Organisation de Coopération et de Développement Économiques
Organisation for Economic Co-operation and Development

17-Jan-2012

English - Or. English

STATISTICS DIRECTORATE

STD/QFS(2011)1
Unclassified

Cancels & replaces the same document of 24 November 2011

QUALITY FRAMEWORK AND GUIDELINES FOR OECD STATISTICAL ACTIVITIES

Version 2011/1

For more information, please contact:

Trevor Fletcher, Statistics Directorate (trevor.fletcher@oecd.org)

JT03314466

English - Or. English

FOREWORD

OECD statistics have a high reputation for quality and integrity throughout the world. They serve both as a key input for analytical work that informs policy recommendations and policy-making and also as an output in their own right in the form of publications and other products. OECD statistics cover a wide range of domains, from economics to health and education, from tax and agriculture to the environment, from migration and innovation to quality of life, just to name a few. Importantly, OECD statistics contribute to the OECD mission of promoting better policies for better lives.

Maintaining the high level of quality of OECD statistics is vital for the Organisation. Indeed, this level of 'data excellence' needs to be constantly monitored and improved along its main dimensions of relevance, accuracy, credibility, timeliness, accessibility, interpretability and coherence. This does not just apply to existing statistical activities but also to any of the new areas of research and analysis that are regularly added to the work programmes of OECD committees and working parties and that result in new statistical data collections. These must also adhere to the strict OECD quality standards.

To manage these quality standards in a systematic manner, the first version of the OECD Quality Framework (and the associated Quality Review process) was produced in 2003 as part of a wider reform of OECD statistical activities. Much of the drive for the review process in the initial period was geared towards the adoption of an OECD-wide corporate IT data management and dissemination system (the OECD Statistical Information System). These goals have now been achieved. And it was thus considered necessary to adapt the OECD Quality Review process to meet more recent and future developments.

To that end, the OECD has undertaken a review of its existing Quality Framework and Review process. The main recommendations from the review were that the main quality dimensions remained valid but that the Framework required improvement to: (i) better take into account the users' perspective; (ii) make review procedures more effective and less burdensome for statistical programme managers; and (iii) better reflect statistical, technological and institutional changes that have taken place since the Framework was first issued.

The OECD is now in the process of implementing these recommendations. I am confident that the newly adopted Quality Framework will ensure that the OECD continues to maintain its strong reputation as an international producer of statistics of the highest quality in years to come.

Martine Durand
OECD Chief Statistician and Director of the Statistics Directorate
August 2011

TABLE OF CONTENTS

INTRODUCTION

1. Improvement in the quality of OECD statistics is one of the main objectives of the OECD Statistics Strategy (OSS). OECD statisticians devote a significant part of their effort to quality improvement at an individual level, and the Organisation already applies techniques that are used in total quality management frameworks without having adopted a formalised approach to quality. Internal quality assurance processes are already in place in several Directorates. These include treatment and validation of questionnaire replies, cross-checking with national publications, compilation of additional information from OECD and other international sources, preparation of draft publication, referral of data queries and other quality issues back to source agency, interpretation, etc.

2. Several projects were launched in the framework of the OSS and considerable improvements achieved. These initiatives will enhance quality, though there is still need for a common framework which can be used to systematically assess, compare and further improve OECD statistics. A quality framework that is applied across the Organisation will:

- provide a systematic mechanism for the ongoing identification and resolution of quality problems;
- significantly increase the transparency of processes used by the OECD to assure quality; and
- reinforce the political role of the OECD in the context of an information society.

3. Much work has been done in recent years to apply the concept of quality to statistical data. For example, the IMF, Eurostat, Statistics Canada and other national statistical offices (NSOs) have identified various sets of data quality components and have adopted quality frameworks to improve their organisations and the quality of data produced. The OECD quality framework benefits from this work and we have avoided "reinventing the wheel" by adapting existing definitions and approaches to the OECD context.

4. For an international organisation, the quality of statistics disseminated depends on two aspects, the quality of national statistics received, and the quality of internal processes for collection, processing, analysis and dissemination of data and metadata. In several fields, national statistics are developed closely in accordance with international standards. On the other hand, statistical processes at international level are often derived from best practices developed at national level. Thus, there is a clear inter-dependence between the two aspects.

5. The OECD Quality Framework outlined in this document focuses on improving the quality of data collected, compiled and disseminated by the OECD through an improvement in the Organisation's internal statistical processes and management, though there will also be a positive knock-on effect on the quality of data compiled at the national level. Thus, in a sense, the OECD quality initiative is similar to those developed by Statistics Canada and other NSOs, which also encompass statistical managerial and technical processes.

6. The OECD Quality Framework has four elements:

- a definition of quality and its dimensions;
- a procedure for assuring the quality of proposed new statistical activities;
- a procedure for evaluating the quality of existing statistical activities on a regular basis;
- a set of broad principles on which OECD statistical activities are to be conducted and quality guidelines covering all phases of the statistical production process.

7. This document is organised as follows. Part 1 provides definitions of quality dimensions, together with a description of procedures for assuring the quality of new statistical activities and for evaluating the quality of existing statistical activities. In addition, it presents a set of broad basic principles on which OECD statistical activities have to be conducted. Part 2 outlines quality guidelines for different phases of the statistical production process.

1. QUALITY DIMENSIONS, CORE VALUES FOR OECD STATISTICS AND PROCEDURES FOR PLANNING AND EVALUATING STATISTICAL ACTIVITIES

1.1 Definition and dimensions of data quality

8. Quality is defined as "fitness for use" in terms of user needs. This definition is broader than has been customary used in the past when quality was equated with accuracy. It is now generally recognised that there are other important dimensions. Even if data is accurate, they cannot be said to be of good quality if they are produced too late to be useful, or cannot be easily accessed, or appear to conflict with other data. Thus, quality is viewed as a multi-faceted concept. The quality characteristics of most importance depend on user perspectives, needs and priorities, which vary across groups of users.

9. Given the work already done by several statistical organisations, the OECD drew on their experience and adapted it to the Organisation's context. Several statistical organisations have already identified the dimensions of quality. These have also been adapted to the OECD context. Thus, the OECD views quality in terms of seven dimensions: relevance; accuracy; credibility; timeliness; accessibility; interpretability; and coherence. These are discussed below. Another factor is that of cost-efficiency, which though is not strictly speaking, a quality dimension, is still an important consideration in the possible application of one or more of the seven dimensions cited previously to OECD statistical output.

Relevance

10. The relevance of data products is a qualitative assessment of the value contributed by these data. Value is characterised by the degree to which the data serves to address the purposes for which they are sought by users. It depends upon both the coverage of the required topics and the use of appropriate concepts. Value is further characterised by the merit of users' purposes in terms of the OECD mandate, the agreements with Member Countries and the opportunity costs of producing the data.

11. Measuring relevance requires the identification of user groups and their needs. There are multiple uses and users, and they may change over time. New needs may arise that require new data. Relevance may be indirectly assessed by ascertaining whether there are processes in place to determine the views of users and the uses they make of the data.

12. OECD context: Users include the Secretariat, Committees, Member governments and other external users. The Secretariat and Committees are primary users and determine priorities, but data are also produced for external users according to the political role of the Organisation vis-à-vis the civil society. The OECD Programme of Work provides the mandate for collecting and treating data for analytical purposes. The OECD Publishing Policy mandates that to maximise the impact of OECD work, it is essential that the content and format of all published outputs be adapted to the different target audiences. Core data (and related metadata) are widely disseminated in the interests of the public good.

Accuracy

13. The accuracy of data products is the degree to which the data correctly estimate or describe the quantities or characteristics they are designed to measure. Accuracy refers to the closeness between the values provided and the (unknown) true values. Accuracy has many attributes, and in practical terms there

is no single aggregate or overall measure of it. Of necessity these attributes are typically measured or described in terms of the error, or the potential significance of error, introduced through individual sources.

14. In the case of sample survey-based estimates, the major sources of error include coverage, sampling, non-response, response, processing, and problems in dissemination. For derived estimates, such as for national accounts or balance of payments, sources of error arise from the surveys and censuses that provide source data; from the fact that source data do not fully meet the requirements of the accounts in terms of coverage, timing, and valuation and that the techniques used to compensate can only partially succeed; from seasonal adjustment; and from separation of price and quantity in the preparation of volume measures.

15. An aspect of accuracy is the closeness of the initially released value(s) to the subsequent value(s) of estimates. In light of the policy and media attention given to first estimates, a key point of interest is how close a preliminary value is to subsequent estimates. In this context it useful to consider the sources of revision, which include (1) replacement of preliminary source data with later data, (2) replacement of judgmental projections with source data, (3) changes in definitions or estimating procedures, and (4) updating of the base year for constant-price estimates. Smaller and fewer revisions is an aim, however, the absence of revisions does not necessarily mean that the data are accurate.

16. OECD context: The accuracy of the data published is largely determined by the accuracy of the data received from the contributing organisations. On the other hand, the activities carried out by the Secretariat can influence the overall accuracy of data published. This influence can be positive because the quality checks adopted by the OECD may detect errors and result in improvements to the estimates previously provided by national agencies. Or it can be negative, due to errors that may result from the collection, processing, derivation, or dissemination procedures adopted by the Secretariat.

Credibility

17. The credibility of data products refers to the confidence that users place in those products based simply on their image of the data producer, *i.e.*, the brand image. Confidence by users is built over time. One important aspect is trust in the objectivity of the data. This implies that the data are perceived to be produced professionally in accordance with appropriate statistical standards, and that policies and practices are transparent. For example, data are not manipulated, nor their release timed in response to political pressure.

18. Credibility is determined in part by the integrity of the production process. Principle 2 of the UN Principles of Official Statistics[1] (1994) states: "*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"*.

19. OECD context: The Secretariat has to decide if the publication of poor quality data received from countries affects the overall credibility of the OECD as a high quality data provider. If the answer is yes, the Secretariat should refuse to publish the data. Furthermore, it must ensure that, once agreement between the Secretariat and countries has been reached on collection of specified data, the data subsequently collected cannot be withdrawn in response to political pressure.

[1] Provided in Annex 1 of this document.

Timeliness

20. The timeliness of data products reflects the length of time between their availability and the event or phenomenon they describe, but considered in the context of the time period that permits the information to be of value and still acted upon. The concept applies equally to short-term or structural data; the only difference is the timeframe.

21. Closely related to the dimension of timeliness, the punctuality of data products is also very important, both for national and international data providers. Punctuality implies the existence of a publication schedule and reflects the degree to which data are released in accordance with it. A publication schedule may comprise a set of target release dates or may involve a commitment to release data within prescribed time period from their receipt. Here "release date" refers to the date on which the data are first made publicly available, by whatever medium, typically, but not inevitably the web site.

22. OECD context: The timeliness of the data published by the OECD is largely determined by the timeliness of the data it receives from the contributing organisations. The Secretariat itself is also a potential source of delay, which may occur during collection, processing, derivation, or dissemination. A publication schedule would help:

- external users, by improving their capacity to make timely use of OECD statistics;
- internal users, by enhancing their capacity to plan their work based on the release dates;
- the Secretariat, by enhancing its capability to resist pressure to tamper with release dates for political reasons.

On the other hand, there may be occasions where the OECD cannot adhere to its schedule, for example, due to changes in priorities. Such changes should be clearly communicated to users.

Accessibility

23. The accessibility of data products reflects how readily the data can be located and accessed from within OECD data holdings. The range of different users leads to such considerations as multiple dissemination formats and selective presentation of metadata. Thus, accessibility includes the suitability of the form in which the data are available, the media of dissemination, and the availability of metadata and user support services. It also includes the affordability of the data to users in relation to its value to them and whether the user has reasonable opportunity to know that the data are available and how to access them.

24. OECD context: Internal and external users might have quite different perceptions of accessibility because of the differences in access methods. According to the OECD Publishing Policy, it is the responsibility of the Secretary General to determine and submit an annual strategic overview of the publications programme of the Organisation to Members at the beginning of each year (by 31 r January). Under the authority of the Secretary General, the Director of the Public Affairs and Communications Directorate decides on the selection, quality, form, timing and external distribution of all OECD publications, including statistics. The Director of PAC consults the Directorates and the Committees at the time of the development and the updates of the Publishing Programme for maintaining and improving the quality of OECD publications.

Interpretability

25. The interpretability of data products reflects the ease with which the user may understand and properly use and analyse the data. The adequacy of the definitions of concepts, target populations, variables and terminology, underlying the data, and information describing the limitations of the data, if any, largely determines the degree of interpretability.

26. The range of different users leads to such considerations as metadata presentation in layers of increasing detail. Definitional and procedural metadata assist in interpretability: thus, the coherence of these metadata is an aspect of interpretability.

27. OECD context*:* Where statistical processes are carried out following a decentralised model, the co-existence of different dissemination mechanisms should be minimised in order to avoid confusing users. Furthermore, where there are alternative definitions available for different uses, the Secretariat should help users in selecting those that are most appropriate to their needs. The OECD Publishing Policy mandates that to maximise the impact of OECD work, it is essential that the content and format of all published outputs be adapted to the different target audiences.

Coherence

28. The coherence of data products reflects the degree to which they are logically connected and mutually consistent. Coherence implies that the same term should not be used without explanation for different concepts or data items; that different terms should not be used without explanation for the same concept or data item; and that variations in methodology that might affect data values should not be made without explanation. Coherence in its loosest sense implies the data are "at least reconcilable." For example, if two data series purporting to cover the same phenomena differ, the differences in time of recording, valuation, and coverage should be identified so that the series can be reconciled. Coherence has four important sub-dimensions: within a dataset, across datasets, over time, and across countries.

29. *Coherence within a dataset* implies that the elementary data items are based on compatible concepts, definitions, and classifications and can be meaningfully combined. Incoherency within a dataset occurs, for example, when two sides of an implied balancing statement, such as assets and liabilities, or inflows and outflows, do not balance.

30. *Coherence across datasets* implies that the data are based on common concepts, definitions and classifications, or that any differences are explained and can be allowed for. An example of incoherency across datasets would be if exports and imports in the national accounts could not be reconciled with exports and imports in the balance of payments.

31. *Coherence over time* implies that the data are based on common concepts, definitions, and methodology over time, or that any differences are explained and can be allowed for. Incoherence over time refers to breaks in a series resulting from changes in concepts, definitions, or methodology.

32. *Coherence across countries* implies that from country to country the data are based on common concepts, definitions, classifications and methodology, or that any differences are explained and can be allowed for.

33. OECD context: Ensuring coherence across countries is one of the major sources of value added provided by the OECD. The role of metadata in explaining possible changes in concepts or methodologies over time and across countries is absolutely fundamental. Unexplained inconsistencies across datasets can seriously reduce the interpretability and credibility of OECD statistics.

Cost-efficiency

34. The cost-efficiency with which a product is produced is a measure of the costs and provider burden relative to the output. Provider burden is a cost that happens to be borne by the provider, but is a cost nevertheless. Whilst the OECD does not regard cost-efficiency as a dimension of quality, it is a factor that must be taken into account in any analysis of quality as it can affect quality in all dimensions. If a product can be produced more efficiently with the same quality, then resources released can be used to improve the quality of that product or other products.

35. OECD context: Respondent burden is usually placed on the NSOs and other government agencies to provide the data, as the OECD rarely, if ever, collects data directly from enterprises or households, and only infrequently requests data that the NSO or other government agency has not already collected for its own purposes.

