For Official Use **STD/QFS(2003)1**

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

17-Oct-2003

English - Or. English

STATISTICS DIRECTORATE

STD/QFS(2003)1
For Official Use

QUALITY FRAMEWORK AND GUIDELINES FOR OECD STATISTICAL ACTIVITIES

Version 2003/1

JT00151752

English - Or. English

FOREWORD

The OECD has a high reputation for both the quality of its analytical work and for the statistics that underpin that work. In some areas, OECD statistics are internationally recognised as the "best" in terms of coverage, timeliness, and comparability. The network of national and international experts created, through the Committees and Working Parties, gives to the Organisation the possibility to launch research activities (and data collections) on new topics, play a leading role in producing new statistics and develop new definitions, classifications and standards.

Nevertheless, the improvement of quality is a continuous process and one of the main objectives of the Statistics Strategy launched by the OECD in 2001 is the enhancement of the quality of OECD statistics. All OECD statisticians devote a large part of their time to improve quality, but, as the experience of other national and international organisations clearly has demonstrated, the adoption of a formalised approach to quality for statistics can bring important benefits. First, it provides a systematic mechanism for ongoing identification and resolution of quality problems, maximising the interaction between experts in different fields (statisticians, IT experts, final users, etc.); second, it gives greatly increased transparency to the processes used by the OECD to assure quality; and third, it reinforces the political role of the OECD in the context of an information society, increasing the credibility of the Organisation as one of the "first class" data providers.

It is important to note that to develop and implement a successful approach to quality, independently from its technical characteristics, the human factor is absolutely crucial. From this point of view I would like to stress how the OECD approach to quality in statistics has developed through the deep involvement of statisticians working in all Directorates, and a process that has achieved, step by step, a growing consensus on its characteristics, usefulness and importance. This process has also promoted a common view among OECD statisticians of the dimensions of quality, the core values of OECD statisticians and their role vis-à-vis stakeholders and civil society at large. A stronger sense of partnership between OECD statisticians is already an important result of the "quality initiative".

As already said, the improvement of quality is a "never ending story", also partly because of the continuous evolution of statistical and IT techniques. This means that the current version of the quality framework and guidelines (version 2003/1) will necessarily be reviewed in the near future, as such developments take place. The framework already contains a procedure for its periodic review and updating, in particular, significant revisions are expected in 2004, after the conclusion of the first wave of quality reviews and the full development of the new OECD statistical information system. The "dynamic" nature of the document is an important characteristic, necessary to provide OECD statisticians with a useful and up-to-date tool for improving the efficiency and effectiveness of the statistical activities they manage.

Even though the development of the "Quality framework and guidelines for OECD statistical activities" has been supported by the contribution of a number of OECD statisticians, I would like to acknowledge the role played by M. Colledge, D. Ward and G. Salou, who have worked with me in designing the system and writing this document. Special thanks also goes to E. Capponi, E. Espinasse, M. Linster, P. Lübkert, P. Richardson and C. Webb, who took part in the task-force established to develop the quality framework. A. Signora and E. Matthews contributed to the preparation and the finalisation of the

document. Important comments on an earlier version of the framework were provided by G. Brackstone (Statistics Canada) and other national experts. Finally, I would like to thank Donald Johnston, the Secretary General of the OECD, for his decision to give high priority to the Statistics Strategy and for his continuous support to its implementation.

Enrico Giovannini
Chief Statistician of the OECD
September 2003

TABLE OF CONTENTS

INTRODUCTION

1. Improvement in the quality of OECD statistics is one of the main objectives of the current 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 have been launched in the framework of the OSS and considerable improvements have been achieved in 2001-2002. 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. A lot of 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 spillover 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 brief 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. Finally, Part 3 presents a description of specific steps to be followed in planning and carrying out statistical activities, together with examples of good practice currently in use in OECD Directorates.

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. Normally these data (and related metadata) are widely disseminated in the interests of 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 major sources of error.

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.

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.

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 born 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 eight 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 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 must meet three different objectives, to:

- disseminate as widely as possible 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, the statistics dissemination policy has to be conducted:

- in the most cost-effective manner, in accordance with OECD general 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 "basic" 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[2]. 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;

[2] As outlined in "OECD Statistics Strategy: an Update", January 2002.

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 below 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 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.

1.4 The development and the update of Quality Guidelines

45. As described in the Introduction, Parts 2 and 3 of this document 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.

46. 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[3]

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. UN/ECE 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[4]	1. Brief note about the proposed activity to SPG
(c). Planning and design involving all stages of the statistical activity[5]	1. Assess resource requirements and time frame - IT aspects - skills required - financial implications 2. Design activity[6] 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[7]	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

[3] 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 below outlines processes for an existing ongoing statistical activity.

[4] OECD.Stat is the new 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.

[5] "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.

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[8] 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[9] 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 series	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] 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.

[7] 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.

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

[9] OECD.Statworks is a set of IT tools 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[10]

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[11]	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)[12]	Statistical activity manager and staff	3 months	1. Consulting major users, including Committees and experts in capitals 2. Consulting appropriate national and international agencies[13] 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[14]	Statistical activity manager and staff	1 month	1. Asking major users, including Committees and/or experts in capitals, 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[15]	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

[10] 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.

[11] 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).

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

[13] Not only national statistical offices, but also other data providers.

[14] Activities (c) and (d) are normally carried out in parallel.

[15] 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.

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

WHAT[16]	HOW	POTENTIAL PROBLEMS	INSTRUMENTS
1. Relevance	1. Identifying policy needs that require changes in data already collected or new data developments 2. Analysing feedback from marketing activities 3. Taking into account general strategies of the OECD	1. Changes in priorities that interrupt production of statistics 2. Overlapping with initiatives of other Directorates and/or organisations 3. Tensions between different priorities coming from various parts of the Organisation	1. UN/ECE Integrated Presentation 2. Gateway to external sources 3. OSWP 4. OECD.Stat 5. Gateway to OECD statistical databases 6. Assistance by PAC
2. Accuracy	1. Evaluating accuracy problems in original sources 2. Evaluating statistical treatments currently used to manage data and metadata and to improve coherence (over time, across countries, etc.)	1. Use of non-optimal sources from accuracy viewpoint 2. Inappropriate checking of data and metadata 3. Insufficient metadata for evaluating accuracy 4. Use of inappropriate definitions or classifications 5. Inappropriate method for improving coherence	1. OECD Glossary of Statistical Terms 2. Assistance by STD and other Directorates 3. OECD Quality guidelines
3. Credibility	1. Evaluating the way in which data quality is currently assessed 2. Assessing how scientific principles and professional ethics are implemented and how political pressures are managed 3. Evaluating the transparency of procedures used for producing statistics	1. Undermining the OECD image as a professional organisation 2. Undermining the confidence of users in OECD statistics	1. OECD Quality guidelines
4. Timeliness and punctuality	1. Evaluating the efficiency and the quality of data collection, verification, management and dissemination procedures 2. Identifying a calendar for data releases	1. Inefficient or inappropriate data capture and verification processes 2. Inefficient data dissemination processes 3. Missing deadlines 4. Inappropriate use of nowcasting procedures	1. OECD Quality guidelines 2. Assistance by STD, ITN and PAC
5. Accessibility and interpretability	1. Evaluating data and metadata management, dissemination procedures 2. Evaluating user needs for different channels for accessing data and metadata 3. Evaluating product integration in OECD statistical information system	1. Inefficient or inappropriate data and metadata management and dissemination systems 2. Use of non-corporate software for data and metadata management and dissemination 3. Inappropriate quality of corporate software	1. OECD Quality guidelines 2. OECD.Statworks 3. OECD.Stat 4. OECD Glossary of Statistical Terms 5. Assistance by STD, ITN and PAC
6. Coherence: within a dataset, across datasets, over time, and across countries	1. Identifying overlap between already existing series 2. Analysing how to meet different user needs in various parts of OECD 3. Analysing and/or developing good practices for improving coherence	1. Incorrect or inefficient statistical treatment for improving coherence 2. Overlap between existing estimates 3. Insufficient availability of metadata to interpret inconsistencies 3. Weakness of policy conclusions based on incoherent data	1. OECD.Stat 2. Gateway to external sources 3. OECD Quality guidelines 4. OECD Glossary of Statistical Terms 5. Gateway to OECD statistical databases 6. Assistance by STD and other Directorates

[16] The OECD quality dimensions used for the self assessment template are defined in Section 1.1 of the OECD Quality Framework.

2. QUALITY GUIDELINES FOR OECD STATISTICS

2.1 Introduction

47. The quality guidelines presented in Part 2 of this document 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.

48. The quality guidelines provided below 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.

49. 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.

50. 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.

51. As stated in para. 8 above, 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.

52. 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, contribution to the formulation of the OECD statistics strategy and the development of an *esprit de corps* among OECD statisticians (e.g. through the semi-annual OECD statisticians meetings).

53. 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[17]. 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.

54. 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.

55. The Guidelines provided below 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

56. 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, recognising that phases may overlap to some extent or may be carried out in parallel or iteratively rather than consecutively. Also, for some statistical activities, some phases may not be applicable. The Guidelines presented below incorporate best practice already in place within the OECD and in other statistical agencies[18].

[17] As outlined in OECD Statistics Strategy: an Update, January, 2002

[18] 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

The following references contain a wide range of information relevant for OECD statistical activities. As usual, some links can be changed over the last few months. Updated links can be found visiting the OECD Intranet for statistics and the OECD Statistics Portal.

OECD intranet for statistics http://asap2.oecd.org/statistics/oecdstats/index.asp The Intranet contains information on: data and metadata, statistical meetings and activities, methodological documents, technical guides, guidelines for statistical activities, etc.

OECD 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.

Questionnaire for compiling the OECD Statistical Programme of Work (OSWP) http://asap2.oecd.org/std/Activity_list.asp.The questionnaire is used for compiling the Programme of Work.

Gateway to OECD statistical databases http://asap2.oecd.org/statistics/dataset/index.htm and http://asap2.oecd.org/statistics/oecdstats/index.asp (data) The Gateway can be browsed by several dimensions to identify statistical activities carried out by OECD Directorates and the related databases.

Gateway to national and international web sites http://asap2.oecd.org/statistics/source/index.htm The Gateway contains links to web sites of a wide range of national and international data and metadata providers.

OECD Glossary of Statistical Terms http://asap2.oecd.org/statistics/glossary/index.htm The Glossary contains about 6.000 statistical terms, definitions and context information where appropriate.

Activities of the Statistical Training Advisory Group (STAG)
http://web.oecd.org/hrmweb/learning-centre/statistical/stats-page-en.htm The STAG is in charge of statistical training initiatives.

Procedures for seasonal adjustment http://asap2.oecd.org/statistics/oecdstats/index.asp (tools) Statistical software is available for seasonal adjustment (X12-Arima, Tramo-Seats, Demetra).

OECD Guidelines for preparing editorial products http://web.oecd.org/pac/guidelines.htm

UN/ECE 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://asap2.oecd.org/statistics/oecdstats/index.asp (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.

Information on GESMES/TS http://www.gesmes.org/ The site contains technical information on the standard for batch exchange of statistical data and metadata.

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

2.2.1 Phase 1: Definition of data requirements in general terms

Aims

57. 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

58. 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.

59. 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.

60. Other users include individuals belonging to the OECD Secretariat who are not members of the Committee or working group responsible for the data and external users within the governments of OECD Member 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.

61. 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

62. 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

63. 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.

64. 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.

65. Metadata describing the content and the structure of databases (see Phase 6 below) 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.

66. 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 potentially useful internal OECD 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 and remedied where possible.

d) Each potentially useful international organisation database is reviewed to determine whether any required data items are available. If so, arrangements are made to access the database (if they do not already exist) and to extract and use such data items, or the reasons for not doing so are documented and justified.

e) Sufficient metadata are acquired for each potentially useful international organisation database to determine whether or not that database can provide any required data items. Deficiencies in the metadata available should be documented and discussed with the international organisation concerned.

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

g) 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.

h) 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.

i) 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 part 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[19]*:* OECD Statistical Work Programme (OSWP); Gateway to OECD statistical databases; UN/ECE Integrated Presentation of Statistical Work and Internet sites of international organisations; OECD Glossary of Statistical Terms; OECD.Stat.

[19] 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

67. 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 on 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

68. 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.

69. 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.

70. 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 within and outside the Organisation.

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 (OECD.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 UN/ECE 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[20]*:* Toolbox for IT solutions; OECD.Statworks; Training programme for statisticians; OECD Glossary of Statistical Terms.

[20] 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

71. 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

72. 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.

73. 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 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 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:

- 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”;
- 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 several variables estimated in a specific statistical context (i.e. national accounts, health accounts, R&D statistics, etc.) 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) When data have to be extracted from a database managed by other international organisations:

- maximum use is made of data contained in officially published databases;

- 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.);

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

- if several variables estimated in a specific statistical context (i.e. national accounts, health accounts, R&D statistics, etc.) 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 OECD publications. In particular, useful information to be acquired can concern:

  - 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;

  - procedures available to automatically extract data and metadata.

