interested persons and organisations can acquire this CD-ROM in the context of research contracts stipulating the strict conditions of use. These prohibit any attempt to identify individual units or any claim by the user to have done so.

Redistributing the data to unauthorised third parties is also not permitted. Various sanctions and disincentives are put in place to discourage breaches of contract.

National Data Collection Units, as part of the network collecting ECHP data, receive the Users' Data Base for free.

Publications by Eurostat:

Statistics in Focus (4- to 8-page reports aimed for the public at large):

1997/5 "Family responsibilities: How are they shared in European households?"

1997/6 "Income distribution and poverty in EU12 - 1993"

1997/9 "Housing conditions in the Europe of Twelve in 1994"

1997/12 "Self-reported health in the European Community"

1998/6 "Low income and low pay in a household context (EU-12)"

1998/11 "Analysis of income distribution in 13 EU Member States"

1998/12 "Lone-parents: A growing phenomenon"

Horizontal publications:

"Youth in the European Union: From education to working life" (1997)

"Statistical Yearbook" (1998)

"Europe in figures" (1998)

"Social Portrait of Europe" (1998)

“Facts through figures” (1998)

Methodological documents:

“The European Community Household Panel (ECHP): Volume 1 - Survey methodology and Implementation”, Theme 3, Series E, Eurostat, OPOCE, Luxembourg, 1996.

“The European Community Household Panel (ECHP): Methods, Volume 1 - Survey questionnaires: waves 1-3”, Theme 3, Series E, Eurostat, OPOCE, Luxembourg, 1996.

Publication of results

A publication presenting 1995 wave results is foreseen during the first semester of 1999. Annual publication tabulating the main results is envisaged thereafter.

Technical Documentation

Full technical documentation on ECHP methodology is available in the form of "PAN documents".

Clarity of the ECHP

Contents of accompanying documentation

In addition to the various documents presenting the results, Eurostat prepares also technical documents, which are finalised after presentation for discussion at the ECHP Working Groups. To date, over 100 PAN documents have been completed. These cover diverse topics such as survey definitions and draft questionnaires; guidelines on survey organisation, sampling, data collection and quality control; follow-up rules, computation of response rates and other aspects of panel implementation; technical specification of the weighting and imputation procedures, etc.

In addition, full detailed documentation on the Production and the Users' Data Bases has been produced. Data alerts pointing out possible problems with the data set are issued from time to time to facilitate the work of data users.

Information services

Eurostat offers the following information services to users:

Help users to better know, understand and analyse the statistics available in the ECHP and their potentialities. This includes many meetings with users inside National Statistical Institutes and outside the Commission, the follow-up of a few TSER projects (including the membership of Advisory Groups), etc.

Meetings with statistical users in order to better know what their needs are, e.g. with a view to building up new derived variables.

Computation of statistics (provided on paper or diskettes) to Commission DGs, National Data Collection Units, data shops, OECD, etc.

Users’ feedback on clarity

Not available, as no user satisfaction survey on the ECHP had been carried out.”

Quality and Performance Indicators

AC1. Number of subscriptions/purchases of each of the key paper reports

General definition: obvious.

To be further defined for subject-matter domain: the paper reports considered to be key.

AC2. Number of accesses to on-line databases

General definition: number of query’s made to an on-line database (retrieve information for visualization or download by subject-matter domain).

Remark 1. The number of queries would more appropriate than the number of hits or downloads since it means effective access to a database instead of an occasional visit to the context page. However, it has to be developed in collaboration with IT experts.

AC3. Rate of completeness of metadata.

General definition: the ratio of the number of metadata elements provided to the total number metadata elements applicable.

Remark. Different rates of completeness can be calculated describing, respectively, (1) statistical outputs, (2) statistical processes, (3) institutional environment, legal aspects etc., and (4) further metadata (contact information, updates etc.)

ESS level

(i) Individual values and aggregates of AC1 over Member States.

(ii) Subscriptions/purchases of ESS reports.

(iii) Individual values and aggregates of AC2 over Member States.

(iv) Web hits and downloads from ESS level websites.

(v) Presentation of AC3 over Member States and of an overall AC3.

Summary

What should be included on Accessibility and Clarity

- A description of the conditions of access to data: media, support, pricing policies, possible restrictions, etc.
- A summary description of the information (metadata) accompanying the statistics (documentation, explanation, quality limitations, etc).
- The description should refer to both less sophisticated and more advanced users and how their needs have been taken into account.
- A summary of user feedback on accessibility and clarity.

6 Coherence and Comparability

ESS Quality Definition

*The **coherence** of two or more statistical outputs refers to the degree to which the statistical processes by which they were generated used the same concepts - classifications, definitions, and target populations – and harmonised methods. Coherent statistical outputs have the potential to be validly combined and used jointly. Examples of joint use are where the statistical outputs refer to the same population, reference period and region but comprise different sets of data items (say, employment data and production data) or where they comprise the same data items (say, employment data) but for different reference periods, regions, or other domains. **Comparability** is a special case of coherence and refers to the latter example where the statistical outputs refer to the same data items and the aim of combining them is to make comparisons over time, or across regions, or across other domains.*

Typically, different sets of data elements are gathered by different processes, for example employment data are obtained by a monthly survey of employing enterprises and production data by monthly survey of manufacturing enterprises. Thus, the term *coherence* is usually used when assessing the extent to which the outputs from *different statistical processes* have the potential to be reliably used in combination, whereas *comparability* is used when assessing the extent to which outputs from (nominally) the *same statistical process* but for different time periods and/or for different regions have the potential to be reliably used in combination. More specifically in the example above, the validity of the combined use of employment data and production data for the same population and time period is said to depend upon their *coherence*, whereas the validity of the combined use of employment data for the same population and region but different time periods depends upon their *comparability*.

It is worth reiterating that although the coherence/comparability is considered a property of statistical outputs, it depends upon, and is assessed entirely in terms of, the statistical processes that produce those outputs.

Crossover between Definitions of Coherence/Comparability and Accuracy

When bringing together outputs from two statistical processes, or the same process over time or across regions, the errors occurring (i.e., lacks of accuracy) in the processes have the potential to cause *numerical inconsistency* of the corresponding estimates. This can easily be confused with a lack of coherence/comparability. In other words, accuracy and coherence/comparability can easily be confounded.

The distinction made in this document is that coherence/comparability refers to, and is measured in terms of *descriptive (design) metadata* (i.e., concepts and methods) about the processes, whereas accuracy is measured and assessed in terms of *operational metadata* (sampling rates, data capture error rates, etc.) associated with the actual operations that produced the data. With this understanding coherence/comparability may be assessed in terms of the design metadata and accuracy in terms of operational metadata. It is also quite clear that the differences between *preliminary, revised and final* estimates generated by the same basic process relate to accuracy problems rather than coherence.

Where the error profiles of the statistical processes are known and included within the description of accuracy there is no need for further reference to them under coherence/ comparability. For example, suppose sampling error bounds are published for two values of the same data item for adjacent time periods indicating the range within which a movement from one period to the next may be due to chance alone and not reflect any actual change in the phenomenon being measured. If and only if the measured movement is larger than this, is there any point in discussing whether the movement is real or due to lack of comparability.

However, where the error profiles are not fully and precisely known (and they rarely are), errors in the estimates may be confounded with the effects of lack of coherence/ comparability. Thus, in so far as the accuracy descriptions do not take into account the errors that may occur, the possibilities of these errors have to be included in the account of coherence/comparability. For example, if there is no assessment of non-response error then the assessment of coherence/comparability has to include the possible consequences of differential non-response rates and patterns.

Another way of viewing the relationship between coherence/comparability and accuracy is to note that the numeric consistency of estimates depends upon two factors:

- the logical consistency (which we call coherence/ comparability) of the processes that generated those estimates; and
- the errors that actually occurred in those processes in generating the estimates.

Thus, coherence/comparability is a prerequisite for numerical consistency. The degree of coherence/ comparability determines the potential for numerical consistency. It does not guarantee numerical consistency as this depends also upon the errors.

6.1 Types of Coherence/ Comparability

In reviewing the characteristics of the statistical outputs produced by a particular statistical process, there are a variety of types of coherence/ comparability that are worth distinguishing.

- *Comparability over time:* for example, of monthly data from the labour force survey in a Member State.
- *Comparability over region:* for example, of data for the same month from the labour force survey in two Member States.
- *Comparability over other domain.* Domains over which comparisons are often made include economic activity group, occupational group, and sex. An example would be annual structural data for agriculture with annual structural data for manufacturing collected by a different survey.
- *Internal coherence:* referring to data produced by a (single) statistical process (but possibly comprising several different segments) for a single time period and region.
- *Coherence between sub-annual and annual statistics:* for example monthly and annual production data for the same industries in the same region

- *Coherence with the National Accounts.* For the economic surveys that feed into the national accounts, coherence is vital and, in so far as it is lacking, the National Accounts compilation process will detect it.
- *Coherence with other statistics:* for example, coherence between *employment* produced by a labour force survey of members of households and *numbers of employees* produced by an economic survey of enterprises.

6.2 Reasons for Lack of Coherence/Comparability

The possible reasons for lack of coherence/ comparability of the outputs of statistical processes may be summarised under two broad headings - differences in *concepts*, and differences in *methods*. Either or both of these may be a result of changes in the statistical process(es) as they are modified over time. Modifications may occur for a whole variety of reasons – introducing improved questionnaires, methods, automation, new technology, more up to date classifications, or in response to changes in legislation, or as a result of contractions or expansions in budget and hence in sample size or follow-up capacity, etc . For example, when Finland changed the data collection medium of the Labour Force Survey from postal enquiries to personal interviewing in 1983 the result was an increase of 100,000 in the estimate of employed people.

The possible reasons for lack of coherence/ comparability may be further broken down by type as described and exemplified below.

Concepts: target population – units and coverage

The target populations may differ for two statistical processes, or for the same process over time, in a variety of different ways, as illustrated in the following examples.

- The definition of economically active population used in the labour force survey may differ from one Member State to another. In one country it might be all persons aged 16-65 who are employed or seek employment, in another country all persons aged 15-70 who are employed or seek employment.
- Persons waiting to start a new job are counted as unemployed in the EU standard Labour Force Survey but as employed in the US Current Population Survey. This has resulted in a difference of 0.23% in unemployment rate (Sorrentino, 2000).
- Monthly statistics of industry might include just manufacturing enterprises whereas another statistical output with the same name might include manufacture and electricity, gas and water production as well.
- An annual structural business survey might use an *enterprise* as its target statistical unit whereas a monthly production survey might use a *kind of activity unit*.

Concepts: geographic coverage

For example, rural areas may be included in one country's labour force survey and excluded from another's.

Concepts: reference period

For example:

- in a survey of employees the enterprise might be asked for the number of full time employees *as of third Monday in the month* or *as of the first of the month*;
- an annual survey may refer to a fiscal year beginning in March, another to a calendar year.

Concepts: data item definitions, classifications

As an example of a difference in definitions, the labour force survey definition of *unemployed person* might be:

- *Any economically active person who does not work, is actively looking for a job and is available for employment during the survey;* or
- *Any economically active person who does not work, is actively looking for a job and is or will be available for employment in the period of up to two weeks after the survey's reference week*

Changes in classification schemes, in particular *revisions* in accordance with new versions of international standards, are a very common cause of coherence/ comparability problems. An example would be adoption of the latest version of NACE in place of an older classification of economic activities.

In addition, even without a change in classification, the procedures for assigning classification codes may be different or change over time, for example with improved training of staff or the introduction of an automated or computer assisted schemes.

Methods: frame population

Whatever the survey target units, the actual coverage of a survey depends upon the frame used for the survey. Possible examples of differences are as follows.

- In one case, enterprises with less than 5 employees might be excluded, in another case all enterprises included.
- A more substantive difference would occur where one frame was based on value added tax, i.e., a source covering all enterprises paying VAT, whereas another frame was based on employment deductions, i.e., a source covering all enterprises with employees subject to tax deductions.
- The legal requirements for VAT registration may change, resulting in more or fewer enterprises in survey frames.
- Surveys may be designed as cross sectional or longitudinal with significant difference in estimates of change as a result. Even within a longitudinal survey, panels or rotation patterns may change over time or between countries.
- Even without any nominal difference in statistical units, the procedures by which statistical units for large enterprises are actually delineated may differ, or change over time in accordance with better training or new methods. For example the procedures for treatment of the creation, amalgamation, merger, split, or cessation of an enterprise may change.