1.2 Core values for OECD statistics

36. The Fundamental Principles of Official Statistics adopted initially by the Economic Commission for Europe of the United Nations, and subsequently endorsed by the United Nations Statistical Commission in April 1994 (refer Annex 1), provide a set of principles through which many of the quality dimensions outlined in the previous Section of this document are applied. The actual implementation of the quality dimensions and the UN Principles is undertaken through the guidelines and procedures provided for all OECD statistical activities.

37. The UN Fundamental Principles have been formally endorsed and adopted by the national statistical organisations of OECD Member countries and several other international organisations involved in the collection and dissemination of statistics. The OECD too is committed to implement the UN Fundamental Principles in its own statistical activities. In addition, OECD statisticians are committed to carrying out their work according to the International Statistical Institute declaration on professional ethics (see Annex 2).

38. The main principles of OECD statistical activities are:

a) *OECD statistics are compiled and made available on an impartial basis. OECD statistics are produced according to strictly professional considerations, including scientific principles and professional ethics with regard to methods and procedures used for the collection, processing, storage and dissemination of statistical data.*

b) *The OECD presents statistical information according to scientific standards on the sources, methods, and procedures adopted to produce its statistics.*

c) *Individual data collected by the OECD for statistical compilation are considered to be strictly confidential and used exclusively for statistical purposes. Specific measures are taken to ensure the full protection of confidential data from any potential disclosure.*

d) *Internal rules and measures under which the OECD statistical system operates are made public.*

e) *The OECD is committed to carrying out its statistical activities in co-ordination with national statistical agencies and with other international organisations.*

f) *The OECD is committed to developing bilateral and multilateral co-operation in statistics in order to contribute to the development of systems of official statistics in all countries.*

g) *Within the constraints of resource availability, OECD data products are of the best possible overall quality in terms of each of the seven quality dimensions outlined in the Organisation's Quality Framework. Effort involved in assuring quality is commensurate with the scale of the statistical activity, the purpose of the activity and its frequency (i.e. whether it is intended to be repeated regularly or occasionally, or is one-off).*

Basic principles adopted in data and metadata collection

h) Data and metadata necessary for the work of the OECD must be provided to the Organisation within the context of the formal obligations undertaken by each Member country when it joined the Organisation. This means that requests for information by the OECD should have the same importance attached to them as requests from other international organisations including those required by Regulation.

i) The OECD explains and justifies requests for data that impose additional burden upon data providers. To this end, national agencies responsible for the co-ordination of statistical relations with international organisations at the national level have the right to receive well structured information about the on-going and ad hoc data needs of the OECD.

j) The burden imposed on data providers is minimised. Data and metadata collection has to be conducted using the most efficient instruments, taking into account the technologies and human resources available in both the OECD and source organisations.

k) Data and metadata collection activity must be co-ordinated as much as possible across the OECD, taking into account user needs and available technologies and resources. This implies co-ordination of OECD requests for data and metadata and automated access by the OECD to data wherever possible.

l) Where there are "minor" differences in the data requested by different users within the OECD, a specific analysis of real user needs should be conducted in order to avoid unnecessary duplication.

m) Where appropriate, the OECD is committed to collect data disaggregated by gender.

n) The OECD helps, develops, promotes and uses international standards for data and metadata and associated statistical processes.

Basic principles adopted in data and metadata management

o) Data sets containing "final" data and metadata have to be designed and implemented according to established OECD standards in order to allow easy access by authorised users. Information on the characteristics of all data sets containing final data and metadata have to be located in a specific database.

p) Data sets containing "preliminary" data and metadata can be developed according to the needs of experts in charge of different work phases.

q) Corporate statistical and publishing IT and software infrastructures are used for storing, accessing and disseminating data and metadata in preference to activity specific applications.

r) Definitional, procedural and operational metadata describing each phase of each statistical activity are documented and readily available to internal users.

s) If confidential data are managed, all necessary technical and organisational measures are undertaken to ensure that only authorised people working in the Secretariat have access to the data and to prevent any confidentiality disclosure.

Basic principles adopted in data and metadata dissemination

t) The OECD dissemination policy for statistics are governed by the OECD Publishing Policy, endorsed by Council, and must meet the objectives, to:

- disseminate as widely as possible and in the most cost-effective manner
- the statistics collected and elaborated by the Organisation, adopting high quality standards to facilitate their accessibility and interpretability;
- enhance the credibility of the OECD as a source of high quality statistics reflecting economic, environmental and social performance in Member countries and in selected non-Member countries;
- contribute to the development of a culture of "informed decision making" at national and international levels, both in government and non-government bodies.

u) In meeting these objectives, statistics dissemination has to be conducted:

- in the most cost-effective manner, in accordance with OECD Publishing policy and with the financial and budgetary regulations and requirements of the OECD Council. Maximum use should be made of online dissemination possibilities;
- ensuring that the general user community can have free access to "core" statistical information and all metadata collected and/or originally produced by the Organisation;maximising co-operation with other national and international data providers. In particular, free access to all statistical products has to be given to all national governmental bodies (including national statistical offices), as well as, subject to reciprocal arrangements, to international organisations.

1.3 Procedures for assuring the quality of OECD statistical activities: an overview

39. The broad set of principles and guidelines through which many of the quality dimensions outlined above are implemented across the OECD are provided in Part 2 of this document. Specific procedures to ensure that new and existing statistical activities are conducted in accordance with these principles are presented in Part 3. This Section provides an overview of procedures for assuring the quality of proposed new statistical activities and for reviewing the quality of the output of existing statistical activities. In addition, the promotion of best practice used in-house and elsewhere is designed to help OECD statisticians adopt the most effective approaches to data and metadata collection, management and dissemination.

40. The procedures presented below are aimed at a single statistical activity. However, statistical activities do not exist in isolation. They are bound together by the same statistical infrastructure and the fact that their outputs may be viewed and used in combination. Improvements to the OECD's statistical infrastructure are the subject of other OECD initiatives. However, whilst the guidelines do not explicitly cover the infrastructure supporting statistical activities – computing, methodology, and publication – they do take into account the interaction between statistical activities through consideration of the data they jointly make available.

Procedure for assuring the quality of new activities

41. The main steps in the development of a new statistical activity were defined as:

a) definition of the data requirements in general terms;

b) evaluation of other data currently available;

c) planning and design of the statistical activity;

d) extraction of data and metadata from databases within and external to OECD;

e) implementation of specific data and metadata collection mechanism;

f) data and metadata verification, analysis and evaluation; and

g) data and metadata dissemination.

42. For each step the quality concerns and the instruments available to help in addressing them were identified. In particular, a set of guidelines and concrete procedures have been prepared for each step, taking into account good existing practices within the OECD and in other statistical agencies. In order to minimise the burden placed on activity managers, a simplified version of the procedure would be appropriate for statistical activities planned to be once rather than repeated. Table 1 provides more details about the procedure.

Procedure for reviewing the quality of existing activities

43. The procedure for reviewing the quality of existing statistical activities conducted across the OECD takes into account the fact that the review will be carried out on a rotation basis over a number of years. The stages envisaged are as follows:

a) identification by the OECD Statistical Policy Group (SPG) of the statistical activities for review during the course of the year, following a biannual rolling calendar;

b) self-assessment by the statistical activity manager and staff, resulting in a report that includes a brief summary of quality problems and a prioritised list of possible improvements, together with an assessment of additional resources required for their implementation. A self-assessment template is provided in Table 3, while more detailed guidelines on how to carry out the assessment are described in Part 3.

c) review of and comments on the self-assessment report by major users;

d) review of and comments on the self-assessment report by statistical, information technology, and PAC dissemination staff, co-ordinated by an expert designated by the SPG;

e) preparation of the final quality report, combining all comments, jointly by the activity manager and designated expert, and tabling of the report to the SPG;

f) discussion and resolution of any concerns about the report by the SPG, and transmission of the report to the relevant director;

g) assignment of resources for selected quality improvement initiatives by the directors and through the Central Priorities Fund;

h) feedback by the Chief Statistician to stakeholders on the quality improvement initiatives proposed and the plans for their implementation.

44. Stages (b), (c), (d) and (e) are the core of the procedure. They involve the preparation of a quality self-assessment by the activity manager, its review by users and experts, and the blending of all comments into a final report. More details are provided in Table 2, whilst Table 3 contains a template to assist the self-assessment. Given that there are about 70 OECD activities potentially subject to such reviews, it is vital that the procedure is flexible. In this context it is recognised that:

- there should be an initial round of reviews to pilot test the procedures;
- the procedure itself should be reviewed and fine-tuned each year;
- the scale of the reviews and resources invested in them should be commensurate with the benefits that can be envisaged – in particular, a simplified process may be appropriate for small scale/low profile activities;
- the review schedule over a four-year period can be provisionally announced, allowing activity managers to express their wishes regarding the most appropriate year and time of year for each review.

Conducting the self-assessment

45. The overall purpose of the Quality Review process is to demonstrate to stakeholders that data quality is being properly managed within the self-assessment, and to identify strengths and weaknesses in the quality of the self-assessment's output data. The self-assessment aims to provide the self-assessment's assessment of these strengths and weaknesses, based on the staff's own knowledge of the self-assessment and the feedback being obtained from its stakeholders. To guide the self-assessment manager in conducting this assessment, Table 3, *Self-assessment template for existing statistical activities*, illustrates the kind of issues a self-assessment should normally address for each dimension of quality. The first two columns provide the definitions of each dimension of quality as already described in Part I of the Quality Framework. The factor of Cost-efficiency, though not strictly a dimension of quality, has been included at the bottom of this Table since it is an important consideration in any assessment of data quality. The third column, *Elements of Quality Management*, summarizes the major functions that a self-assessment needs to undertake continuously in order to ensure the ongoing quality of the Activity's statistical outputs. The fourth column, *Possible Issues for Assessment*, suggests particular issues or aspects of these functions that might be considered at the time of a Quality Review. The issues listed are those thought to apply to most of OECD's self-assessments, but some may be inappropriate for some activities. On the other hand, the list is not exhaustive; managers may also consider and include other issues important to data quality for their particular Activity.

46. In addition, the self-assessment should identify any instances where the self-assessment is not able to follow OECD guidelines or best practices as described in Part 2 of the Quality Framework, Quality Guidelines for OECD Statistics, as well as any situations where the Fundamental Principles of Official Statistics (Annex 1) or the principles of the ISI Declaration on Professional Ethics (Annex 2) came under threat.

User consultation

47. Many of the issues in Column 4 of Table 3 imply a need for user views or feedback. With *data quality* defined as "fitness for use", user input is clearly essential to any assessment of data quality. Most self-assessments should have mechanisms in place for obtaining user views and feedback on a regular basis, and these will provide some of the required information needed for the self-assessment. However, the Quality Review is the occasion when stakeholders (users, clients, suppliers) must have an opportunity to provide their views, if they wish to, in a broader context than their regular liaison or interaction with the self-assessment.

48. The user consultation required for a Quality Review would normally consist of two parts: a general invitation, and a survey. The general invitation is a public announcement that the Quality Review will be taking place, a description of its purpose and scope, and an invitation to provide comments on any aspect of the quality of data from the self-assessment by a specified date. This invitation could be included in OECD communications that statistical users are likely to see, and be visible on websites where relevant statistical data reside. (A draft of such an announcement is included in the Appendix).

49. The directed survey would be addressed to known major users of the self-assessment's outputs. These would typically include some or all of the following: OECD Committees or their statistical working groups; statistical offices of member countries; other contacts in member countries; users within the OECD Secretariat; relevant international agencies; subscribers to the outputs of the self-assessment; other known major users. The method of surveying these users may depend on the lines of contact available for each user group, in collaboration with PAC for external users.

50. This survey should provide the same background information as the general invitation, but then ask for views on specific issues relevant to the particular self-assessment. Based on the information requirements implicit in Table 3 the following general issues could be addressed in the survey:

- satisfaction with current content of the self-assessment (topics covered, frequency, detail, country coverage, etc.);
- satisfaction with the accuracy of the self-assessment's data, including the ability to make accurate comparisons between countries;
- satisfaction with the timeliness of the self-assessment's outputs;
- adequacy of metadata describing the published statistical data;
- satisfaction with the statistical outputs of the self-assessment;
- adequacy of the systems for locating and accessing the self-assessment's outputs;
- concerns about compatibility between the self-assessment's data and other sources;
- satisfaction with the self-assessment's regular consultation process with users;
- suggestions for future direction of the self-assessment.

Specific questions addressing these issues in the context of the particular self-assessment will need to be formulated.

The Self-assessment report

51. The output of the self-assessment is a report to management and stakeholders. While self-assessment managers should feel free to convey their conclusions in the most effective way they can, the following outline has worked well for some managers and is suggested for these reports:

- An Introductory section that describes the overall self-assessment, its purpose, its main features, its scope, and its history or origins;

- A body of the report that covers each dimension of quality in turn, describing for each the current practice of quality management, an assessment of its strength or weakness, and recommendations for action (if any);

- A concluding section that summarizes the strength or weakness of each quality dimension in a tabular form (*e.g.* on a 5-point scale from very weak to very strong), and lists the primary recommendations for action - with a sense of priority, including a sense of which can be accomplished within existing budgetary constraints.

The concluding section should be considered as an Executive summary that could serve as the basis for eventual management decisions about future changes to the self-assessment.

1.4 The development and the update of Quality Guidelines

52. As already described in the Introduction and Part 2 of this document which contain technical quality guidelines for developing statistical activities and checklists for helping managers in developing a new activity or reviewing an already existing activity. Quality guidelines are based on the most advanced statistical methods and procedures, best practice already adopted by the OECD and other national and international statistical organisations, as well as on IT tools currently identified by the OECD as standards for the Organisation. Therefore, they tend to evolve over time and need to be periodically reviewed and updated.

53. Before the end of each year, the Chief Statistician of the OECD proposes to the SPG necessary changes, if any, to the Quality Framework and/or to Quality Guidelines. The SPG discusses and finally endorses proposed amendments. The updated Quality Frameworks and Guidelines are made available on the Intranet statistics site.