- when ratios and indicators have to be calculated, is strongly suggested to extract individual series and calculate them according to the methodology established by the activity manager;

- data extraction procedures are automated. Manual imputations are 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, the 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 necessary, formal agreements are established between the OECD and the organisation concerned;

- 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. If necessary, formal agreements are established between the OECD and the organisation concerned.

e) 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[21]*:* OECD Glossary of Statistical Terms; Gateway to OECD Statistical Databases; OECD.Stat; Corporate procedures to extract data and metadata from existing sources.

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

2.2.5 Phase 5: Implementation of specific data collection mechanism

Aims

74. 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

75. 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.

76. 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.

77. 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.

78. 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[22];

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

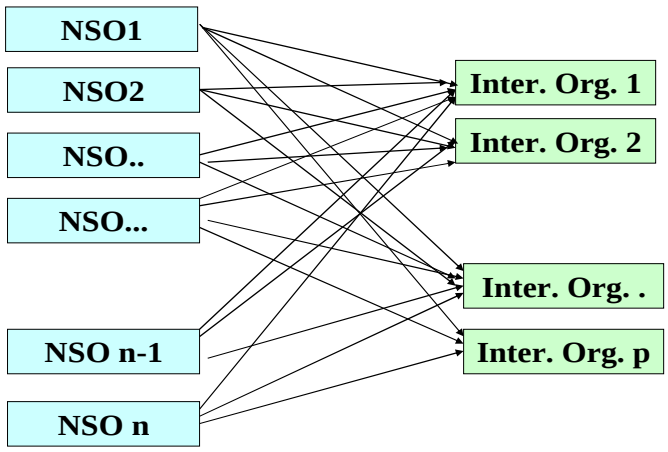
c) become available only because a statistical survey is carried out by national sources following an OECD special request.

79. 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

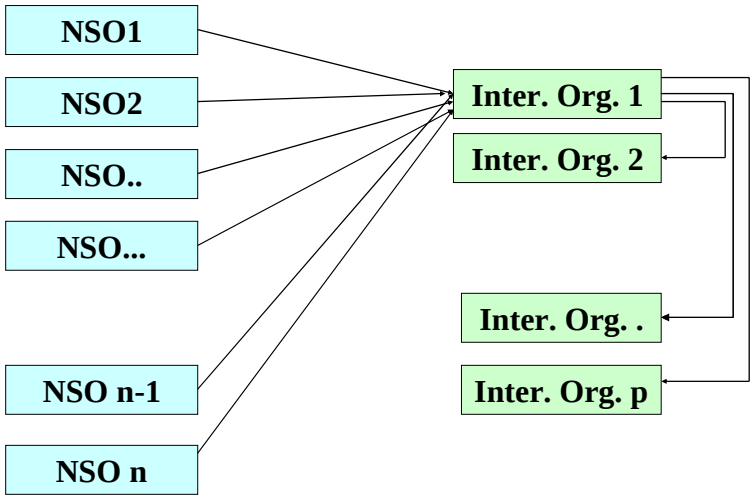
Many to many



- 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



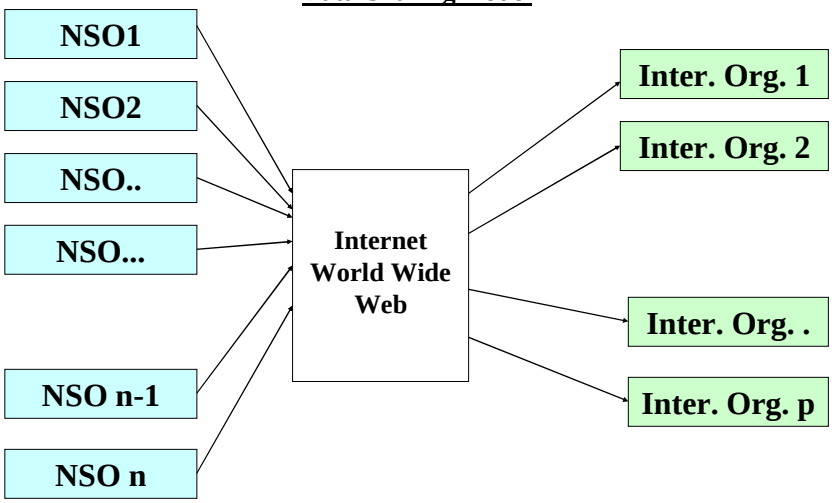
- The third model is the "data-sharing" model where the source agency loads their data into a repository accessible to a number of international organisations. The repository could simply be the internet. The data made available by each source agency need not use the same format or even be in one physical 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.

80. The first two models are normally used for cases (a) and (b) outlined in para. 78 above, 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[23]. 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).

81. 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

[23] 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

data 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 (i.e., time series, cross sectional data, etc.);

- 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:

i) 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 *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;

ii) 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;

iii) 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;

iv) 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;

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

vi) 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.

g) Whenever possible and appropriate, corporate tools provided in the context of OECD.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

Table 4 – Taxonomy of tools for data and metadata transmission 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 or other statistical standard format	Very useful option both for data and metadata. **Preferred option** for exchanging large files with other international organisations.	**Second best option** both for data and metadata	Useful both for data and metadata	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	**Preferred option** both for data and metadata
On-line extractions from web-sites	**Preferred option** for data, also for metadata	**Preferred option** only for data, also for metadata	**Second best** for data, and for metadata	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	**Second best option** for data, sometimes also for metadata	Very useful option only for data, rarely for metadata	Useful option for data, and for metadata	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**

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[24]:

- 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.

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

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

Importance for cost efficiency: Very high

Corporate tools available for program managers[25]*:* 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 (OECD.Statworks).

[25] 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

82. 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

83. 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.

84. 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.

85. 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.

86. 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.

87. 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.

88. 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.

89. 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.

90. 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 (i.e. national accounts, health accounts, R&D statistics, etc.) 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 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 OECD.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.

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[26];
- 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 above) 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.

[26] 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**			
PC Express	Not supported					
MS-Access databases					**Second best option**	
Spreadsheet fie (or .csv)			**Second best option**			
SAS				**Second best option**	Possible option	**Second best 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				**Preferred option**	**Preferred option**
OECD Central database(s)	**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[27]: OECD.Stat; OECD.Statworks

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

2.2.7 Phase 7: Data and metadata dissemination

Aims

91. 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

92. 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;
- *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.

93. 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 (free for internal users and members of the Secretariat, charged for the majority of external users). In addition to paper publications, the main electronic distribution mechanisms are:

- *OLISnet*: 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;
- *SourceOECD*: primarily aimed at non-governmental institutional users who are willing to purchase OECD products. It provides access to published data;
- *OECD web site*: for all classes of external user. It provides access to published data mainly through the Statistics Portal[28].

[28]. For the moment, SourceOECD only contains data files in Beyond 2020 format, while OLISNet contains all data files published by the OECD (Beyond2020 files, MS Excel files, Access to SQL databases, MS Access databases and simple.csv files). OLISNet in particular permits the download of entire files in a format suitable for the bulk updating of a database. In addition, OLISNet permits the presentation of OECD data according to a thematic classification while SourceOECD simply lists the data files. Also,

94. The OECD has publishing and pricing policies within which dissemination must take place. In April 2002, the OECD revised its publishing policy[29]. The revised 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.

95. 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[30].

Guidelines

(a) Statistical products are prepared according to internationally recognised standards[31]. In particular:

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

- statistical data have to be interpretable by the average user;

- data are published according to a calendar of release; all products include a contact name, e-mail address, and telephone number 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

---

the OLISNet interface provides keyword and full text searching for statistics, both leading the user directly to the statistical tables/information requested.

[29] OECD Publishing Policy C(2002)80.

[30] A wide range of purchasing options are available, including single product, range of products, and individual series or group of series (through resellers). For 2002, the Organisation's gross revenues for publications and other products have been around €12 M (excluding International Energy Agency). Statistical products (publications, databases, CD-ROMs, etc.) yielded a gross revenue of around €4 M (33% of the total). In particular, 25% of this revenue derived from SourceOECD, 25% from CD-ROMs and diskettes, 30% from STD paper publications and the remaining 20% from data re-sellers (Bloomberg, DRI-WEFA and others) and paper publications from other OECD Directorates.

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

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:

- 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[32]. 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 will be provided with complimentary copies of statistical publications and other products, according to the general 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 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. Table 7 summarises the pricing policy for statistical products.

(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) The OECD allows all users to have free electronic access to basic statistical information. The Chief Statistician, the Director of PAC and the manager in charge of the statistical activity establish the precise definition of "basic information" in each domain. In any case, no more than 10% of any commercialised database can be included in the category of "basic information". All methodological documents (manuals, handbooks, etc.), even though they are sold as commercial products in hardcopy form, are made available electronically for free on the OECD web site, as well as other statistical products from a specified time after their publication.

[32] 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.

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.

Table 7 – Summary table of OECD publishing policy for statistical products

Type of data	Updates	Channels	Pricing policy
Basic statistics	Concurrent	Statistics portal	Free
Standard statistics	Concurrent	SourceOECD	Priced (variable for different categories of users)
	Concurrent	OLISnet (only for the OECD network[33])	Free
	Old issues	Statistics portal	Free
	Concurrent	Paper publications	Priced
Methodological documents, manuals and handbooks	Concurrent	Statistics portal, OLISnet and SourceOECD	Free
		Paper publications	Priced
Customised statistics	Concurrent	Eurostat Datashops[34]	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;

[33] Including NSOs and other international organisations.

[34] A project has commenced aiming at having the Eurostat network of "datashops" promote OECD products and to provide OECD data in the context of their customised products and services.

- 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, prior to first release;

- a help service supporting the data is well identified and available during working hours.

(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 Chief Statistician approves 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.

(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[35]*:* OECD.Stat; OECD Statistics portal.

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

3. PROCEDURES AND CHECKLISTS FOR OECD STATISTICAL ACTIVITIES

96. As described in Part 1 (paras. 39-44) the OECD Quality Framework outlines specific procedures to ensure that existing and new statistical activities are conducted in accordance with the guidelines outlined in Part 2 for each of the seven phases of the statistical processing cycle (outlined in para. 54 above). The specific procedures outlined in this Part of the Quality Framework take the form of separate Quality Checklists for existing and new statistical activities. The checklists are intended primarily as an aid memoir of the issues to be considered by statistical managers when completing their reviews and will help ensure that all guidelines are taken into consideration by different Directorates across the OECD. The checklists also provide links to specific OECD guidelines in key areas (such as: treatment of confidential data; data and metadata dissemination; identification of existing IT tools; data presentation standards; etc) and examples of good practice, both within the OECD and in other organisations (on issues such as: data extraction; questionnaire design; presentation of new activity to the relevant working party or committee; etc).

3.1 Quality checklist for existing statistical activities

97. As stated in Part 1 (para. 43) above, the procedure for reviewing the quality of existing statistical activities embodied in the related Quality Checklist (see Questionnaire 1) envisages the following stages:

(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.;

(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 dissemination staff, co-ordinated by an expert designated by the SPG;

(e) preparation of the final quality report, combining all the 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 on the plans for their implementation.

98. The Quality Checklist questionnaire is intended for use by the manager in charge of the activity. The manager should first look carefully through the questionnaire to identify the main domains of the review and key areas where specific additional analyses or other work need to be conducted during the

review. Second, the manager should carry out such analyses, etc, eventually involving other staff participating in the activity, and draw the main conclusions[36]. Third, the manager should complete the questionnaire, eventually adding information necessary for an understanding of the results. Finally, the manager should prepare a short note, along the lines of the template outlined at the end of the questionnaire. The completed questionnaire and the brief note are then sent to the Chief Statistician, who will forward them to the team in charge of the discussion of the review.

99. Above all, the Quality Checklist questionnaire has to be seen as a flexible document, which can be expanded as the manager prefers. For example, more lines can be added for comments, where questions are not applicable the manager can indicate this, comments can be added to qualify YES-NO questions even though this is not foreseen in the template, etc. Questions on completion of specific parts of the questionnaire can be referred directly to STD, PAC and ITN.

3.2 Quality checklist for new statistical activities

100. 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 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.

101. 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

[36] Information about revenues from publications, accesses and downloads of electronic products, etc. can be obtained from PAC.