Methods: source(s) of data and sample design

An example of a difference might be that in one statistical structural survey financial data for small enterprises was obtained from income tax data whereas in another it was obtained by direct survey.

Methods: data collection, capture and editing

In one case there might be intensive follow-up of non–response and consequential reduction of non-response rate to 10%, in another there are no resources for follow-up and the non-response rate is 40% thus giving rise to a substantially increased probability of non-response bias. As noted above, if the probable biases due to non-response errors have been reported under Accuracy for both surveys then this would not need to be repeated as a comparability/coherence issue

Methods: imputation and estimation

Different imputation methods may be used for dealing with missing data items. For example in one survey zeroes may be imputed for missing financial items whereas in another survey non-zero values may be imputed based on a "nearest neighbouring" record. Likewise in dealing with missing records in an enterprise survey there are various options, such as assuming the corresponding enterprises are non-operational or assuming they are similar to enterprises that have responded.

6.3 Assessment and Reporting

Methods for assessing and reporting coherence/ comparability are presented in the following paragraphs in two groups, first general methods that apply to all type of coherence/ comparability and subsequently those methods that are specific to a particular type.

General Approach

The cause of any lack of coherence/comparability whether due to changes in concepts or methods or both should be clearly explained. In this situation the producer should facilitate reconciliation of the estimates by quantifying, at least approximately, the effects of the main sources of incoherence/ incomparability. The minimum requirement is that each instance is indicated in the quality report and that the reason for it and its order of magnitude is stated according to the producer's best knowledge.

Any general changes that have occurred that may have impact on coherence/ comparability should be reported, for example changes in legislation affecting data sources or definitions, re-engineering or continual improvement of statistical processes, changes in operations resulting from reductions or increases in processing budget, etc.

Deviations from relevant ESS legislation and other international standards that could affect coherence/ comparability should be reported

As previously noted, statistical outputs that describe the same phenomenon may be coherent, and yet not present identical values because of the errors that occur.

Concepts and methods should be presented in the quality report in the introduction or relevance chapters or as part of the error profile descriptions in the accuracy chapter. Under coherence/ comparability the report should make as clear as possible to what causes a given problem can be attributed. Ideally, the sources of incoherence/ non-comparability should be quantitatively decomposed by each possible source. If this is possible the corresponding sets of statistical outputs are *reconcilable*. Although this is usually not fully attainable, the quality report should as be as informative as possible with this goal in mind.

More precisely, for the statistical process(es) in question, the first step is to conduct a systematic assessment of possible reasons (as listed in Section 6.3) for lack of coherence/ comparability. The assessment should be based primarily on examination of the key metadata elements, and identification and analysis of differences. Looking at the data themselves may provide some indication of the likely magnitude of any lack of coherence/ comparability but not of its causes.

For each difference in metadata, for example differing frame populations, the next step is to deduce the likely effect of such a difference on the statistical outputs. The final step is to aggregate and summarise in some way the total possible effect, in other words to form an impression of the degree of (lack of) coherence/ comparability.

Comparability over Time

Comparability over time is a crucial quality aspect for all statistical outputs published on a number of consecutive occasions. For many users, changes over time of economic or social phenomena are the most interesting aspects of the statistics, and comparability over time is essential if the data are to reflect the actual economic or social changes that occurred.

Regardless of whether statistics are directly published in time series form or whether the users have to construct their time series themselves from basic data, users need to be informed about possible limitations in the use of data for comparisons over time. This information also has to be included in the quality report.

In assessing comparability over time the first step is to determine (from the metadata) the extent of the changes in the underlying statistical process that have occurred from one period to the next. There are three broad possibilities:

1. There have been no changes, in which case this should be reported;
2. There have been some changes but not enough to warrant the designation of a *break in series*;
3. There have been sufficient changes to warrant the designation of a *break in series*.

In the second and third cases, the changes and their probable impacts should be reported.

In the second case, the effect of the change may be sufficiently small that is has negligible effect on outputs. The NSO may simply note it in metadata describing the process. Sometimes an effect may not be negligible but be too small to warrant a series break. In this case the NSI may *wedge in* the changes to the outputs over a period of time so that, between any two periods, the adjustments being made to move from old to new values is less than the sampling error and thus cannot by itself be detected and interpreted as a real change.

In the third case, users have to be informed that there has been break in series and provided with the information they need to deal with its consequences. The information provided may range from very complete to minimal depending upon the NSO resources available and the size of the break.

- The most comprehensive treatment is to carry forward both series for a period of time and/or to backcast the series, i.e., to convert the old series to what it would have been with the new approach by duplicating the measurement in one time period using the original and the revised definitions/methods.
- A less expensive treatment is to provide the users with transition adjustment factors giving them the means of dealing with the break for example by doing their own backcasting.
- The least expensive treatment is to simply describe the changes that have occurred and provide only qualitative assessments of their probable impact upon the estimates. Obviously this is the least satisfactory from the user perspective

Comparability over Region

Comparability over region may be assessed in two different ways: pair-wise comparisons of the metadata across regions; and comparison of metadata for the region with a standard, in particular an ESS standard or, in its absence, an example of best practice from one of the NSIs.

Two broad categories of situation can be identified:

- where essentially the same statistical processes are used, e.g., a labour force survey designed in accordance with ESS standard, and differences across regions are expected to be quite small; and
- where a different sort of statistical process is used, for example a direct survey in one case and a register based survey in another. In such cases the differences are likely to be more profound.

To assess the overall impact of all the possible differences it may be worthwhile summarising the differences in terms of a scoring system. The most simple scoring scheme is to define the key metadata elements for which a difference could be significant and for each one to assign a binary score: no difference; difference. An overall impression of comparability can then be obtained by assigning a weighting to each key metadata element according to its potential effect on comparability and computing a weighted score across all metadata elements. Such an overall score would be useful not only in cross country comparisons but also in tracking a single process over time.

ESS level

Using a scoring scheme as described above is one way of summarising the results for all Members States in a matrix.

Comparability over Other Domains

As previously noted, the domains over which comparisons may be made include economic activity group, occupational group, and sex. The methods of assessment and reporting are similar to those used for comparability over region. Again there is a useful distinction to be made between situations where essentially the same statistical instrument is used, for example a direct survey, and those where different instruments are used.

Internal Coherence

Based on a given statistical process, statistical outputs are published. Each set of outputs should be internally coherent, meaning that all the appropriate arithmetic and accounting identities should be observed. However, this is not always the case. For example, some otherwise efficient estimation methods have this drawback. It may also occur where the process actually comprises more than one segment with data from different sources or for different units in each segment. In these circumstances a brief explanation should be given to users and also be reflected in a quality report, with the reasons for publishing non-coherent results explained.

Coherence between sub-annual and annual statistics

Coherence between subannual and annual statistical outputs is a natural expectation on the part of users and yet the statistical processes producing them are often quite different. Thus, reasons for lack of coherence need to be assessed and explained.

A starting point for assessing the likely magnitude of differences due to lack of coherence is to compare subannual and annual estimates:

- If both annual and subannual series measure levels then annual aggregates can be constructed from subannual estimates and compared to totals from the annual series;
- If one or other of the series produces only growth rates not levels, then comparison can be made of year over year growth rates.

If the differences thereby observed cannot be fully explained in terms of sampling error or other measure of accuracy then their explanation requires assessment of the possible causes by metadata comparison, as for all forms of coherence assessment.

Coherence with the National Accounts

As previously noted, the National Accounts compilation process is a method for detecting lack of coherence in data received from its various source statistical processes, whether they be direct surveys, register based surveys or indexes. Feedback from the National Accounts on the degree of incoherence and the adjustments that had to be made in order to bring the accounts into balance are excellent indicators of the accuracy and/or coherence of the statistical outputs received. They should be reported and should be a trigger for further investigation.

Mirror statistics

As previously noted in Section 3.8.6, for certain selected statistical outputs from a Member State, notably in trade, balance of payments, migration and tourism, it may be possible to find counterpart statistical output in another Member State or country. For example the UK ONS may publish emigration from the UK to Australia and the Australian Bureau of Statistics may publish immigration from the UK.

Mirror statistics involve coherence, geographical comparability as well as accuracy issues. Having assessed the degree of lack of coherence, any difference in outputs that cannot be explained in terms of coherence are an indication of the lack of accuracy in either or both of the outputs and/or may reflect lack of comparability between the countries for the same data items.

For example, if the UK estimate of emigration to Australia in a particular year exceeds that of the Australian estimate of immigration from the UK for the same year by 10% then this could reflect lack of accuracy in the form of overcounting in the UK or undercounting in Australia, and/or it may be the result of lack of comparability of UK and Australian definitions of immigration, or emigration, or both.

Coherence with other statistics

There may be other statistical outputs in combination with which the statistical output in question can be used. For each such output the report should contain an assessment of incoherence in terms of possible sources and their impacts as noted in previous paragraphs.

6.4 Examples

Example 6.4.A Coherence Analysis in Quality Report of 1996 UK Annual Production and Construction Inquiries (UK ONS 1999b, page 41)

The Index of Production (IoP) is not benchmarked to the Annual Business Inquiry (ABI). However, some comparisons are made at unit (respondent) level. Checks are also carried out at unit level between ABI and the quarterly capital expenditure and stocks inquiries.

For 1995 and 1996 aggregate level comparisons were made of deflated value added with IoP measure of production. The table below shows the 1995-1996 growth in manufacturing industries of deflated unweighted and weighted gross output (GO and WGO) and gross value added (GVA and WGVA) against the IoP growth. (Here, *unweighted* means simply added, *weighted* means using IoP weights. IoP deflators are not strictly suitable for value added but are used for the lack of an alternative).

1995-1996 Growth of Gross Output and Gross Value Added against IoP at Constant Prices

SIC 92 subsection	IoP Weights	Unweighted and weighted growth				IoP
		GO	GVA	WGO	WGVA	
DA	108	-3.44	-0.20	-2.86	3.49	1.13
DB	38	0.17	3.11	-3.21	0.85	-1.16
DC	5	0.75	8.40	3.62	12.44	-0.11
DD	11	-2.61	-3.45	-4.02	-7.10	-2.03
DE	99	2.37	-0.37	1.50	-0.25	-1.34
DF	18	7.74	-20.61	8.35	-20.22	-10.86
DG	91	2.50	-0.64	4.30	3.24	2.12
DH	40	4.82	5.51	4.96	5.19	-1.02
DI	30	-6.08	-4.31	-5.42	-8.45	-3.84
DJ	93	-0.51	-7.26	-0.86	-10.48	-0.03
DK	72	-0.24	-0.68	-1.39	-0.90	-1.76
DL	101	8.08	1.19	9.27	1.31	2.64
DM	77	6.88	5.94	8.18	6.93	3.86
DN	29	0.41	6.14	1.95	9.56	1.69
Total manufacturing	812	2.02	-0.38	4.55	-0.13	0.27

It can be seen from the table that the growth in IoP is closer to that in GVA than in GO although one would expect that the IoP data to be closer to the latter. The difference between WGO and IoP values is over 4% for the total, whilst GO is much closer, with the difference being around 2%.

Clearly there are large differences for subsection DF (Manufacture of Coke, refined Petroleum Products and Nuclear Fuel) and subsection DL (Manufacture of Electrical and Optical Equipment). In the former case, WGO and IoP differ by over 19%, with the IoP suggesting a substantial fall and the actual ABI results suggesting a large growth. In the latter, there is also a large difference, but not to the same extent as in DF. It is evident that further investigation is required at the 4-digit level in order to ascertain reasons for these differences.

Example 6.4.B Coherence Analysis of Gross Income from ECHP and HBS in France

The table below shows the distribution of gross income in France by component as it was estimated from waves 1 and 2 of the European Community Household Panel (ECHP) in 1993 and 1994 and from the Household Budget Survey (EBF) for 1994/95.

