



STATISTICS DIRECTORATE

STD/NAES/TASS/ITS(2003)4
For Official Use

National Accounts and Economic Statistics - International Trade Statistics

**OECD DATA QUALITY FRAMEWORK AND METADATA MANAGEMENT
FOR TRADE STATISTICS**

Paper prepared by Andreas Lindner - OECD

4TH INTERNATIONAL TRADE STATISTICS EXPERT MEETING

Château de la Muette, Paris

7 April 2003 - 9 April 2003 (morning)

Beginning at 10.00 a.m. on the first day

JT00142200

Document complet disponible sur OLIS dans son format d'origine
Complete document available on OLIS in its original format

English - Or. English

OECD'S DATA QUALITY FRAMEWORK AND METADATA MANAGEMENT FOR TRADE STATISTICS

1. Introduction

1. Improvement in the quality of OECD statistics is one of the main objectives of the current OECD Statistics Strategy (OSS) and the Organisation already applies techniques that are used in total quality management frameworks without having adopted a formalised approach to quality. The “OECD Quality Framework” which is available already to all OECD staff through the Intranet provides an extremely useful tool to systematically assess, compare and further improve OECD statistics.

2. For OECD's Trade Statistics this general framework will be applied to test the quality of data and to correct or improve data held wherever necessary. It is obvious that such a process can only be successful with close and continuous interaction with national data providers. A key element of assessing data quality is the degree of completeness and pertinence of the accompanying Metadata and the way this qualitative information is accessible.

3. This document is divided into two parts. The first part is a selection of the most relevant guidelines, taken directly from the “OECD Quality Framework” which are of direct importance for dealing with trade statistics. The second part sets out the Metadata approach in place for Trade Statistics.

2. The OECD Quality Framework

4. Much work has been done in recent years to apply the concept of quality to statistical data. For example, the IMF, Eurostat, Statistics Canada and other national statistical offices (NSOs) have identified various sets of data quality components and have adopted quality frameworks to improve their organisations and the quality of data produced. The OECD quality framework benefits from this work and is not “reinventing the wheel” since it adapts existing definitions and approaches to the OECD context.

5. The OECD Quality Framework has four elements:

- a definition of quality and its dimensions;
- 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;
- 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

In the context of this paper, only the first two elements will be briefly described because of their direct impact on the trade database.

a) *Definition and dimensions*

6. Quality is defined as “fitness for use” in terms of user needs. This definition is broader than has been customary 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.

7. Given the work already done by several statistical organisations, the OECD drew on their work and adapted it to the Organisation's context. Several statistical organisations have already identified the dimensions of quality. These have been adapted to the OECD context. Thus, the OECD views quality in terms of eight dimensions: relevance; accuracy; credibility; timeliness; punctuality; 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 eight dimensions cited previously to OECD statistical output.

1) Relevance

8. 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 characterized 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.

9. 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.

2) Accuracy

10. The accuracy of data products is the degree to which the data correctly estimate or describe the quantities or characteristics that 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.

11. 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.

12. In the context of OECD, 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.

3) *Credibility*

13. The credibility of data products refers to 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,

14. Credibility is determined in part by the integrity of the production process. Principle 2 of the UN Principles of Official Statistics (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”.

4) *Timeliness*

15. 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.

16. At the OECD, timeliness of the data published by the OECD is largely determined by the timeliness of the data it receives from the contributing organisations.

5) *Punctuality*

17. The punctuality of data products implies the existence of a publication schedule and reflects the degree to which the 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 the 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.

6) *Accessibility*

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.

7) *Interpretability*

18. 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

19. 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

20. At OECD, 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.

8) *Coherence*

21. 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

22. Coherence has four important sub-dimensions: within a dataset, across datasets, over time, and across countries.

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

b) *Some basic principles adopted in data and metadata...*

1) collection

- 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.
- 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.

- 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
- 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.