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 2 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 “Post mortem” analysis of the activity and proposals (if any) for future changes

QUESTIONNAIRE 1 - QUALITY CHECKLIST FOR EXISTING OECD STATISTICAL ACTIVITIES

Phase 1. Definition of data requirements in general terms

After having evaluated whether the statistical activity is able to meet existing and future user needs, answer the following questions:	
1.1 Have you conducted, during the last biennium, a comprehensive analysis or review of current and emerging statistical needs linked to analytical and policy issues that the relevant Committee has identified as priorities in the medium term? If YES, please attach the document (if any) where results of such analysis are described. If the document is not available, please describe briefly the main statistical needs. …….. …….. …….. .	YES NO
2.1 Have you conducted, during the last year, a comprehensive analysis or review of activities carried out and/or planned by other international organisations relevant for the activity?	YES NO
3.1 Have you recently identified that other Directorates of the OECD have carried out and/or planned projects potentially related to this activity? 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) …….. …….. …….. …….. …….. …….. …….. ……..	YES NO
4.1 Did you involve other Directorates in the definition of data requirements from the beginning of the activity? If YES, which Directorates? ……..	YES NO

5.1 Do you think that external users' needs expressed at international meetings, committee or working party meetings over the last few years have been taken into account in the design of the activity? If NOT, what are the main concerns that are not reflected in the activity and why? …… …… ……	YES NO
6.1 What are the main policy "drivers" for the activity? …… ……	
7.1 Are other international organisations normally involved in defining the characteristics of the activity?	YES NO
8.1 What are the main strengths and weaknesses of the activity, in relation to similar initiatives undertaken by other international organisations? …… …… ……	
9.1 Did you receive, over the last biennium, comments and criticisms by internal and external users on the outputs of the activity? If YES, what are the <u>main</u> issues or concerns expressed by users? …… …… ……	YES NO
10.1 Were international statistical guidelines and recommendations (definitions, classifications, etc.) developed by the OECD or other international organisations adopted in carrying out the activity? If YES, to what extent? HIGH MEDIUM LOW If NOT, please explain the reasons for this decision (non-relevance of definitions, classifications, etc.) …… …… …… What possibilities exist for future harmonisation with existing statistical standards? …… …… ……	YES NO

11.1 What is the degree of coherence between concepts, definitions and classifications used for compiling OECD data and those used at national level, by the majority of countries concerned? HIGH MEDIUM LOW What possibilities exist for future improvements? …… …… ……	
12.1 What possibilities exist for improving the consultation process with stakeholders and other OECD Directorates? …… …… ……	
13.1 What are the 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) …… …… …… …… …… …… ……
14.1 How many copies of <u>paper</u> publications have been sold to all customers over the last 12 months? What is the percentage change over the same period of the previous year?	…… …… %
15.1 What were the overall revenues for <u>all</u> statistical products related to the activity over the last 12 months? What is the percentage change over the same period of the previous year?	…… …… %
16.1 How many downloads from SourceOECD of the <u>paper</u> publication have been registered over the last 12 months? What is the percentage change over the same period of the previous year?	…… …… %
17.1 How many downloads from SourceOECD of the <u>database</u> containing statistical data have been registered over the last 12 months? What is the percentage change over the same period of the previous year?	…… …… %

18.1 How many downloads from OLISnet of the database containing statistical data have been registered over the last 12 months? What is the percentage change over the same period of the previous year?	…… …… %
19.1 How many hits for statistical products related to the activity have been registered on the OECD web over the last 12 months? What is the percentage change over the same period of the previous year?	…… …… %
20.1 Do you have a regular survey of user satisfaction? If YES, what are the main characteristics of the survey and its key results? …… …… ……	YES NO
21.1 What are the main changes in the activity envisaged in the future for meeting user needs, including main changes expected to existing output? …… …… ……	

Phase 2. Evaluation of other data currently available

Having evaluated possible duplication(s) between data collection related to activity carried out by OECD and those of other international organisations, answer these questions:	
1.2 Is there any overlap in data content with other internal OECD or external databases that would require a substantial change in the way in which the activity (or other OECD activities) is currently conducted in order to be resolved?	YES NO
2.2 Could changes in concepts, definitions or classifications equally satisfy users' needs, allowing the usage of data already collected by the OECD and/or other international organisations?	YES NO
3.2 If you replied YES to either questions 1.2 or 2.2, please indicate what changes could be envisaged to the activity, identifying pros and cons of various alternatives and under what conditions such changes could be implemented. …… …… ……	
4.2 Please, attach the list of variables already collected for conducting the activity in the past and/or those identified for future data collections.	

Phase 3. Design and planning of statistical activity

When the conclusions of previous analyses (for Phases 1 and 2) are drawn, the way in which the activity has been designed/planned in the past and the appropriateness of statistical methods and technical procedures currently adopted for conducting the activity are evaluated. Following this review, and taking into account the last cycle of the activity:	
1.3 Have you been obliged to devote significant resources to increase the response rate for specific countries, or to agree with the latter on a substantial reduction of the originally planned data and metadata transmission? If YES, why? …… …… ……	YES NO
2.3 Have you been obliged to redesign the questionnaire used for data and metadata collection? If YES, why? …… …… ……	YES NO
3.3 Have you been obliged to change the system of data and metadata collection because of its inefficiency or the inaccuracy or incompleteness of data and metadata collected? If YES, why? …… …… ……	YES NO
4.3 Have you been obliged to substantially change IT solutions originally designed for data and metadata management and dissemination? If YES, why? …… …… ……	YES NO
5.3 Have you been obliged to substantially change statistical methods and techniques initially identified to process data and conduct necessary analyses? If YES, why? …… …… ……	YES NO

6.3 Have you been obliged to redesign the structure of the output (paper publication, database, etc.) because of significant differences in qualitative terms between expected and actual results? If YES, why? …… …… ……	YES NO
7.3 Did you pay unexpected costs for data and metadata collection, management and dissemination? If YES, why? …… …… ……	YES NO
8.3 Did you face shortage in statisticians' skills? If YES, why? …… …… ……	YES NO
9.3 Did you encounter unexpected financial/resource, managerial and organisational problems? If YES, why? …… …… ……	YES NO
10.3 Did you experience limited support from other parts of the OECD? If YES, why? …… …… ……	YES NO
11.3 Did you face difficulties due to unexpected changes in Directorate's (Division, Unit) priorities, reduction in resources previously allocated to the activity, external "shocks", insufficient support from other Directorates, although agreed well in advance, etc. If YES, what situation(s) did you encounter? …… …… ……	YES NO
12.3 What statistical and technical training has been given to staff over the last one or two years? …… …… ……	

Phases 4-5. Data and metadata collection

Having carefully analysed data and metadata collection systems, taking into account available IT tools and Quality guidelines, answer the following questions:	
Are data and metadata extracted/collected: From OECD databases? From databases managed by other international organisations? Directly from national data sources?	
If data and metadata are extracted from OECD databases, please complete Section 4.A below. If data and metadata are extracted/collected from databases managed by other international organisations go to Section 4.B. If data and metadata are extracted/collected from national data providers go to Section 4.C. Please answer these questions considering the overall situation. If you wish to provide a precise assessment of the quality of data and metadata collection procedures and to identify proposals for their improvement, you may alternatively answer questions 27.4-33.4 for each country concerned. If data and metadata are extracted/collected from multiple sources, complete each of the relevant sections 4 A, B, C.	

SECTION 4.A – DATA AND METADATA EXTRACTED FROM OECD DATABASES	
1.4 If data and metadata are already extracted from OECD databases, are they collected from either the "original" database or from the set of "OECD references series"? If NOT extracted from either source, please provide a brief statement on the positive or negative impact on the activity if either source were used (e.g., in terms of revision of historical time series, etc) …… …… ……	YES NO
2.4 Are data always extracted from databases together with appropriate metadata? If NOT, why? …… …… ……	YES NO
3.4 Are data extracted from different databases coherent each other? If NOT, would it be possible to improve coherence changing the source database or the way in which derived measures (i.e. ratios, shares, etc.) are calculated? YES NO	YES NO

4.4 Did you encounter, in the past or during the quality review, relevant problems (in terms of availability, timeliness, interpretability, coherence, etc.) in data and metadata stored in OECD databases? If YES, describe the most relevant problems. Name of database Main problems …… …….. …… …….. …… ……..	YES NO
5.4 Have you discussed with managers of the databases concerned possible improvements in the latter that would allow you to improve quality and cost-efficiency of the statistical activity? If YES, describe the most important conclusions of these contacts. Name of database Main conclusions …… …….. …… …….. …… ……..	YES NO
6.4 Are procedures used for data and metadata extraction/collection automated and efficient, given the availability of the most recent ICT tools? If YES, were they reviewed over the last year, in co-operation with ITN and STD (at some stage)? YES NO If NOT, did you carry out such a review during the quality assessment? YES NO If YES, what are the main conclusions of the review? …… …… ……	YES NO
7.4 Are confidential data accessed, extracted or received? If YES, describe briefly measures and procedures established to avoid privacy disclosure or unauthorised utilisation of these data. …… …… ……	YES NO
8.4 Did you verify that all requested information about data and metadata collection systems is correctly stored in the OSWP database?	YES NO

SECTION 4.B – DATA AND METADATA EXTRACTED/COLLECTED FROM DATABASES MANAGED BY OTHER INTERNATIONAL ORGANISATIONS	
9.4 Are the organisations concerned fully aware of the data use by the OECD? If YES, is there any formal Memorandum of Understanding (or similar agreement) signed between them and the OECD? YES NO If YES, please attach a copy of it to the report and indicate whether it has to be updated or modified. …… …… ……	YES NO
10.4 Are data and metadata extracted directly from databases? If NOT, are they normally received by (tick one or more as required): E-MAIL CD-ROM or DISKETTE OTHER Is this solution fully satisfactory? YES NO If NO, why? …… …… ……	YES NO
11.4 If data and metadata are extracted directly from a database, is it a "dissemination or output" database or an internal database? Is this solution fully satisfactory? YES NO If NO, why? …… …… ……	Dissemination Internal
12.4 Are data always extracted/collected with appropriate metadata? If NO, why? …… …… ……	YES NO
13.4 Are data extracted/collected from different databases coherent each other? If NO, would it be practicable to improve coherence by changing the source database or the way in which derived measures (i.e. ratios, shares, etc.) are calculated? YES NO	YES NO

14.4 In the past, have you encountered significant problems (in terms of availability, timeliness, interpretability, coherence, etc.) in data and metadata stored in relevant databases managed by other international organisations? If YES, describe the most relevant problems. Name of the database (Organisation) Main problems …… …….. …… …….. …… ……..	YES NO
15.4 Have you discussed with managers of the databases concerned possible improvements in the latter that would allow you to improve quality and cost-efficiency of the statistical activity? If YES, describe the most important conclusions of these contacts. Name of the database Main conclusions …… …….. …… …….. …… ……..	YES NO
16.4 Are procedures used for data and metadata extraction/collection automated and efficient, given the availability of the most recent ICT tools? If YES, were they reviewed over the last year, in co-operation with ITN and STD (at some stage)? YES NO If NOT, did you carry out such a review during the quality assessment? YES NO If YES, what are the main conclusions of the review? …… …… ……	YES NO
17.4 Is confidential data accessed, extracted or received? If YES, describe briefly measures and procedures established to avoid privacy disclosure or unauthorised utilisation of the data. …… …… ……	YES NO
18.4 Did you verify that all requested information about data and metadata collection systems is correctly stored in the OSWP database?	YES NO

SECTION 4.C – DATA AND METADATA EXTRACTED/COLLECTED FROM NATIONAL DATA PROVIDERS	
19.4How do you make organisations concerned fully aware of their obligations to provide OECD with necessary data and metadata? …… …… …… Please attach a copy of the most recent letter used to ask national agencies to provide necessary data and metadata to the OECD.	
20.4Taking into account quality guidelines, does the letter contain the following information? • Presentation of the initiative • Future use of information to be transmitted • Characteristics of information requested • Clear reference to the expected treatment of confidential data • Contact name for answering questions and providing relevant information	YES NO YES NO YES NO YES NO YES NO
21.4Is the letter normally copied to the: • OECD Chief Statistician? • National agencies responsible for co-ordination of statistical relationships with international organisations?	YES NO YES NO
22.4Did you encounter particular difficulties (timeliness, accuracy, coherence, etc.) in collecting data and metadata? If YES, please report the most frequent problems encountered, eventually indicating countries (or groups of countries) concerned. …… …… ……	YES NO
23.4Did data providers express concerns, over the last two years, about the burden imposed on them or about other aspects (transmission mechanisms, structure of the questionnaire, technical capacity available in the institution, support provided by the Secretariat, etc.)? If YES, please report the most frequent problems mentioned, eventually indicating countries (or groups of countries) concerned. …… …… ……	YES NO
24.4What are the procedures currently adopted to ensure the quality and the efficiency of the data and metadata collection mechanism? In particular, report about the ways in which data providers are involved in developing and testing questionnaires, in identifying the best transmission mechanism, in verifying whether data available at national level satisfy the needs of the activity will be described. …… ……	

……	
25.4 Is the data and metadata collection carried out in co-operation with other international organisations? If YES, what is the "transmission model" adopted for the statistical activity? MANY-TO-MANY MANY-TO-ONE DATA SHARING SPECIAL AGREEMENT Is the model currently adopted satisfactory for both sides of the transmission process? YES NO If NO, why? …… …… …… Do you envisage any possibility of applying the "data sharing" model? YES NO If YES, under what conditions? …… …… ……	YES NO
26.4 Are tools currently used for data and metadata collection coherent with those suggested by OECD Quality Guidelines? If NO, why? …… …… ……	YES NO
27.4 If data and metadata are extracted from databases managed by national agencies, are they normally "dissemination or output" databases or internal databases? Is this solution fully satisfactory? YES NO If NO, why? …… …… ……	Dissemination Internal
28.4 Are data always extracted/collected together with appropriate metadata? If NO, why? …… …… ……	YES NO