	ECHP - 1993	**ECHP – 1994**	**ECHP - 1994**	**EBF – 1994/95**
	Income per household	**Income per household (households existing in 1993)**	**Income per household (all households)**	**Income per household**
Total income before tax	170,228	174,952	171,147	168,590
%	100%	100%	100%	100%
Income from labour	65.2%	64.4%	64.6%	63.7%
-salaries and wages	55.2%	55.2%	55.5%	55.8%
-self-employment income	9.3%	8.4%	8.3%	7.4%
-income from second job	0.7%	0.8%	0.8%	0.5%
Pensions	21.1%	21.6%	21.3%	21.6%
Other social transfers	7.8%	8.1%	8.1%	8.5%
. Unemployment benefits . Family related benefits . Invalidity benefits . Social assistance . Housing allowances	2.4% 3.0% 1.1% 0.2% 1.1%	2.5%¨ 2.8% 1.4% 0.3% 1.1%	2.5%¨ 2.8% 1.4% 0.3% 1.1%	2.6% 3.1% 1.2% 0.3% 1.3%
Capital income	**4.7%**	**4.1%**	**4.3%**	**5.2%**
Private transfers	**1.0%**	**1.4%**	**1.5%**	**0.9%**

Clearly, the two sources produce very close results, albeit with some slight differences. For instance, self-employment income and income from a second job are lower in the EBF, while social transfers seem higher.

Example 6.4.C Analysis of Coherence of Level of Educational Attainment from ECHP, LFS and OECD 'Education at a Glance'

The following table indicates the percentage of the population aged 25 to 64 by the highest completed level of education as estimated from the ECHP and the LFS and as in the OECD publication.

Attainment level (%)	**EU and Member States**												
	EU	**B**	**DK**	**D**	**EL**	**E**	**F**	**IRL**	**I**	**L**	**NL**	**P**	**UK**
ISCED 0-2													
ECHP94	43	38	26	26	56	67	40	51	59	53	19	84	42
LFS94	45	47	24	18	59	73	42	55	64	53	22	79	48
EAG96(1994)		51	40	16	55	74	33	55	67		40	81	26
ISCED 3-4													
ECHP94	37	31	38	51	24	15	39	35	32	29	60	9	34
LFS94	37	30	50	60	28	12	40	27	28	25	56	10	31
EAG96(1994)		27	40	62	27	11	50	27	27		38	8	54
ISCED 5-7													
ECHP94	20	31	36	23	20	19	21	14	9	18	21	7	24
LFS94	17	23	26	22	14	15	17	18	7	23	22	11	21
EAG96(1994)		22	20	23	18	15	17	19	8		21	10	21

Clearly, there are some notable discrepancies between these different sources. These are partly related to classification difficulties. Overall, the ECHP tends to show higher results at the higher educational levels than the other two sources.

Example 6.4.D Analysis of Coherence between Swedish Short Term Production Volume Index (STPVI) and National Accounts (UKONS 1999b)

"The National Accounts (NA) receives a special quarterly version of the STPVI, which is adjusted for the change in stock, as well as the regular unadjusted index. These data are not officially published elsewhere. However, quarterly data can be derived from the monthly index-series of the STPVI but as these are not adjusted for the change in stock, the officially published STPVI is not fully coherent with the quarterly NA.

To illustrate what this lack of coherence means in practice the following table lists the quarterly changes from 1994 to 1995 for the STPVI and the National Accounts.

NACE	NA Q=1	STPVI Q=1	NA Q=2	STPVI Q=2	NA Q=3	STPVI Q=3	NA Q=4	STPVI Q=4
C	9,9	9,2	5,4	7,7	14,2	10,5	15,5	6,9
D	13,6	12,6	12,2	12,3	11,5	11,4	5,8	4,6
15+16	1,7	0,3	3,2	4,2	2,0	2,6	3,8	4,6
17-19	18,0	15,3	13,1	13,1	16,6	18,4	4,4	4,7
20	23,6	20,5	11,0	9,7	11,3	11,1	2,7	2,0
21	6,9	6,1	2,0	2,1	-1,0	-1,2	-12,8	-13,4
21.11	12,9	9,3	7,8	4,3	4,8	1,6	-11,9	-14,3
22	-3,3	-7,5	2,5	-2,2	0,4	-5,0	3,4	-2,6
23	-6,6	-6,3	-8,0	-7,4	17,5	17,7	2,2	2,4
25	4,3	5,7	-1,8	0,0	-6,3	-4,5	-9,7	-9,1
26	19,8	24,8	5,4	6,5	5,2	5,1	-3,3	-4,3
27	0,4	0,1	-0,1	-0,3	0,3	0,0	3,4	3,2
28	28,5	26,3	16,2	17,0	16,6	15,3	8,1	7,1
29	20,4	20,0	23,9	25,5	18,8	16,7	7,4	11,4
30	17,1	8,3	7,6	13,3	10,8	12,4	10,0	10,8
33	4,4	6,0	11,5	8,4	-7,4	-13,8	3,3	-2,9
36+37	10,8	8,4	-1,8	-1,0	1,3	2,9	-5,4	-5,3

6.5 Quality and Performance Indicators and Summary

Quality and Performance Indicators

CC1. Lengths of comparable time series.

General definition: number of reference periods in time series from last break.

Remark. The unit of measurement depends on the reference period of the survey (month, quarter, year etc.).

CC2. Asymmetries for statistics mirror flows.

General definition: **discrepancies between data related to flows, e.g. for pair of countries.**

Remark. In domains where there are mirror statistics it is possible to assess geographical comparability measuring the discrepancies between inbound and outbound flows for pairs

of countries. Being a really quantitative measurement and not only the result of a count, this indicator represents an important assessment of the level of quality of the country data for the domains in which mirror statistics are available.

Summary

What should be included on <u>Coherence and Comparability</u>

<u>General</u>

- Brief descriptions of all conceptual and methodological metadata elements that could affect coherence/ comparability.
- An assessment (preferably quantitative) of the possible effect of each reported difference on the output values.
- Differences between the statistical process and the corresponding European regulation/ standard and/or international standard (if any).

<u>Comparability over Time</u>

- Reference periods at which series breaks (if any) occurred, the reasons for them and treatments of them.

<u>Comparability over Region</u>

- A quantitative assessment of comparability across regions based on the (weighted) number of differences in metadata elements.
- At ESS level, a coherence/comparability matrix summarising by region the possible sources of lack of comparability relative to a specified standard.

<u>Internal Coherence</u>

- Any lack of coherence in the output of the statistical process itself.

<u>Coherence with National Accounts</u>

- Where relevant, the results of comparisons with National Account framework and feedback from National Accounts with respect to coherence and accuracy problems.

<u>Coherence with Other Statistics</u>

- Where the statistical outputs were combined with those from other processes, the limitations if any set by coherence considerations

<u>Mirror Statistics</u>

- Assessment of discrepancies (if any).

7 Trade-offs between Output Quality Components

As previously noted, the output quality components are not mutually exclusive in the sense that there are relationships between the factors that contribute to them. There are cases where the factors leading to improvements with respect to one component result in deterioration with respect to another. This chapter deals with the trade-offs that have to be made in such circumstances. It assumes the budget is fixed and that all possible efforts have been made to improve quality in terms of each of the components individually, i.e., where this can be done without negatively affecting quality in terms of another component.

There are five output quality components, three of which have two subcomponents. Hence there is a minimum of 28 pairs of quality trade-offs to consider. However, improvements to accessibility and clarity are more on less independent of improvements to other quality components except in so far as they require resources that have to be shared. This leaves around 15 pairs of trade-offs for consideration. The following sections outline the ones that are likely to be the most significant.

For each case the quality report should indicate the trade off decision and on what basis it was made. In some cases the decision may have been based on an analysis and, in this case, the results of the analysis should be reported. In other (frequently occurring) cases the decision may have been made without any analysis, on an ad hoc basis, and, if so, this should also be reported.

7.1 Trade-off between Relevance and Accuracy

One way to improve relevance may be to provide additional breakdowns, for example of national totals by geographic region or by economic activity. If the sample is redesigned to support such breakdowns and the overall sample size stays the same, then the sampling errors of the national totals will likely increase as the sample is no longer optimised for national estimates. In other words, accuracy will be adversely affected. Likewise, if the number of data elements is increased without additional funding, the amount of time available for editing the individual data cells is reduced and the potential for measurement errors increased, again adversely affecting accuracy.

Conversely, reduction in the number of data elements, and/or in the detail with which the results are broken down, may well increase accuracy but to the detriment of relevance.

7.2 Trade-off between Relevance and Timeliness

Timeliness can be improved by reducing the number of data elements collected and processed and/or by replacing some of those that are more difficult to collect or process by ones that are easier. This will have a negative effect on relevance. Conversely improvements in relevance, for example collecting regional as well as national data, may well introduce more, and/or more complex, data elements and hence reduce timeliness.

7.3 Trade-off between Relevance and Coherence

Improvements in relevance of a particular statistical process in response to user requirements, for example by fine-tuning the definitions of some variables or classifications may reduce the coherence of its outputs with those of other processes. Conversely, the desire to retain coherence between the outputs from two statistical processes may inhibit the changes required to improve the relevance of either one to its particular users.

7.4 Trade-off between Relevance and Comparability over Time

Improvements in relevance in response to user requirements, for example by redefining the variables for which data are collected, or moving to a later version of a classification, may well reduce comparability over time, perhaps to the point of requiring a series break. Conversely, the desire to retain comparability may inhibit changes in content required to improve relevance.

7.5 Trade-off between Comparability over Region and Comparability across Time

In a similar fashion, the desire to have more comparability across region may well result in changes that reduce comparability over time.

7.6 Trade-off between Accuracy and Timeliness

This is probably the most frequently occurring and important of the trade-offs. Improvement in timeliness can be obtained by reducing collection and processing time, in particular by terminating collection earlier, compiling outputs based on a smaller number of responses and/or by reducing the amount of editing. However, this reduces accuracy.

A compromise situation may be taken for important statistics by publishing an early set of estimates and then one, two or more subsequent revisions. As previously noted, the releases in a three part sequence are typically termed *preliminary, revised, and final*.

Although the final figures are not completely accurate, they are normally more accurate than figures in earlier outputs. Thus, the size of these revisions is an indicator of degree of accuracy that is being sacrificed in order to produce the increased timeliness of the outputs.

7.7 Quality and Performance Indicators and Summary

Quality and Performance Indicators

None specifically defined.

Summary

What should be included in the Quality Report on <u>Trade-offs</u>

- A description of each important trade-off that has been analysed and the basis on which the trade-off decision has been made.
- A statement concerning any trade-offs that should have been analysed but have not been.

8 Assessment of User Needs and Perceptions

ESS Quality Declaration: User Focus

We provide our users with products and services that meet their needs. The articulated and non-articulated needs, demands and expectations of external and internal users will guide the ESS, its members, their employees and operations.

Assessment of user needs and perceptions typically involves consideration of all the output quality components, not simply relevance (although this is most important). This section deals with those aspects of assessment that are not confined to a single aspect of output quality.

8.1 Understanding and Classifying Users

The starting point for design and conduct of a statistical process is user needs. Such needs are expressed not only in terms of data content but also in terms of the degree of accuracy required, the timing, the dissemination arrangements, the metadata required for interpretation, and the relationship to other relevant statistical outputs. In other words, they cover the whole range of the output quality components.

Assessment of user needs is not trivial, first because there are many types of users, second, because there are many different uses for which the users want the outputs.

The first step is to assemble information about the *users* - who they are, how many they are, and how important they are individually and collectively from the perspective of the NSO. Based on information available from advisory committees, lists of paying recipients, Internet accesses, the usual approach is to develop a *classification of users* and to estimate the number of each type

The second step is to determine the *needs* of each class of users, and, in the case of important users, their individual needs. For users, acquiring output data is a means to an end, not an end in itself, and the uses to which these data are put are relevant. Quite frequently users may not fully understand what data they actually need nor what is available. By understanding the uses of data, the NSO is in a better position to determine the actual needs. Furthermore, these needs have to be interpreted in the statistical context in which they are to be addressed – concepts, accuracy, timing, etc., have to be aligned with what can actually be delivered. Information about user needs is typically accomplished through subject matter advisory committees, user groups, ad hoc focus groups, requests, complaints and other user feedback.