Table 1: Procedures for a proposed new statistical activity[1]

WHAT	HOW	POTENTIAL PROBLEMS	INSTRUMENTS AVAILABLE WITHIN QUALITY FRAMEWORK	CONTRIBUTION TO CORPORATE TOOLS
(a). Initial definition of output data requirements in general terms: coverage, content, users, uses	Obtain initial views of data requirements through: 1. Discussion with users, including Committees and internal users; 2. Discussion with other Directorates	1. Difficulties in evaluating relevance		
(b). Evaluation of data currently available within OECD and from other international and national organisations, and identification of needs for data	1. Review literature 2. Review data currently available within the OECD 3. Review data currently available from other international organisations 4. Review data currently available from national organisations	1. Difficulties in identifying and accessing data available within the OECD 2. Difficulties in identifying and accessing data available outside the OECD 2. Difficulties in interpreting data and metadata available	1. OECD Statistical Work Programme (OSWP) 2. Gateway to OECD statistical databases 3. UNECE Integrated Presentation of Statistical Work and internet sites of international organisations 4. OECD Glossary of Statistical Terms 5. Consultation with SPG members 6. OECD Quality guidelines 7. OECD.Stat[2]	1. Brief note about the proposed activity to SPG
(c). Planning and design involving all stages of the statistical activity[3]	1. Assess resource requirements and time frame - IT aspects - skills required - financial implications 2. Design activity[4] in terms of: - definitional content and coverage - statistical methodology, IT needs - marketing and dissemination 3. Establish contacts with experts in national and international statistical organisations	1. Underestimating resources required 2. Underestimating time required 3. Poor choice of statistical methods 4. Lack of communication with and involvement of national statistical experts responsible for coordination with international organisations 5. Inefficient IT solution[5]	1. Contacts through the Analytical Statistical Task Force (ASTF) with ITN, STD, PAC and other experts working in the Secretariat 2. Toolbox for IT solutions 3. OECD. StatWorks 4. Training program for statisticians 5. OECD Quality guidelines 6. OECD Glossary of Statistical Terms	1. Completion of OSWP entry for the activity 2. Information about activity to relevant international and national statistical organisations

1. In accordance with the terminology of the OECD Statistical Work Programme, a statistical activity is interpreted as an activity that produces at least one statistical output, such as a dataset or database available to internal or external users through Internet, Intranet, OLISNet, CD-ROM, etc., or a publication (whether classified or not) that is statistical or is analytical with extensive statistical content. A new statistical activity can be proposed as ongoing, *i.e.*, to be repeated at regular intervals, or one-off. This table is intended primarily for activities that are proposed to be ongoing, but can be used, possibly in abbreviated form, for an activity that is one-off. Table 2 outlines processes for an existing ongoing statistical activity.

2. OECD.Stat is part of the OECD statistical information system. It comprises a data warehouse, where final statistical data are stored, and a set of procedures for extracting data and metadata from OECD statistical databases. Among other things, the system provides users with a catalogue of variables available in individual databases. See Annex 4.

3. "All stages" implies the complete data life cycle - definition, feasibility study, collection, management, dissemination, etc. The problems uncovered and the design decisions made during this step are re-examined and elaborated in subsequent steps, *i.e.*, there is interaction between steps.

4. This includes: selection of software, design of the database, definition of data and metadata storage needs, definition of a new survey at the national level (if required), definition of rules for treatment of confidential data, etc.

5. For example, leading to difficulties in database access by internal users, difficulties in data and metadata exchange with other databases, disclosure of confidential data, use of non-corporate software, etc.

Table 1: Procedures for a proposed new statistical activity (continued)

WHAT	HOW	POTENTIAL PROBLEMS	INSTRUMENTS AVAILABLE WITHIN QUALITY FRAMEWORK	CONTRIBUTION TO CORPORATE TOOLS
(d). Extract data from databases within and external to OECD	1. Direct access to data, *i.e.* without the need to involve data providers in data collection or transmission	1. Inefficiencies in accessing internal and external databases 2. Difficulties in interpreting data and metadata 3. Incoherence across databases	1. OECD Glossary of Statistical Terms 2. Gateway to OECD Statistical Databases 3. OECD Quality guidelines 4. OECD.Stat 5. Corporate procedures to extract data and metadata from existing sources	
(e). Implement new data collection mechanism	1. Contacts with data providers 2. Preparation and test of questionnaire[6] 3. Dissemination of questionnaire 4. Data and metadata collection/ transmission	1. Insufficient contact with national data providers 2. Incorrect or inefficient questionnaire design 3. Use of inappropriate definitions 4. Inefficient choice of systems for data, metadata transmission	1. OECD Glossary of Statistical Terms 2. International statistical guidelines and recommendations 3. OECD Quality guidelines 4. OECD.Statworks[7] 5. Corporate procedures to extract data and metadata from external sources	1. Update OECD Glossary of Statistical Terms 2. Update OSWP
(f). Data and metadata verification, compilation, storage, analysis and evaluation	1. Verification of individual data 2. Evaluation of coherence of data: - across data items within datasets - over time - across countries - with other data sources 3. Overall evaluation of data relative to objectives	1. Inappropriate or inefficient statistical methods 2. Different methods across countries for the same subject	1. OECD Glossary of Statistical Terms 2. Gateway to OECD Statistical Databases 3. Statistical and econometric software for dealing with series breaks 4. Advice from STD and other OECD experts 5. OECD. StatWorks 6. OECD.Stat 7. OECD Quality guidelines	1. Update Data Catalogue 2. Update OECD Glossary of Statistical Terms
(g). Data and metadata dissemination	1. Paper publications 2. Offline databases 3. Online databases 4. Through the OECD Statistics Portal	1. Inefficient dissemination procedures 2. Inconsistency across databases 3. Inappropriate presentation of metadata 4. Disclosure of confidential data 5. Inappropriate data release procedures, affecting credibility	1. OECD Style Guide 2. OECD Quality guidelines 3. Assistance from ITN and PAC 4. OECD StatWorks	1. Update OSWP 2. Update OECD Glossary of Statistical Terms 3. Contribute to OECD.Stat

6. The questionnaire may be designed to collect macro or micro level data from national data providers or micro level data from enterprises, households, etc.

7. StatWorks is an IT tool for collecting, storing, validating and disseminating data and metadata. It also comprises a tool for designing electronic questionnaires.

Table 2: Procedures for existing statistical activities[1]

WHAT	BY WHOM	TARGET DATE / TIME SPAN	HOW	POTENTIAL PROBLEMS	INSTRUMENTS AVAILABLE	OUTPUTS
(a). Identification of statistical activities for review on a rolling biannual calendar[2]	SPG	By the end of January (year t)	1. Discussing review proposals and schedules presented by Directorates	1. Directorates slow to agree on schedule for quality reviews	1. OECD Statistical Work Programme (OSWP)	Set of statistical activities to be reviewed by end of year
(b). Self-Assessment (self assessment template on subsequent sheet)[3]	Statistical activity manager and staff	3 months	1. Consulting major users, including Committees and country experts 2. Consulting appropriate national and international agencies[4] 3. Comparing current practices with guidelines 4. Identifying cost-efficiency of currently adopted procedures	1. Operational concerns take priority away from quality review 2. Inadequate evaluation of all quality dimensions 3. Poor identification of quality improvements 4. Available resources	1. Quality checklist for self-assessment 2. OECD Quality guidelines	Self-assessment report including summary of quality problems, prioritised list of possible improvements and an assessment of additional resources (if any) required for implementation (included new data developments)
(c). User review of the self-assessment report[5]	Statistical activity manager and staff	1 month	1. Asking major users, including Committees and/or experts, to comment on the self-assessment	1 Major users do not have time or resources to make detailed comments		Additional potential improvements and priority assignment from user perspective
(d). Horizontal review of the self-assessment report	Statistical activity manager and designated expert[6]	1 month	1. Commenting on the self-assessment from a "corporate" perspective and suggesting improvements	1. Incorrect evaluation of quality dimensions 2. Incorrect identification of proposed improvements	1. OECD Quality Guidelines	Additional potential improvements and priority assignment from horizontal perspective and evaluation of resource assessments

1. In accordance with the terminology of the OECD Statistical Work Programme, a statistical activity is interpreted as an activity that produces at least one statistical output, such as a dataset or database available to internal or external users through Internet, Intranet, OLISNet, CD-ROM, etc., or a publication (whether classified or not) that is statistical or is analytical but with extensive statistical content. Table 1 outlines procedures for proposed new statistical activities.

2. All statistical activities would be reviewed over a time frame of four years. A review should be conducted when main technical or organisational changes are envisaged (for example, when the software used to maintain the database has to be changes/revised).

3. The scale of self-assessment should be commensurate with scale and significance of activity. A simplified approach is appropriate for small scale activities.

4. Not only national statistical offices, but also other data providers.

5. Activities (c) and (d) are normally carried out in parallel.

6. For each activity, or group of activities, the SPG will designate an expert to be responsible for conducting the horizontal review and for drafting the final quality report, in co-operation with the manager of the statistical activity. The horizontal review will be done with the assistance of STD, PAC and ITN experts and other statisticians.

Table 2: Procedures for existing statistical activities (continued)

WHAT	BY WHOM	TARGET DATE / TIME SPAN	HOW	POTENTIAL PROBLEMS	INSTRUMENTS AVAILABLE	OUTPUTS
(e). Preparation of the final quality report	Statistical activity manager and designated expert	1 month	1. Merging the self-assessment and comments received through the reviews 2. Identifying a final list of proposals for potential quality improvements	1. Conflicting views from managers, users and horizontal Directorate experts	1. OECD Quality Guidelines	Final quality report including summary of quality problems, prioritised list of possible improvements and an assessment of resources required for implementation tabled with SPG
(f). Review by SPG and transmission of official report to relevant Director	SPG	1 month	1. SPG members may comment on conclusions, discuss in detail or raise their concerns. 2. After resolution of any concerns, or in absence of comments, the report is regarded as official.	1. SPG members slow to react		Final quality report including summary of quality problems, prioritised list of possible improvements and an assessment of resources required for implementation sent to relevant Director
(g). Assignment of resources for quality improvement initiatives	Relevant Director, Budget Committee, Secretary General, Chief Statistician	By the end of December	1. Evaluating priorities at Directorate level 2. Identifying initiatives to be financed by the CPF	1. Improvements are not made because of lack of resources		Quality improvement initiatives embedded in Programme of Work
(h). Feedback to stakeholders on initiatives to improve the quality of OECD statistics	Chief Statistician	By the end of January (year t+1)	1. Proposing changes (if any) to quality framework and guidelines 2. Summarising proposed quality improvement initiatives 3. Indicating which proposed improvements are being implemented and how	1. Credibility of the OECD data is affected if quality problems not solved		Annual report to the SG and to Council on the implementation of the quality framework

Table 3: Self-assessment template for existing statistical activities

Dimension of Quality	Definition	Elements of Quality Management	Possible Issues for Assessment
Relevance	The extent to which the statistical outputs of the self-assessment satisfy the information needs of the client community - OECD Committees, the Secretariat, Member governments, and external users.	1. Keeping informed of the data needs of the client community. 2. Assessing user feedback on current products and services. 3. Adaptation of the self-assessment in response to client needs or dissatisfaction.	1. Effectiveness of mechanisms for staying abreast of the needs of each client group. Pertinence of the self-assessment to OECD strategies and Committee priorities. 2. Dependency of other Secretariat activities on the outputs of the self-assessment. 3. Effectiveness of mechanisms for obtaining user feedback. 4. Measures of product use and usefulness - internal and external. 5. Changes made to the self-assessment to maintain relevance and their success. 6. Coordination with related activities by other international organisations. 7. Important statistical needs of clients currently unmet.
Accuracy	The degree to which the statistical outputs correctly estimate or describe the characteristics they are designed to measure. For OECD, the provision of data that are comparable across countries is a crucial aspect of the accuracy of statistical data (see also coherence below).	1. Ongoing liaison with data suppliers to ensure accurate inputs using common standards. 2. Design and quality control of statistical treatment of input data. 3. Assessment of accuracy and provision of accuracy measures.	1. Effectiveness of mechanisms for maintaining liaison with data suppliers. 2. Extent of agreement on common standards for reporting statistical data. 3. Availability and quality of metadata describing input data. 4. Degree of conformity to common reporting standards by member countries. 5. Availability of description of adjustments and treatments applied to input data. 6. Checks and controls to avoid error in producing output data. 7. Availability of indicators of accuracy provided to users. 8. Availability of analytic or comparative studies that shed light on data accuracy. 9. Incompatibilities between country data, or over time, that confound comparisons.

Table 3: Self-assessment template for existing statistical activities (continued)

Dimension of Quality	Definition	Elements of Quality Management	Possible Issues for Assessment
Timeliness	The time lag between the end of the reference period and the release of data. Also encompasses the achievement of target dates announced for data release.	1. Ongoing liaison with data suppliers to ensure timely inputs. 2. Balancing timeliness and accuracy in designing the statistical treatment process 3. Use of pre-announced release date targets.	1. Effectiveness of mechanisms for maintaining liaison with data suppliers. 2. Analysis of trends in the timeliness of input data reporting. 3. Analysis of trends in the time between arrival of input data and output release. 4. Methods used to compensate for missing data in order to achieve timely release. 5. Performance in achieving release date targets. 6. Improvements made in timeliness of outputs. 7. User satisfaction with timeliness of outputs.
Accessibility	How readily the data can be located and accessed from within OECD data holdings - includes the suitability of the form in which data are available, the media of dissemination, and the availability of metadata and user support services.	1. Maintenance of data descriptions, catalogues and searching facility. 2. Delivering data in formats and media suitable for users. 3. Publicising the availability of data and promoting its use.	1. Timely inclusion of data in OECD.Stat. 2. Timely documentation of data in MetaStore and elsewhere. 3. Availability of data in alternative media or formats. 4. Facility of downloading data into user systems. 5. Measures used for announcing and publicising the release of data. 6. User awareness of available data. 7. Measures of data usage and analysis of user-reported access problems. 8. Access controls for confidential data (where applicable). 9. Data and metadata are provided in a structured manner such that they are human and machine readable.

Table 3: Self-assessment template for existing statistical activities (continued)

Dimension of Quality	Definition	Elements of Quality Management	Possible Issues for Assessment
Interpretability	The ease with which the user may understand and properly use and analyze the data, addressing both specialists and non-specialists.	1. Provision of metadata covering concepts, definitions and methods. 2. Provision of assessments or analysis of data accuracy. 3. Provision of interpretative analysis based on the data.	1. Adequacy of metadata from source data providers. 2. Adequacy of metadata describing OECD processes. 3. Inclusion and accessibility of metadata in MetaStore. 4. User feedback on adequacy of metadata. 5. Usefulness of the indicators of accuracy provided to users. 6. Examples of analytic work that show how the data might be properly used. 7. Occurrences of data misuse or misinterpretation and how they were handled.
Coherence	The degree to which data items and products are mutually comparable. This includes coherence within datasets, across datasets, and over time. Ensuring coherence across countries is a major value added by the OECD	1. Use of common concepts, definitions and classifications for subject content across member countries and within OECD. 2. Use of common methods and systems for processing data within OECD. 3. Confrontation and reconciliation of datasets that might conflict.	1. Degree of conformity to common standards by member countries. 2. Methods used for handling cases where countries cannot conform. 3. Methods used for handling changes in the standards used over time. 4. Degree to which SIS components are used where applicable. 5. Efforts made to identify and resolve potential data conflicts across datasets. 6. Design changes made to prevent future data conflicts. 7. Outstanding data conflicts that require attention.

Table 3: Self-assessment template for existing statistical activities (continued)

Dimension of Quality	Definition	Elements of Quality Management	Possible Issues for Assessment
Credibility	The confidence that users place in data products based on their image of the data producer. While credibility depends on OECD's overall statistical performance, individual self-assessments need to avoid actions that might undermine corporate credibility.	1. Transparent and professional decision making on statistical matters. 2. Avoiding political interference in the timing or content of data releases. 3. Policies and procedures for handling suspect data from respondents.	1. Availability of descriptions and rationales for the chosen methodology. 2. Consistency with the ISI Declaration on Professional Ethics. 3. Consistency with UN Fundamental Principles of Official Statistics. 4. Existence of, and adherence to, pre-announced release dates, coordinated with OECD Public Affairs and Communications. 5. Occurrences of attempted political interference. 6. Effective and open liaison with suppliers to address and resolve reporting problems. 7. Occurrences of data not being published for quality reasons.
Cost-efficiency	Though not regarded as a dimension of quality, cost-efficiency is a factor that must be taken into account in any analysis of quality. If less costly or less burdensome methods can be found, resources may be released to improve quality.	1. Avoidance of duplication in statistical activities. 2. Management of costs to suppliers by minimizing the burden to respond. 3. Management of processing costs through use of common methods and software for processing or analysis.	1. Effective liaison with other international agencies to coordinate similar statistical activity. 2. Effective internal liaison with STD and other Directorates before launching or expanding a self-assessment. 3. Liaison with providers to minimize the amount of data to be reported. 4. Liaison with providers to establish the least burdensome method of response. 5. Use of common formats (*e.g.* SDMX) for transmitting data. 6. Use of common software (*e.g.* SIS components) for processing or analysis.

