United Nations E/CN.3/2004/26

# Economic and Social Council

Distr.: General
17 December 2003

Original: English

**Statistical Commission**
**Thirty-fifth session**
2-5 March 2004
Item 6 (e) of the provisional agenda*
**Activities not classified by field: presentation of statistical data and metadata**

## Report of the Organisation for Economic Cooperation and Development on the presentation of statistical data and metadata

### Note by the Secretary-General

In accordance with a request of the Statistical Commission at its thirty-fourth session,** the Secretary-General has the honour to transmit the report of the Organisation for Economic Cooperation and Development on the presentation of statistical data and metadata. The report outlines proposals for the preparation of a manual containing guidelines and recommended best practice for the presentation of statistical data and metadata disseminated by national agencies and international organizations on various dissemination media. The necessity for such a manual, consolidating existing standards and developing new recommendations where necessary, stems from the need to further improve data quality (especially interpretability and coherence) and to minimize the burden of reporting data and metadata to international organizations.

Specific issues where Commission guidance is sought are outlined in section VII of the report.

* E/CN.3.2004/1

** See *Official Records of the Economic and Social Council, 2003, Supplement No. 4* (E/2003/24), chap. I.A.

03-66675 (E) 140104

# Report of the Organisation for Economic Cooperation and Development on the presentation of statistical data and metadata

## Contents


| | | *Paragraphs* | *Page* |
|---|---|---|---|
| I. | Introduction | 1–3 | 3 |
| II. | Background on the need for clearer articulation of data and meta-data presentation standards | 4–11 | 3 |
| | A. Improved data interpretability and coherence | 5–8 | 4 |
| | B. Minimization of reporting burden | 9–11 | 4 |
| III. | Relevance of existing presentation standards at national and international levels | 12–14 | 5 |
| IV. | Possible scope of presentation guidelines manual | 15–19 | 6 |
| V. | Links to related international initiatives | 20–27 | 8 |
| | A. Statistical data and meta-data exchange project | 21–22 | 8 |
| | B. Development of common lists of variables at the international level | 23–25 | 8 |
| | C. National accounts world wide exchange project | 26–27 | 9 |
| VI. | Draft timelines for the development of a guidelines manual | 28 | 10 |
| VII. | Issues for which Commission guidance is sought | 29 | 10 |
| Annex | | | |
| | Possible content of proposed manual on data and meta-data presentation | | 14 |

# I. Introduction

1. At its 2003 meeting, held in Geneva from 8 to 10 September 2003, the Committee for the Coordination of Statistical Activities (CCSA) discussed a document prepared by the Organisation for Economic Cooperation and Development (OECD) outlining draft proposals for the preparation of a manual outlining guidelines and recommended best practice for the presentation of statistical data and metadata disseminated by national agencies and international organizations on various dissemination media.

2. To facilitate discussion on issues related to data and meta-data presentation by CCSA in September 2003, OECD undertook preliminary work on data and meta-data presentation standards. The need for the articulation of such standards from the OECD context was initially prompted by the requirements to reduce the reporting burden of national providers of data to the OECD and to improve data quality. From that initial work, OECD concluded that the preparation of any presentation manual would require the active involvement of other international organizations as well as a number of national agencies currently active in the area. CCSA welcomed OECD proposals for the preparation of an international data and meta-data presentation manual and requested OECD to prepare a submission on the proposal for discussion at the thirty-fifth session of the Statistical Commission, highlighting the need for the articulation of a comprehensive set of international presentation guidelines and some initial thoughts on what such a manual could contain.

3. The present report:

(a) Provides background information, emphasizing the need for guidelines on data and meta-data presentation at both the national and international levels;

(b) Outlines the possible scope of any presentation manual;

(c) Provides links to other related initiatives currently under way at the international level;

(d) Provides draft timelines for the development of a guidelines manual, aiming at the preparation of a preliminary version for the 2005 Commission meeting;

(e) Indicates several areas where comment and direction from the Commission is sought.