The third step is to determine in general terms the *priorities* to be given to the various classes of users in satisfying their needs. For example the needs of government policy makers may be set ahead those of academic researchers. Some needs are important but transient. Some users may also be respondents and their requirements merit special consideration.

The fourth step is to determine the associated metadata needs of users, i.e., what explanatory and quality related material should accompany the data and how it should be presented. For this purpose it is convenient to classify users into groups according to the complexity of their data, and associated metadata, needs. For example, the Australian Bureau of Statistics uses three groups, which it refers to as *tourists, harvesters and miners*, reflecting increasing levels of demand, as previously mentioned in Section 5.

In summary the quality report should contain a classification of users, an indication of the uses for which they want the outputs, the priorities in satisfying their needs, and an account of how this information was obtained, for example through general or domain specific advisory committees, other regularly convened user groups, ad hoc focus groups, feedback/complaints from users. The following paragraphs provide some examples.

Example 8.1.A Classification of Users from the UK Office of National Statistics

Holt and Jones (1998) present the following diagram and accompanying description, which illustrate the range of customers for official statistics in the United Kingdom.

"At the top of the tree, in the UK, is government and Parliament. This includes HM Treasury, who is our principal paymaster and primary customer, especially for macro-economic statistics. It also includes the other major government departments (OGDs) who depend upon statistical information to monitor and develop policy and to allocate resources. With all of our Government customers, there has to be a lot of direct contact, of dialogue about needs and the same applies to our customers in the wider community for which we have established a wide range of user groups and advisory committees. In the UK in the 1980s, Government considered itself to be the sole customer. Others could benefit only incidentally. This has now changed, and we have made great strides in recent years in understanding the wider market and seeking to meet its needs."

This example not only provides a classification of users it indicates a change in priorities.

Example 8.1.B Classification of Users from Measuring the Non-Observed Economy: A Handbook

In Measuring the Non Observed Economy: A Handbook prepared by the OECD, IMF and other international organisations, there is a grouping of users under seven broad headings:

- internal statistical office users, specifically including the national accounts area;
- national government – the national bank, and the ministries dealing with economic affairs, finance, treasury, industry, trade, employment, environment;
- regional and local governments;
- business community – individual large businesses and business associations;
- trade unions and non-governmental organisations;
- academia – universities, colleges, schools, research institutes, etc;
- media – newspapers, radio and TV stations, magazines, etc;
- general public;
- international organisations.

Example 8.1.C Classification of Users from the Australian Bureau of Statistics Economic Statistics Strategy 1990

The major uses to which users put the data can be classified as macro-economic analysis (structural and short term), micro-economic analysis (short term, industry based, activity based, and business dynamics), or regional analysis, and summarised by user, as indicated in the following table.

	Macro-Economic Structural	**Macro-Economic Short Term**	**Micro-Economic Industry**	**Micro-Economic Activity**	**Micro-Economic Short Term**	**Micro- Economic Dynamics**	**Regional**
Principal Users/Uses	National Bank, Treasury, M. Finance, National Accounts	National Bank, Treasury, M. Finance, National Accounts	National Bank, Treasury, M. Finance, M. Industry, Industry Associations	M. Industry, M. Environ, M. Technology, Industry Associations	Industry Associations, Marketing Agencies	M. Small Business, Monopolies Commission, Enterprises	M. Regional Develop, Regional Government, Development Enterprises
Reference period	Year	Month/ Quarter	Year	Year	Month/ Quarter	Year Longitudinal	Year
Data Items	turnover, expenses, purchases, stocks, earnings, hours, employment, labour costs, capital expenditure, operational surplus, assets, liabilities	retail sales, stocks, earnings, hours, employment, capital expenditure, business opinions	retail sales, stocks, earnings, hours, employment, capital expenditure	commodities produced, services rendered, research & development expenditure, environmental impacts & expenditure	retail sales, commodities produced, services rendered, motor vehicle registrations	as for second col. + counts by enterprise/ establishment of births, deaths & organisation changes (by type)	turnover, employment, earnings
Industrial Breakdown	division/	division	class	varies	none	branch	branch
Geographic Breakdown	none	none	region	region	region	region	locality
Size Breakdown	none	none	2-6 classes	2-6 classes	none	2-6 classes	2-6 classes
Output Frequency	annual	monthly/ quarterly	annual/ occasional	annual/ occasional	monthly/ quarterly	annual/ occasional	annual/ occasional
(Note: M. = Ministry of)							

Example 8.1.D Classification of Users as presented in [Data Communication - Emerging International Trends and Practices of the Australian Bureau of Statistics, 2006 Catalogue 1211.0]

"Broadcasting...is defined as the proactive ("push") dissemination of information using the web site to suit a diverse range of user interests in a manner that facilitates communication. To do this effectively, we must ensure the information provided on the ABS web site is relevant to the diverse range of web users e.g. "visitors", "harvesters" and "miners".

The layered approach is fundamental to the ABS broadcasting strategy. "Tourists" who have limited knowledge of the types of statistical information available from the ABS web site, can browse the Statistical Headline News to look for interesting leads that will entice them to read more. On the other hand, experienced users, "harvesters"/"miners", can bookmark the relevant web page, thereby bypassing the common navigation paths and reducing the number of clicks required. Note that an expert user in a particular field of statistics may well be a "tourist" in another field."

Example 8.1.E User Needs from Le Monde, 27 June 2001

On 25 June 2001, the French Prime Minister expressed in a public meeting his need for "a new statistical tool to measure unsafeness regarding juvenile delinquency". He said that the tool should "go beyond the simple statement of infractions toward a true measure of unsafeness and of results obtained to reduce it"

This quotation illustrates how volatile and unpredictable user needs may be, that user needs can sometimes be deduced from media reports, and that they have to be interpreted in statistical terms.

ESS level

In an ESS level quality report, an overview of the users and uses of national outputs should be given as well as the additional, specific uses of the ESS level aggregations and comparisons.

In this context, Eurostat uses (or has used) the following classification of users, which could also be readily adapted to national level.

Example 8.1.F Classification of Users by Eurostat

Institutions	*European level*: Commission (DGs, Secretariat General), Council, European Parliament, ECB, other European agencies. *Member States, at the national or regional level*: Ministries of Economy/Finance, other ministries (for sectoral comparisons), National Statistical Institutes and other statistical agencies (for norms, training, etc). *International organisations*: ECB, OECD, UN, IMF, ILO, etc.
Social actors	Employers' associations, trade unions, lobby groups, etc, at European, national and regional level.
Media	International or regional media – specialised or for the general public – interested both in figures and analyses or comments. The media are the main channels of statistics to the general public.
Researchers and students	Researchers and students need statistics, analyses, ad hoc services, access to specific data.
Enterprises/ businesses	For their own market analysis, marketing strategy (large enterprises) or because they offer consultancy services in the information sector.
Internal to Eurostat	Statistics produced by a unit sometimes rely on results produced by other units (for example, for calculation of ratios or more complex indicators, such as aggregated accounts).

8.2 Measuring User Perceptions

User satisfaction is the number one priority. The most effective method of evaluation is a full scale user satisfaction survey, conducted in accordance with normal survey best practices - drawing a representative sample of users from an appropriate frame, designing and testing a suitable questionnaire, collecting, processing and analysing the results, etc.

Conducting a user satisfaction survey is not always affordable, particularly for small statistical processes where it would represent a significant share of the operation's total budget. Other methods of assessment include analysis of publication sales, user comments, requests and complaints received, web site accesses, etc., and feedback from advisory committees and focus groups.

The quality report should present the main results regarding user satisfaction, preferably broken down by the most important classes of users. It should also indicate the methods used for assessment and the measures taken to improve user satisfaction. The same comments apply for ESS level as for national level.

The following paragraphs provide some examples.

Example 8.2.A User Satisfaction Survey at Statistics Sweden (1997)

Statistics Sweden conducted an interview survey of just over 1000 of the agency's most important customers, asking them to rank its services and products according to a number of different quality dimensions on a scale from 1 to 10. Analysis of the results indicated that:

- Customers have a high level of confidence in the agency's ability to produce impartial and reliable statistics.
- The agency's personnel are given high marks in all the surveys. This applies to telephone reception, service-mindedness, statistical and subject matter competence as well as respecting agreements.
- The timeliness of its data has been given a grade that is relatively lower than other quality factors.
- Offers, negotiation of prices, delivery of consignments on time, and handling of complaints are important questions for clients. This is also an area where the agency can make improvements.

Example 8.2.B User Satisfaction Assessment for Euro-SICS database

Eurostat conducts an evaluation of user satisfaction for the *Euro-SICS database* containing Euro-zone short-term indicators. It is undertaken mainly through continuous dialogue with its two main users, DG ECFIN and the European Central Bank (ECB). The January 2001 Quality Report noted that users requested “more indicators but less breakdowns”. This is obviously the type of information that helps give an idea of the relevance of the output and to orient future developments.

Example 8.2.C Eurostat Satisfaction Survey 2007

Eurostat conducted a user satisfaction survey during the months of June and July 2007. The survey covered three main topics:

- information on the type of user and type of use of Community statistics;
- quality aspects of Community statistics; and
- aspects of disseminating statistics.

8.3 Quality and Performance Indicators and Summary

Quality and Performance Indicators

US1. User satisfaction index.

General definition: the degree of satisfaction with services and products.

Remark. To be calculated for different segments of users.

US2. Length of time since most recent user satisfaction survey.

General definition: obvious.

To be further defined for subject-matter domain: what is deemed to constitute a “user satisfaction survey”.

What Should Be Included in Quality Report on User Needs and Perceptions

- Means of obtaining information on users and uses.
- Description and classification of users.
- Uses for which users want the outputs.
- Users and uses given special consideration.
- Means of obtaining user views.
- Main results regarding user satisfaction.
- Date of most recent user satisfaction survey.

9 Performance, Cost and Respondent Burden

European Statistics Code of Practice

Principle 9. Resources must be effectively used.

Principle 10. Respondent burden should be proportional to the needs of users and not excessive for respondents. Respondent burden should be measured and targets set for its reduction over time.

Performance, cost and respondent burden are aspects of process quality that cannot be covered under any of the output quality components. However, there are trade-offs to be considered between cost and response burden and the output quality components, or, expressed differently, cost and respondent burden are constraints on output quality.

The capacity to calculate costs is essential for efficient management in general, and for quality and performance assessment in particular. Cost benefit analyses are required in order to determine the appropriate trade-off between costs on the one hand and benefits in terms of the output quality components on the other. Likewise, respondent participation must be viewed as a cost (to respondents) that has to be balanced against the benefits of the data thus provided.

In some specific statistical domains, ESS legislation highlights the need to consider the relationships between output quality, cost and respondent burden, as indicated in the following examples. In addition, Eurostat has a rolling review programme.

Example 9.A Regulation N°295/2008 of European Parliament and Council of 11 March 2008 concerning Structural Business Statistics

Article 6 states "Quality evaluation shall be carried out comparing the benefits of the availability of the data with the costs of collection and the burden on business, especially on small enterprises".

Example 9.B Council Regulation N° 1165/98 of 19 May 1998 concerning Short-term Statistics.

Article 14 states "The Commission shall … submit a Report … on the statistics compiled … and in particular on their relevance and the burden on business".

Example 9.C Eurostat Rolling Review Programme

Rolling Reviews are systematic reviews of Eurostat's statistical work looked at together with main users and partners in the Member States. They are based on several assessment tools such as an assessment checklist, user surveys and partner surveys and aim at evaluating issues such as:

- Are the requirements of Eurostat's statistical programme met?
- Are the production processes organised in an efficient way?
- Are the partners satisfied with Eurostat's guidance and way of working?
- Do the users get adequate and satisfactory information and service?
- What are the costs to Eurostat and the Member States?
- Could the work be done more efficiently?
- Are the data disseminated by Eurostat of good quality?

Rolling Reviews are in-house evaluations concerned with examining ways of improving and enhancing the implementation and management of interventions. They are conducted for a number of statistical processes in given intervals and have as a main purpose the improvement of Eurostat's performance by finding possible ways to improve the functioning within each statistical area. They involve a thorough review of users' satisfaction, partners' satisfaction, Eurostat's and Member States' resources and costs.