29.4 Is the meaning of metadata coherent with that contained in the SDMX Common Metadata Vocabulary? If NO, why? …… …… ……	YES NO
30.4 Is the metadata normally requested and received satisfactory with regard to providing adequate transparency to the data? If NO, why? …… …… ……	YES NO
31.4 Are data extracted/collected from different national sources coherent with each other? If NO, would it be practicable to improve coherence by changing the sources or the way in which derived measures (i.e. ratios, shares, etc.) are calculated? YES NO	YES NO
32.4 If data and metadata are not extracted directly from databases, are they normally received by: E-MAIL CD-ROM or DISKETTE OTHER Is this solution fully satisfactory? YES NO If NO, why? …… …… ……	
33.4 Are historical data normally transmitted together with the most recent data? If NO, why? …… …… ……	YES NO
34.4 Was the questionnaire used for data and metadata collected designed using best practice? YES NO Were standard tools (for example, BLAISE) used to design the questionnaire? …… …… ……	

What consultation process was used in the development of the questionnaire (e.g. committee/working party consultation, ISWG consultation, etc)? …… …… …… Was the questionnaire tested prior to use and evaluated after use? YES NO If NO, why? …… …… …… Is the questionnaire accompanied by clear and easily understood explanatory notes containing instructions for the compilation of the latter (definitions, classifications, etc.) and its transmission to the OECD? YES NO If NO, why? …… …… ……	
35.4 Have you recently discussed with national data providers possible improvements in the way in which data and metadata are collected/extracted? If YES, describe the most important improvements envisaged in the future or the main difficulties that prevent the latter. …… …… ……	YES NO
36.4 Are procedures used for data and metadata extraction/collection automated and efficient, given the availability of the most recent ICT tools? If YES, were they reviewed over the last year, in co-operation with ITN and STD (at some stage)? YES NO If NO, did you carry out such a review during the quality assessment? YES NO If YES, what are the main conclusions of the review? …… …… ……	YES NO

37.4	Are confidential data accessed, extracted or received? If YES, describe briefly measures and procedures established to avoid privacy disclosure or unauthorised utilisation of these data. …… …… ……	YES NO
38.4	Are confidential data accessed, extracted or received according to special agreements? If YES, please attach a copy of these agreements (if there are different agreements with various countries, attach all of them).	YES NO
39.4	If an "OECD survey" were conducted, do you think that activities carried out by the Secretariat for designing the survey (frame, questionnaire, etc.) were well conducted? If YES, please attach the document where details about the characteristics of the survey are described. If NO, why? …… …… ……	YES NO
40.4	Were metadata transmitted by data providers adequate, in terms of accuracy measures, description of tasks undertaken, statistical procedures adopted, etc.? If NO, why? …… …… ……	YES NO
41.4	Were systems used for data and metadata transmission efficient and appropriate? If NO, why? …… …… ……	YES NO
42.4	Did you verify that all requested information about data and metadata collection systems is correctly stored in the OSWP database?	YES NO

Phase 6. Data and metadata verification, compilation, storage, analysis and evaluation

Once data and metadata are collected, they are verified, compiled, processed, stored and analysed. All these steps need to be evaluated, to see whether procedures adopted are able to ensure the requested accuracy, interpretability, accessibility and coherence according to the quality guidelines. Therefore, after having carefully analysed all these characteristics and evaluated whether they require change, taking into account current standards of the OECD statistical information system (*OECD.Stat*), answer the following questions:	
1.5 Who is in charge of checking that the methods used for verifying, compiling and processing data are appropriate? THE ACTIVITY MANAGER OTHER ADMINISTRATORS EACH STATISTICIAN INVOLVED IN THE ACTIVITY OTHERS (please, specify…….)	
2.5 Are automated procedures used for identifying outliers or unexpected differences between data linked by identities, revisions of historical time series, etc.? If YES please briefly describe system used …… …… ……	YES NO
3.5 Are cross-checks carried out with data coming from other (internal and external) sources? If NO, why? …… …… ……	YES NO
4.5 Do procedures used to check data incorporate the latest statistical developments (e.g. time series, degree of volatility, etc.) and are they tested before use on simulated series? If NO, why? …… …… ……	YES NO
5.5 Is the same procedure used for similar data collected from different countries, for similar topics, etc.? YES NO If NO, have you evaluated whether this approach can affect data accuracy and coherence? YES NO	YES NO

6.5 Are procedures used to check data well documented? If NO, why? …… …… ……	YES NO
7.5 When possible errors are identified through various data verification procedures, what are the procedures used to investigate the reason for unusual values, to accept them as outliers, or to proceed to imputation? …… …… ……	
8.5 Are statistical methods used for imputing missing or incorrect values? If YES, what methods are used? …… …… …… Is the publication of results subject to any specific agreement with the source agency? YES NO Is the overall impact of the editing process on the accuracy of final results measured? YES NO If YES, what method is used? …… …… ……	YES NO
9.5 Do you currently systematically provide feedback to original sources about problems encountered while carrying out the activity, especially when errors are too frequent or reflect inaccurate preparation of data transmitted to the OECD?	YES NO
10.5 What kind of information is provided to users about methods used for data imputation, time series linkage, etc.? …… …… ……	
11.5 Are data estimated by the Secretariat regularly flagged with the label “OECD estimate”?	YES NO
12.5 Do you verify that supra-national aggregates are computed using internationally accepted standards and practices?	YES NO
13.5 Are information about weights used for aggregating national values and necessary information to explain possible inconsistencies with other sources provided in associated metadata? If NO, why? …… ……	YES NO

……	
What procedure is used for producing seasonally adjusted or trend-cycle data? TRAMO-SEATS X12-ARIMA OTHER NOT APPLICABLE Is the information about parameters used in those procedures provided in associated metadata? YES NO	
14.5 Are adjustments made to original figures to improve their comparability over time and across countries well documented through appropriate metadata? If NO, why? …… …… ……	YES NO
15.5 Are data and metadata stored according to OECD Quality Guidelines? If NO, what are the suggested changes to current practices for the activity? …… …… …….	YES NO
16.5 Are confidential data stored in a way that prevents any possible confidentiality disclosure? If YES, describe briefly measures and procedures established to avoid privacy disclosure. …… …… …… If NO, what are the suggested changes to current practices for this activity? …… …… ……	YES NO
17.5 Is there any overlap between data and metadata contained in the production database and those stored in other OECD databases? If YES, have data and metadata contained in the production database been copied from other OECD databases or are they extracted through a dynamic link to the latter? COPIED DYNAMICALLY LINKED Do you envisage any change in the way data and metadata are obtained from other OECD databases? YES NO If YES, please explain the envisaged changes. …… ……	YES NO

……	
18.5 Are automated procedures used to monitor in real-time how data storage is progressing, in order to compare how actual data transmissions meet plans or expectations? If YES, do you normally compute measures about the number and amount of data punctually received? YES NO If NO, do you envisage any advantage for the production process in having such real-time monitoring? YES NO	YES NO
19.5 Have metadata been compiled following definitions used in the SDMX Common Metadata Vocabulary?	YES NO
20.5 Are all relevant variable definitions stored in the OECD Glossary of Statistical Terms?	YES NO
21.5 How do you evaluate the quantity and quality of metadata stored in the relevant database, taking into account internal user needs?	APPROPRIATE INSUFFICIENT REDUNDANT
22.5 Do you consider metadata easily available to and interpretable by internal users?	YES NO
23.5 Do you consider the final non-confidential data to be easily accessible to the Secretariat staff? If restrictions apply to unauthorised Staff, please describe why such limitations are in place and evaluate whether current practices need to be changed. …… …… ……	YES NO
24.5 What are the procedures currently used to report to stakeholders and data providers, the methods and processes adopted for data and metadata verification, compilation, storage, etc.? …… …… ……	
25.5 What possibilities exist for improving documentation and procedures adopted for checking, editing, compiling and storing data and metadata? …… …… ……	
Did you verify that all requested information about data and metadata storage is correctly stored in the OSWP database?	YES NO

Phase 7. Data and metadata dissemination

When final data and metadata are compiled, they are disseminated to external users following various formats and through different media. Therefore, quality aspects, with reference to the content of the publication(s) the media used for its dissemination and price policy, are analysed. The analysis is carried out for each product released as output of the activity (paper publication, databases, etc.). After having completed the analysis, answer the following questions:	
1.6 Were OECD editorial guidelines correctly adopted in preparing the publication? If NO, why? …… …… ……	YES NO
2.6 How do you evaluate the clarity of data presentation (titles, headings, acronyms, footnotes, structure of tables, explanatory notes) in the publication? If INSUFFICIENT, what measures can be taken to improve the clarity of the presentation? …… …… ……	APPROPRIATE INSUFFICIENT
3.6 How do you evaluate the interpretability of data and metadata (strengths and weaknesses of data, comparability issues, breaks in time series, etc.) in the publication? If INSUFFICIENT, what measures can be taken to improve the clarity of the presentation? …… …… ……	APPROPRIATE INSUFFICIENT
4.6 In particular, do you think that the metadata provided in the publication are adequate with respect to: - main characteristics of sources (statistical survey, administrative data, etc.) - conceptual issues covered by data - key concepts, definitions and classifications used - evaluation of accuracy - coherence over time and across countries - revision policy - statistical methods adopted for data compilation (seasonal adjustment, statistical editing, etc.)	 YES NO YES NO YES NO YES NO YES NO YES NO YES NO
5.6 Are apparent inconsistencies (if any) with data contained in other OECD products correctly explained?	YES NO

6.6 Is the related database available on SourceOECD? If NO, why? …… …… …… **If NO, go to question 12.6**	YES NO
7.6 Were guidelines correctly adopted in preparing the SourceOECD files? If NO, why? …… …… ……	YES NO
8.6 How do you evaluate the clarity of the data presentation (titles, headings, acronyms, footnotes, structure of tables, explanatory notes) in SourceOECD? If INSUFFICIENT, what measures can be taken to improve the clarity of the presentation? …… …… ……	APPROPRIATE INSUFFICIENT
9.6 How do you evaluate the interpretability of data and metadata (strengths and weaknesses of data, comparability issues, breaks in time series, etc.) in SourceOECD? If INSUFFICIENT, what measures can be taken to improve the clarity of the presentation? …… …… ……	APPROPRIATE INSUFFICIENT
10.6 In particular, do you think that metadata available in SourceOECD cover in an appropriate way: - main characteristics of sources (statistical survey, administrative data, etc.) - conceptual issues covered by data - key concepts, definitions and classifications used - evaluation of accuracy - coherence over time and across countries - revision policy - statistical methods adopted for data compilation (seasonal adjustment, statistical editing, etc.)	 YES NO YES NO YES NO YES NO YES NO YES NO YES NO
11.6 Are apparent inconsistencies (if any) with data contained in other OECD databases correctly explained?	YES NO
12.6 Is the related database available on OLISnet? If NO, why? …… …… ……	YES NO

If NO, go to question 18.6	
13.6 Were guidelines correctly adopted in preparing the OLISnet files? If NO, why? …… …… ……	YES NO
14.6 How do you evaluate the clarity of the data presentation (titles, headings, acronyms, footnotes, structure of tables, explanatory notes) in OLISnet? If INSUFFICIENT, what measures can be taken to improve the clarity of the presentation? …… …… ……	APPROPRIATE INSUFFICIENT
15.6 How do you evaluate the interpretability of data and metadata (strengths and weaknesses of data, comparability issues, breaks in time series, etc.) in OLISnet? If INSUFFICIENT, what measures can be taken to improve the clarity of the presentation? …… …… ……	APPROPRIATE INSUFFICIENT
16.6 In particular, do you think that the metadata available in OLISnet cover in an appropriate way: - main characteristics of sources (statistical survey, administrative data, etc.) - conceptual issues covered by data - key concepts, definitions and classifications used - evaluation of accuracy - coherence over time and across countries - revision policy - statistical methods adopted for data compilation (seasonal adjustment, statistical editing, etc.)	 YES NO YES NO YES NO YES NO YES NO YES NO YES NO
17.6 Are apparent inconsistencies (if any) with data contained in other OECD databases correctly explained?	YES NO
18.6 Is the related database available on a CD-ROM? If NO, why? …… …… …… **If NO, go to question 24.6**	YES NO
19.6 Were guidelines correctly adopted in preparing the CD-ROM? If NO, why? …… …… ……	YES NO

20.6How do you evaluate the clarity of the data presentation (titles, headings, acronyms, footnotes, structure of tables, explanatory notes) in the CD-ROM? If INSUFFICIENT, what measures can be taken to improve the clarity of the presentation? …… …… ……	APPROPRIATE INSUFFICIENT
21.6How do you evaluate the interpretability of data and metadata (strengths and weaknesses of data, comparability issues, breaks in time series, etc.) in the CD-ROM? If INSUFFICIENT, what measures can be taken to improve the clarity of the presentation? …… …… ……	APPROPRIATE INSUFFICIENT
22.6In particular, do you think that the metadata available in the CD-ROM cover in an appropriate way: - main characteristics of sources (statistical survey, administrative data, etc.) - conceptual issues covered by data - key concepts, definitions and classifications used - evaluation of accuracy - coherence over time and across countries - revision policy - statistical methods adopted for data compilation (seasonal adjustment, statistical editing, etc.)	 YES NO YES NO YES NO YES NO YES NO YES NO YES NO
23.6Are apparent inconsistencies (if any) with data contained in other OECD databases correctly explained?	YES NO
24.6Are any statistical outputs related to the activity available free of charge on Internet? If NO, why? …… …… …… **If NO, go to question 28.6**	YES,COMPLETELY YES, PARTIALLY NO
25.6If statistical outputs are entirely or partially available on Internet, are they correctly accessible through the statistics portal? If NO, why? …… …… ……	YES NO