2. QUALITY GUIDELINES FOR OECD STATISTICS

2.1 Introduction

54. The quality guidelines presented here reflect known good practice adopted in the OECD Secretariat or in other national and international bodies active in statistics. They are designed to fill an information gap existing across OECD Directorates, by helping statisticians to identify the highest quality and the most effective solutions for carrying out their activities. They are intended for internal use within the OECD and should be implemented by all parts of the Organisation, according to the nature of individual activities and available resources.

55. The quality guidelines provided underpin the OECD Quality Framework by providing a basis for the evaluation of a proposed new statistical activity and for the self-assessment of an existing statistical activity. They may evolve and be refined over time as better practices are identified and/or the guidelines are streamlined.

56. For an international organisation such as the OECD, the quality of statistics produced depends on two aspects - the quality of data and metadata obtained from national agencies and the quality of the internal processes for collection, verification, compilation, analysis and dissemination of these data. The quality framework and guidelines focus on enhancing the quality of data used and disseminated by the OECD through improvements in the Organisation's statistical processes and management.

57. Whilst the quality of data obtained from national agencies is a very important aspect of the quality of OECD output, it is not the subject of these guidelines, which are primarily concerned with those activities that are completely under OECD control. The quality of data obtained from national agencies should be assessed by the activity element managers, and the relevant OECD Committees, Working Groups or user Directorates, in partnership with the national agencies in accordance with quality assessment procedures specifically designed for that purpose.

58. As stated in para. 8, quality is defined as "fitness for use" in terms of user needs. In accordance with the quality framework, it is viewed as having seven dimensions, namely, relevance, accuracy, credibility, timeliness, accessibility, interpretability, coherence. The quality characteristics of most importance depend on user perspectives, needs and priorities, which vary across groups of users. In these guidelines, particular focus is given to those aspects of a statistical activity that reflect the value added to the data by the OECD, in particular, those features that ensure accessibility, interpretability and coherence. In addition, the adoption of best available practices would help OECD statisticians to adopt the most effective approaches to data and metadata collection, management and dissemination.

59. Success in the management of quality requires a common concern for, and pride in, quality among statisticians and managers at all levels of the OECD. High quality OECD statistics cannot be achieved only through the imposition of rules. It also requires attention to client needs coupled with the application of sound statistical and managerial knowledge at all levels, as well as the application of advanced statistical techniques and IT technologies. To achieve this target a motivated and qualified staff is essential. Therefore, on-going initiatives are required to improve the expertise of statistical staff (*e.g.* through the training programme for statisticians), sharing good practice, formulation of the OECD statistics strategy and the development of an *esprit de corps* among OECD statisticians (*e.g.* through OECD statisticians meetings).

60. The guidelines presented below are aimed at a single statistical activity. However, statistical activities do not exist in isolation. They are bound together by a common statistical infrastructure and the fact that their outputs may be viewed and used in combination. Potential improvements to statistical infrastructure are the subject of other OECD initiatives. However, whilst the guidelines do not explicitly cover the infrastructure supporting statistical activities – computing, methodology, and publishing – they do take into account the interaction between statistical activities through consideration of the data they jointly make available.

61. A statistical activity is considered in terms of seven phases:

- definition of the data requirements in general terms;
- evaluation of other data currently available;
- planning and design of the statistical activity;
- extraction of data and metadata from databases within and external to the OECD;
- implementation of a specific data and metadata collection mechanism;
- data and metadata verification, analysis and evaluation; and
- data and metadata dissemination.

Each phase is considered within the context of the seven quality dimensions described in Part 1 of this document. Particular attention is given to those aspects of a statistical activity that reflects the value added to the data provided by the OECD, in particular, the features that ensure accessibility, interpretability and coherence.

62. The Guidelines provided in Section 2.2 are given separately for each of the seven phases listed above, recognising that phases may overlap to some extent or may be carried out in parallel or iteratively rather than consecutively. Furthermore, for some statistical activities, some phases may not be applicable.

2.2 Guidelines for different phases of statistical activities

63. The following Guidelines are provided for each of the seven phases outlined in para. 54 above for each statistical activity undertaken within the Organisation. The Guidelines for each phase comprise: aims, context and a list of the guidelines to be applied. Guidelines are provided for each phase separately. The Guidelines presented below incorporate best practice already in place within the OECD and in other statistical agencies[2].

[2] See "Statistics Canada – Quality Guidelines" (1998); "Statistics Canada's Quality Assurance Framework" (2002); "US Census Bureau Section 515 Information Quality Guidelines" (2002).

Useful linkages to statistical web pages[3]

The following references contain a wide range of information relevant for OECD statistical activities. In case of any changes, updated links can be found visiting the OECD Statistics Portal.

OECD intranet for statistics http://intraweb1/STD-SIMS-OECDSTATS-WA/index.asp
The Intranet contains information on: data and metadata, statistical meetings and activities, methodological documents, technical guides, guidelines for statistical activities, etc. (***internal only***).

OECD 2011 Statistical Programme of Work
http://www.oecd.org/document/16/0,2340,en_2825_293564_2496336_1_1_1_1,00.html
The Programme provides an overview of the implementation of the OECD statistics strategy, a summary description of OECD statistical activities and more detailed information on each activity (***internal only***).

Questionnaire for compiling the OECD Statistical Programme of Work (OSWP)
http://intraweb1/STD_SIMS_OSWP_WA/sa_activity_list.asp
The questionnaire is used for compiling the Programme of Work (***internal only***).

OECD Glossary of Statistical Terms http://stats.oecd.org/glossary/
The Glossary contains about 6 000 statistical terms, definitions and context information where appropriate.

OECD Publishing Policies and Guidelines
http://oecdshare.oecd.org/pac/Lists/Best%20Practices%20%20How%20To/Best%20Practice%20%20How%20To.aspx
(***internal only***).

UNECE Integrated Presentation of Statistical Work http://www.unece.org/stats/stats_e.htm
The Integrated Presentation gives a detailed description of statistical activities undertaken by the OECD and other international organisations.

International Classifications http://unstats.un.org/unsd/methods.htm
The site contains all main international statistical classifications.

International standards for statistical activities
http://unstats.un.org/unsd/methods/statorg/Principles_stat_activities/principles_stat_activities.htm (metadata) and
http://unstats.un.org/unsd/methods.htm
Several manuals and handbooks developed by international organisations are available here.

IMF data dissemination standards http://dsbb.imf.org/Applications/web/dsbbhome/
The site describes the IMF standards for disseminating statistics, as well as quality standards for statistics.

SDMX http://www.sdmx.org
The site contains information on the OECD-Eurostat-IMF-ECB-BIS-World Bank initiative for establishing standards for statistical data and metadata exchange. This site also describes the SDMX Common Metadata Vocabulary (MCV) which contains definitions of metadata terminology, the aim being to ensure a common understanding of metadata elements by metadata authors.

[3] As the Guidelines are for OECD use, a number of the references contained here are only available to OECD employees, and therefore cannot be accessed from outside the OECD.

2.2.1 Phase 1: Definition of data requirements in general terms

Aims

64. The aims of the guidelines are to:

- ensure that information about expected new (or revised) data requirements is shared across the Organisation, minimising the risk of duplication of effort and waste of resources;
- provide an opportunity to interested staff in different parts of the Organisation to contribute to the development of the activity;
- help horizontal directorates (STD, ITN and PAC) to perform their roles, maximising internal co-ordination and improving their capacity to support new developments (through appropriate statistical inputs, IT infrastructures, dissemination tools, etc.).

Context

65. At the beginning of each project, potential data requirements are normally specified in terms of data concepts, coverage, and content. These in turn are determined by data users and uses to which the data are put. Thus, specifying data requirements implies identification of user groups and their data uses. In general, there are multiple users and uses, both of which change over time.

66. In the OECD context, the primary users are OECD Committees and working groups reporting to Committees or Directorates. They use data for analytical and policy making purposes. They determine the basic data requirements and set priorities for their collection, as reflected in the OECD Programme of Statistical Work.

67. Other users include individuals belonging to the OECD Secretariat who are not members of a Committee or working group responsible for the data and external users within the governments of OECD countries and in other organisations. These users do not determine content or priorities except indirectly though their influence on the Committee or working group responsible for the data.

68. The analysis of data requirements in general terms is carried out when a new activity has to be launched, or an existing activity has to be reviewed.

Guidelines

a) The uses to which the data are put are well articulated and documented by the Committee, Directorate or working group responsible for the statistical activity.

b) The essential data requirements are specified and documented in terms of required concepts, content, coverage, frequency and timeliness by the Committee, Directorate or working group responsible for the statistical activity.

c) Existing potential internal users and statistical experts are consulted in determining the characteristics of data requirements in general terms.

d) In determining the data requirements, consideration is given to the trade-offs between content, coverage, frequency, accuracy, timeliness, costs and provider burden.

e) Wherever available, existing international statistical guidelines and recommendations are used for concepts, definitions, units, classifications, nomenclatures and compilation methods. Divergences from these international standards are documented and justified.

f) In the absence of existing international standards or Committee recommendations and with a view to minimising provider burden, concepts, definitions, classifications, nomenclatures are harmonised with commonly used national practices to the maximum extent possible. Lack of harmonisation with prevailing national practice is documented and justified.

g) On-going statistical activities are reviewed at regular intervals, in close partnership with stakeholders, in order to identify new needs, to adopt the most appropriate statistical methods and to utilise the most effective technical solutions, etc.

Relevant quality dimensions: Relevance and coherence

Importance for cost efficiency: Low

2.2.2 Phase 2: Evaluation of other data currently available and integration with other OECD datasets

Aims

69. The aims of the guidelines are to:

- minimise costs associated with the statistical activity;
- minimise unnecessary burden on respondents;
- improve the coherence between sources used by the OECD;
- improve the coherence of OECD datasets;
- adopt as much as possible, concepts, definitions and classifications already in use by the international statistical community.

Context

70. After defining the data requirements of a statistical activity (Phase 1), the next step is to determine which data and metadata are already available in other databases accessible by the OECD. Such databases may provide some or all of the data and metadata required, thus reducing or eliminating the need for additional data collection. They may be divided into three categories:

- internal, created by other OECD statistical activities;
- external, belonging to other international organisations;
- external, belonging to national agencies.

71. The extent to which other databases can meet the data needs of the statistical activity depends not only upon their nominal content and quality (timeliness, accuracy, etc.), but also on the degree of harmonisation between them and the statistical activity. Statistical activities within the OECD do not take place in isolation from one another. They should be harmonised to the extent that unnecessary differences in concepts, content, coverage, frequency, accuracy, and timeliness are eliminated and that the data available can be integrated, *i.e.*, viewed and used collectively. Similarly, the statistical activities of the OECD and other international organisations should be harmonised to the extent that unnecessary differences are eliminated and that the data available can be integrated.

72. Metadata describing the content and the structure of databases (see Phase 6) is essential, not only to enable an assessment of whether the data can (partly) satisfy the needs, but also to describe the relevant data items to users. The absence of readily available metadata as well as the absence of data may be a limiting factor in the use of other databases.

73. In some cases, the statistical activity can imply the development of new definitions, classifications, etc. Therefore, methodological aspects have to be developed maximising integration with other international standards.

Guidelines

a) Statistical activities are carried out minimising the overall cost for the Organisation and the burden on respondents. Maximum use is made of data already existing in OECD databases or in databases managed by other international organisation.

b) Each internal OECD or international organisation database is reviewed to determine whether any required data items are available. If so, arrangements are made to extract and use the data items, or the reasons for not doing so are documented and justified.

c) Sufficient metadata are acquired for each potentially useful internal OECD database to determine whether or not that database can provide any required data items. Deficiencies in the metadata available are documented, discussed with the corresponding database manager at the OECD or international organisation and remedied where possible.

d) If necessary, potentially useful national agency databases are analysed to determine whether any required data or metadata items are available.

e) Consideration is given to changes in the statistical activity, concepts, content, coverage, frequency, accuracy, and timeliness that would allow more use of data available from internal or external databases. Such changes are made or planned within future work programmes, or the reasons for not doing so are documented and justified.

f) Consideration is given to adjustments that could be made to data items available from internal or external databases that would make them usable. Such adjustments are made or planned within future work programmes, or the reasons for not doing so are documented and justified.

g) Summary information about a new planned activity is provided, as soon as possible, to the Chief Statistician and other members of the Statistical Policy Group. In particular, a note drafted according to the outline contained in Phase 3 is sent to the Chief Statistician once this phase is completed and the decision to undertake the statistical activity is taken.

Relevant quality dimensions: Coherence, timeliness, accuracy, accessibility and interpretability.

Importance for cost efficiency: High

Corporate tools available for program managers[4]*:* OECD Statistical Work Programme (OSWP); Gateway to OECD statistical databases; UNECE Integrated Presentation of Statistical Work and Internet sites of international organisations; OECD Glossary of Statistical Terms; OECD.Stat.

[4] All tools are accessible through the OECD Intranet site for statisticians and analysts.

2.2.3 Phase 3: Design and planning of statistical activity

Aims

74. The aims of the guidelines are to:

- ensure that the most accurate and appropriate concepts, definitions, classifications and statistical methodologies are used;
- meet deadlines established by stakeholders;
- minimise costs associated to the statistical activity;
- ensure that, in carrying out the activity, the best available IT and statistical tools are used;
- involve appropriate expert networks in the development of the activity;
- help horizontal Directorates to perform their roles, maximising internal co-ordination and improving their capacity to support new developments (through appropriate statistical inputs, IT infrastructures, dissemination tools, etc.).

Context

75. Following the definition of data requirements in general terms (Phase 1) and the establishment of the extent to which they can be satisfied from other databases (Phase 2), Phase 3 concerns the design, planning and control of the statistical activity. This involves finalising the requirements with regard to concepts, content and coverage, definitions and classifications, reference period, frequency, accuracy and timeliness, and then designing, planning and controlling data collection, processing and dissemination mechanisms to satisfy these requirements efficiently and effectively.

76. Appropriate planning is crucial to the quality of the final output, as well as its cost-effectiveness. The planning phase takes into consideration existing and expected resources, IT opportunities and constraints, and potential external users relevant for marketing initiatives, etc. In this phase is important to evaluate all main aspects of new activities, in particular, those that can be defined as "horizontal", involving as early as possible, experts in STD, ITN and PAC, in order to identify the best tools available and to minimise costs.

77. In this phase, several contacts are normally established with experts working in other international organisations and in national agencies. In this respect, it is fundamental to ensure that the best statistical experts in the issue are involved. On the other hand, national agencies in charge of co-ordination of statistical relations with international organisations have to be informed about future OECD plans. Similar information has to be provided to all interested parties across the OECD, in particular, in STD, in order to make possible any necessary co-ordination, within the Secretariat and with other international organisations.