In summary, a quality report should contain measurements of cost and respondent burden and an account of the considerations in determining appropriate levels. Whilst there are no universal standards or guidelines to follow the following sections provide some ideas.

9.1 Cost

A comprehensive assessment of the costs associated with a statistical process is complicated because it requires a mechanism for allocating shared costs (for example, the costs of the business register) and overheads (office space, utility bills etc). This approach is the so-called full-cost approach. A simple assessment of the principal direct costs is also feasible and would be mainly based on time spent for a given statistical process.

The choice of using a full-cost approach or an approach based on direct costs depends on the utilised cost accounting system in each Member State's administration.

Some examples are provided in the following paragraphs. Note that, in this context, the actual sources of funding are irrelevant as they have nothing to do with efficiency.

Example 9.1.A Cost Model Proposed by Eurostat Unit for planning and reporting

The total cost for a statistical process is computed as the sum of:

- staff cost;
- data collection cost (printing and mailing); and
- other costs.

Costs may be divided into two groups: those associated with national reporting obligations, whether or not an EU legal act specifies them; and those exclusively associated with EU reporting obligations, i.e., that would not be incurred in the absence of EU legislation. The latter are defined as costs related to EU reporting obligations

Example 9.1.B Proposal for Cost Measurement for SBS March 1999 (Eurostat/D2/SSE/MAR99/3.3)

Eurostat proposed use of the following table to measure the costs associated with implementation of the SBS Regulation

1. Number of staff involved

	Units
Total	
of which, professional and managerial	

2. Costs to NSO

		Thousands	% sub-contracted
Staff cost	+		
Data collection costs (printing and mailing)	+		%
Costs for the treatment of non-response (post, telephone, interview)	+		%
Costs for data control (checking, editing) and compilation of results (extrapolation, tabulation, formatting)	+		%
Other costs	+		%
Total costs	=		

Notes

- Number of staff involved should distinguish between National Statistical Institutes, ministries, banks, regional bodies and other agencies or institutes.
- Ideally, numbers should be restricted to directly employed staff and should be presented as full-time equivalents.
- Ideally, the breakdown of non-staff costs should be provided. Where this is not possible, please indicate the total costs and add footnotes regarding omissions, deficiencies, alternative definitions and so on.
- The total costs should be full economic costs, in other words including staff costs, computing, materials, equipment and services used, whether purchased externally or provided from within the organisation.
  Staff costs should include employer's social security costs where this applies.
- Capital expenditure (investment) should be representative of capital consumption not capital expenditure for the financial year for which information is given.
- Estimates should cover all current activity in the chosen area, but footnotes should indicate where this exceeds EU requirements.
- Under % subcontracted indicate the shares of the figures given in the first column that are accounted for by payments to private firms or other Government agencies".

In summary, the quality report should indicate the cost model used, the sources of cost information, and cost estimates, where available.

9.2 Respondent Burden

Over the past decade the EC has been making substantial efforts to reduce the administrative burden placed by legislation, and accompanying regulations, on businesses. A summary of progress is provided in European CommissionDocument on Measuring Administrative Costs 2008. Associated with these activities is the *EU Standard Cost Model* for measuring costs imposed by legislation on businesses. This is the starting point for defining respondent burden, whether imposed on individuals, household members or businesses, by statistical processes.

The overall cost of delivering the information requested by a particular questionnaire depends on three components:

- the number of respondents (R);

- the (average) time (T) required to provide the information, including time spent assembling information prior to completing a questionnaire or taking part in an interview and the time taken up by any subsequent contacts after receipt of the questionnaire; and
- the average hourly cost of a respondent's time (C).

Start-up costs associated with creating systems to comply with the survey and computing costs, etc., are not included.

The total respondent burden for a questionnaire is computed as R*T*C. Summing over all questionnaires for all repetitions of the statistical process over a year, usually a calendar year, provides the annual cost.

The average hourly cost is likely the most difficult of the three parameters to measure, thus response burden carried by respondents is often measured simply in hours spent (R*T) rather than in financial terms.

Sometimes the *number of questionnaires* is used in place of the *number of respondents*, thus giving a (maximum) *design level* measure of respondent burden rather than the burden associated with the actual respondents.

The following paragraphs provide two concrete examples.

Example 9.2.A Measurement of Respondent Burden at the Australian Bureau of Statistics

For every business survey, respondent burden is measured as the product of the number of questionnaires multiplied by the average completion time. For most surveys the final question in the questionnaire asks the respondent for an estimate of the completion time. The average completion time for the survey is then based on the responses received, with outliers being removed. For some surveys, including proposed new surveys, estimates are obtained from focus groups and by in house simulations. The ABS computes the total annual burden over all surveys of businesses and sets targets for reduction.

Example 9.2.B Respondent Burden Measurement within Framework of UKONS Compliance Plan 1998/2000.

This method “is based on consultations with respondents, pilot tests or surveys, or other ways of obtaining evidence of the impact of the survey. Estimation of the compliance costs involves three main components:

- R = the number of respondents;
- T = the time required to provide the information, including time spent assembling information prior to completing a form or taking part in interview and the time taken up by any subsequent contacts after receipt of the questionnaire;
- C = the typical hourly cost of a respondent’s time.
- Cost = RxTxC”

“The measurement of compliance cost should not in itself add significantly to the compliance cost burden. For this reason large-scale surveys are not run to establish the costs of other surveys. A selection of the following methods should be employed, at least every three years:

- Data should be collected from a random sample of contributors, each form-type being collected separately;
- Data should usually be collected by a telephone survey as this is likely to be less burdensome and contributors can be guided through the process;
- An extra question can be added to the questionnaire, particularly for the new form types; although this may be less flexible than direct contact;
- The effect of new initiatives should be measured directly through contacting a sample of contributors.

- The surveys should approach all size-bands and average completion times specific to those size-bands should be calculated wherever possible. […]"

The UKONS also notes that estimation of the same components for non-respondents is useful as it cannot be assumed a priori that the cost to non-respondents is zero.

Methods to reduce respondent burden include limiting the range and detail of data collected to what is absolutely necessary, using administrative sources to the fullest extent possible, ensuring that the data sought from businesses are readily available from their accounts and that approximations are accepted where exact details are not available, using electronic means to facilitate data collection, and limiting the burden on individual respondents by minimizing overlap of respondents with other surveys.

In summary the quality report should indicate the measure taken to minimise respondent burden, the respondent burden measurement model, respondent burden estimates and the sources of this information.

9.3 Quality and Performance Indicators and Summary

Quality and Performance Indicators

PCR1. Annual operational cost, with breakdown by major cost components.

Definition: direct costs of staff involved in data collection (questionnaires, distribution, capture), reducing non-response, processing, and compilation of estimates.

To be further defined for subject-matter domain: precise breakdown by cost component.

PCR2. Annual respondent burden in hours and/or financial terms.

Definition: respondent burden in hours is defined as number of respondents/questionnaires * average time per respondent summed over all repetitions of statistical process within a year; respondent burden in financial terms defined as respondent burden in hours * average hourly cost to respondents.

To be further defined for subject-matter domain: precise cost model.

Summary

What should be reported on <u>Performance, Cost and Respondent Burden</u>

<u>Performance and Cost</u>

- Annual operational cost with breakdown by major cost component.
- Recent efforts made to improve efficiency.
- The procedures for internal assessment and for independent external assessment of efficiency.
- The extent to which routine operations, in particular data capture, coding, validation and imputation, are automated.

- The extent to which ICT is effectively used for used for data collection and dissemination and the improvements that could be made.

Respondent Burden

- Annual respondent burden in financial terms and/or hours.
- Respondent burden reduction targets.
- Recent efforts made to reduce respondent burden.
- Whether the range and detail of data collected by survey is limited to what is absolutely necessary.
- Whether administrative and other survey sources are used to the fullest extent possible.
- The extent to which data sought from businesses is readily available from their accounts.
- Whether electronic means are used to facilitate data collection.
- Whether best estimates and approximations are accepted when exact details are not readily available.
- Whether reporting burden on individual respondents is limited to the extent possible by minimizing the overlap with other surveys.

10 Confidentiality, Transparency and Security

European Statistics Code of Practice

Principle 5. The privacy of data providers (households, enterprises, administrations and other respondents), the confidentiality of the information they provide and its use only for statistical purposes must be absolutely guaranteed.

Principle 6. Statistical authorities must produce and disseminate European statistics respecting scientific independence and in an objective, professional and transparent manner in which all users are treated equitably.

10.1 Confidentiality

Typically confidentiality protection is required by law and survey staff have legal confidentiality commitments. The quality report should confirm such arrangements or report on any exceptions. It should also outline the procedures for ensuring confidentiality during collection, processing and dissemination. These include protocols for ensuring that individual data are accessed strictly on a need to know basis, rules for defining confidential cells in output tables, and procedures for detecting and preventing residual disclosure. In addition, the arrangements, if any, under which users outside the NSO may access microdata for research purposes, and the associated confidentiality provisions, should be described.

10.2 Transparency

Respondents should be informed of the mandate under which a survey is being conducted, the uses to which the data requested are to be put, and the confidentiality provisions.

Statistical announcements and statements made in press conferences should be objective and non-partisan. Users should be informed about the accuracy and coherence of the statistical outputs and corresponding limitations in their use. Sources and methods metadata should be published with the data so that users are in a better position to determine the utility of the data for themselves. Mistakes discovered in published statistics should be corrected and publicised.

In so far as they are not already covered in Chapter 5, the quality report should describe the arrangements to ensure these requirements are met and should draw attention to any deficiencies.

10.3 Security

The report should describe the provisions in place to ensure the security of:

- data acquisition processes, in particular data collected through the Internet;
- storage of survey forms, how long they are kept and how they are disposed of ;
- encryption of data on laptops and other devices;

- storage of micro-data in databases;
- storage of aggregate data in databases and on output CD’s.

There are trade-offs to consider between confidentiality and security on the one hand and accessibility on the other, and these should be described.

10.4 Quality and Performance Indicators and Summary

Quality and Performance Indicators

None explicitly defined.

Summary

What should be included on <u>Confidentiality, Transparency and Security</u>

<u>Confidentiality</u>

- Whether or not confidentiality is required by law and if so whether survey staff have signed legal confidentiality commitments.
- Whether external users may access micro-data for research purposes, and, if so, the confidentiality provisions that are applied.
- The procedures for ensuring confidentiality during collection, processing and dissemination, including rules for determining confidential cells in output tables and procedures for detecting and preventing residual disclosure.

<u>Transparency</u>

- The ways in which the uses to which the data are to be put and confidentiality provisions are made known to respondents.
- Whether statistical announcements and statements made in press conferences are objective and non-partisan.
- Whether errors discovered in published statistics are corrected and publicised.

<u>Security</u>

- The provisions in place to ensure the security of data acquisition processes, in particular data collected through the Internet.
- The provisions that are in place to ensure the security and integrity of completed questionnaires, micro and macro databases and data outputs.

11 Conclusion

The quality report should conclude with a section referring to the principal quality problems, each of which should be accompanied by recommendations for improvements. The intended users and uses of the quality report and the follow-up action items should be indicated.

Quality and Performance Indicators

None explicitly defined.

Summary

What should be included in the <u>Conclusion</u>

- Principal quality problems.
- Recommendations for improvements.
- Follow-up action items.

PART III: REFERENCE MATERIAL

1 Bibliography

1.1 International Quality Standards

ISO 9000 series

- *ISO 9000: 2005 Quality Management Systems - Fundamentals and Vocabulary* describes the fundamentals of quality management systems, including the principles associated with total quality management, quality related terminology and quality models.
- *ISO 9004: 2000 Quality Management Systems – Guidelines* is designed to assist organisations having mature quality management systems to obtain performance improvements.
- *ISO 9001: 2000 Quality Management Systems – Requirements* specifies the required characteristics of a quality management system and is the standard with respect to which organisations may seek to be certified. The requirements are presented under five headings as follows.