26.6If statistical outputs are partially available on the Internet, does their presentation follow OECD guidelines prepared for the dissemination of "basic statistics"? If NO, why? …… …… ……	YES NO
27.6If the entire statistical output is available on the Internet, how do you evaluate the interpretability of data and metadata (strengths and weaknesses of data, comparability issues, breaks in time series, etc.)? If INSUFFICIENT, what measures can be taken to improve the clarity of the presentation? …… …… ……	APPROPRIATE INSUFFICIENT
Are statistical procedures adopted to ensure that released products do not contain confidential data? If YES, describe briefly these procedures. …… …… ……	YES NO NOT APPLICABLE
28.6Was appropriate information given to PAC and STD to enable them to plan the dissemination of the product (e.g. inclusion of the product in the annual plan of publications, in the OSWP database, in the KAPPA database, etc.)? If NO, why? …… …… ……	YES NO
29.6Was the expected release period communicated to the public through the OECD advance calendar of releases? If NO, why? …… …… ……	YES NO
30.6Was the product released to the public on schedule? If NO, why? …… …… ……	YES NO
31.6Did political pressures contribute in any way to delays in the dissemination of data to the public? If YES, how were such pressures managed? …… …… ……	YES NO

32.6 Were the ways in which the publication has been promoted/publicised by the OECD or presented to the media satisfactory? If NOT, why? …… …… ……	YES	NO
26.5 Did you encounter any significant problems in the dissemination phase? If YES, please describe them. …… …… ……	YES	NO
27.5 What measures need to be taken to improve accessibility to statistical outputs of the activity? …… …… ……		
28.5 Did you verify that all requested information about dissemination is correctly stored in the OSWP and KAPPA databases?	YES	NO

General issues

When all technical phases of the activity have been reviewed, an overall evaluation of the degree of coherence between the approach currently used in carrying out the activity and basic principles defined for OECD statistical activities is carried out. In particular:	
1.7 Are there any aspects of the activity that do not accord with the Fundamental Principles of Official Statistics adopted by the UN Statistical Commission? If YES, what should be done for overcoming these differences? …… …… ……	YES NO
2.7 Are there any aspects of the activity that need to be reviewed or improved in the light of any aspect of the ISI declaration on Professional Ethics? If YES, what should be done for overcoming these differences? …… …… ……	YES NO
3.7 Are methods and procedures used for compiling, managing and disseminating data chosen under any political influence? What measures can be taken to avoid this influence? …… …… ……	YES NO
4.7 Is there any evidence that OECD requests for data and metadata are not treated by some Member countries with the same importance attached to those of other international (or supranational) organisations? What measures can be taken to improve this aspect? …… …… ……	YES NO
5.7 Is there evidence that these statistics are used in inappropriate ways? If YES, quote some of these cases and actions (if any) undertaken to avoid it. …… …… ……	YES NO

After having completed the questionnaire, the activity manager is asked to write a brief note (three-four pages) summarising, from his/her perspective, the main issues identified in the above checklist and proposed solutions. In particular, the manager, using the structure described in Table 3 as a general reference, will include the following elements:

- **short description of the activity and of resources involved in the activity;**
- **main strengths and weaknesses (if any) of current practices, according to each quality dimension;**
- **main actions that need to be undertaken to resolve problems identified through the quality review;**
- **evaluation of resources necessary for implementing recommendations.**

An example of the quality assessment carried out for the STAN database is contained in Annex 5.

QUESTIONNAIRE 2 - SUMMARY INFORMATION TO SPG MEMBERS ABOUT A NEW STATISTICAL ACTIVITY

NAME OF THE PROJECT: ……

……..……..……

NAME OF THE ACTIVITY MANAGER: ……

DIRECTORATE/DIVISION……..…….

DATE: ……..……

1.1	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? If YES, is the proposed activity fully coherent with such priorities?	YES NO
2.1	Have you conducted a comprehensive analysis of similar activities carried out and/or planned by other international organisations?	YES NO
3.1	Have you identified other Directorates of the OECD that have carried out and/or planned projects potentially related to this activity? 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) …… …….. …… …….. …… …….. …… ……..	YES NO
4.1	Did you involve other Directorates in the definition of data requirements from the beginning of the activity? If YES, which Directorates? ……	YES NO

5.1 Are other international organisations involved in the activity? If YES, which organisations (indicate acronyms)? ……	YES NO
6.1 Are other Directorates involved in the activity? If YES, which Directorate (indicate acronyms)? ……	YES NO
7.1 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
8.1 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	
9.1 What are the expected main groups of users?	Importance (increasing magnitude 1-7)
- 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	……
10.1 What are the main policy "drivers" for the activity? …… …… ……	
11.1 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.	
12.1 What is the expected time horizon for the activity? Starting date ……	

Closing date ……	
13.1 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
14.1Does the activity imply a new data and metadata collection?	YES NO
15.1Has 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

ANNEXES

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.

ANNEX 3 - TOWARDS THE NEW OECD STATISTICAL INFORMATION SYSTEM

1. Introduction

The OECD has a decentralised statistical structure where statisticians and analysts work closely together on a subject by subject basis. Because of the decentralised decision process, past decisions involving IT implementation for statistics have given priority to the efficiency of individual processes and to requirements of statisticians in individual subject matter areas. This has created a fragmented system focused on individual activities, developing optimal solutions from the producers' point of view, but sub-optimal solutions for users in the Organisation and outside.

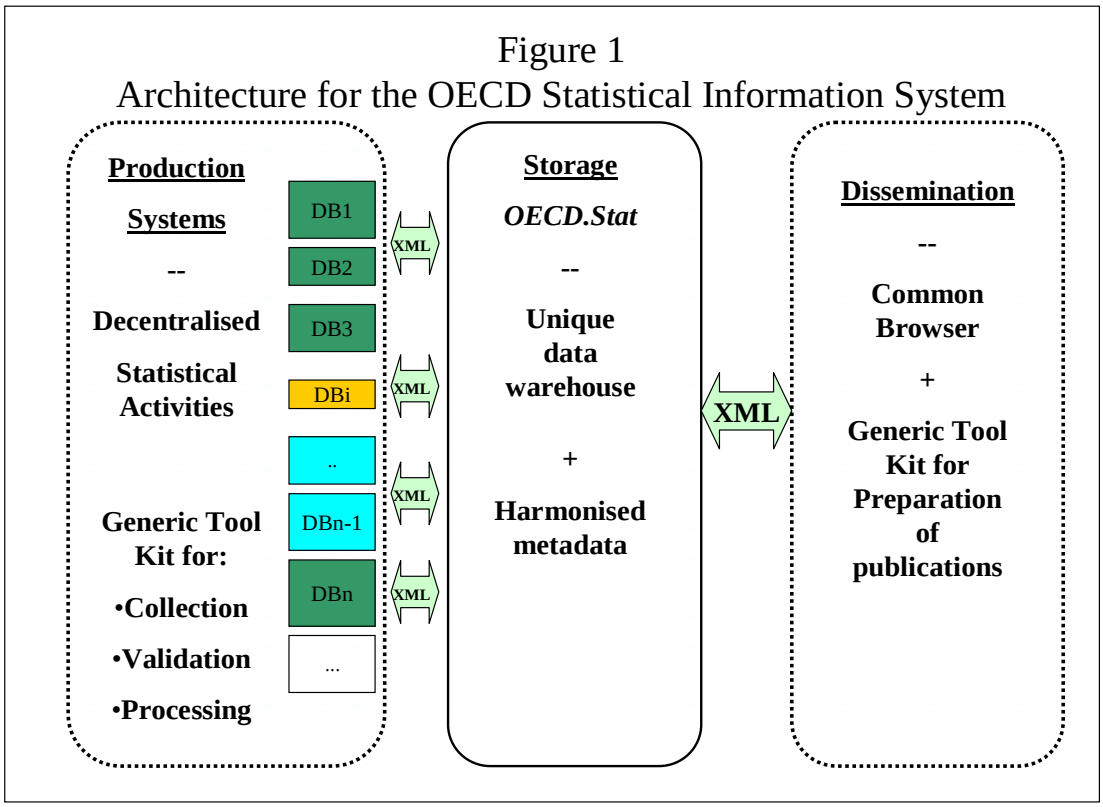
A "vision" for a modernised OECD Statistical Information System has been developed in the context of the OECD new strategy for statistics launched in 2001. In that vision, the OECD statistical information system encompasses tools for data collection, manipulation, storage, dissemination and user tools for data discovery and retrieval. The main idea is to preserve the independence of data producers while making their data and metadata part of a coherent and seamless corporate system. Figure 1 below shows a schematic view of the architecture of the OECD statistical information system. It comprises three layers, as follows:

- a production layer for data collection, validation and processing;
- a storage layer for providing data and metadata to internal users; and,
- a dissemination layer for producing publications.

New developments in information technology offer advanced tools for managing a decentralised statistical system. In particular, IT architectures based on the use of web technologies provide the possibility for integrating data and metadata among different groups of statisticians, increasing the efficiency and overall coherence of statistical activities and products, maintaining a decentralised and flexible system. However, given the constraints of the regular production of statistics and of limited resources, the OECD needs to have a pragmatic approach and to implement the vision through incremental improvements to the existing individual production processes. Those systems need to be maintained, further developed and eventually changed in order to take advantage of new technologies or of changes in the overall technical environment. In the context of the new strategy for statistics, those tasks are conducted with users' needs and corporate considerations in mind.

- This Annex describes the progress made in that area over the last two years. In particular, new developments in the areas of data collection, data processing and database management and dissemination are described. The OECD is also using managerial levers to implement its vision of a modern statistical system and to fully realise the potential efficiency gains offered by technical changes. Finally, the Annex shows how the OECD Quality Framework and the implementation of the new infrastructures support each other to improve the overall quality of OECD statistics.

Figure 1
Architecture for the OECD Statistical Information System



2. Improving the existing production systems

2.1 Data collection

In most cases the OECD collects its data from national statistical organisations (NSOs)[37] or other international statistical organisations. Data are collected via specialised questionnaires or from the data already disseminated at national level. The nature of the details and processes involved depends on a number of criteria:

- nature of data to be collected, e.g. whether or not they are part of a regular production process;
- frequency of the collection, which varies from weekly to annual or even occasional;
- regularity of the content over time;
- homogeneity of content regarding the internal structure of the NSO (e.g. subject matters);
- quantity and level of data (e.g., disaggregated versus aggregated data);
- method of collection;
- security and confidentiality requirements

[37] In this paper NSOs comprises all national statistical organisations, including the National Statistical Office, the Central Bank and the various ministries producing statistics.

- technology and infrastructure.

In addition to the diversity generated by those criteria, processes for data collection have evolved over the years as the result of bilateral arrangements between the OECD and national agencies or other international organisations. Of course, progress in information technology and communication has been an important driver of such evolution. Also, policies regarding arrangements vary from accepting any format and medium suggested by the data provider to imposing a fixed format and fixed medium. This Section reviews successively, progress in the areas of data collected through off-line questionnaires, through on-line databases and describes an experiment for a new way of providing data to the international statistical system.

2.1.1 Collection through off-line questionnaires

Regarding electronic questionnaires, current experience shows that good practices in data collection developed by some Directorates are applicable to other parts of the OECD. In 2002-2003, new electronic questionnaires have been designed, incorporating automatic data checks and the management of metadata. In 2003, a review of current practices will be carried out in the context of the quality reviews for several statistical activities that imply large data and metadata collections. The results will be used for further improving the tool for preparing new questionnaires, which is part of the **statistical toolkit** developed for modernising individual statistical production processes. This initiative should reduce the cost of data collection and minimise the burden on data providers. Also, since 2003 all new questionnaires should be submitted to the Chief Statistician before being sent. A central collection of those questionnaires is going to be developed to make them available to all statisticians in the Organisation.

Improving internal co-ordination inside the Organisation (through the OECD Glossary of Statistical Terms, the statistical toolkit, etc.) should improve the harmonisation of definitions of variables across questionnaires sent to countries. Organising information on data collection consistently across collections will further increase their corporate value and avoid the risks of duplication, or inconsistencies in data collection.

2.1.2 Extractions from on-line databases

For the collection of data that are already part of the regular production of NSOs (e.g. in the area of short-term statistics), the OECD has been very flexible and has preferred to adjust to formats readily available to individual data providers in order to gain on timeliness. New ICT has permitted very spectacular gains in efficiency in this domain, for both the OECD and its data providers.