Guidelines

a) The design and planning phase takes the following aspects into consideration:

- the content of the statistical activity (definitions, classifications, frequency, geographical breakdown, etc.);
- types of outputs;

- statistical methodologies envisaged for conducting necessary analyses;
- statistical processes and IT related aspects, such as tools for collecting data and metadata, characteristics of the database that will contain data and metadata, dissemination tools, etc.;
- adequacy of the current skills of statisticians and other people involved in designing, planning and conducting the activity;
- managerial and financial aspects.

b) Consideration is given to changes in the statistical activity, concepts, content, coverage, frequency, accuracy, and timeliness that would allow more use of data available from internal or external databases.

c) Wherever available, existing international standards are used for concepts, definitions, units, classifications, nomenclatures and compilation methods. Divergences are documented and justified. In the absence of existing international standards, and with a view to minimising provider burden, concepts, definitions, classifications, and nomenclatures are harmonised with commonly used national practices to the maximum extent possible. Lack of harmonisation with either international standards or prevailing national practice is documented and justified.

d) Where the statistical activity entails the development of new concepts, classifications, etc., new developments are discussed with experts working in relevant national and international organisations. These discussions are carried out according to normal OECD procedures.

e) New concepts, definitions and classifications adopted by or agreed within the OECD are widely publicised.

f) The design of the statistical activity is based, as much as possible, on OECD statistical guidelines on data and metadata collection, management and dissemination. Divergences from these guidelines are documented and justified. If necessary, existing guidelines are revised to take into account more advanced or more effective solutions.

g) In this phase, special consideration is given to the planning of measures that ensure the maximum possible application of the UN Fundamental Principles of Official Statistics (refer Annex 1).

h) Financial and managerial planning is carried out according to general OECD principles and rules.

i) Particular attention is paid to the analysis of training needs on statistical and IT issues. Training needs are discussed with the Statistical Training Advisory Group (STAG), which is in charge of planning and organising (in close co-operation with the OECD Learning Centre) training activities for statisticians.

j) To plan the activity, early contacts are established with ITN, STD and PAC to identify the best technical and methodological solutions, as well as to plan necessary training activities.

k) The design of the activity has to take into account, as much as possible, the availability of corporate tools for conducting statistical activities (StatWorks etc).

l) Necessary information about future activities is transmitted to STD, in order to prepare the OECD Statistical Programme of Work and to contribute to the UNECE Integrated Presentation of the statistical activities of international organisations.

Relevant quality dimensions: relevance, accuracy, credibility, timeliness, accessibility, interpretability and coherence.

Importance for cost efficiency: Very high

Corporate tools available for program managers[5]*:* Toolbox for IT solutions; StatWorks; Training programme for statisticians; OECD Glossary of Statistical Terms.

[5] All tools are accessible through the Intranet site for statisticians and analysts.

2.2.4 Phase 4: Usage of data from databases within and external to OECD

Aims

78. The aims of the guidelines are to:

- minimise OECD resources required for the usage of data and metadata;
- ensure that, in carrying out the activity, the best available IT and statistical tools are used;
- minimise the risk of errors in extracting and interpreting data and metadata.

Context

79. In Phase 2 the existence of required data items within one or more internal and/or external databases managed by other international organisations is established. In addition, everything possible is done to harmonise data requirements. This phase refers to the efficiency and effectiveness of the data and metadata extraction process rather data content.

80. As for Phase 2, the databases accessed fall into one of three categories: they can be internal, created by other OECD statistical activities; or external, either belonging to other international organisations; or belonging to national agencies. Given that data available from international organisation databases have mostly been collected from national agencies in the first place, there are often two access options available:

- data can be extracted from the international agency database. This has the advantage of benefiting from any editing or compilation already undertaken by the international agency collecting the data. However, the identification of the source of errors and their correction is made more difficult;
- data can be obtained directly from original national sources. This option may produce more timely data, but is more resource intensive. However, in this case, as the data are transmitted by the national agencies rather than extracted by the OECD, the operation is described in Phase 5, as well as all technical aspects related to data extractions from external databases.

Guidelines

a) If the same data are contained in both internal and external databases, *ceteris paribus* (*i.e.* they have the same detail, timeliness, etc.), OECD databases are used.

b) If the same data are contained in both internal and already published OECD databases, *ceteris paribus* (*i.e.* they have the same detail, timeliness, etc.), databases already published by the OECD are used.

c) When data have to be extracted from an OECD database or a database managed by other international organisations:

- the data to be used for preparing publications are extracted from the original database (or data set) containing necessary data and not from derived databases;
- maximum use is made of "OECD reference series" or in officially published databases;

- data are always (wherever possible) used with related metadata. If metadata are stored separately from data, files containing metadata are always collected and analysed;

- if data disseminated by other international organisations are used to produce on-going OECD publications, a communication is officially sent to the organisations concerned. A formal agreement (Memorandum of Understanding) between the OECD and the organisation concerned should be signed, in order to clearly establish data and metadata characteristics necessary for the OECD activity (timeliness, error free, etc.); OECD Publishing Editorial and Rights provide policies and guidelines, which should be adhered to.

- if several variables estimated in a specific statistical context (*e.g.* national accounts, health accounts, R&D statistics) have to be used, they are preferably extracted from the same database. If different databases must be used and metadata are not exhaustive, direct contact with managers of the databases from which data are extracted is established;

- in all cases, liaison with manager(s) of the database(s) from which data are extracted is strongly suggested, especially if data are used to produce regular publications. In particular, useful information to be acquired may concern:

  - presence of provisional or non-verified information in available data and metadata;

  - expected future revisions in data and metadata (calendar, content, etc.);

  - potential problems in data quality (mainly accuracy and coherence) for data not completely documented by related metadata;

  - any other information useful to evaluate aspects related to data quality;

  - coherence between the required data and those contained in other similar sources;

- ratios and indicators can be used as such only if they are considered "OECD reference series". Otherwise, individual series are extracted and ratios are calculated according to the methodology established by the activity manager;

- data extraction procedures are automated. Manual imputations are always avoided;

- data extraction procedures are efficient. In particular, for automated extraction, the data requirements are embedded in software and thus extraction can be conducted at the press of a button with minimal intervention. In any case, data extraction procedures are documented;

- data extraction procedures are error proofed. In particular, possible sources and types of error are analysed and provisions are in place to check and correct for errors that occur;

- if complex extractions are needed and/or generalised extraction procedures are not available, the procedure to extract data is agreed to by the activity manager and manager(s) of the database(s) from which data have to be extracted;

- if the extraction requires the treatment of confidential data not accessible to unauthorised OECD officials, the procedure to extract data is agreed to by the activity manager and manager(s) of the database(s) from which confidential data have to be extracted. Temporary authorisations to access confidential data can be attributed to Secretariat officials according to the rules established by the Organisation;

- information about the OECD database used for data extraction is reported in the OSWP.

d) Summary information about OECD source and external databases used to carry out the activity is reported in the OSWP.

Relevant quality dimensions: accuracy, timeliness, accessibility, and interpretability.

Importance for cost efficiency: Very high

Corporate tools available for program managers[6]*:* OECD Glossary of Statistical Terms; Gateway to OECD Statistical Databases; OECD.Stat; Corporate procedures to extract data and metadata from existing sources.

[6] All tools are accessible through the Intranet site for statisticians and analysts.

2.2.5 Phase 5: Implementation of specific data collection mechanism

Aims

81. The aims of the guidelines are to:

- minimise OECD resources required for data and metadata collection;
- minimise the burden on data and metadata providers;
- minimise the risk of errors in collecting data and metadata;
- provide to national agencies in charge of co-ordination of statistical activities summary information about data and metadata flows between the country and the OECD;
- provide the OECD community of statisticians information on data and metadata flows between each country and the OECD.

Context

82. The requirements for data are established in Phase 1; the possibilities for extracting these data from other databases are determined in Phase 2 and elaborated in Phase 4. Phase 5 refers to the acquisition of data that cannot be extracted from other databases within the OECD and therefore require the establishment and operation of data collection (or selection) and transmission mechanisms specifically to service the statistical activity.

83. In general, this imposes a burden on the data provider, who has to obtain (or select) the data and transmit them to the OECD. For Member countries, provision of data by national agencies is a condition of membership. For co-operating non-Member countries and international organisations, provision of data is a condition of co-operation with the OECD. Nevertheless, in all cases, the OECD has an obligation to minimise provider burden.

84. OECD data collection activities are a reflection of the decentralised nature of the OECD statistical system and collection processes have, in the main, evolved over the years from bilateral arrangements between individual areas within the OECD Secretariat and national agencies or other international organisations. Furthermore, the diversity of data management systems used at the OECD has contributed to the wide variety of such arrangements. As a result, the OECD currently uses a large number of individual data flows, formats and technologies to collect data and metadata. In the context of this phase, the guidelines focus on issues of data and metadata exchange rather than data and metadata content issues.

85. In general, data collection activities may be divided between those that:

a) collect data that are part of an individual source agency on-going dissemination programme established at national level;

b) are already available at national level, but are especially requested by the OECD (for example, following a more detailed breakdown or a different classification) and are not part of an on-going dissemination programme established at national level[7];

[7] For example, national accounts are transmitted to international organisations adopting a specific questionnaire, which also contains data not regularly disseminated at national level.

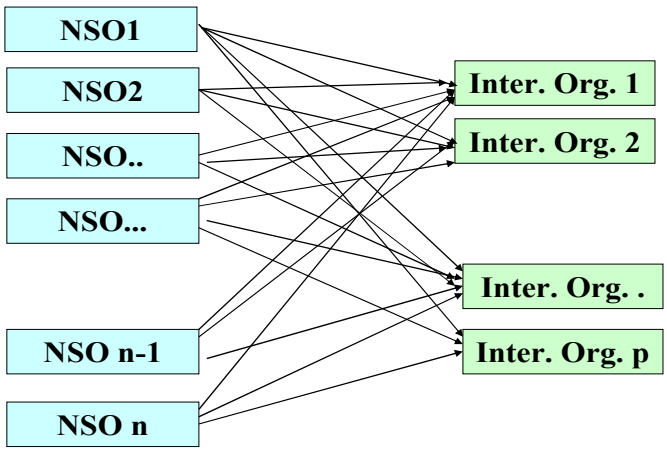
c) become available only because a statistical survey is carried out by national sources following an OECD special request.

86. There are four main possible data and metadata transmission models:

- The first model "many to many" corresponds to the situation where the source agency sends different files containing very similar (or the same) data and metadata to all international organisations involved in the exchange. In this case the reporting burden on the source agency is maximised. This model represents the most common situation and is the only feasible model where data requirements, transmission formats, data presentation formats and media are not common across international organisations, or where the national source is not able to adopt the standards proposed by the international organisation. In some cases, common questionnaires are defined by international organisations. In these instances the source agency sends the same file at the same time to all international organisations involved. This is the most efficient model from the source agency's point of view but presupposes the co-ordination of content, data presentation format and medium between international organisations.

Chart 1

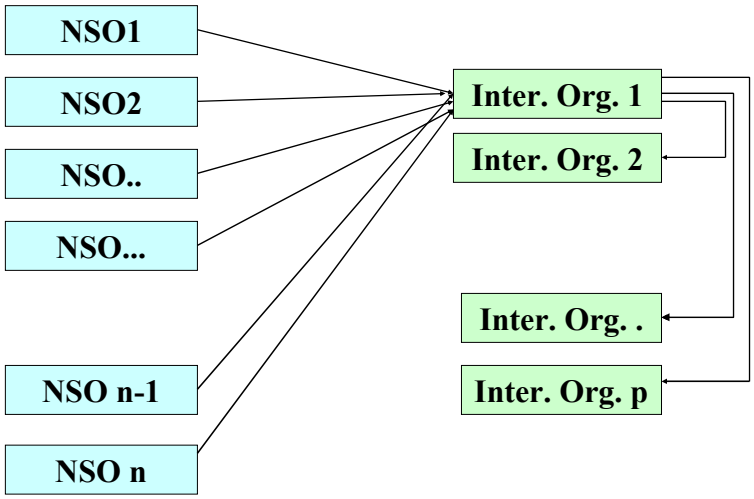
Many to many



- The second model "many to one" is the situation where data is sent to one international organisation that redistributes the data to other organisations. This model also presupposes co-ordination of content, format and medium between international organisations. In this model, the data provider, this time different from the source agency, is the most active participant of the exchange. A major drawback of this model is the absence of direct contact between international organisations and national providers, other than the organisation acting as the gateway. Another problem occurs when a data error is discovered by one of the international organisations, corrected, but changes are not communicated to the others.

Chart 2

Many to one



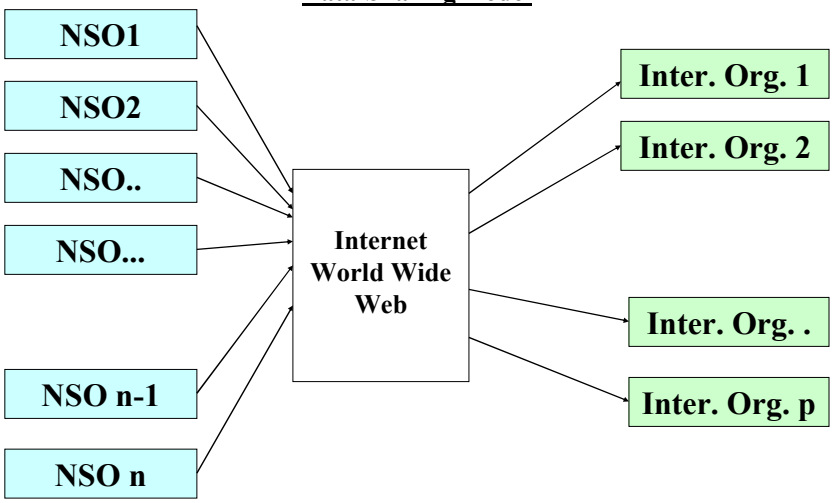
- The third model is the "data-sharing" model where the source agency loads their data into a repository accessible to a number of international organisations. The repository could simply be the internet. The data made available by each source agency need not use the same format or even be in one physical database, but, if data are in the same format, it is possible to present the data as a common database. A simple form of the data-sharing model is the situation where national statistical organisations place all required data on the internet in their web databases. This model is very efficient from the data provider's point of view and the data user is the more active participant. Furthermore, the receiving organisation benefits from receiving data to the same quality standards the source agency extends to their regular web products. Other important advantages of this model are that it offers the opportunity for automating the extraction process and that it could be used as a common repository. The main obstacle to the implementation of this model by a large number of source agencies in the short-term is the high degree of co-operation and agreement needed amongst data providers and data collectors. Nevertheless, use of technologies such as XML would greatly facilitate the implementation of this model.

- Finally, the "special agreement" model is normally used when data are transmitted as part of a wider agreement, which normally oblige the national source to conduct an "OECD survey". In this model, the transmission mechanism is established on a case by case basis.

87. The first two models are normally used for cases (a) and (b) outlined in para. 78, while the fourth model is used in special cases or when microdata have to be transmitted. Nowadays, "data sharing" is still rare, even if it seems very promising and perhaps represents the future of data and metadata

Chart 3

Data-Sharing model



transmission[8]. It is used by the OECD for data collections of type (a) for a few countries. In fact, international organisations are currently co-operating to establish new standards for data and metadata descriptions. In particular, OECD, Eurostat, the UN Statistical Division, the European Central Bank, the IMF and the Bank for International Settlements have launched the SDMX project which is a joint initiative to explore common e-standards and on-going standardisation activities (more information about these type of initiatives are presented in Annex 4 of this document).