# II. Background on the need for clearer articulation of data and meta-data presentation standards

4. There are two broad imperatives relevant at both the national and international levels that justify the need for the articulation of a comprehensive set of standards for the presentation of data and metadata. These concern the need to improve data quality and minimize the data-reporting burden in the provision of data and metadata to international organizations.

## A. Improved data interpretability and coherence

5. The first imperative relates to the need to improve the quality of statistics presented to users at both the national and international levels, in particular with respect to interpretability and coherence (within data sets, across data sets, over time and between countries). These are important dimensions of quality that are already included in one form or other in each of the quality frameworks developed by national agencies and international organizations (by Statistical Office of the European Communities (Eurostat), the International Monetary Fund (IMF), OECD and Statistics Canada etc.). However, beyond stating the case for improvement, such frameworks seldom go into much detail about how these quality dimensions would be implemented in the context of data and meta-data presentation.

6. Interpretability 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 possible limitations of the data largely determines the degree of interpretability. Interpretability is assisted by the presentation of metadata that is appropriate to the needs of a range of different users and uses of the data and is both well structured and readily accessible.

7. With respect to coherence, users are often confronted by three broad problems when comparing statistics compiled over time within the one agency and by agencies in different countries and by different international organizations:

(a) Conceptual differences arising from the use of different variable definitions, units and classifications;

(b) Operational differences flowing out of differences in data collection and processing practices countries;

(c) Different practices in the presentation of data. Such differences include: the type of presentation (absolute figures, indices, growth rates), the form of the data (raw, seasonally adjusted, trend-cycle); revision practices; presentation of sampling errors; base years; and methodological transparency afforded through ready access to appropriate statistical metadata.

8. The proposed manual would focus on the third issue and attempt to present the main presentation practices in the context of a framework, together with draft recommendations, guidelines and best practice for use by both international organizations and national agencies in their various forms of disseminated output.

## B. Minimization of reporting burden

9. The second imperative refers to the need to minimize the reporting burden of national agencies in their provision of data and metadata to international organizations. Discussions at recent international forums (such as the 2002 Conference of European Statisticians (CES) (OECD/IMF, 2002) and the 2003 meeting of the OECD High-Level Group for Statistics (OECD, 2003)) outlined the benefits of using a data-sharing model in the transfer of data and its associated metadata between national sources and the various international organizations. Such a model envisages the extraction of common data requirements by international organizations from data located on national agency web sites.

10. The evolution of new technologies over the last five years, particularly web-based technologies, has provided the technical possibility for the implementation of the data-sharing model. Prerequisites for such adoption involve not only the resolution of a number of technical information technology (IT) issues but also agreement between national agencies and international organizations on a number of data "content" issues, including:

(a) The identification of a set of common data requirements for key statistics (see paras. 23-25 below). A brief outline is provided in section V below of a number of related initiatives designed to further the coordinated collection of data and metadata by international organizations from national sources, and either directly or indirectly to contribute to the evolution of the data-sharing model referred to above;

(b) Agreement on key data presentation practices that would facilitate both the identification of identical series disseminated by national agencies and international organizations and the dissemination of consistent data, in particular by international organizations.

11. The development of the required guidelines in these areas is the responsibility of international organizations in cooperation with national agencies. Obviously, the implementation of the data-sharing model will only occur with the active participation of national agencies in whose databases the shared data and metadata would reside. Data-sharing implies a fundamental change in data dissemination with respect to coordination between international organizations and the role of national agencies in disseminating data to international organizations through their implementation of data and meta-data presentation guidelines that are designed not only to improve the interpretability and coherence of data but also to facilitate dissemination of data and ultimately minimize their reporting burden.