The latest technique developed and implemented by the OECD in the last two years is called “Web Queries”. This technique is based on the ability of Microsoft Excel to save, on the user’s PC, queries made to a database through the web. When it is possible to use that technique, there is no extra development needed on the NSO side. On the contrary, web queries add at no cost a functionality that had been developed by some NSOs for their output database systems. Also, preselections are stored on the user side, which does not increase the burden on the producer side and does not require giving users the right to write into the NSO database. On the OECD side, one click on a button automatically retrieves the data needed from the NSO database. Also, since Web queries can be parameterized, modifications are easily made. The required conditions for Web queries to be used to retrieve data from one NSO database are that data can be seen on a web browser. The technique allows going through password protection. Unfortunately, some database systems in NSOs have extra layers, used for security or for commercial purposes, that prevent the technique to work.

2.1.3 Data sharing and NAWWE

The example of Web queries described above shows how efficient data collection is for the OECD when the data collected are published by the data provider on the Web. The "data sharing model" pushes the idea one step further in suppressing the collection by international organisations. In such a model, data available from international organisations' web sites are not extracted from international organisations' databases but directly retrieved from NSOs web sites on users' request. Of course, for the model to work a central data catalogue must be maintained and some statistical as well as IT standards must be used by NSOs. This concept was presented at the 2002 Conference of European Statisticians meeting in Paris[38] and is further developed in the context of the Statistical Data and Metadata Exchange initiative (SDMX)[39]. The 2002 OECD meeting of National Accounts experts proposed that an experiment be started to test the implementation of the data sharing model in the National Accounts area amongst NSOs and the OECD. The project is called National Accounts World Wide Exchange (NAWWE). This section describes the progress on the experiment so far.

The idea of the NAWWE project is to implement a model in which data are not transferred across organisations but, rather, published on the web in such a form that users can access them by simply using the country and variable references. A constraint in the project is to start from the Excel tables already produced by NSOs for transmitting national accounts data to International Organisations.

Another objective of the NAWWE project is to have the data collected by international organisations to be the data officially disseminated by NSOs. The two advantages of this model are that the burden of reporting to international organisations is minimised, and data quality is maximised for the international statistical community since the data they use are those officially disseminated and not specially compiled for and transmitted to international organisations.

In technical terms, the approach is based on XML and the associated technology standards. In succinct terms, NSOs would post data on their web sites using a standard XML based format. The files can be password protected or not. Then a directory of the corresponding XML files URLs for each country is made available centrally on the web, for example on the OECD site. The OECD would make available a "web service" permitting simple extraction of data from the system using parameters, the country identifier and the national accounts standard SNA93 code. In this model, each NSO can also use the same XML file as central source of data for its own dissemination for all formats and media. The NSO can easily mix, in their own tables, data from other countries and from international organisations for international comparisons.

OECD developed an Excel macro that allows NSOs to produce standard XML files from the existing Excel files they send to International Organisations. The first draft XML schema proposed by SDMX in April 2003 has been used to produce the NAWWE XML files. A simple Web application has been developed by the OECD which exploits the XML files from several countries together. The system simulates a database where in fact there is no database. In the next step the OECD would like to deploy the system amongst the statistical community in order to generate a discussion on the political implications of the data sharing model.

[38] See the IMF and OECD, "Progress Report on New Developments in Data and Metadata Collection for International Organisations", Conference of European Statisticians, Paris, June 2002. http://www.unece.org/stats/documents/ces/2002/8.e.pdf

[39] See www.sdmx.org

2.2 Databases and production systems

Data at the OECD is stored in databases, dataset by dataset. Because of the decentralised decision process, datasets have been implemented over the years in a variety of systems. Major elements in the database software infrastructure used across OECD statistical areas today are:

- MS SQL-Server 2000 for data and metadata storage and cataloguing;
- MS OLAP component for multi-dimensional data-manipulation;
- ORACLE Express for multi-dimensional data-manipulation, 4GL programming and some data storage;
- FAME for time-series data manipulation, graphics, 4GL programming and data storage;
- SAS for analysis of disaggregated data;
- MS-Excel/Access for common data manipulation and some data storage.

This diversity of systems has resulted in a variety of different implementations all of which require continued maintenance and, sometimes, further developments. In fact, at present, because of a lack of resources, it is impossible to abandon existing solutions and to develop a completely new production system. It has been decided to continue maintaining existing production systems and to develop an intermediate layer between original production systems and final users, to allow the latter to navigate across datasets using a common software package. This new system is called *OECD.Stat* and will be described in section 3 of this Annex.

In the meantime, recent evolutions in the software market, notably the phasing out of the current version of Express software by Oracle, have made it necessary to replace a range of statistical production systems at the OECD. In the context of the new strategy for statistics, a generic toolkit for statistical database management has been developed (*OECD.Statworks*). This toolkit has been assembled into an application and uses standard OECD software tools (MS SQL Server 2000 relational database, MS OLAP, Excel, .NET development environment). The generic SQL Server database is also open to access via existing statistical software (SPSS, SAS etc) and 3rd party analytical tools as required. This toolkit ensures the full integration of statistical data with the corporate statistics warehouse *OECD.Stat.*

2.3 Systems for data dissemination

On the dissemination side of the process a lot of progress has been made to increase the global value of OECD statistics. Most electronic dissemination is standardised on a unique software package[40] and data format, used on the web as well as on CD-ROMs. However, the implementations still vary and the production processes used to create the necessary input files have been developed independently to each other. The content of data files is not harmonised, with consistent data and metadata presentations. Furthermore, because data files are independent, some data series need to be duplicated and some apparent inconsistencies can occur because of, for example, differences in data vintage and of lack of metadata.

To address these issues several actions have been undertaken, to develop corporate tools for improving the efficiency and the coherence of statistical activities. In particular, it is worthwhile to mention the new OECD statistics portal. The first OECD Statistics Portal (www.oecd.org/statistics) was launched in September 2001, in the context of a major revision of the OECD web site. The Statistics

[40] The package presently used by the OECD for data dissemination is Beyond20/20 TM.

Portal provided, for the first time, an OECD web site exclusively dedicated to statistical information. The Portal has been revised and a new version has been made public at the beginning of July 2003.

This new version has been designed taking into accounts results of usability tests conducted since the launch of the previous version. The main conclusions of those tests were that the navigation and the jargon were still complex for most users. In the new version navigation is simplified. In particular, the thematic classification is complemented by a short description with standard keywords for all themes. Also, the number of pages has been reduced and data files are available on the first click on the selected theme.

3. The new output database: OECD.Stat

3.1 General description

The preceding sections have described how the Organisation is working on improving its existing systems to incrementally implement the vision of a corporate statistical system. This section describes how the Organisation is putting together the information necessary to seamlessly integrate the output, to facilitate usage of statistics across the Organisation and to help collaboration between individual statistical units. The name of this corporate system is *OECD.Stat*.

The aims of the project are to:

- improve accessibility to, and global visibility of, OECD statistical databases, enabling internal users to conduct quick searches and extractions from various datasets, to develop analytical and horizontal studies and create new "integrated" products;

- minimise the number of duplications of data and metadata and to improve coherence across different databases;

- provide a corporate technical framework in which the Organisation can facilitate the implementation the quality standards required by the OECD Quality Framework

In *OECD.Stat*, each Directorate contributes to a corporate "data warehouse", where all relevant data are stored with their metadata, in order to give full and easy access to all authorised internal users. Each Directorate remains independent in determining its own statistical and analytical processing of raw and final data under the guidelines of the OECD Quality Framework, but the outputs of these processes are part of the corporate data warehouse. In practice, because of the complexity and variety of statistical activities, the system is based on a "constellation" of datasets, connected to each other and parts of a single data warehouse. The intention is to make *OECD.Stat* the unique corporate source of statistical information.

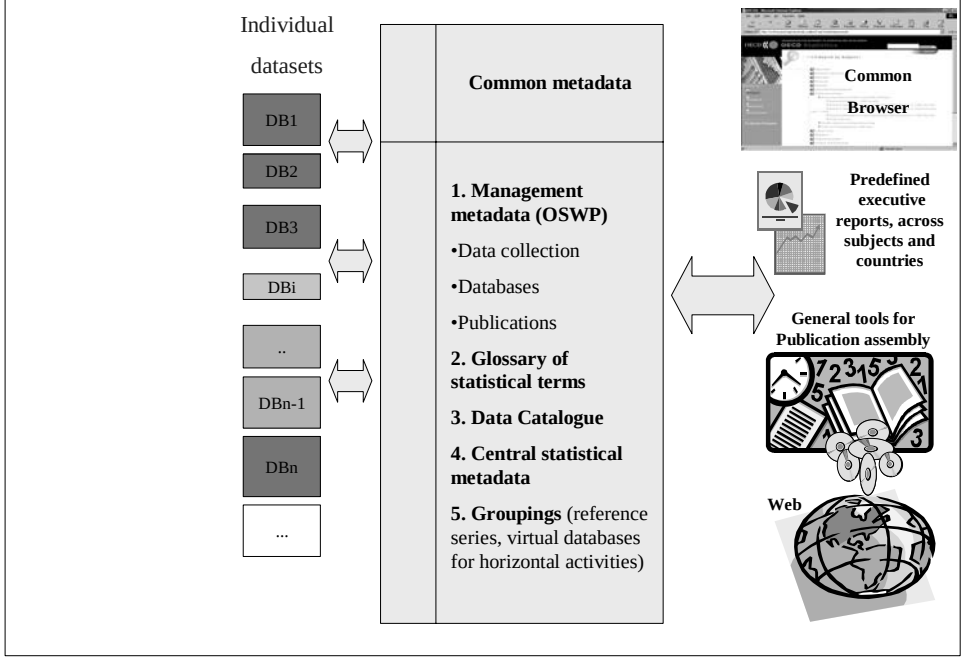
Integration of datasets in this environment is realised through metadata. The OECD Statistical Work Programme (OSWP) database, described below, and the "OECD Glossary of Statistical Terms" are parts of *OECD.Stat* and they play the role of catalysts for the integration process. The OSWP represents the top part of the metadata system, containing descriptions of datasets. The Glossary permits association of target definitions to variables in the system, improving accessibility, interpretability and coherence. Figure 2 below shows a conceptual view of *OECD.Stat*. Metadata are represented in the central box and include:

- management metadata giving detailed information on statistical activities, the OECD Statistical Program of Work (OSWP);

- a glossary of statistical terms for the harmonisation of terminology and concepts;

- a central data catalogue for the location of data in the collection of OECD datasets;

Figure 2
***OECD.Stat* Statistical Architecture**



- a central metadata repository for the storage of metadata elements that are independent of individual data items; and

- information on groupings of data: publications, the set of most commonly used data series (referred to as Reference Series), and virtual databases for horizontal studies.

Metadata provide the integration and individual datasets remain independent. As stated above, each Directorate remains independent in determining its own statistical and analytical processing of raw and final data under the guidelines of the OECD Quality Framework. *OECD.Stat* is the output database. Each dataset is stored in *OECD.Stat* as a multidimensional object or hypercube. Because of the multiplicity of production systems, it has been decided to create a generic data entry module for *OECD.Stat*. The module accepts generic XML files and updates data in the data warehouse accordingly. XML files are generated independently by each production process based on a standard acceptable schema. This makes *OECD.Stat* and the existing production systems technically independent, but linked in real time to the latter.

This project is technically complex and has required experimenting with several IT solutions, and a wide analysis of data and metadata models. The sections below describe successively, various parts of the system, the way in which *OECD.Stat* and the production areas are linked, the "groupings" concept (in particular, the group of "Reference Series"), the metadata structure and the technical implementation.

3.2 OECD Statistical Programme of Work (OSWP) and associated database

The OECD Statistical Programme of Work (OSWP) was developed as a tool for internal co-ordination and external communication. The Statistics Directorate is responsible for preparing the Programme, using information provided by relevant Directorates through an electronic questionnaire for

each line of statistical activity[41]. The two main uses of the Programme are: for users of OECD statistics to have information on what is going on in the OECD and to enable the OECD to take decisions about existing activities (e.g. launching of new activities, or co-ordinating work on activities across international organisations). The classification used to present activities in the OSWP has been derived from the UN classification of statistical themes and is currently used in the OECD web "Statistics Portal".

A further use of the database behind the OSWP is to facilitate access to the products and outputs of the statistical work of the Organisation and to provide information on other parts of the OECD statistical information system. Information collected by the OSWP on individual datasets is used to allow users to search for datasets by Directorate, theme, activity, related publication or keywords. Information collected for the OSWP is also used to conduct activity reviews within the context of the OECD Quality Framework. Therefore, the OSWP represents the main entry point into the OECD statistical information system.

In 2003 the OSWP database containing detailed information was made available on *OLIS.net*, the network that is normally used by all Organisations' governmental bodies to access all OECD documents and outputs (publications, statistics, etc.). This will represent an important tool for agencies responsible for co-ordinating statistical activities, as well as for all public bodies using OECD statistics.

3.3 Glossary of statistical terms

Definitions of statistical data elements and concepts are an essential part of international standards in the area of statistics. Because of the diversity of the institutional, economic and cultural climate within different countries these definitions are often the result of compromises and by necessity, are fairly broad in some instances. To a large extent they can be viewed as "target" definitions requiring elaboration and adaptation to national circumstances.

Notwithstanding these limitations, it is important to promote the use of these concepts and definitions by increasing their visibility by making them readily accessible via the Internet. For this reason, the OECD has developed a Glossary of Statistical Terms, available at http://cs3-hq.oecd.org/scripts/stats/glossary/index.htm.