88. The OECD currently makes greater use of the first type of collection process (the many-to-many model), with arrangements generally being made on a bilateral basis between individual Directorates and the external source agency. The use of on-line access to databases is increasing, though there is considerable variation in the range of data contained on those databases and the efficiency of search and extraction facilities provided by the source agency.

Guidelines

a) The data and metadata transmission procedure is designed to be:

- as automated as possible at both ends of the transmission process. If this is not possible, the reasons why are regularly reviewed and documented;

- as efficient as possible for both organisations. The costs associated with the transmission comprise both those derived from the development of the procedures for data and metadata transmission, and those associated with the on-going transmissions and the maintenance of the procedures;

- error proofed. In particular, an analysis of possible sources and types of error is carried out before choosing the procedure. Procedures (embedded in electronic questionnaires or used when data

[8] This approach is going to be adopted, as case study, for the transmission of annual national accounts. See the paper on the National Accounts World Wide Exchange (NAWWE) project at http://www.oecd.org/doc/M00033000/M00033849.doc

are received) are put in place to check whether such errors have occurred and to correct for them if they have;

- timely, *i.e.* transmission occurs as soon as the data are available.

b) Multiple data and metadata transmission procedures can be used for the same statistical activity. They are established according to:

- the technical capacity of the agency that provides data and metadata. The OECD promotes the use of the most advanced transmission tools, encouraging national agencies to adopt international standards available for data and metadata exchange;

- the amount of information to exchange and its characteristics (*e.g.*, time series, cross sectional data;

- the frequency of the transmission (monthly, annual, etc.);

- the participation of other international organisations in the transmission agreement or in similar exchanges.

c) If different options are available, the data and metadata transmission process is selected in agreement with the source agency, taking into account the needs of both sides of the transmission process.

d) Preliminary information is acquired from the data providers to establish whether the data satisfy the concepts, content, coverage, frequency, accuracy, timeliness, and coherence specified by the primary users and to enable other users to interpret data and determine whether they are fit for their intended use(s). This information can be acquired through national representatives sitting in the body (committee, working party, etc.) interested in the statistical activity.

e) For each statistical activity, data and metadata transmission procedures are documented. Detailed documentation about these procedures is updated and stored under the supervision of the manager in charge of the activity, whilst summary information is stored in the OSWP database.

f) Because data collections carried out across the OECD have very different characteristics, specific guidelines have been developed for each category of data collection. In particular, OECD data requirements may be summarised as falling within six broad categories:

  - Monthly and quarterly short-term economic (aggregate) data and indicators (including quarterly national accounts, labour force survey data, international trade, etc.) used mainly by the OECD Economics Department for the preparation of the *OECD Economic Outlook* and country surveys. At first sight, these data may appear comparable across countries, though closer examination frequently reveals extensive differences in definitions, classifications, coverage, etc, thus highlighting the need for the collection of appropriate metadata for each indicator. Such data are normally part of the regular dissemination programme of national statistical institutions;

  - Annual "basic" economic and social statistics (annual national accounts, structural business statistics, labour force survey data, international trade by commodity, etc.) required for both specific country analyses and for international comparisons. These data are generally compiled by national agencies following international standards (and questionnaires agreed to by several international organisations) though some adjustments to the data by the OECD Secretariat are

often required to produce more comparable figures. Such data are normally compiled by national statistical organisations but not necessarily as part of their "normal" dissemination programme;

- Other statistical data on specific economic, social and environmental variables collected either annually or on an ad hoc basis. These data are widely produced in the absence of any existing international standards, using national definitions and classifications. Extensive work is required by the OECD Secretariat to make such data more comparable and in the collection of metadata to permit evaluation of accuracy and comparability. Such data are also normally compiled by national statistical organisations or other public agencies;

- Statistics compiled annually or occasionally by agencies other than national statistical organisations. These data may be by-products of administrative activities compiled by agencies without a well established "statistical culture" and/or a commitment to adherence to existing international statistical standards. Metadata is frequently not available for such data;

- Data produced annually or occasionally by national agencies (often national statistical organisations or other government agencies) conducting surveys specifically designed by the OECD;

- Microdata, normally produced by national statistical organisations transmitted to the OECD or accessed by OECD research staff through electronic networks.

Table 4 below provides a synthetic view of suggested options for each of the six categories presented above with reference to transmission processes. For each category, the preferred option is indicated, in addition to (at least) one "second best" solution.

Table 4 – Taxonomy of data and metadata transmission methods to be used for different categories of data and metadata collections

TOOL	**Category A** *Short-term, part of national dissemination*	**Category B** *Annual basic NSO statistics, standard*	**Category C** *Annual basic NSO statistics non-standard*	**Category D** *Annual non-NSO statistics*	**Category E** *Annual, occasional OECD specific*	**Category F** *Microdata*
Paper form	**To be avoided both for data and metadata**	**To be avoided both for data and metadata**	**To be avoided both for data and metadata**	Possible only for a very small amount of data and metadata	**To be avoided both for data and metadata**	**To be avoided both for data and metadata**
Spreadsheet file (or .csv)	Possible for small amount of data or for non-member countries. Useful for metadata	Possible for data (with internal checks) and metadata.	**Preferred option** both for data (with internal checks) and metadata (in particular, for one-off data collections)	**Preferred option** for data (with internal checks) and metadata (especially for one-off data collections)	**Preferred option** both for data (with internal checks) and metadata (in particular, for one-off data collections)	Possible for data. Useful for metadata
GESMES/TS	Useful option both for data and metadata if SDMX not implemented.	**Second best option** both for data and metadata if SDMX not implemented.	Useful both for data and metadata if SDMX not implemented.	Normally not feasible because of initial investment and the limited statistical capacity of national sources	**Second best option** both for data and metadata but might require high initial investment if SDMX not implemented.	**Preferred option** both for data and metadata
SDMX	**Preferred option** for data, also for metadata	**Preferred option** for data, also for metadata	**Preferred option** for data, also for metadata	**Preferred option** for data, also for metadata	**Preferred option** for data, also for metadata	Not yet possible
On-line extractions from web-sites	Useful option both for data and metadata (via SDMX web service)	Useful option both for data and metadata (via SDMX web service)	Useful option both for data and metadata (via SDMX web service)	Normally not feasible because of the limited statistical capacity of national sources	**To be avoided both for data and metadata**	Not feasible for data because of confidentiality issue. Sometimes possible for metadata
On-line extractions from DBs internal to national sources	Useful option both for data and metadata (via SDMX web service)	Very useful option only for data, rarely for metadata	Useful option both for data and metadata (via SDMX web service)	Normally not feasible because of the limited statistical capacity of national sources	Very useful option both for data and metadata	Normally not feasible for data because of the security-confidentiality issue. Sometimes possible for metadata
Ad-hoc formats specified case-by-case	**To be avoided**	**To be avoided**	Possible option both for data and metadata	**Second best option** both for data and metadata	Possible option both for data and metadata	**Second best option both for data and metadata**

g) Whenever possible and appropriate, corporate tools provided in the context of StatWorks have to be used. In particular, when the data collection is undertaken using a spreadsheet file, the generalised tool for designing the questionnaire has to be used.

h) When time series have to be collected, each transmission should cover the full length of the series (*i.e.*, not only the most recent values, but also the historical values). The origin of the time series should be the initial period (year, quarter, month, etc.) when the most recent methodological rules (definition, coverage, classification, etc.) are applied, in order to incorporate all revisions to historical data, if any.

i) If confidential data are collected, they must be clearly flagged by the source agency.

j) If a specific (electronic or paper) questionnaire is issued, it is designed according to internationally recognised standards, is tested prior to use and (for on-going data collections) is evaluated after use. The request transmitted to the source agency has to contain the following elements:

- a letter containing a brief presentation of the purpose of the initiative and the political/analytical context in which it has been launched, a description of the future use of data to be transmitted and the general characteristics of data and metadata requested, and a clear reference to the expected treatment of confidential data, if any. A contact name must be indicated for answering questions and providing relevant information to source agencies. A copy of the letter is always sent to the OECD Chief Statistician and to the national agency (normally the national statistical office) in charge of co-ordination of statistical relations with international organisations;

- explanatory notes containing instructions for the compilation of the questionnaire and for the transmission to the OECD of the compiled questionnaire. Ensure that the instructions are concise, clear and easy to read, for example, all acronyms should be explained.

- the questionnaire;

- detailed description of definitions and classifications to be adopted in compiling the questionnaire.

k) In designing the questionnaire:

- use words and concepts that have the same meanings for both respondents and the questionnaire designers. When appropriate, reuse questions from other surveys;

- ensure that the questionnaire has a professional appearance;

- pre-test the questionnaire before it is sent out officially. The test can be conducted informally using two or three respondents with different degrees of statistical/technical capacity;

l) The metadata required from data providers are expressed using internationally or commonly accepted formats. The metadata are guaranteed by the providers to be accurate and up to date. The initial collection and subsequent on-going maintenance of statistical metadata is a costly exercise. For this reason, careful thought needs to be given to the amount of necessary metadata before starting the collection.

m) If an "OECD survey" has to be explicitly designed, clear and precise guidelines have to be issued to the national agency in charge of the survey. At a minimum, these guidelines should cover the following aspects[9]:

- the expected results in terms of final tables;

- the survey frame, which has to be derived using the most appropriate source of target statistical units. Typically, the business register is used in the case of an economic survey, the farm register in the case of an agricultural survey, and the labour force or household expenditure area frame in the case of a household survey;

- the required coverage and content, typically comprising a stratified simple random sample in the case of an economic survey and a multi-stage area sample in the case of a household survey;

- the minimum sample size required to produce estimates of the required degree of accuracy. In the case of a survey that is repeated annually or more frequently, sample rotation is considered as a means of reducing the burden on individual respondents;

- the instructions for compiling the report to be transmitted to the OECD with the results of the survey. The report has to contain summary information about the way in which the survey has been conducted, statistical measures of accuracy, data editing procedures adopted, etc.

Relevant quality dimensions: accuracy, timeliness, accessibility, and interpretability.

Importance for cost efficiency: Very high

Corporate tools available for program managers[10]: OECD Glossary of Statistical Terms; corporate procedures to extract data and metadata from existing sources (web sites and databases) and to collect data from external providers through electronic questionnaires (StatWorks).

[9] For more detailed survey guidelines, see *Statistics Canada Quality Guidelines*, October, 1998

[10] All tools are accessible through the Intranet site for statisticians and analysts.

2.2.6 Phase 6: Data and metadata verification, compilation, storage, analysis and evaluation

Aims

89. The aims of the guidelines are to:

- minimise OECD resources required for data and metadata verification, compilation and storage;
- ensure that the best statistical tools are used to conduct data verification, compilation and analysis;
- ensure that appropriate documentation about data verification, compilation, storage and analysis processes is produced and made accessible to interested users;
- maximise accessibility to and interpretability of data and metadata;
- allow easy integration of data and metadata compiled by various statistical activities, in order to maximise the coherence of OECD statistics.

Context

90. This phase involves:

- the application of checks that identify missing, invalid or inconsistent data and metadata or that point to data and metadata that are potentially in error (*data editing*);
- the compilation of final data and metadata, as well as the production of statistical data or indicators calculated using information collected from external or internal sources;
- the storage of data and metadata in datasets from which final statistical products are compiled or users can extract data and metadata;
- the usage of data for producing statistical and/or analytical results;
- the review of how data and metadata are useful for the original purpose for which they were collected and compiled.

91. The goals of data verification are threefold to tidy up data and metadata; to provide information to users about the accuracy of the data and metadata disseminated; and to provide the basis for future improvements to the statistical activity process. Data verification involves data item edit checks to ensure values are within valid ranges and respect accounting identities. It also includes comparisons over time, across data items, across datasets and across countries. Similar verification can be made for metadata. In this case, verification mainly involves the coherence of the information received over time, across datasets and across countries. Data and metadata editing is likely to be the single most expensive phase of a statistical activity. Not only is the practice of editing costly in terms of resources used, timeliness and increased response burden, but it can also lead to severe bias resulting from an incorrect design or application of editing procedures.

92. Data compilation involves the production of additional figures or indicators derived from data originally collected by the OECD. These comprise ratios, supra-national aggregates, seasonally adjusted or trend-cycle values, composite indicators or composite leading indicators. Metadata compilation involves

the preparation of summary or detailed metadata about data received or autonomously compiled by the Secretariat, in a format useful for the uses for which they have been collected.

93. In some cases, the compilation phase also requires the estimation (imputation) of missing values, adjustments to harmonise data across countries and to overcome series breaks over time caused by changes in concepts or methodology. Some problems can be eliminated through contact with the data and metadata provider agency, but in several cases the OECD has to impute (or estimate) some values or time series. In this case, the most appropriate technique has to be used, in order to improve the overall quality of data disseminated.

94. Storage involves placing the data and associated metadata in a database from where they can be extracted and/or disseminated. Different types of metadata are commonly used to describe data:

- *Table headings and footnotes*. They are an integral part of statistical tables published by the OECD and their aim is to make table headings clear and as brief as possible. Footnotes are also kept to a minimum and are restricted to those essential for an understanding of the data. In most instances, tables and footnotes are the only metadata accessed by users.

- *Explanatory notes*. They generally provide a brief general description of the indicator and an outline of key issues that can affect the use of the data, target definitions, etc. Explanatory notes do not necessarily have to provide much detail on individual country methodology/practices.

- *Summary sources and definitions*. They provide a brief outline of current national practices summarised under a restricted number of broad headings or topics (*e.g.* definition, coverage, collection and calculation).

- *More detailed sources and methods*. Such metadata contain more detailed methodological information on individual country practices normally collected and disseminated on the basis of a detailed metadata model, template or prompt points. These metadata can encompass the whole range of methodologies involved in describing the source, concepts and coverage, data collection, data manipulation, etc., for the compilation of an indicator. Such metadata are generally very costly to compile and maintain.

- *Methodological information disseminated by national sources in publications and/or on websites*. These are potentially the source of the most detailed methodological information available. Some (though not all) OECD Member countries publish very detailed concepts, sources and methods for a number of their statistics. The provision of more extensive methodological information, and the need to make it readily accessible to users through its dissemination on the web, is now receiving greater recognition.

95. Analysis involves use of data to examine the issues for which the data collection was initiated. In essence, an analysis comprises a statement of the issue, the evidence, and the answer based on the data, while evaluation involves review of the data relative to the original data requirements.

96. The way in which these sub-phases are conducted is absolutely crucial for minimising "non-sampling errors", which are the most relevant source of errors for a "secondary" data provider such as the OECD. These errors can be produced by:

- *coverage errors*, due to divergence between the target population and definitions established by the OECD and those used by the source agency. They include both over-coverage, under-coverage (*i.e.* the inclusion/exclusion of not requested items) and misclassification;

- *processing errors*, due to the incorrect application of processes to code, classify, transform and tabulate data and metadata.

97. Verification, compilation, storage, analysis and evaluation activities are carried out by OECD Directorates using a wide range of methodological tools, IT infrastructures and procedures. In some cases, this variety of approaches can reduce the efficiency of the statistical activity, the accuracy of the results, as well as the accessibility to and interpretability of results. In addition, insufficient documentation about procedures used to verify and compile data can affect the coherence of data produced, especially when staff in charge of these phases change over time.