(more)

2) *management*

- 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.
- Data sets containing “preliminary” data and metadata can be developed according to the needs of experts in charge of different work phases.
- Corporate statistical IT and software infrastructures are used for storing, accessing and disseminating data and metadata in preference to activity specific applications
- Definitional, procedural and operational metadata describing each phase of each statistical activity are documented and readily available to internal users.
- 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.

3) *dissemination*

23. 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.

24. 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 (included national statistical offices), as well as, subject to reciprocal arrangements, to international organisations.

c) **Metadata storage**

25. 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.* Generally provide a brief general description of the indicator and an outline of key issues that can impact on use of the data, target definitions, etc. Explanatory notes may not necessarily provide much detail on individual country methodology/practices.
- *Summary sources and definitions.* 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 contains more detailed methodological information on individual country practices normally collected and disseminated on the basis of a detailed 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.

Minimal list of metadata:

26. 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 adjustments made to the data and their impact.

Metadata storage:

Table 1: Metadata storage guidelines

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

Source: Draft OECD Quality Framework Table 6, page 51

3. The Metadata system in place for OECDs International Trade Statistics

27. This part will not focus on the technical aspects of metadata organisation, storage and availability, which as been already presented, but rather summarise the structure and characteristics of the qualitative and methodological information available. The source of this information is the OECD Metadata Questionnaire which has been sent out to OECD member countries in preparation of the 1st OECD Trade Statistics meeting in June 1999 (attached as Annex).

28. The OECD follows the *International Merchandise Trade Statistics: Concepts and Definitions* published by the United Nations Department of Economic and Social Affairs, Statistics Division [Studies in Methods, Series M, No 52, Rev.2 , United Nations New York 1998 (ST/ESA/STAT/SER.M/52/Rev2)].

29. The Metadata structure at present is organised as follows:

- 1) **General characteristics and contents of the ITCS Database**
- 2) **Definitions**
- 3) **Country notes**

1) *General Characteristics and contents of the ITCS Database*

30. This part provides information common to all countries in the database, namely:

- The list of Reporting Countries
- The time dimension
- Flows
- Values and Quantities
- List of Partner Countries
- Commodity classifications used

31. It summarises the statistical treatment of reporting countries, including the treatment of EU countries and the calculation of zone aggregates.

32. The availability and general characteristics of the different commodity classifications are described, including the data availability over time and the relationships between these classifications in terms of conversion possibilities and limitations.

33. Trade conversion factors from national currencies and € to US \$ are given. The different quantity units used are given as well and a description is given as the standard rules applied for aggregation and conversion. For quantity measurement, the use of the *Standard Units of quantity recommended by the World Customs Organisation (WCO)* is encouraged .

2) *Definitions*

34. This part provides guidelines for the rules to be applied for the compilation of trade statistics.

It includes:

- The statistical territory and the system of trade used : the differences between General Trade and Special Trade
- General coverage of trade : rules for inclusion or exclusion of trade to be reported
- Incoterms used for imports and exports
- UN Methodology versus EU methodology
- Geographical and Economic Classifications
- Overview of changes in OECD country coverage
- Reporting country aggregates used
- Valuation principles applied
- Aggregation of commodities and confidentiality rules

3) *Country Notes*

35. These notes on an individual country basis are the core element of the metadata accompanying the ITCS database and are of particular importance for an understanding of the “value” of collected trade statistics and degree of comparability across countries.

36. In these country-specific notes, all available detail for the two preceding points are given. The particular value of these notes is that they allow to draw a “quality profile” along the lines summarised in the first part of this document.