# III. Relevance of existing presentation standards at national and international levels

12. Where available, the recommendations and guidelines to be presented in the proposed manual would draw on the extensive range of existing international statistical standards that have been developed by international organizations in cooperation with national agencies (United Nations Statistics Division, 2002a; Eurostat, 2003b). The focus of these standards is primarily conceptual and also encompasses definitional issues, classifications, coverage and best practice for the collection of data. In the main, international standards are largely silent, or give only brief mention to data presentation issues[1]. Even the IMF special data dissemination standards (SDDS) (IMF, 2003b) give only general coverage to presentation standards, focusing on the provision of metadata to enhance interpretability and the adoption of good practice with respect to data revision. The Eurostat manual on short-term business statistics cites the need for greater harmonization of EU member State presentation of indices and growth rates, which it believes would assist Eurostat in checking that data disseminated by Eurostat are consistent with nationally released series (Eurostat, 2002). In the light of the absence of existing international standards, it may be necessary in some instances to develop new recommendations.

13. Almost all agencies at both the national and international levels have documents containing technical guidelines that touch on organizational standards with respect to data presentation, layout of tables, citation etc. for use by authors involved in the preparation of statistical publications. In the main, these tend to focus on the preparation of paper publications and coverage of data presentation issues is often very general. Ideally, key elements of the proposed international manual on data and meta-data presentation would be included in such documents and/or be linked to it.

14. One aim of the proposed manual would be to bring together in the one source a comprehensive set of guidelines for the presentation of statistical data and for a range of practices and processes that impact on data presentation, such as those outlined in paragraph 16 below. Because such practices may differ according to the dissemination medium used, it may be necessary to provide guidelines specific to a range of different dissemination media, such as online databases, data disseminated on web sites, paper publications, other electronic products and press releases. It is envisaged that future international statistical standards would either include presentation practices that are consistent with such an international manual or be linked to it.

# IV. Possible scope of presentation guidelines manual

15. It is envisaged that any set of presentation guidelines formulated at the international level would comprise a number of specific recommendations covering the two broad dimensions in which all data may be specified, namely:

(a) Types of data — absolute figures, indices, growth rates. Absolute figures may be either stock series, which are measures of activity at a point in time, or flow series, which comprise measures of activity to a date. Indices, growth rates and ratios are further transformations of absolute figures. Absolute figures may be presented either in terms of physical units (numbers, tonnes), or in value terms expressed at current or constant prices. The dissemination of absolute figures is common for statistics published at annual or less frequent intervals. Such data are also disseminated for many short-term indicators, such as monthly or quarterly data on motor vehicle registrations, construction permits etc. However, it is more common to disseminate short-term statistics in the form of indices or growth rates which more readily allow conclusions to be made on changes over time in economic phenomena. There are a number of different types of growth rates;

(b) Form of data — raw (original or non-seasonally adjusted series), working day adjusted, seasonally adjusted, trend-cycle.

16. In addition, there are a small number of key data presentation practices that also have a significant impact on data interpretability and in which the different approaches used by national and international agencies complicate the implementation of the data-sharing model and comparisons of national data etc. Such practices include:

(a) Data revision;

(b) Presentation of series breaks;

(c) Sampling errors;

(d) Use of common practices in rebasing indices;

(e) Presentation of related but not identical series/variables;

(f) Citation practices;

(g) Practices for maintaining classifications, including the historical records of classifications;

(h) Availability and presentation of metadata (including explicit templates for meta-data presentations).

17. It is envisaged that any future work on these issues would provide examples of current good practice and identify a small set of key recommendations. Such practices are outlined in a number of instances including IMF reports on the observance of standards and codes (IMF, 2003a), which summarize the extent to which countries observe internationally recognized standards, including those related to data dissemination. The manual would also include extensive bibliographic references to relevant work (including the relevant academic body of work). Consideration could also be given to developing adequate training materials based on the guidelines presented in the manual for use in workshops at the national level, especially for developing countries.

18. A factor complicating the development of international presentation standards has been the use of different terminology, particularly in reference to the various forms of growth rates used by different countries and in the same country for different series. The use of inconsistent labels frequently leads to misunderstanding. Problems associated with the inconsistent application of terminology also apply more generally to both data collection and the actual preparation of meta-data text containing definitions, outlining national practices with respect to data collection, manipulation etc. Such inconstancies severely limit the use of much existing metadata in comparing national data. To help overcome these problems, a number of international organizations have developed extensive glossaries containing definitions of key concepts and variables derived largely from existing international standards. The OECD glossary of statistical terms (OECD, 2002a and 2002b) is just one example of such glossaries, though others have been developed (see Eurostat, 2003a; and United Nations Statistics Division, 2002b). Also, as will be further discussed in section V below, another glossary (the meta-data common vocabulary is also being developed in the context of the statistical data and meta-data exchange (SDMX) project.