EFQM Excellence Model

The European Foundation for Quality Management has produced the EFQM Excellence Model, which is in effect a quality standard somewhat similar to the ISO 9000 series.

ISO 20252:2006 Market, Opinion and Social Research – Vocabulary and Service Requirements

This new international standard establishes terminology and service requirements for organizations conducting market, opinion and social research and sets a common level of quality for market research globally.

UN Principles

The UN Fundamental Principles of Official Statistics for national statistical systems were promulgated by the UN Statistical Commission in 1994. (Copy annexed for ease of reference.)

The UN Principles Governing International Statistical Activities were adopted by the UN Committee for the Coordination of Statistical Activities in 2005. (Copy annexed for ease of reference.)

1.2 ESS Quality Reference Documents

The ESS Quality Declaration was adopted by the Statistical Programme Committee in 2001 as a formal step towards total quality management in the ESS, in line with the EFQM Excellence

Model. It is just one and a half pages long, comprising the ESS mission statement, the ESS vision statement and ten principles (copy annexed for ease of reference).

The Statistical Law (Regulation 322/97 of 17 Feb. 1997) makes provision for the production of Community Statistics. Article 10 states that "*In order to ensure the best possible quality in both deontological and professional aspects, Community statistics shall be governed by the principles of impartiality, reliability, relevance, cost-effectiveness, statistical confidentiality and transparency*" and it then defines these terms.

With promulgation of the European Statistics Code of Practice (CoP) by the European Commission in 2005, Eurostat and the NSIs of the EU Member States are committed to fifteen principles covering the *institutional environment, statistical processes* and *outputs*. For each principle the ESCP defines a set of indicators reflecting good practice and providing a basis for assessing implementation. (Copy annexed for ease of reference.)

The ESS Quality Definition was articulated in a document presented at the October 2003 meeting of the ESS Working Group *Assessment of Quality in Statistics*. Although never given a more formal status, this document has provided the basis for defining output quality components in all subsequent quality related documents, including the CoP. The document would benefit from a minor revision to clarify the concepts in the light of recent experience.

How to make a Quality Report (HtMaQR) was presented to the same ESS Working Group meeting in October 2003. Its purpose was to provide guidelines in reporting quality in the ESS. In addition to presenting reporting guidelines, it discusses quality concepts, costs, and EU legislation relating to the quality of statistics. It makes extensive use of examples drawn from Member States and includes a substantial bibliography. It is superseded by the ESQR.

ESS Standard Quality Report (ESQR) was presented at the same ESS Working Group meeting in October 2003. Its purpose was to provide abbreviated guidelines on reporting quality. It is a distillation from the more comprehensive HtMaQR. It is superseded by the ESQR.

The ESS Quality Glossary was also presented at the same ESS Working Group meeting in October 2003 as a companion document to the HtMaQR. It covers the main technical terms used in the HtMaQR, providing a short definition of each term and indicating the source of the definition. The document would benefit from revision, first, as a number of key terms are not included and, second, in the light of other glossaries that have subsequently appeared.

Standard Quality Indicators was presented to the May 2005 meeting of the ESS Working Group *Assessment of Quality in Statistics.* Its purpose was to present a standard set of indicators for use by producers in summarizing the quality of their statistical outputs. It is aligned with the ESQR. It will be essentially superseded by the ESQR and EHQR as they will incorporate all the indicators that have proven useful based on experience.

The European Self Assessment Checklist for Survey Managers (DESAP) is a tool for ESS survey managers that enables the conduct of quick but systematic and comprehensive quality assessments of statistical processes and outputs and identification of potential improvements. It is compliant with the ESS Quality Definition and is generic in the sense that it applies to all statistical processes irrespective of the specific subject matter area or the survey methodology. It is totally compatible with, and, in essence, a checklist version of, the EHQR.

There is a Condensed Version of Checklist for survey managers (DESAP) that contains only selected key questions and there is an Electronic Version of Self Assessment Checklist with a Electronic Version User Guide.

The ESS Handbook on improving quality by analysis of process variables describes a general approach and useful tools for identifying, measuring and analysing key process variables. It includes practical examples of the application of the approach to various statistical processes.

The ESS Handbook on Data Quality - Assessment Methods and Tools details the full range of methods for assessing process and output quality and the tools that support them. It makes recommendations on how these methods and tools can be implemented. It is primarily targeted at ESS quality managers, enabling them to introduce, systematise and improve the work carried out in the field of data quality management.

1.3 Other References

Other References

AAPOR (2008): *Standard Definitions. Final Dispositions of Case Codes and Outcome Rates for Surveys,* The American Association for Public Opinion Research.

Andersson, C., Lindström, H. and Polfeldt, T. (1999): *Att mäta statistikens kvalitet,* (in Swedish), Statistics Sweden, R&D Report 1999:3.

Berthier, C. and Néros, B. (1998) *Une méthode de mesure de l'effet enquêteur*, (in French), paper presented at the Journées de méthodologie Statistique de l'INSEE, Paris, France, March 17-18.

Biemer, P. and Stokes, L. (1991): *Approaches to Modeling Measurement Error,* in Biemer, P. Groves, R.M. Lyberg, L.E. Mathiowetz, N.A. and Sudman, S., (eds.) (1991) *Measurement Errors in Surveys*. John Wiley & Sons, New York.

Bureau of the Census (2002): Census 2000 Testing, Experimentation and Evaluation Program.

Chenu, A. and Guglielmetti, F. (2000) *Coder la profession, nouvelles procédures, nouveaux enjeux,* (in French), paper presented at the *Journées de méthodologie Statistique de l' INSEE*, Paris, France, December 4-5, 2000.

Dalén, J. (1981): *Metoder för evalvering av noggrannheten i SCBs statistik. En översikt,* (in Swedish), Statistics Sweden, Promemorior från P/STM nr 6.

Djerf, K. (1997). *Effects of Post-stratification on the Estimates of the Finnish Labour Force Survey*, *Journal of Official Statistics*, Vol. 13, No. 1, 29-39.

European Central Bank (2006): *Euro Area Balance of Payments and International Investment Position Statistics.*

EU-LFS Quality Report (2007): Quality Report of the European Union Labour Force Survey 2005.

Eurostat (2005): Serious accidents at work.

Eurostat (2007): *Greenhouse gas emissions, Metadata in SDDS format: Summary Methodology*

Eurostat (2007): *Healthy Life Years Expectancy, disability-free life expectancy – DFLE*

Eurostat (2007): Quality Report on International Trade Statistics, Eurostat – International Trade

Fixler D. and Grimm B. (2007): *Reliability of the NIPA Estimates of U.S. Economic Activity, Survey of Current Business*, February 2005, pp. 8-19.

Holt T., and Jones T. (1998): Quality work and conflicting quality objectives, paper presented at the *84th DGINS conference*, Stockholm, Sweden, May 28-29, 1998.

Hörngren, J. (1992): The Use of Registers as Auxiliary Information in the Swedish Labour Force Survey, Statistics Sweden, R&D Report 1992:13.

International Labour Office (2004) Consumer Price Index Manual: Theory and Practice, International Working Group on Price Statistics.

Meader R. and Tily G. (2007): *Monitoring the quality of the National Accounts*, Economic and Labour Market Review, Vol 2, No 3, 24-33.

Medin, K. and Wilson, B. (1994): *Quality assurance and quality assessment in official statistics: the case of Sweden's panel census of agriculture*, Statistical journal of the United Nations, **11**, 125-145.

Morgenstern, O. (1965): *On the Accuracy of Economic Observations*, Second Edition, Princeton University Press, Princeton, New Jersey.

Nordberg, L. (2000): *On Variance Estimation for Measures of Change. When Samples are Coordinated by the Use of Permanent Random Numbers*, Journal of Official Statistics, 16, 363-378.

OECD et al (2002): Measurement of the Non-Observed Economy: A Handbook.

Prewitt, K. (2000): *Statement on the feasibility of using statistical methods to improve the accuracy of census 2000*, US Bureau of the Census. Washington DC. Available online at:

Särndal, C.-E., Swensson, B. and Wretman, J. (1992): *Model assisted survey sampling*, Springer-Verlag, New York.

Sorrentino, C. (2000): *International Unemployment Rates: How Comparable Are They?* Monthly Labor Review, 123, 6, pp.3-2

Statistics Canada (2003): Statistics Canada Quality Guidelines, 4th edition, October 2003.

Statistics Finland (2007): Quality Guidelines for Official Statistics.

Statistics Sweden (2001): *The Swedish Consumer Price Index. A handbook of methods*.

Statistikersamfundet (2005): *STANDARD FÖR BORTFALLSBERÄKNING*, (in Swedish).

Swedish Environment Protection Agency (2006): *Quality Report for Statistics on Generation and Recovery and Disposal of Waste in Sweden 2004 According to EU Regulation on Waste Statistics 2150/2002.*

UK Office for National Statistics (1999a): *Comparison of variance estimation software and methods,* Model Quality Report in Business Statistics Vol 2, Eds. Davies P. and Smith P.

UK Office for National Statistics (1999b): *Model Quality Reports,* Model Quality Report in Business Statistics Vol 3, Eds. Davies P. and Smith P.

UK Office for National Statistics (2007): *United Kingdom National Accounts The Blue Book 2007.*

United Nations Statistical Commission and Economic Commission for Europe (1994): *Statistical Data Editing Volume No 1 Methods and Techniques*

US Office of Management and Budget (2001): *Measuring and Reporting Sources of Error in Surveys* Statistical Policy Working Paper 31, Statistical Policy Office.

Vale, S. (2008): *Accessibility and clarity: The most neglected dimensions of quality?* Presented at Conference on Data Quality for International Organizations, Rome, 7-8 July 2008.

Wallgren A. and Wallgren B. (2007): *Register-Based Statistics. Administrative Data for Statistical Purposes*, Wiley.

Wood, J. (2008): *On the Covariance Between Related Horvitz-Thompson Estimators*, Journal of Official Statistics, 24, 53-78.

Zarkovich S. (1966): *Quality of Statistical Data*, Food and Agriculture Organisation of the United Nations, Rome.

Zilhao, M. J. (2001) *Quality Indicators Panel.* Available from the “Quality in Statistics”, CIRCA Interest Group

2 Copies of Key Documents

2.1 UN Fundamental Principles of Official Statistics

The Statistical Commission

- Bearing in mind that official statistical information is an essential basis for development in the economic, demographic, social and environmental fields and for mutual knowledge and trade among the States and peoples of the world,
- Bearing in mind that the essential trust of the public in official statistical information depends to a large extent on respect for the fundamental values an principles which are the basis of any society which seeks to understand itself and to respect the rights of its members,
- Bearing in mind that the quality of official statistics, and thus the quality of the information available to the Government, the economy and the public depends largely on the cooperation of citizens, enterprises, and other respondents in providing appropriate and reliable data needed for necessary statistical compilations and on the cooperation between users and producers of statistics in order to meet users' needs,
- Recalling the efforts of governmental and non-governmental organizations active in statistics to establish standards and concepts to allow comparisons among countries,
- Recalling also the International Statistical Institute Declaration of Professional Ethics,
- Having expressed the opinion that resolution C (47), adopted by the Economic Commission for Europe on 15 April 1992, is of universal significance,
- Noting that, at its eighth session, held in Bangkok in November 1993, the Working Group of Statistical Experts, assigned by the Committee on Statistics of the Economic and Social Commission for Asia and the Pacific to examine the Fundamental Principles, had agreed in principle to the ECE version and had emphasized that those principles were applicable to all nations,
- Noting also that, at its eighth session, held at Addis Ababa in March 1994, the Joint Conference of African Planners, Statisticians and Demographers, considered that the Fundamental Principles of Official Statistics are of universal significance

Adopts the present principles of official statistics:

Principle 1. Official statistics provide an indispensable element in the information system of a democratic society, serving the Government, the economy and the public with data about the economic, demographic, social and environmental situation. To this end, official statistics that meet the test of practical utility are to be compiled and made available on an impartial basis by official statistical agencies to honour citizens' entitlement to public information.

Principle 2. To retain trust in official statistics, the statistical agencies need to decide according to strictly professional considerations, including scientific principles and professional ethics, on the methods and procedures for the collection, processing, storage and presentation of statistical data.