These notes provide detailed and most complete information about

- The Source agency (or agencies), including contact points and Website access
- The Trade system used
- Valuation practice of trade flows including any deviation from proposed standard practice
- Publication/dissemination release schedule. Possibility for OECD to have at the earliest possible moment
 - Provisional aggregates and detail
 - Final aggregates and detail
- Conversion routines in place for national and international purposes
- Calculation routines used to adjust and estimate data
- Treatment of confidential data
- Commodity coverage (any deviation from recommended practice is to be signalled) and specific remarks concerning any national change in practice over time to allow to determine the degree of consistency and comparability of commodity time series
- Treatment of re-imports and re-exports

37. A very important element of these notes is that they form the ground stock of individual country detail in an international context, hence enabling OECD to draw up country profiles and comparisons, including the “Tracking System” which has been developed internally to give at any moment in time a complete and fully detailed picture of data arrival, processing status and dissemination. This system has already proven its great usefulness in managing a complex monthly and yearly database.

4. Concluding remark

38. Data quality and Metadata can not be dissociated from each other. Both become increasingly important in a context of an ever increasing amount of data made available through different media. In a globalised world, international comparability and the respect of maintaining certain data quality and data measurement standards are a pre-condition for sound economic analysis. Also data integration is virtually impossible without clear standards and rules.

39. Uniform transmission standards for data and qualitative information, such as GESMES/CB, are offering new possibilities for a more effective information management. As in the past, OECD is very interested in learning from country delegates about their opinion and national experience in this field.

Annex:

OECD Metadata Questionnaire sent out for the first ITS Meeting in June 1999
--

COUNTRY (please complete): _____

Report prepared by: _____

I. GENERAL CHARACTERISTICS

Please describe in general, but precise, terms the main characteristics of your country's trade statistics system. You may wish to append more complete descriptions in an Annex.

- Trade system used and statistical territory
- Incoterms used for imports and exports (c.i.f./f.o.b.)
- Exceptions to international standard rules
- Statistical body (or bodies) responsible for collecting trade statistics
- **Only for EU countries:** Intrastat and national system

II. DATA AVAILABILITY

a) Yearly data

- Classifications used (for internal and external purposes, national classification), level of data detail available, length of time series
- Data collection characteristics
- Conversion routines in place (for national and international purposes)
- Availability of volume and value data (please provide details)
- Calculation routines to derive data (unit values, etc.)
- Treatment of re-imports and re-exports (inclusion or exclusion)
- Earliest availability of complete data
 - Provisional status (aggregates and detailed data)
 - Final status
- Treatment of confidential data and estimations
- To which agencies outside your country is trade data made available? What is the data dissemination policy?
- From which international agencies do you receive trade data?

Please feel free to attach any existing documents/descriptions in support (English or French)

b) Monthly data

- National classifications used
- Availability of data by SITC sections
- Calculation routines to derive data (unit values, indices, seasonal adjustments, etc.)
- Earliest availability of data

- Totals
- Detail: provisional status
- Detail: final status
- Treatment of confidential data and estimations
- Use of Internet as dissemination tool and possibilities of downloading
- To which agencies outside your country is trade data made available? What is the data dissemination policy?
- From which international agencies do you receive trade data?

III. THE DATABASE SYSTEM IN PLACE

- Please describe briefly which database platform(s) you use for both yearly and monthly statistics. Mainframe and/or client/server? A combination of both? Main software package(s) used.
- Media used for data collection
- Media used for data dissemination
- The inclusion and extent of metadata
 - Linked to data
 - Available on separate support
- The use of Internet
 - As data dissemination tool
 - As data collection tool

IV. DEVELOPMENTS AND PLANS

a) Data developments

- What main developments have taken place?
- Which classifications do you intend to discontinue/maintain/add?
- Which methodological studies or plans exist to improve data quality?
- How do you see the future of custom-based trade data?

b) System developments

- Modernisation plans
- Data dissemination policy
- Internet policy

V. USE OF OECD DATA

- Which international trade database do you use for data from other countries? Please specify
 - OECD
 - Eurostat
 - UN
 - IMF
- Do you use OECD trade data publications/CD-ROM's?
- Would you be interested in accessing OECD data via Internet? If so, what type of data and what level of detail would you like to access?