19. A final key issue related to data-presentation standards concerns the need for national agencies and international organizations to prepare adequate metadata describing their presentation practices and for this metadata to be readily accessible and understood by users with different degrees of statistical expertise. Ideally, such metadata should be expressed by different organizations within different countries on the basis of a common terminology. The main terminological problem areas would also be discussed in the proposed manual, which would include a glossary of concepts related to data presentation.

# V. Links to related international initiatives

20. There are several initiatives currently under way at the international level that would benefit either directly or indirectly from the development and adoption (by international organizations and national agencies) of a common set of data-presentation practices. The three projects described below have been brought together under recent initiatives to develop the data-sharing model. As stated in paragraph 10 above, the data-presentation standards to be included in the proposed manual are key elements in the implementation of this model.

## A. Statistical data and meta-data exchange project

21. The statistical data and meta-data exchange (SDMX) project[2] is a consortium of seven international bodies (Bank for International Settlements, European Central Bank, Eurostat, IMF, OECD, the United Nations Statistics Division and the World Bank), working to develop a set of common business practices in the field of statistical information that would allow more efficient processes for the exchange and sharing of data and metadata within the current scope of their collective activities. The aim of the project is to explore common e-standards and ongoing standardization activities that could allow those bodies to gain efficiencies and avoid duplication of effort in their own work and possibly for the work of others in the field of statistical information. (SDMX will be discussed in more detail under agenda item 6 (f)).

22. One of the four current SDMX projects is the development of a glossary (the meta-data common vocabulary (OECD/Eurostat, 2003)) as a tool to help ensure the consistency of metadata prepared by authors at the national and international levels, both with respect to content and the range of methodological issues covered by the metadata. The common vocabulary is designed to encompass the range of meta-data terms used in the different meta-data models that have been developed by national and international agencies. In the context of the SDMX project, particular care is being taken to ensure coverage by the common vocabulary of terms in the IMF/SDDS meta-data model, though it is also intended for use in meta-data models developed by other international organizations and national agencies.

## B. Development of common lists of variables at the international level

23. The most notable example of a common list of variables in operation at the international level is the questionnaire used by OECD, Eurostat, IMF, the World Bank, and the United Nations Statistics Division for the collection of annual national accounts data. The questionnaire outlines a set of common national accounts data requirements that have been identified by these international organizations, the primary purposes of which are to reduce the reporting load of national agencies and the dissemination of consistent data at the international level. The questionnaire comprises a very detailed set of national accounts variables that have been specified to meet the requirements of international agencies. These variables are identified in an extensive set of Excel spreadsheets by means of a common code and specific presentation format (e.g., in national currency at current price/constant prices). The questionnaire forms the basis of the national accounts world wide exchange project described below.

24. However, the annual national accounts questionnaire is by and large an exception with respect to the development of common lists of variables, though the approach used could be applied to other fields of statistics that require common ongoing collection by a number of international organizations, for both structural and short-term indicators. For short-term economic indicators, OECD and Eurostat have developed independent lists of variables for short-term economic indicators that could be used as starting points for the formulation of a common list of variables in a key area of reporting burden of particular concern to national agencies. These lists comprise:

(a) The OECD list of "target" indicators sought for inclusion in its monthly main economic indicators (MEI) database. The main purpose of this list is to provide a focus for OECD requests to member country agencies and other international organizations for MEI data and methodological information. Such focus is necessary to ensure the collection of a range of indicators common to as many member countries as possible. The list is revised at regular intervals as priorities change and new topics of interest to users emerge. No one OECD member country compiles all the indicators in the list;

(b) The list of variables specified in the European Commission short-term statistics regulation (European Commission, 1998), which specifies both the reference period and the form of data to be transmitted to Eurostat by European Union (EU) member States. A further list containing a subset of priority short-term economic indicator requirements for the European Statistical System, the Principal European Economic Indicators (PEEI), was set up in 2001. The list, which will be refined over time, also includes target release dates and other quality objectives. Eurostat will compile and release PEEIs based on member State contributions on a common dissemination platform accessible via the Euroindicators site, and will cover both the EU/Euro area and national indicators compiled according to EU standards.