Principle 3. To facilitate a correct interpretation of the data, the statistical agencies are to present information according to scientific standards on the sources, methods and procedures of the statistics.

Principle 4. The statistical agencies are entitled to comment on erroneous interpretation and misuse of statistics.

Principle 5. Data for statistical purposes may be drawn from all types of sources, be they statistical surveys or administrative records. Statistical agencies are to choose the source with regard to quality, timeliness, costs and the burden on respondents.

Principle 6. Individual data collected by statistical agencies for statistical compilation, whether they refer to natural or legal persons, are to be strictly confidential and used exclusively for statistical purposes.

Principle 7. The laws, regulations and measures under which the statistical systems operate are to be made public.

Principle 8. Coordination among statistical agencies within countries is essential to achieve consistency and efficiency in the statistical system.

Principle 9. The use by statistical agencies in each country of international concepts, classifications and methods promotes the consistency and efficiency of statistical systems at all official levels.

Principle 10. Bilateral and multilateral cooperation in statistics contributes to the improvement of systems of official statistics in all countries.

2.2 ESS Quality Declaration

The mission of the European Statistical System

"We provide the European Union and the World with high quality information on the economy and society at the European, national and regional levels and make the information available to everyone for decision-making purposes, research and debate."

The vision of the European Statistical System

"The ESS will be a world leader in statistical information services and the most important information provider for the European Union and its member states. Based on scientific principles and methods, the ESS will offer and continuously improve a programme of harmonized European statistics that constitutes an essential basis for democratic processes and progress in society."

User focus

We provide our users with products and services that meet their needs. The articulated and non-articulated needs, demands and expectations of external and internal users will guide the ESS, its members, their employees and operations.

Continuous improvement

The needs and demands of users will change as will the environment we operate in. Globalisation and advances in methods and technology will avail new possibilities. It is imperative that we actively strive to improve our work methods to take advantage of the new possibilities and to better meet the demands of our users.

Product quality commitment

We produce high quality statistical information according to scientific methods in accordance with objectivity and confidentiality. We provide information on the main quality characteristics of each product so that users are able to assess product quality.

Accessibility of information

We provide statistical results in a user-friendly and accessible form. Utilizing the possibilities of new media ensures easy access to the information. As far as possible, we will enhance user awareness of the strengths and limitations of the produced statistics. Consulting on how to use data is an integral part of dissemination.

Partnership with and beyond the European Statistical System

The cooperation between current and future members of the ESS as well as with other organisations will be encouraged. Only by working together, can we learn from others and gradually develop our system. The broad knowledge of staff and our users, suppliers, partners and other parties must be combined for us to excel in our purpose.

Respects for the needs of data providers

The suppliers of data for statistics – the respondents – are an especially important group with which a mutually rewarding partnership must be established. The producers of statistics should strive to always minimise the respondent burden, both the objective and the perceived burden.

Commitment of leadership

The leaders of the organisations in the ESS exercise a personal, active, and visible leadership to create and sustain a culture of quality. By providing a clear overall direction, prioritizing improvement activities and stimulating empowerment and innovation, leaders enable the staff to perform a successful job and to continuously strive for improvement.

Systematic quality management

We systematically and regularly identify strengths and weaknesses in all relevant areas to continuously identify and implement improvements where needed. A long-term strategic orientation is vital for the development of the ESS. The long-term effects in all situations must be considered with the more obvious short-term effects.

Effective and efficient processes

ESS activities should be seen as processes that create value for the users. We work efficiently to produce output with as little resources as possible and to prevent errors in the processes and products. The processes and their quality are continuously reviewed and improved.

Staff satisfaction and development

To attract and keep competent staff, it is vital to satisfy staff needs. The ESS members should treat their employees as the key resources they are.

2.3 European Statistics Code of Practice

For the national and community statistical authorities

Adopted by the Statistical Programme Committee on 24 February 2005 and promulgated in the Commission Recommendation of 25 May 2005 on the Independence, Integrity and Accountability of the National and Community Statistical Authorities

The European Statistics Code of Practice is based on 15 principles. Governance authorities and statistical authorities in the European Union commit themselves to adhering to the principles fixed in this code covering the institutional environment, statistical processes and outputs. A set of indicators of good practice for each of the 15 principles provides a reference for reviewing the implementation of the Code.

Institutional environment

Institutional and organisational factors have a significant influence on the effectiveness and credibility of a statistical authority producing and disseminating European statistics. The relevant issues are professional independence, mandate for data collection, adequacy of resources, quality commitment, statistical confidentiality, impartiality and objectivity.

PRINCIPLE 1: PROFESSIONAL INDEPENDENCE

The professional independence of statistical authorities from other policy, regulatory or administrative departments and bodies, as well as from private sector operators, ensures the credibility of European statistics.

Indicators

- The independence of the statistical authority from political and other external interference in producing and disseminating official statistics is specified in law.
- The head of the statistical authority has sufficiently high hierarchical standing to ensure senior-level access to policy authorities and administrative public bodies. He/she should be of the highest professional calibre.
- The head of the statistical authority and, where appropriate, the heads of its statistical bodies have responsibility for ensuring that European statistics are produced and disseminated in an independent manner.
- The head of the statistical authority and, where appropriate, the heads of its statistical bodies have the sole responsibility for deciding on statistical methods, standards and procedures, and on the content and timing of statistical releases.
- The statistical work programmes are published, and periodic reports describe progress made.
- Statistical releases are clearly distinguished and issued separately from political/policy statements.
- The statistical authority, when appropriate, comments publicly on statistical issues, including criticisms and misuses of official statistics.

PRINCIPLE 2: MANDATE FOR DATA COLLECTION

Statistical authorities must have a clear legal mandate to collect information for European statistical purposes. Administrations, enterprises and households, and the public at large may be compelled by law to allow access to or deliver data for European statistical purposes at the request of statistical authorities.

Indicators

- The mandate to collect information for the production and dissemination of official statistics is specified in law.
- The statistical authority is allowed by national legislation to use administrative records for statistical purposes.
- On the basis of a legal act, the statistical authority may compel response to statistical surveys.

PRINCIPLE 3: ADEQUACY OF RESOURCES

The resources available to statistical authorities must be sufficient to meet European statistics requirements.

Indicators

- Staff, financial, and computing resources, adequate both in magnitude and in quality, are available to meet current European statistics needs.
- The scope, detail and cost of European statistics are commensurate with needs.
- Procedures exist to assess and justify demands for new European statistics against their cost.
- Procedures exist to assess the continuing need for all European statistics, to see if any can be discontinued or curtailed to free up resources.

PRINCIPLE 4: QUALITY COMMITMENT

All ESS members commit themselves to work and cooperate according to the principles fixed in the ‘Quality declaration of the European statistical system’.

Indicators

- Product quality is regularly monitored according to the ESS quality components.
- Processes are in place to monitor the quality of the collection, processing and dissemination of statistics.
- Processes are in place to deal with quality considerations, including trade-offs within quality, and to guide planning for existing and emerging surveys.
- Quality guidelines are documented and staff are well trained. These guidelines are spelled out in writing and made known to the public.
- There is a regular and thorough review of the key statistical outputs using external experts where appropriate.

PRINCIPLE 5: STATISTICAL CONFIDENTIALITY

The privacy of data providers (households, enterprises, administrations and other respondents), the confidentiality of the information they provide and its use only for statistical purposes must be absolutely guaranteed.

Indicators

- Statistical confidentiality is guaranteed in law.
- Statistical authority staff sign legal confidentiality commitments on appointment.
- Substantial penalties are prescribed for any wilful breaches of statistical confidentiality.
- Instructions and guidelines are provided on the protection of statistical confidentiality in the production and dissemination processes. These guidelines are spelled out in writing and made known to the public.
- Physical and technological provisions are in place to protect the security and integrity of statistical databases.
- Strict protocols apply to external users accessing statistical microdata for research purposes.

PRINCIPLE 6: IMPARTIALITY AND OBJECTIVITY

Statistical authorities must produce and disseminate European statistics respecting scientific independence and in an objective, professional and transparent manner in which all users are treated equitably.

Indicators

- Statistics are compiled on an objective basis determined by statistical considerations.
- Choices of sources and statistical techniques are informed by statistical considerations.
- Errors discovered in published statistics are corrected at the earliest possible date and publicised.
- Information on the methods and procedures used by the statistical authority are publicly available.
- Statistical release dates and times are pre-announced.
- All users have equal access to statistical releases at the same time and any privileged pre-release access to any outside user is limited, controlled and publicised. In the event that leaks occur, pre-release arrangements should be revised so as to ensure impartiality.
- Statistical releases and statements made in press conferences are objective and non-partisan.

Statistical processes

European and other international standards, guidelines and good practices must be fully observed in the processes used by the statistical authorities to organise, collect, process and disseminate official statistics. The credibility of the statistics is enhanced by a reputation for good management and efficiency. The relevant aspects are sound methodology, appropriate statistical procedures, non-excessive burden on respondents and cost effectiveness.

PRINCIPLE 7: SOUND METHODOLOGY

Sound methodology must underpin quality statistics. This requires adequate tools, procedures and expertise.

Indicators

- The overall methodological framework of the statistical authority follows European and other international standards, guidelines and good practices.
- Procedures are in place to ensure that standard concepts, definitions and classifications are consistently applied throughout the statistical authority.
- The business register and the frame for population surveys are regularly evaluated and adjusted if necessary in order to ensure high quality.
- Detailed concordance exists between national classifications and sectorisation systems and the corresponding European systems.
- Graduates in the relevant academic disciplines are recruited.
- Staff attend international relevant training courses and conferences, and liaise with statistician colleagues at international level in order to learn from the best and to improve their expertise.
- Cooperation with the scientific community to improve methodology is organised and external reviews assess the quality and effectiveness of the methods implemented and promote better tools, when feasible.

PRINCIPLE 8: APPROPRIATE STATISTICAL PROCEDURES

Appropriate statistical procedures, implemented from data collection to data validation, must underpin quality statistics.

Indicators

- Where European statistics are based on administrative data, the definitions and concepts used for the administrative purpose must be a good approximation to those required for statistical purposes.
- In the case of statistical surveys, questionnaires are systematically tested prior to the data collection.
- Survey designs, sample selections, and sample weights are well based and regularly reviewed, revised or updated as required.
- Field operations, data entry, and coding are routinely monitored and revised as required.
- Appropriate editing and imputation computer systems are used and regularly reviewed, revised or updated as required.
- Revisions follow standard, well-established and transparent procedures.

PRINCIPLE 9: NON-EXCESSIVE BURDEN ON RESPONDENTS

The reporting burden should be proportionate to the needs of the users and should not be excessive for respondents. The statistical authority monitors the response burden and sets targets for its reduction over time.

Indicators

- The range and detail of European statistics demands is limited to what is absolutely necessary.
- The reporting burden is spread as widely as possible over survey populations through appropriate sampling techniques.
- The information sought from businesses is, as far as possible, readily available from their accounts and electronic means are used where possible to facilitate its return.
- Best estimates and approximations are accepted when exact details are not readily available.
- Administrative sources are used whenever possible to avoid duplicating requests for information.
- Data sharing within statistical authorities is generalised in order to avoid multiplication of surveys.

PRINCIPLE 10: COST EFFECTIVENESS

Resources must be effectively used.

Indicators

- Internal and independent external measures monitor the statistical authority's use of resources.
- Routine clerical operations (e.g. data capture, coding and validation) are automated to the extent possible.
- The productivity potential of information and communications technology is being optimised for data collection, processing and dissemination.
- Proactive efforts are being made to improve the statistical potential of administrative records and avoid costly direct surveys.

Statistical output

Available statistics must meet users' needs. Statistics comply with European quality standards and serve the needs of European institutions, governments, research institutions, business concerns and the public generally. The important issues concern the extent to which the statistics are relevant, accurate and reliable, timely, coherent, comparable across regions and countries, and readily accessible by users.

PRINCIPLE 11: RELEVANCE

European statistics must meet the needs of users.

Indicators

- Processes are in place to consult users, monitor the relevance and practical utility of existing statistics in meeting their needs, and advise on their emerging needs and priorities.
- Priority needs are being met and reflected in the work programme.
- User satisfaction surveys are undertaken periodically.