Guidelines

a) Individual data items are subject to automated checks to ensure that values are within valid ranges and respect accounting identities. To identify outliers individual data items are subject to automated comparisons within each country, across countries and over time, using appropriate statistical techniques. Verification processes are to be as repeatable as possible and are to be objective. This is best accomplished through automation and preliminary test of automatic procedures on simulated data.

b) Specific documentation about the procedures used for verification and the rules to be used to accept outliers is prepared before starting the verification phase. Regular assessments of verification and imputation procedures are carried out over time.

c) If applicable, comparisons of data and metadata items are made across related datasets and with OECD statistical glossaries. Inconsistencies are identified for further investigation.

d) Unusual values identified by verification programs are investigated, accepted as outliers or queried with data providers and documented so that users are informed.

e) Edits to data and metadata are internally consistent (*i.e.* not self-contradictory) and ensure that reapplying the verification procedure to edited values yields no further errors.

f) A data item or series of items that are missing or deemed to be inconsistent or illogical by the editing process may be subject to imputation processes that substitute valid values for the missing or illogical values with the permission of the source agency. Imputation without such permission can only be done when:

   - supra-national aggregates have to be produced (without publishing values for the country involved) and the country concerned does not represent a large portion of the aggregate;

   - data are not produced at all by the source agency, but good proxies or basic data are available (for example, in the case of time series breaks due to changes in classifications or when data can be calculated starting from other data provided by the national source).

When data are estimated or imputed, the label “OECD estimate” is to be used and be clearly identifiable by users.

g) Supra-national aggregates are computed using internationally accepted standards and practices. The weights used to calculate aggregates are always made available to users. If, for methodological reasons, an inconsistency exists between the elementary data and the aggregates, an explanation should be provided to users.

h) Seasonally adjusted and trend-cycle values are computed using internationally accepted standards and practices. Parameters used are documented. If nationally adjusted data are available, they are published only if the procedures used are compatible with those adopted by the OECD.

i) Particular attention is directed to making the figures as comparable as possible across countries and over time. Harmonisation adjustments to achieve such comparability are computed using internationally accepted standards and practices. In particular, adjustments in definitions, coverage and classifications are made wherever possible to improve international comparability, even if national data remain important for within-country analyses and assessments. All adjustments made by the OECD are clearly described in the framework of metadata associated with data produced in order to clearly distinguish them from data published by national agencies.

j) If variables estimated in different statistical contexts (*e.g.* national accounts, health accounts, R&D statistics) have to be used to compile derived measures and indicators, they are preferably extracted from the "OECD reference series" environment. If different databases must be used, and metadata are not exhaustive, direct contact with managers of the databases from which data are extracted is established.

k) Liaison with manager(s) of the database(s) from which data are extracted is strongly suggested, especially if data are used to produce regular OECD publications. In particular, useful information to be acquired may concern:

- expected future revisions in data and metadata (calendar, content, etc.);
- potential problems in data quality (mainly accuracy and coherence) of data not completely documented by related metadata;
- any other information useful to evaluate aspects related to data quality;
- coherence between concerned data and those contained in other similar sources;
- procedures available to automatically extract data and metadata.

l) Data are stored in accordance with OECD statistical and IT standards. Table 5 contains an overview of preferred options for storing data with different characteristics. Tools available in the context of StatWorks have to be used as much as possible.

m) When confidential data are received, they are stored and analysed in a manner that avoids risk of disclosure. The manager in charge of the database containing confidential data must adopt all necessary precautions and measures to avoid any unauthorised access to them. The existence of confidential data and procedures adopted to avoid the confidentiality disclosure is described in a special register, managed by the Chief Statistician of the OECD.

n) When new data are to be stored and compiled, they are preferably managed in the framework of already existing databases, according to their main characteristics. The number of databases is minimised, as well as the use of new software similar to those already managed by the Organisation. Final data are stored in OECD.Stat, according to technical standards defined by STD and ITN and according to publishing standards defined by PAC.

o) Automatic procedures are adopted to constantly monitor how data storage is progressing, how the coverage (in terms of variables, countries, etc.) of data received is compared to expectations and deadlines, and how significant are revisions in data, in order to take timely action to solve any emerging problems.

p) Metadata are compiled using internationally accepted standards and practices. In particular:

- metadata are compiled following definitions contained in the SDMX Metadata Vocabulary[11];
- all relevant definitions used in conducting statistical activities are stored in the OECD Glossary of Statistical Terms;.

q) The OECD collects and maintains minimal summary methodological information on its databases, consistent with the primary objective of providing appropriate transparency to the statistics they describe, except for activities whose primary purpose is the collection of metadata on practices adopted by national and international data providers. The OECD makes maximum use of more detailed sources and methods type metadata compiled and disseminated by other international organisations and national agencies through the creation of URL links with metadata stored in OECD databases in lieu of direct collection by the OECD.

r) Metadata are stored according to the general rules presented in Table 6. In particular, all metadata (summary sources and definitions and more detailed sources and methods metadata described in para. 87) should be readily accessible to all users. Final metadata are stored, as much as possible, in the forthcoming common repository of metadata.

s) Non-confidential final data and metadata are made easily accessible to internal users. If specific restrictions apply to internal users other than people involved in the preparation of publications, the reasons for these restrictions are communicated to users.

t) The manager of the statistical activity is responsible for the preparation and on-going maintenance of documentation about procedures adopted for data and metadata verification and compilation. The documentation is prepared following OECD standards and made available to internal users.

u) Data are subject to issue-oriented analyses providing insight into the issues for which the data collection was initiated or that are topical.

v) The data and the processes for their collection, processing and dissemination, are evaluated with respect to the objectives for which the statistical activity was initiated or is being continued.

w) A global evaluation is conducted at the end of each production cycle to evaluate the procedures adopted to design and plan the statistical activity, the data and metadata in terms of accuracy, timeliness and relevance, and the processes adopted during the implementation of the activity. For activities that involve “continuous” (monthly, quarterly, etc.) production cycles, the assessment is made annually. The results of this evaluation are transmitted to national and international data and metadata providers, as well as to OECD stakeholders.

x) When data and metadata are collected or compiled in the framework of a project mainly financed through voluntary contributions, at the end of the project a careful evaluation is carried out to determine if, in the interest of the OECD, it should be continued. To this purpose, the manager of the activity will prepare a brief document illustrating pros and cons of continuing the activity for consideration by the SPG.

[11] Available on the SDMX project website at http://www.sdmx.org

Table 5 – Taxonomy of tools for data storage to be used for different categories of data

TOOL	**Category A** *Time series, large databases for on-going activities*	**Category B** *Time series, small databases for on-going activities*	**Category C** *Time series, small databases for one-off activities*	**Category D** *Non-time series, large databases for on-going activities*	**Category E** *Non-time series, small databases for on-going activities*	**Category F** *Microdata*
SQL database	**Preferred option**	**Preferred option**		**Preferred option**	**Preferred option**	**Preferred option**
Fame database	**Second best option**	**Second best option**	**Preferred option**			
MS-Access databases					**Second best option**	
Spreadsheet file (or .csv)			**Second best option**			
self-assessments				Possible option	Possible option	Possible option

Table 6 – Metadata storage guidelines

TOOL	**Category A** *Table headings and footnotes*	**Category B** *Explanatory notes*	**Category C** *Sources and definitions*	**Category D** *Sources and methods*	**Category E** *Methodological information disseminated by national sources in publications and/or on websites*
Web document				Possible option	Possible option
OECD Central database(s)	**Preferred option**	**Preferred option**	**Preferred option**	**Preferred option**	**Preferred option**
OECD Local database	**Second best option**	**Second best option**			

Relevant quality dimensions: accuracy, timeliness, accessibility, interpretability and coherence.

Importance for cost efficiency: Very high

Corporate tools available for program managers[12]: OECD.Stat; StatWorks, MetaStore

[12] All tools are accessible through the OECD Intranet site for statisticians and analysts.

2.2.7 Phase 7: Data and metadata dissemination

Aims

98. The aims of the guidelines are to:

- ensure that all statistical products are disseminated in accordance with OECD Publishing and statistical policies;
- maximise the accessibility to and interpretability of OECD statistical products;
- minimise resources necessary to disseminate statistical products through different media;
- ensure that statistics disseminated are timely and punctual;
- maximise the coherence of data disseminated by the OECD in terms of content and presentation.

Context

99. Dissemination means making data and accompanying metadata and analyses available to users. The main considerations typically involve the following:

- *content*: the degree of aggregation and the need to respect privacy and confidentiality constraints;
- *dissemination media*: with advances in information technology, statistical agencies are adding new formats, especially using internet technologies, including making data and metadata available in a structured manner so that it is machine readable
- *pricing policy*: data and other statistical products are a potential source of revenue. On the other hand, organisations that have acquired data using public funds have a responsibility to civil society and an obligation to disseminate (at least some) data as a public good. While some data may be available for free of charge, it may be in parallel offered for fee with value-added services.

100. In the OECD context, users may be divided into three broad categories: internal users, external users and Secretariat staff. These categories use different media and are subject to different pricing policies. In addition to print publications, the main electronic distribution mechanisms are:

- *OLISNet*: freely available for authorised users in OECD Member governments. It provides access to the majority of data, including preliminary and unreleased data;
- *Intranet*: for Secretariat staff only. It provides access to the majority of data, including preliminary, unreleased and confidential data;
- *OECD iLibrary*: primarily aimed at institutional users who are willing to purchase OECD value-added online services. It provides access to published data;
- *OECD web site*: for all classes of external users. It provide access to published data.

101. The OECD Publishing Policies within which dissemination must take place. The policy states that the main target of OECD Publishing is to disseminate as widely as possible and in the most effective

manner the results of work carried out within the OECD on issues of significant, recognised interest and relevance so as to:

- help build support and understanding in Member countries and, as appropriate, in other countries for policy approaches and standards that Members are pursuing within the framework of the Organisation's programme of work; and

- enhance the credibility of the OECD as a source of timely, relevant analyses, high quality statistics and policy prescriptions broadly reflecting the economic, environmental and social performance within the scope of the Organisation's work programme.

102. In addition, in acknowledging that dissemination of statistics requires special consideration, the OECD Council recognised the need to improve the value of statistical publications in order to safeguard the reputation and credibility of the Organisation.

Guidelines

(a) Statistical products are prepared according to internationally recognised standards[13]. In particular:

- the presentation of data and metadata has to be clear and unambiguous;

- statistical data have to be interpretable by thethe various audiences; ;

- data are published according to a calendar of release; all products include a contact name, e-mail address, in order to answer user queries;

- statistical data contained in different products have to be as coherent as possible. Apparent inconsistencies should be avoided as much as possible.

(b) Data is not manipulated, nor is their release timed in response to political pressure. The Secretariat has to decide if the publication of poor quality data received from countries affects the overall credibility of the OECD as a high quality data provider. If the answer is yes, the Secretariat should refuse to publish the data.

(c) When data validation by external individuals or organisations (in particular, by national experts participating in OECD bodies) is necessary to significantly enhance data quality, unreleased information may be provided to selected individuals or organisations. The latter must take adequate measures to ensure the confidentiality of the information received. Once agreement between the Secretariat and countries has been reached on collection and elaboration of specified data, the data that are going to be published cannot be withdrawn in response to political pressure.

(d) Users must be provided with information necessary to understand both the strengths and limitations of the data being disseminated. Users must be provided with information necessary to avoid inconsistencies between OECD statistical products. The documentation provided to users on data quality should engender an awareness of quality to ensure the proper use of the data.

(e) Documentation on methodology must permit users to assess whether the data adequately approximate what they wish to measure and whether data are produced with tolerances acceptable for their intended use. The documentation provided should be clear, well organised and accessible. It should cover, at least:

[13] PAC is developing editorial guidelines for statistical paper products.

- the type of data sources used;
- the nature and purpose of the product, as well as the intended uses of the data;
- the conceptual universe covered by data;
- key concepts, variables (or characteristics) and classifications used;
- a statement of key accuracy issues, as well as an acknowledgement that the data are subject to error and (if applicable) that the level of error may vary geographically and by other characteristics;
- any variation in accuracy and coherence over time and across countries. The issue of coherence is especially relevant for OECD statistics;
- if applicable, a statement advising that the data are subject to revision;
- if applicable, a description of benchmarking and seasonal adjustment made to the data and their impact.

(f) For statistics derived from administrative sources, the following topics should be also covered:

- the purposes for which the data were originally collected;
- the merits and shortcomings of the data for the statistical purpose for which they are being used (*e.g.*, in terms of conceptual and coverage bias);
- how the data are processed after being received and what, if anything, is done to correct problems in the original data set;
- the reliability of the estimates, including caveats where necessary.

(g) For products that include primarily, or only, analytical results, documentation should be provided on both the source data and the method of analysis. The requirements for documentation on data source are similar to those for other statistical products. Documentation of the methods of analysis may be incorporated into the product either as part of the presentation of the analytical results in the body of the report, or in separate "text boxes". Such "text boxes" should also include summary information on:

- the data sources;
- key features of the methodology and accuracy of the source data pertinent to the analysis;
- analytical objectives, concepts and variables;
- analytical methods used, their assumptions and caveats;
- statistical significance of the results and any relevant conflicting or corroborating results;
- appropriate use of the results;
- a brief description of the possible impact of accuracy issues, assumptions and caveats

(h) The OECD regularly publishes an advance release calendar for all statistical products. The publication schedule may comprise a set of target release dates (for example, for press releases of key economic indicators) or may involve a commitment to release data within a prescribed time period from their receipt[14]. The calendar is issued monthly, covering the following three months. If there are occasions where the OECD cannot adhere to its schedule, the reasons for that (for example, technical reasons or changes in the priorities of the Organisation) are clearly communicated to users.

(i) The OECD encourages public accessibility to its statistics through the media. To this end, media representatives are provided with access to online resources and complimentary copies of statistical publications and other products, according to the OECD Publishing Policy established by the OECD Council.

(j) To promote a wide knowledge and use of statistics produced by the OECD, press releases are normally issued to present new data. Statistical press releases can contain:

- Monthly and quarterly data and indicators published on a regular basis that can have an impact on market behaviour. A list of these releases and the procedure for disseminating these data is established by the Chief Statistician in consultation with PAC and made public. In any case, these data must be simultaneously disseminated to all users.

- Annual data and indicators produced on a regular basis or one-off. The procedure for disseminating these data is established by PAC and can comprise their advanced dissemination to media under embargo.

(k) All statistical products are published in accordance with OECD Publishing Policy and standards. See Table 7 below.

(l) To maximise co-operation with other national and international data providers, free access to all statistical products is given to all national government bodies (including national statistical offices), as well as, subject to reciprocal arrangements, to international organisations.

(m) Core statistical information and methodological documents (manuals, handbooks, etc.) are freely available.

Customised products and services requested by external clients (tabulations, extraction of data from databases, etc.) are priced according to the "marginal cost" rule. The current policy is to not provide customised products and services but rather to encourage users to use OECD standard products and services.

[14] Here "release date" refers to the date on which the data are first publicly made available, by whatever medium, typically, but not inevitably the web site.