The Glossary contains both target definitions of the main variables (data elements) collected by the Organisation for use in its statistical and analytical output (e.g. for industrial production, services, unemployment) and definitions of terminology/concepts (data element concepts) used in OECD publications (e.g. accrual accounting, metadata). The 5,600 definitions included in the OECD Glossary cover a very broad range of statistical subjects. The main elements of the Glossary are:

- unique title for the definition;
- the actual definition;
- for some definitions, text providing further background on the definition, its application and relation to similar or related concepts. This field may also contain URLs to relevant documents describing appropriate use of the variable defined, etc;
- detailed source information;
- classification of each definition to a broad statistical theme;

[41] A line of activity is defined 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 an analytical publication with extensive statistical content*".

- internal cross-links to related definitions, etc., contained elsewhere in the Glossary;

- URL links to the complete source document containing the definition where this is currently located on the websites of international organisations or national agencies.

The Glossary also includes search and interrogation facilities and information describing each of the fields. One of the powerful features of the Glossary is the linkage (via URLs where available) of definitions to the full standard/source containing the definition. This enables the user to readily obtain further information on the definition. The Glossary is an essential part of *OECD.Stat*. It will permit to relate variables, to navigate the system and to improve the overall coherence.

3.4 Groupings: reference series and horizontal products

In the present situation, users have still to navigate the various hypercubes in order to locate their data. In order to deliver rapidly tangible benefits to analysts in the Organisation it has been decided to create, from data residing in various hypercubes, a specific dataset comprising data series that are the most frequently accessed by users who are not involved in the corresponding statistical activity. An initial list of Reference Series has been obtained through consultation with analysts throughout the OECD. The list contains the following statistics: main aggregates of national accounts, with history and forecasts; labour force and population data; exchange rates; purchasing power parities and price indices.

Analysts who are not experts in a subject matter area from which they need data have difficulties in locating those data. This is because the data they are searching are part of complex datasets containing thousands of data series. For example, GDP data are stored with all the rest of national accounts data in a complex accounting framework. In some cases, there is a risk that the wrong data is used. In the past, data corresponding to "Reference Series" were duplicated in individual databases to facilitate their use as background data to calculate ratios, per capita, etc. This was another factor for confusion and risk of inconsistencies. Reference series are also meant to define the standards in terms of associated documentation, with the objective to provide information to users who are not experts in the particular domain of the corresponding Reference series.

In technical terms, Reference Series are contained in a specific virtual hypercube taking its data from actual hypercubes corresponding to output datasets. The same technique can be used to create virtual cubes for horizontal or any analytical study.

3.5 Metadata structure

A very difficult part of the project is the management of metadata. Following on from the development of the OECD Glossary of Statistical Terms, which includes all main statistical definitions used by various OECD Directorates, the SDMX Common Metadata Vocabulary developed by the OECD and Eurostat, under the auspices of the SDMX project (see below), will be released in 2003. The definitions contained in the SDMX Vocabulary will be used for developing the OECD common repository of metadata, which should be completed by the end of 2003.

In the *OECD.Stat* data model, harmonized metadata are present at all levels of the data structure, including dataset level, dimensions, elements in dimensions, crossing of dimensions and their elements. Dataset information come from the OSWP, relevant target definitions are linked to the Glossary and, in the future, more corporate metadata will be added, making *OECD.Stat* the central source for all OECD statistical information. The OECD will align the metadata structure and content to those suggested by SDMX in the context of the work on metadata repositories.

3.6 Technical implementation

In technical terms, *OECD.Stat* is developed using Microsoft tools. MS SQL Server 2000 and its OLAP component are used as central data and metadata repository. As stated above, datasets are stored as hypercubes in that structure, with common dimensions taken from a central repository. Each hypercube is related to the relevant item from the OSWP. Elements of dimensions, in general those other than country and time are related to their corresponding entry in the Glossary. An Excel add-in has been developed to permit easy access to individual datasets and to Reference Series using Excel Pivot Tables.

4. The reinforcement of quality through the implementation of the new architecture

The implementation of the new statistical information system is part of the overall strategy for improving the cost efficiency of statistical activities and the quality of OECD statistics. In particular, the new system can improve several of the quality dimensions quoted in the OECD Quality Framework: relevance, accuracy, accessibility, interpretability and coherence. First, as already said, in preserving individual production processes in Directorates, the new architecture preserves the benefits of the decentralised organisation. At the same time, it improves the capabilities for cross-cutting studies by permitting navigation and selections across datasets and by permitting the creation of new virtual datasets (as is already done for Reference Series) according to user needs. At the same time, accuracy is potentially improved through more data confrontation possibilities. Therefore, the new system reinforces both the relevance and the accuracy of OECD statistical products.

Second, accessibility is the most obvious quality dimension immediately improved by the new statistical infrastructure. The architecture has been developed focusing on the needs of the final user, providing the latter with a "user friendly" tool for accessing all datasets in the system. Metadata help data discovery and retrieval.

Interpretability is improved by central tools for data documentation and by improved accessibility to metadata. The Glossary, central metadata items and a common set of metadata are the main instruments for achieving this target. Coherence of metadata is also an important factor for improving interpretability.

Coherence of data has four important sub-dimensions: within a dataset, across datasets, across countries and over time. Coherence within datasets is improved in *OECD.Stat* by the fact that data are related to the glossary which forces data concepts to be related to their official target definition and other attributes, whether internationally agreed to or not. Coherence across datasets is improved by the data confrontation allowed by the data warehouse in the improved accessibility and interpretability of data.

Finally, the OECD Quality Framework will facilitate the implementation of the new architecture which in turn will increase quality. Quality guidelines encourage the use of the new tools provided in the new statistical system and quality reviews will encourage and facilitate the full implementation of all statistical activities into the new architecture.

ANNEX 4 - A LAYOUT FOR THE NOTES SECTION OF OECD DATA SETS PUBLISHED ON SOURCEOECD

Since the end of 2001, OECD statistics have been available on-line from three sources:

- to paying clients via the SourceOECD internet service hosted by Ingenta;
- to OLISnet users via the WDS service;
- to all users via the statistics portal.

In particular, both SourceOECD and WDS are based on Beyond 20/20 technology and use the same underlying files ("IVT" format) that are currently generated for diskette and CDROM products.

Early tests on these services, carried out at OECD, highlighted apparent inconsistencies across OECD data products, particularly when using the search engine facilities. For example, *(i)* searches using the key words "GDP" and "Employment" can yield many possible data products for the user to choose from; and *(ii)* comparisons of key variables (e.g. GDP) across different data products sometimes show different values when the user might expect them to be identical. Also, identical searches on *SourceOECD* and WDS can give very different results. Finally, there was some concern about the general 'look-and feel' and ease of navigation within the Statistics section of *SourceOECD* as well as the general technical environment.

The main causes for apparent inconsistencies across OECD statistics products on *SourceOECD* and WDS were identified as follows:

- different data products with different vintages of the same time-series without the necessary documentation explaining reasons and/or giving the last update date of the corresponding file;
- different data products containing similar variable names and/or classifications without the necessary documentation explaining underlying sources and definitions and relations to other data products;
- missing or incomplete minimal metadata elements in data file (units, power of ten, etc.);
- different types of search engines used;
- misleading results from the search engines. On *SourceOECD*, this can be due to superfluous information within the abstracts for certain data products. On the WDS, this can be due to the wide scope of the searches across IVT files.

Therefore, recommendations were issued to maximise the information to users, improving the metadata associated both to individual statistical databases. In particular, it is necessary to provide a standard presentation of general notes within OECD data products (IVT files) published on *SourceOECD* and WDS. Within the framework of B20/20 technology, it makes extensive use of links to PDF, HTM and Excel files to enhance the provision of metadata. It would be highly desirable for the *SourceOECD* abstract for each product to match the corresponding description in the IVT general notes section.

The box below describes a standard layout for the notes section of OECD data sets published on SourceOECD and uses STI's STAN industrial database as an example. Note that this does not refer to the

SourceOECD abstract - the first point of contact by clients and the page interrogated by Ingenta's search engine. It would be desirable for the abstract to closely match the description in the notes section.

Basically, the information is divided into two columns with the general description on the left and links to more detailed information / metadata on the right. Notes on the various elements suggested follow.

Last update : **15th May 2001** The *STAN database for Industrial Analysis* has been developed to provide analysts and researchers with a comprehensive tool for analysing industrial performance at a detailed level of activity. It includes annual measures of output, labour input, investment and trade that allow users to construct a wide range of indicators to focus on areas such as productivity growth, competitiveness and general structural change and make comparisons across countries. A standard industry list is used that provides sufficient detail to enable users to highlight high-technology sectors. STAN is primarily based on member countries' annual National Accounts tables and uses data from other sources, such as national industrial surveys/censuses, to estimate any missing detail. Since many of the data points are estimated, they do not represent official member country submissions. Notes for previous users of STAN : This new version of STAN is based on ISIC Rev. 3 (compatible with NACE Rev.1) and has been expanded to cover all activities (including services) and a wider range of variables - it has effectively been merged with OECD's International Sectoral Database (ISDB) which is no longer updated. STAN is now updated on a 'rolling basis' (i.e. new tables are posted as soon as they are ready) rather than annually - thereby improving timeliness. Recommended citation : *OECD, STAN database, 2001*	• News • Corrections [Optional] • Full documentation • Current data coverage [Optional] • Recommended uses and limitations • Relationships to other OECD data sets • Examples of use in publications and research papers [Optional] • Compatible data sets for further research [Optional] • Useful link [Optional] • Technical help [Optional] • Contact

1. General notes

- **Last update** - useful to be highlighted as it immediately allows regular users to see if any updates have occurred since their last visit.
- **The general description** - should be clear and concise and approved by Editorial Board for Statistics to ensure that is it up to date and uses terminology consistent with other data sets.
- **Notes for previous users** [Optional] - In cases where major changes have occurred to the data sets, it could be useful to alert regular users.
- **Recommended citation** - useful to ensure that an OECD data product is cited in a consistent and recognisable manner around the world.

2. Links to detailed information

A standard set of links would provide users with easy-to-access information concerning the data set as a whole. (More detailed country and variable notes will be found in the 'dimension notes' within the data

product's 'selection' pages.) Some categories would be optional, though they should have consistent titles across data products. The links generally refer to PDF or HTM files, though in some cases Excel files may be preferable. Although some work will be required in the short-term to set up such lists and regular updates required for some of the categories, the results should make it worthwhile. Suggestions for other categories or different titles for those below are welcomed.

News - could include outlines of revisions and updates since the last version; descriptions of new tables; changes in underlying methodologies; links to recent related publications/ methodological manuals; recent major citations and any other new information that may interest users.

Corrections [Optional] - In the (unlikely) event that erroneous or misleading figures have been published, this link would point to a page outlining the nature of the error(s); an explanation; how long the error has been in the data set; and a humble apology.

Full documentation - This would be a link to a PDF file, which users can print out to read at their leisure, containing the usual detailed sources, definitions, methodological notes, estimation techniques and other detailed metadata. Some of the information will also be accessible separately from the data product notes above (such as recommended uses and limitations and relations with other data sets) and from the dimension notes (such as variable definitions).

Current data coverage [Optional] - Many users find data coverage tables helpful, particularly for quickly determining, for example, to what extent they may make aggregates across countries, or more generally, the usefulness of the data product for the type of analysis they had in mind. [HTM, PDF or Excel file depending on detail provided] *In STAN an Excel file is used with one page per country giving the period for which data is available for each variable x industry combination.*

Recommended uses and limitations - It is important to guide users with limited knowledge of the statistics presented and to help them determine whether the product meets their requirements. It could contain examples of the type of indicators that can be constructed and/or inferences that can be made; the types of analyses that can be performed; the type policy questions it can help answer; the 'shelf-life' of the data product etc. These notes could also explain when caution should be employed, what type of calculations should be avoided and the type of inferences that should not be made.

Relationships to other OECD data sets - If the data product uses other OECD databases as inputs, these should be mentioned here (and possibly linked to). If there is any possibility that duplicated data may not match - reasons should be made clear. If there are other OECD data products containing identically named (and/or defined) series, reasons for differences should be provided. Criticism of other OECD data products is to be avoided! [HTM page or PDF file] *STAN uses STD's ANA, SSIS, IIA and ITCS databases as inputs (as well national sources) and has variable and industry descriptions that to varying degrees match those in ANA, SSIS, IIA and STD's "Services: Statistics on Value Added and Employment" publication. Therefore discussion is required on the difference between National Accounts and Survey/Census data; detailed v. aggregate National Accounts; Short-term v. Annual measures; product v. industry classifications, etc.*

Examples of use in publications and research papers [Optional] - an opportunity to publicise the benefits of the data product by giving titles (and if possible links) to any publications and research papers, (not just OECD) that have made extensive use of the data.

Compatible data sets for further research [Optional] - mention (and provide links to) any other data set (OECD or otherwise) that can be used in conjunction with the data product. *The STAN family of databases also includes databases containing R&D expenditures (ANBERD), bilateral trade by industry (BTD) and Input-Output tables (I-O) linked by compatible industry lists. These would be mentioned here.*

Useful link [Optional] - provides links to websites that may enrich use of the data product. For example, methodological manuals, international standard classifications, academic research, future developments in the field, data for non-OECD countries etc.