PRINCIPLE 12: ACCURACY AND RELIABILITY

European statistics must accurately and reliably portray reality.

Indicators

- Source data, intermediate results and statistical outputs are assessed and validated.
- Sampling errors and non-sampling errors are measured and systematically documented according to the framework of the ESS quality components.
- Studies and analyses of revisions are carried out routinely and used internally to inform statistical processes.

PRINCIPLE 13: TIMELINESS AND PUNCTUALITY

European statistics must be disseminated in a timely and punctual manner.

Indicators

- Timeliness meets the highest European and international dissemination standards.
- A standard daily time is set for the release of European statistics.
- Periodicity of European statistics takes into account user requirements as much as possible.
- Any divergence from the dissemination time schedule is publicised in advance, explained and a new release date set.
- Preliminary results of acceptable aggregate quality can be disseminated when considered useful.

PRINCIPLE 14: COHERENCE AND COMPARABILITY

European statistics should be consistent internally, over time and comparable between regions and countries; it should be possible to combine and make joint use of related data from different sources.

Indicators

- Statistics are internally coherent and consistent (e.g. arithmetic and accounting identities observed).
- Statistics are coherent or reconcilable over a reasonable period of time.
- Statistics are compiled on the basis of common standards with respect to scope, definitions, units and classifications in the different surveys and sources.

- Statistics from the different surveys and sources are compared and reconciled.
- Cross-national comparability of the data is ensured through periodical exchanges between the European statistical system and other statistical systems; methodological studies are carried out in close cooperation between the Member States and Eurostat.

PRINCIPLE 15: ACCESSIBILITY AND CLARITY

European statistics should be presented in a clear and understandable form, disseminated in a suitable and convenient manner, available and accessible on an impartial basis with supporting metadata and guidance.

- Statistics are presented in a form that facilitates proper interpretation and meaningful comparisons.
- Dissemination services use modern information and communication technology and, if appropriate, traditional hard copy.
- Custom-designed analyses are provided when feasible and are made public.
- Access to microdata can be allowed for research purposes. This access is subject to strict protocols.
- Metadata are documented according to standardised metadata systems.
- Users are kept informed on the methodology of statistical processes and the quality of statistical outputs with respect to the ESS quality criteria.

3 ESS Legislation Referring to Quality Reporting

3.1 Labour Force Survey

Council Regulation (EC) N° 577/98 of 9 March 1998 (OJ L77/3 14.3.1998) on the organisation of a labour force sample survey in the Community describes the main features of the survey. Article 3 deals with the representativeness of the sample and states that:

- for certain sub-populations, the relative standard error for the estimation of annual averages shall not exceed a certain percentage, and
- in the case where non response results in missing data, a method of statistical imputation shall be applied where appropriate.

3.2 Short-term Business Statistics

Council Regulation (EC) N° 1165/98 of 19 May 1998 contains only one article dealing with quality, namely Article 10 entitled "Quality", which lays down that:

- Member States shall ensure that the transmitted variables reflect the population of units …
- The quality of the variables shall be measured by each Member State according to common criteria.
- The quality of the variables shall be tested regularly by comparing them with other statistical information. In addition they shall be checked for internal consistency.
- Quality evaluation shall be carried out comparing the benefits of the availability of the data with the costs of collection and the burden on businesses, especially on small enterprises.

3.3 Structural Business Statistics

In this context there is specific regulation concerning quality, namely Commission Regulation N° 1618/1999 of 23 July 1999, (OJ L192, 24.7.1999, p11) entitled "Regulation concerning the criteria for the evaluation of quality of structural business statistics".

The Regulation requires the collection/compilation of coefficients of variation, unit nonresponse rates, and item nonresponse rates for certain data items defined in terms of NACE classes. In additional, it requires that Member States provide "Specific reports" on the survey's strategy and practices for determining the principal activity of the observed units.

3.4 Labour Costs Statistics

Council Regulation (EC) N° 530/1999 concerns statistics on earnings and on labour costs. It is supplemented by Commission Regulation (EC) N° 452/2000 of 28 February 2000 (OJ L55, 29.2.2000, p53) which defines the quality evaluation requirements. It is based on seven quality components identified by Eurostat at that time. For each of these, the Regulation defines a set quality indicators that Member States should collect and transmit to Eurostat. Some indicators are mandatory, others optional. They include the biases due to the estimation methods and the coefficients of variation for certain data items.

3.5 Labour Costs Index

Regulation (EC) No 450/2003 of the European Parliament and of the Council concerns the labour cost index. It is supplemented by Commission Regulation (EC) No 1216/2003 of 7 July 2003 which requires a full assessment for each of the quality components

3.6 Other Regulations.

Other regulations which include an explicit provision for quality reporting include:

- Regulation (EC) No 2150/2002 of the European Parliament and of the Council of 25 November 2002 on waste statistics.
- Regulation (EC) No 91/2003 of the European Parliament and of the Council of 16 December2002 on rail transport statistics.
- Regulation (EC) No 1177/2003 of the European Parliament and of the Council of 16 June 2003 concerning Community statistics on income and living conditions (EU-SILC).
- Council Regulation (EC, EURATOM) No 1287/2003 of 15 July 2003 on the harmonisation of gross national income at market prices (GNI Regulation).

4 Glossary of Terms

Definitions of the terms are in line with the ESS Glossary. Where a term is not in the ESS Glossary its definition is drawn from another international source where available, such as the Metadata Common Vocabulary (MCV), otherwise it is created for this document.

Term	**Definition**	**Source**
Accessibility	A measure of the ease with which users can *obtain* a statistical output.	ESS Glossary, fine tuned for this Handbook.
Accuracy	The degree of closeness of estimates to the true values.	ESS Glossary, fine-tuned for this Handbook.
Administrative statistics	(In this Handbook) same as *Statistical process using administrative source(s).*	
Census	Special case of sample survey where all frame units are covered.	
Clarity	Measure of the ease with which users can understand a statistical output.	ESS Glossary, fine-tuned for this Handbook.
Coherence	The degree to which the statistical processes by which statistical outputs were generated used the same concepts - classifications, definitions, and target populations – and harmonised methods. Coherent statistical outputs have the potential to be validly combined and used jointly.	Based on ESS Glossary, fine-tuned for this Handbook.
Comparability	Special case of coherence where the statistical outputs refer to the same data items and the aim of combining them is to make comparisons over time, or across regions, or across other domains.	Based on ESS Glossary, adapted for this Handbook.
CV (coefficient of variation)	The standard error of the estimator divided by the expected value of the estimator. In practice CVs are estimated by dividing the estimated standard error by the estimate itself.	
Domain =Broad statistical domain	A larger subject-matter area to which a certain statistical process belongs, e.g. tourism, education, agriculture.	
Domain (domain of estimation)=subdomain	A subset of the target population for a statistical process for which estimates are made	
Economic index	A price index, a volume index or (rarely) a productivity index.	
Evaluation	An assessment of a quality component – here usually accuracy or a component of accuracy.	

Indicator	A single estimate of a target statistical concept. For example total employment of women in the restaurant industry in a country.	
Key indicator	An estimate of a target statistical concept that is considered of special importance by the users. For example national unemployment, GDP growth, 12-month inflation rate.	
National Statistical Institute (NSI)	The agency that plays the lead role in a national statistical system.	
National Statistical Office (NSO)	The National Statistical Institute or any other national agency or unit that produces official statistics of relevance to the European Statistical System (ESS).	
Output	The production result, to be accessed by the user.	
Price index	An indicator showing average price change for a defined set of products.	
Punctuality	Time lag between the release date of data and the target date on which they were scheduled for release as announced in an official release calendar, laid down by Regulations or previously agreed among partners.	
Quality component	Quality has many aspects that need to be reported separately. Each such aspect is called a component.	
Quality indicator	An indicator of the degree of quality of a certain quality component.	
Relevance	The degree to which statistical outputs meet current and potential user needs - depends on whether all the statistics that are needed are produced and the extent to which concepts used (definitions, classifications etc.,) reflect user needs.	Based on ESS Glossary, adapted for this Handbook
Sample survey	Survey based on a usually probabilistic sampling procedure involving direct collection of data from respondents.	
Statistical compilation	Aggregates, usually economic, that are compiled using basic statistics from a variety of primary sources.	
Statistical output	Results from a statistical process to be accessed by the final users.	
Statistical process	Sample survey, census, use of administrative data, production of price or other economic index, or any other statistical compilation commonly performed by a national statistical office.	

Statistical process using administrative source(s)	Statistical process that makes use of data collected for other purposes than direct production of statistics.	
Timeliness	The length of time between the event or phenomenon that a statistical output describes and its availability.	ESS Glossary, adapted for this Handbook.
Variable	A target concept to be estimated by the statistical process. Typically there will be a set of indicators for each variable, one for each subdomain.	

5 Quality and Performance Indicators

The following table contains a list of quality and performance indicators, each with brief description, and hyperlinked to more a detailed description and computational formulae (if any) in the main body of document.

Identification and Name (hyperlink)	**Brief Description**
Relevance	
R1. Rate of available statistics.	The ratio of the number of output data elements provided in accordance with a relevant ESS regulation to those required by the regulation.
Accuracy	
A1. Coefficient of variation (CV).	The standard error of the estimator divided by the expected value of the estimator.
A2. Rate of overcoverage.	The proportion of units accessible via the frame that do not belong to the target population.
A3. Edit failure rate.	The proportion of responding units for which an error signal is triggered by a specified checking algorithm.
A4. Unit response rate.	The ratio of the number of units for which data for at least some variables have been collected to the total number of units designated for data collection.
A5. Item response rate.	The ratio of the number of units which have provided data for a given variable to the total number of designated units.
A6. Imputation rate.	The ratio of the number of assigned values (data are missing, invalid or inconsistent or have failed edits) for a given variable to the total number of values.
A7. Number of mistakes made, by type.	The number of serious mistakes in calculation or presentation of aggregates that are not found until after publication.
A8. Average size of revisions.	The average over a time period of the difference between a later and an earlier estimate expressed as the average revision, the average absolute revision, and/or the

Identification and Name (hyperlink)	Brief Description
	corresponding relative quantity(ies).
Timeliness and Punctuality	
T1. Time lag between the end of reference period and the date of the first/provisional results.	The number of days from the last day of the reference period to the day of publication of first results.
T2. Time lag between the end of reference period and the date of the final results.	The number of days from the last day of the reference period to the day of publication of final results.
T3. Punctuality of publication.	Number of days separating a previously announced date of publication and the actual date.
Accessibility and Clarity	
AC1. Number of subscriptions/ purchases of each of the key paper reports.	As stated opposite.
AC2. Number of accesses to on-line databases.	As stated opposite.
AC3. Rate of completeness of metadata.	The ratio of the number of metadata elements provided to the total number of metadata elements applicable.
Coherence and Comparability	
CC1. Lengths of comparable time series.	Number of reference periods in time series from last break
CC2. Asymmetries for statistics mirror flows.	Discrepancies between data related to flows measured in two ways, e.g. discrepancies between inbound and outbound flows for pairs of countries.
Assessment of User Needs and Perceptions	
US1. User satisfaction index.	The degree of satisfaction with services and products, often calculated for different segments of users.
US2. Length of time since most recent user satisfaction survey.	As stated opposite.
Performance Cost and Respondent Burden	
PCR1. Annual operational cost, with breakdown by major cost components.	Direct costs of staff involved in data collection (questionnaires, distribution,

Identification and Name (hyperlink)	Brief Description
	capture), reducing non-response, processing, and compilation of estimates.
PCR2. Annual respondent burden in hours and/or financial terms.	Respondent burden in hours is defined as number of respondents/questionnaires * average time per respondents/questionnaires, summed over all production rounds of the statistical process within a year. Respondent burden in financial terms is defined as respondent burden in hours * average hourly cost to respondents.

European Commission

ESS Handbook for Quality Reports

Luxembourg: Office for Official Publications of the European Communities

2009 — 135 pp. — 21 x 29.7 cm

ISBN 978-92-79-07855-2
ISSN 1977-0375

KS-RA-08-015-EN-N

Publications Office
Publications.europa.eu