Table 7 – Summary table of OECD Publishing Policy for statistical products

Type of data	Updates	Channels	Pricing policy
Core statistics	Concurrent	OECD iLibrary, OECD.Stat Extracts	Free
Complete databases	Concurrent	OECD iLibrary	Priced (variable for different categories of users)
	Concurrent	OLISNet (only for the OECD network[15])	Free
	Archive issues	OECD iLibrary, CDROMs,	Priced
Publications	Concurrent	OECD iLibrary, online statistical table collections, print publications	Priced
Methodological documents, manuals and handbooks	Concurrent	Statistics portal, OLISNet and OECD iLibrary	Free
		Print publications	Priced

(n) To improve accessibility and interpretability of OECD statistics, as well as to improve cost-efficiency of data and metadata dissemination activities:

- all statistical products are listed in the "OECD statistical catalogue";
- data have to be made available with metadata as far as possible. Definitional, procedural and operational metadata describing the statistical activity are to be readily available. Metadata are presented in layers of increasing detail, matching user needs. Table and graph headings reflect the main issues. Seasonal adjustment procedures are clearly indicated. Measures of accuracy are provided where available (for example, coefficients of variation, response rates, etc.), as well as adjustments made for harmonisation;
- full navigability across different datasets is a medium-term target of the OECD;
- clear explanations of the reasons for apparent inconsistencies between different data sources have to be provided to users and support to them in identifying the best data source for their specific purposes;
- the use of different formats for presenting different datasets must be kept to a minimum and only used when they are necessary due to the nature of data concerned;
- usability tests are conducted on new data products, in their various forms,periodically;
- a help service supporting the data is well identified and available during working hours.

[15] Including NSOs and other international organisations.

(o) To improve the efficiency of publishing processes, tools developed by ITN and PAC have to be used as much as possible.

(p) For purposes of recognition of authors, publications are grouped in the following three categories:

- *Publications containing data tabulations which may include explanatory notes and highlights*. The names of employees and contributors who have had significant and non-routine input to the contents and presentation of a publication may be listed in the preface or foreword.

- *Reports, manuals and handbooks intended to explain statistical concepts, methods, definitions and classifications*. These publications should contain a preface or foreword under the authority of the Chief Statistician and the Director of the Directorate responsible for the statistical activity. The preface should acknowledge primary authorship and major contributions together with a short description of such contributions.

- *Studies on statistical and analytical issues.* In this case the author(s) name should appear on the front cover and the title page.

For all categories of statistical publications, decisions regarding attribution of credit are the responsibility of the Director of the Directorate in charge of the statistical activity. Prior to publication, texts of proposed prefaces are sent to the Chief Statistician.

(q) Licensing of the redistribution and use of OECD statistical information by other parties is an important component of the OECD's dissemination strategy. Clients are granted a licence to either redistribute or use the statistical information within terms and conditions stipulated in the license agreement. For statistical products, the Rights Unit manages the related terms and conditions. In any case, they have to oblige resellers to:

- clearly indicate the data source;

- disseminate metadata together with data, in order to ensure the necessary interpretability of statistical tables. In order to do so, authors must store published metadata in a standardised and machine readable format

(r) When a new statistical activity comes to the end of the first production cycle, the activity manager prepares a short report to the SPG, describing the main issues encountered in undertaking the activity and suggestions (if any) to amend the Quality framework (and guidelines), as well as to improve existing corporate statistical tools.

Relevant quality dimensions: timeliness, accessibility, interpretability, coherence and credibility.

Importance for cost efficiency: High

Corporate tools available for program managers[16]*:* OECD.Stat; OECD Statistics portal.

2.3 Quality checklist for new statistical activities

103. When a new statistical activity is envisaged, the manager has to ensure that Quality guidelines are carefully taken into account during the entire process of planning and implementation of the activity. Table

[16] All tools are accessible through the Intranet site for statisticians and analysts.

8 highlights the main steps to be followed in the development of a new statistical activity, together with risks associated with each phase, and proposed actions to minimise the latter.

104. In particular, the following activities have to be undertaken to ensure the optimal level of co-operation with other parts of the OECD:

- Before the activity starts, the activity manager will conduct a careful analysis and evaluation of Quality Guidelines for each step of the process. In addition, the activity manager will involve other staff concerned in analysing Quality Guidelines and evaluating best practices already available within the OECD.

- To ensure an appropriate level of information to other parts of the Organisation, the activity manager is asked to send to SPG members the minutes of the meetings of all statistical working parties. In addition, when new demand for a statistical output is formulated by a Committee, Working group or Directorate, and the preliminary analysis of the topic has been carried out according to Quality Guidelines, the activity manager is asked to compile questionnaire 2 and send it to SPG members for comment. Individual SPG members can provide their comments on the project within 14 working days. In very particular cases, they can ask to discuss the issue at the SPG, explaining the motivation and purpose for such a discussion. Unnecessary plenary discussions should be avoided.

- In the planning phase of the activity the manager is asked to contact ITN, PAC and STD experts in order to find the most appropriate and efficient technical and organisational solutions. Once planning of the activity is concluded, the activity manager has to compile the OECD Statistical Programme of Work (OSWP) questionnaire, where all technical details of the activity are specified.

- Once the first cycle of the activity is finished, the activity manager will prepare a short report on the way in which the project was carried out, highlighting strengths and weaknesses of the activity and proposals for future improvements, as well as for amending the Quality Framework and Quality Guidelines. If the activity has been carried out as a one-off exercise, the report will also illustrate the pros and cons of continuing the activity under different organisational arrangements. In particular, in light of a possible interest for the Organisation as a whole to continue the activity, the manager will quantify the costs associated to such hypothesis.

Table 8 - Main steps to be followed in the development of a new statistical activity

PHASE	POTENTIAL PROBLEMS	ACTIONS REQUIRED
Definition of the data requirements in general terms	Difficulties in evaluating relevance	Involvement of OECD experts
Evaluation of other data currently available and identification of needs for new data	Difficulties in identifying and accessing data available within the OECD Difficulties in identifying accessing data available outside the OECD Difficulties in interpreting data and metadata available	Involvement of OECD experts Compilation of the questionnaire and its transmission to SPG members
Planning and design of the statistical activity	Underestimating resources required Underestimating time required Poor choice of statistical methods Lack of communication with, and involvement of, national statistical experts responsible for coordination with international organisations Inefficient IT solution	Compilation of the OSWP entry for the activity Involvement of ITN, PAC and STD in planning of the activity Maximum use of corporate tools for data collection, management and dissemination Design of training activities involving STAG
Extraction of data and metadata from databases within and external to OECD	Inefficiencies in accessing internal and external databases Difficulties in interpreting data and metadata Incoherence across databases	Maximum use of corporate tools for data and metadata extraction
Implementation of specific data and metadata collection mechanism	Insufficient contact with national data providers Incorrect or inefficient design of the questionnaire Use of inappropriate definitions Inefficient choice of systems for data, metadata transmission	Maximum use of corporate tools for data and metadata collection Verification of definitions with those contained in the OECD Glossary of Statistical Terms
Data and metadata verification, analysis and evaluation	Inappropriate or inefficient statistical methods Different methods across countries for the same series	Maximum use of corporate tools for data verification, storage and analysis
Data and metadata dissemination	Inefficient dissemination procedures Inconsistency across databases Inappropriate presentation of metadata Disclosure of confidential data Inappropriate data release procedures, affecting credibility	Maximum use of corporate tools for data and metadata dissemination Adoption of OECD editorial standards Adherence to OECD Publishing Policy “Post mortem” analysis of the activity and proposals (if any) for future changes

QUESTIONNAIRE
SUMMARY INFORMATION TO STATISTICAL POLICY GROUP (SPG) MEMBERS
ABOUT A NEW STATISTICAL ACTIVITY

NAME OF THE PROJECT: ……..……

……...……...……

NAME OF THE ACTIVITY MANAGER: ……

DIRECTORATE/DIVISION……...…….

DATE: ……..……

1. Have you conducted a careful analysis of current and emerging statistical needs linked to analytical and policy issues that the relevant Committee has identified as a priority in the medium term?	YES	NO
If YES, is the proposed activity fully coherent with such priorities?	YES	NO
2. Have you conducted a comprehensive analysis of similar activities carried out and/or planned by other international organisations?	YES	NO
3. Have you identified other Directorates of the OECD that have carried out and/or planned projects potentially related to this activity?	YES	NO
If YES, what are the main projects carried out and/or planned by other OECD Directorates closely linked to the activity?		
Name of the project Directorate(s)		
…… ……..		
…… ……..		
…… ……..		
…… ……..		
4. Did you involve other Directorates in the definition of data requirements from the beginning of the activity?	YES	NO
If YES, which Directorates?		
……		

5. Are other international organisations involved in the activity? If YES, which organisations (indicate acronyms)? ……	YES NO
6. Are other Directorates involved in the activity? If YES, which Directorate (indicate acronyms)? ……	YES NO
7. Have you analysed if there are any existing international statistical guidelines and recommendations (definitions, classifications, etc.) developed by the OECD or other international organisations useful for conducting the activity? If YES, to what extent they can be used? HIGH MEDIUM LOW	YES NO
8. What is the degree of coherence between concepts, definitions and classifications envisaged for the activity and those used at national level, by the majority of countries concerned? HIGH MEDIUM LOW	
9. What are the expected main groups of users? - Internal users within the Directorate - Internal users working in other Directorates - Experts working in national public administrations - External users: Public bodies Private companies Research institutes Others	Importance (increasing magnitude 1-7) …… …… …… …… …… …… ……
10. What are the main policy "drivers" for the activity? …… …… ……	
11. Describe in one or two pages (to be attached to this questionnaire) the content of the activity, expected outputs and their main characteristics (frequency, variables, geographical coverage, etc.). The note should highlight the main strengths and weaknesses of the activity.	

12. What is the expected time horizon for the activity? Starting date …… Closing date ……	
13. Do you expect that the activity will produce new statistical recommendations and/or standards (definitions, classifications, etc.)? If YES, what kind of standard? …… …… ……	YES NO
14. Does the activity imply a new data and metadata collection?	YES NO
15. Has the decision to carry out the activity already been taken by the relevant body (Committee, Working Party, Directorate)? If YES, have you already compiled the OSWP questionnaire? YES NO	YES NO

ANNEX 1 - THE UNITED NATIONS 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 and 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, has 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 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.*

ANNEX 2 - THE INTERNATIONAL STATISTICAL INSTITUTE DECLARATION ON PROFESSIONAL ETHICS

1. Obligations to society

1.1 Considering conflicting interests: Statistical inquiry is predicated on the belief that greater access to well-grounded information is beneficial to society. The fact that statistical information can be misconstrued or misused, or that its impact can be different on different groups, is not in itself a convincing argument against its collection and dissemination. Nonetheless, the statistician should consider the likely consequences of collecting and disseminating various types of data and should guard against predictable misinterpretations or misuse.

1.2 Widening the scope of statistics: Statisticians should use the possibilities open to them to extend the scope of statistical inquiry, and to communicate their findings, for the benefit of the widest possible community.

1.3 Pursuing objectivity: While statisticians operate within the value systems of their societies, they should attempt to uphold their professional integrity without fear or favour. They should also not engage or collude in selecting methods designed to produce misleading results, or in misrepresenting statistical findings by commission or omission.

2. Obligations to funders and employers

2.1 Clarifying obligations and roles: Statisticians should clarify in advance the respective obligations of employer or funder and statistician; they should, for example, refer the employer or funder to the relevant parts of a professional code to which they adhere. Reports of the findings should (where appropriate) specify their role.

2.2 Assessing alternatives impartially: Statisticians should consider the available methods and procedures for addressing a proposed inquiry and should provide the funder or employer with an impartial assessment of the respective merits and demerits of alternatives.

2.3 Not pre-empting outcomes: Statisticians should not accept contractual conditions that are contingent upon a particular outcome from a proposed statistical inquiry.

2.4 Guarding privileged information: Statisticians are frequently furnished with information by the funder or employer who may legitimately require it to be kept confidential. Statistical methods and procedures that have been utilised to produce published data should not, however, be kept confidential.

3. Obligations to colleagues

***3.1 Maintaining confidence in statistics*:** Statisticians depend upon the confidence of the public. They should in their work attempt to promote and preserve such confidence without exaggerating the accuracy or explanatory power of their data.

***3.2 Exposing and reviewing methods and findings*:** Within the limits of confidentiality requirements, statisticians should provide adequate information to colleagues to permit their methods, procedures, techniques and findings to be assessed. Such assessments should be directed at the methods themselves rather than at the individuals who selected or used them.

***3.3 Communicating ethical principles*:** To conduct certain inquiries statisticians need to collaborate with colleagues in other disciplines, as well as with interviewers, clerical staff, students, etc. In these cases statisticians should make their own ethical principles clear and take account the ethical principles of their collaborators.

4. Obligations to subjects

4.1 Avoiding undue intrusion: Statisticians should be aware of the intrusive potential of some of their work. They have no special entitlement to study all phenomena. The advancement of knowledge and the pursuit of information are not themselves sufficient justifications for overriding other social and cultural values.

4.2 Obtaining informed consent: Statistical inquiries involving the active participation of human subjects should be based as far as practicable on their freely given informed consent. Even if participation is required by law, it should still be as informed as possible. In voluntary inquiries, subjects should not be under the impression that they are required to participate; they should be aware of their entitlement to refuse at any stage for whatever reason and to withdraw data just supplied. Information that would be likely to affect a subject's willingness to participate should not be deliberately withheld**.**

4.3 Modifications to informed consent: On occasions, technical or practical considerations inhibit the achievement of prior informed consent. In these cases, the subjects' interests should be safeguarded in other ways. For example:

- *Respecting rights in observation studies*. In observation studies, where behaviour patterns are recorded without the subject's knowledge, statisticians should take care not to infringe what may be referred to as the 'private space' of an individual or group. This will vary from culture to culture.

- *Dealing with proxies*. In cases where a 'proxy' is utilised to answer questions on behalf of a subject, say because access to the subject is uneconomic or because the subject is too ill or too young to participate directly, care should be taken not to infringe the 'private space' of the subject or to disturb the relationship between the subject and the proxy. Where indications exist or emerge that the subject would object to certain information being disclosed, such information should not be sought by proxy.

- *Secondary use of records.* In cases where a statistician has been granted access to, say, administrative or medical records or other research material for a new or supplementary inquiry, the custodian's permission to use the records should not relieve the statistician from having to consider the likely reactions, sensitivities and interests of the subjects concerned, including their entitlement to anonymity.

- *Misleading potential subjects*. In studies where the measurement objectives preclude the prior disclosure of material information to subjects, statisticians should weigh the likely consequences of any proposed deception. To withhold material information from or to misinform subjects involves a deceit, whether by omission or commission, temporarily or permanently, which will face legitimate censure unless it can be justified.

4.4 ***Protecting the interests of subjects:*** Neither consent from subjects nor the legal requirement to participate absolves the statistician from an obligation to protect the subject as far as possible against the potentially harmful effects of participating. The statistician should try to minimise disturbance both to subjects themselves and to the subjects' relationships with their environment.

4.5 ***Maintaining confidentiality of records:*** Statistical data are unconcerned with individual identities. They are collected to answer questions such as 'how many?' or 'what proportion?'; not 'who?'. The identities and records of co-operating (or non- cooperating) subjects should therefore be kept confidential, whether or not confidentiality has been explicitly pledged.

4.6 ***Inhibiting disclosure of identities:*** Statisticians should take appropriate measures to prevent their data from being published or otherwise released in a form that would allow any subject's identity to be disclosed or inferred.

For more information see http://www.cbs.nl/isi/ethics.htm.