Technical help [Optional] - could provide hints on navigating around the data product within SourceOECD. Useful if advanced features are being used and/or data organised in a 'non-standard' way. It could provide, for example, best practice for downloading large selections of data.

Contact - As well as providing contact details, a note encouraging users to provide comments and suggestions could be included.

ANNEX 5 - AN EXAMPLE FOR THE PREPARATION OF THE SELF-ASSESSMENT IN THE CONTEXT OF QUALITY REVIEWS OF EXISTING ACTIVITIES

A. An overview of the STAN[42] database today

The STAN database is designed to provide analysts and researchers with a comprehensive tool for analysing industrial performance at a relatively detailed level of activity and carry out comparative studies across countries.

It includes annual measures of output, labour input, investment and international trade in goods which allow users to construct a wide range of indicators to focus on areas such as productivity growth, competitiveness and general structural change. The industry list provides sufficient detail to enable users to highlight high-technology, high-growth and 'knowledge-based' sectors and is compatible with those used in related OECD databases such as the other data sets in the "*STAN family"* – notably ANBERD (R&D expenditure by industry), Bilateral Trade Database (BTD) and Input-Output tables[43].

STAN is primarily based on Member countries' annual National Accounts by activity tables and uses data from alternative sources, such as national industrial/business surveys, to estimate missing detail. Trade data by industry are derived from data classified by products using a standard conversion process.

The latest version of STAN is based on ISIC Rev. 3 and covers all activities (including services) - it has effectively been merged with the OECD's International Sectoral Database (ISDB) which is no longer maintained. STAN is now updated on a "rolling basis" rather than published as an annual "snapshot".

The STAN database is managed by the Economic Analysis and Statistics (EAS) division of the Science, Technology and Industry Directorate (STI). Its development since the early 1990s has been overseen by the Statistical Working Party (SWIC) of OECD's Committee for Industry and the Business Environment (CIBE).

Detailed information including full documentation, industry and variable lists, details of data availability and sample tables, can be found at www.oecd.org/sti/stan.

The following summary and outline of main issues were derived from the first part of the quality review process, the "self-assessment" by data managers, carried out in early 2003. Comments and suggestions from SWIC delegates, OECD colleagues and a wide range of STAN users, including those in national statistical offices (NSOs), have also been taken into account.

B. Attributes of STAN identified in the Quality Review

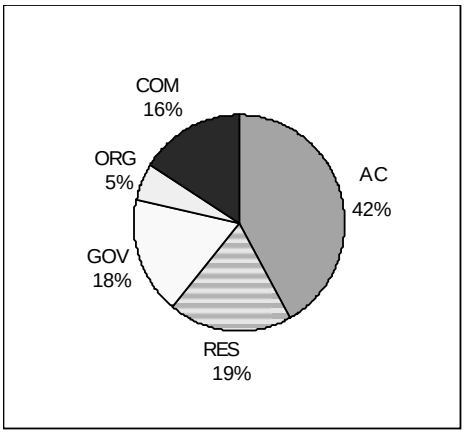
B.1 Perceived strengths

- ***High demand*** – STAN is widely used by OECD, national administrations and other international organisations and by a variety of academic and research institutions – see below for illustration of user profiles based on a sample of over 200 different institutions. There is high interaction with

[42] The name STAN is derived from STructural ANalysis.

[43] The other data sets in the "STAN family" were not considered in this quality review.

users via a STAN users distribution list which encourages regular feedback and ensures ongoing user needs are understood;



COM = commercial enterprises

AC = universities

RES = research institutes

GOV = government agencies

ORG = international organisations.

- ***Proximity to analysts*** - STAN benefits from being maintained within an analysis-oriented environment. The STAN data manager assumes the role of user when contributing to publications such as OECD's *Science, Technology and Industry Scoreboard*[44]. Also, STAN is ultimately overseen by an OECD policy committee, namely CIBE, and has thus developed to address issues forced by this group such as economic growth, globalisation, structural change and competitiveness;

- ***Low demands on national data providers*** – The request for data for STAN, which usually accompanies the standard OECD/Eurostat Annual National Accounts questionnaire, simply asks NSOs to send National Accounts by activity tables for as many measures and years as possible at the most detailed level of activity (according to national classifications) available, in whatever format they find convenient. Much use is also made of tables provided on the Internet;

- ***Timely*** - STAN is released on a "rolling basis" (i.e. new country tables are disseminated as they become available) rather than as an annual snapshot. It is therefore only available via Internet through OECD's commercial service *SourceOECD* and OLISNet;

- ***Variety of dissemination formats*** - To meet varying user requirements, STAN is available either via a query-based application[45] or as a set of ready-to-use Excel tables. An ASCII version is available on request for easy loading into users' local data systems (e.g. SQL, SAS, Fame);

- ***Good provision of metadata*** – A range of information is provided including full documentation, data coverage tables, update history, country-specific notes, recommended uses and links to other datasets. Notes attached to individual data points highlight whether they are official National Accounts data or Secretariat estimates;

- ***Comprehensive data checking and diagnostics*** – A range of summary statistics are supplemented by detailed graphics, both on-screen and printed, of source data, selected ratios, industry shares and growth rates – comparing different data sources and/or data vintages. Feedback is provided to data providers when anomalies are discovered.

[44] See web version at http://www1.oecd.org/publications/e-book/92-2001-04-1-2987/

[45] See http://cs4hq.oecd.org/oecd/selected_table.asp?tableId=591&lang=eng&userid=&password=

B.2 Known weaknesses

- ***Difficult to determine the quality of estimated data*** – Except when NSOs subsequently provide National Accounts data previously estimated within the STAN process. This can also be an issue at the national level since few of the series in National Accounts are directly observed;

- ***Comparability problems*** – Despite the use of common frameworks (such as SNA93, ESA95), there are many problems in comparing National Accounts, and thus STAN, across countries due to the varying underlying data sources and methods used by NSOs to fill the various tables. Such issues are increasingly being addressed in broader discussions of comparability such as the measurement of economic growth and productivity across countries. Reconciling national industrial classifications to ISIC Rev.3 can also be challenging. To address these issues in the short-term, the STAN country notes highlight where national definitions may diverge from target definitions, while general notes attempt to expose areas where analysts should be prudent when making international comparisons;

- ***Data coverage problems*** – Data coverage is weak for many countries for certain measures such as investment and capital stock, output volumes and hours worked – productivity analysts are the most vocal on this point. This mainly reflects difficulties that some NSOs are having in making the necessary estimates by activity under SNA93/ESA95 recommendations. For example, with software now considered an investment good, re-estimating investment series by activity for long time periods in order to generate capital stock by industry is non-trivial. The feasibility of generating OECD estimates of capital stock by activity, the practice in the old ISDB database, is being investigated;

- ***Vintage software*** - The STAN system is based on software (dating from 1995) soon to be phased out by its vendor. Following a technical review towards the end of 2002, first steps have been undertaken to re-develop the STAN data system. Most likely, the solution will involve SAS linked to SQL data storage;

- ***Technical knowledge concentrated on one person*** – An important problem currently being addressed. The proposed technical re-development is an opportunity to simplify and codify processes for others to inherit, while continuing the close involvement of data users;

- ***Heavy data updating processes*** – Cumbersome updating processes are partly due to varying database structures within the OECD (from which STAN draws some of its source data), and partly due to the desire to minimise data providers' burden to maximise their response. Improved data flows within OECD are expected as the work to create a general data warehouse gathers pace. However, continuing to work with country-specific data submissions from NSOs could be considered an advantage since it leads to a better understanding of national practices.

- ***Complex estimation routines*** – During the STAN technical re-development the convoluted, and often difficult to understand, estimation routines (used to fill detail not provided in National Accounts tables) will be re-specified, re-coded and well documented so that the more mechanical aspects of the STAN process can be delegated to other members of staff.

Quality dimensions	Overall evaluation				
	Very weak	Weak	Satisfactory	Strong	Very strong
Relevance					X
Accuracy			X		
Credibility				X	
Timeliness				X	
Accessibility				X	
Interpretability				X	
Coherence			X		

C. Summary of main recommendations and action points

C.1 Main issues for National Statistical Offices

- ***Co-ordination*** – Many NSOs group all available National Accounts by activity tables together either within their on-line data services or when submitting tables to OECD. However for some NSOs, much effort is required to gather the tables for the various measures presented in STAN since their dissemination reflects internal organisational structures. Contact with up to four separate divisions within NSOs may be required - such as I-O division (value added and components), investment division (GFCF, capital stock), employment division and price division (output volumes). In fact, we often get requests from national analysts, within ministries and other government agencies, asking for STAN versions of their own countries' data. Co-ordinating the compilation of National Accounts by activity tables would not only provide easier access for productivity and structural analysts but could encourage NSOs to review the coherence across measures – for example, employment v. labour costs.

- ***Encourage closer links with users*** – It can be useful for statisticians to have a good knowledge of how the statistics they produce are used. This allows them to make judgements on what should be publicly released and/or what metadata are required. Often statisticians are far removed from the users of their data, particularly in an international context. For example, Eurostat's requirement to provide 'back estimates' (pre-1995) according to ESA95 has resulted in some apparently rough estimates being produced by under-resourced national accounts divisions to meet deadlines – and while these estimates may not appear in national publications, the fact that they may be widely used once submitted to Eurostat and OECD is sometimes overlooked. Occasionally, we send data providers with examples of where STAN has recently been used, particularly by their own governments (e.g. White Papers), not only to motivate them to submit updates but to highlight that once sent, their statistics do not just gather dust in the corner of an OECD database. Also, national statisticians are encouraged to review the STAN tables for their country and join the STAN distribution list so they are kept informed of how their contributions are used.

- ***Details of compilation of National Accounts tables*** – As mentioned earlier, much of the content of National Accounts is not directly measured and sources and methods for filling various tables differ across countries. For STAN, more precise knowledge of what combination(s) of sources are used to fill the activity tables (industrial surveys? business registers? income surveys? labour force surveys?) could help improve estimation techniques and help to make them more country specific. Better understanding of national practices could also contribute to more prescriptive chapters of future versions of SNA (for example, suggestions of types of surveys and computational methods to use to fill certain tables).

- ***National STAN tables*** – If NSOs compiled National Accounts by activity tables at a sufficient level of detail going back 20 years or more, OECD's role could then be one of co-ordination and second level quality control from an international perspective (which to some extent is already

undertaken). There is an increasing demand for such tables, and despite STAN shortcomings, analysts around the world have a common data set to work with. Without it, analysts would take National Accounts tables and estimate the detail they require in different ways, with varying levels of understanding and degrees of success.

C.2 Main issues for OECD

The suggestions below could be considered within OECD in order to improve the STAN database. Some concern STAN directly while others relate to the more general OECD statistical environment.

- ***Address data duplication issues*** – Not surprisingly, there is overlap between STAN and the activity tables presented in OECD's annual *National Accounts of OECD Countries* publication (tables 7 to 10 of volume II). However, while the latter is limited to 31 relatively aggregate activities (the A31 list), STAN attempts to exploit the most activity detail countries can provide. Also, tables and content can differ due to differing updating cycles and publication regimes. Efforts could be made to address this issue;

- ***Review quality of source data sets*** – The quality of the contents of STAN is inevitably influenced by the quality of underlying data coming from other OECD data sets. Outcomes of forthcoming quality reviews of OECD's *Structural Statistics of Industry and Services* (SSIS), *International Trade in Commodity Statistics* (ITCS) and annual *National Accounts of OECD Countries* (ANA) should also contribute to improvements in STAN;

- ***Reduce process time*** – As the STAN system matured and managers became more experienced with national practices, process time decreased – a natural occurrence for most data systems (major changes notwithstanding). With the STAN technical development, it is hoped that process time can eventually be further reduced;

- ***Improve online dissemination*** – A query tool is not always appropriate for many analysts who need to download complete data sets, often the case for STAN. There is currently no mechanism for doing this on the commercial on-line data service (*SourceOECD*) – though complete STAN tables by country in zipped Excel format are available on OLISNet. The presentation of metadata could also be improved, for example, making it to easier to compare certain notes across countries. The current metadata structure was designed within the constraints of the software chosen;

- ***Encourage NSOs to publish STAN tables*** – Continuing the theme outlined above, OECD should strive to reduce the number of Secretariat estimates in STAN by encouraging national accounts divisions to provide more activity detail. For many countries the inclusion of just a few (less than 10) extra detailed activities (such as pharmaceuticals, shipbuilding, aircraft, telecommunications services), would meet the needs of a large proportion of users. As mentioned earlier, OECD's role could then be one of co-ordination and second level quality control from an international perspective. Providing advice and proposing common solutions for the more mechanical aspects of the compilation of national accounts tables could also be a function;

- ***Refine quality review process*** – Although having been partially involved in the design of the quality framework and review questionnaire, after participating in the process it seems that some improvements could be made to the 30-page questionnaire to make it easier for data managers to complete and interested parties to interpret. Also, since the OECD quality review summaries may provide information useful to NSOs it could be helpful to define a standard layout.