25. There is a need for the identification of a common set of short-term economic indicator variables akin to the annual national accounts questionnaire by all relevant international organizations. This need was recognized at a June 2003 meeting of the OECD High-Level Group on Statistics, which called for international organizations to work together to develop such a list for short-term statistics. This list would include variable requirements and, ideally, the form in which such data should be presented in the context of a data-sharing model.

## C. National Accounts World Wide Exchange (NAWWE) project

26. A 2002 meeting of OECD national accounts experts proposed an experiment to test the implementation of the data-sharing model in the national accounts area among national agencies and OECD. The idea behind the project (OECD 2002c) is to implement a model in which data are not transferred across organizations but rather published on the World Wide Web in such a form that users can extract them by simply using the country and variable references. The idea is to start from the Excel tables already produced by national agencies for transmitting annual national accounts data to international organizations as described above.

27. Another objective of the project is to have the data collected by international organizations to be the data officially disseminated by national agencies. The two advantages of this model are that the burden of reporting to international organizations would be minimized, and data quality would be maximized for the international statistical community since the data they use are those officially disseminated and not specially compiled for and transmitted to international organizations.

# VI. Draft timelines for the development of a guidelines manual

28. OECD has undertaken some initial work on the preparation of a draft document containing an initial set of data and meta-data presentation guidelines and recommended good practice in key areas. Work will continue over the next 12 months and will incorporate CCSA suggestions on the issues to be covered by the manual and other input from CCSA members, together with guidance provided by the Statistical Commission on the issues outlined in section VII below. The proposed timeline for the future evolution of the manual would be:

| | *Critical Dates* |
|---|---|
| Preparation of a draft for initial exposure to OECD member countries for the 28-29 June 2004 OECD Short-term Economic Statistics Expert Group in Paris | 30 April 2004 |
| Further work and incorporation of Expert Group comments | |
| Submission to CCSA for discussion at its next meeting (New York, 6-8 September 2004) | 30 July 2004 |
| Incorporation of CCSA comments | |
| Submission to the Commission at its thirty-sixth session for comment (2005) | 17 December 2004 |
| Incorporation of Commission comments and preparation of a final version of the manual | 30 June 2005 |

# VII. Issues for which Commission guidance is sought

29. **Comments and guidance from the Commission is sought with regard to**:

- **The justification and need for the preparation of an international manual on data and meta-data presentation (as outlined in sect, II above);**
- **The intended content of the proposed manual and identification of data and meta-data presentation issues not currently covered either above or in the appendix to the present report (outlining the possible content of the proposed manual), particularly those relating to annual statistics and/or social statistics (as outlined in sect. III above);**
- **Proposals to obtain input from international organizations and national agencies in preparation of the manual (as outlined in sect. VI above);**

- **The identification of any relevant reference material on presentation issues prepared by either national agencies or international organizations which the Commission believes would be of use in the preparation of the manual**.

*Notes*

[1] There are exceptions; for example, the European Commission short-term statistics regulation (European Commission, 1998) specifies the reference period, type and form of data to be transmitted to Eurostat, e.g., absolute values, indices, non-seasonally adjusted, trend-cycle etc. However, the regulations do not tend to go into presentation in any detail and specify the provision of data to Eurostat through file transfer.

[2] More detailed information about SDMX and the four projects currently under way under the umbrella of SDMX are available on the SDMX web site at www.sdmx.org.

**References**

European Commission (1998). Council regulation No. 1165/98 of 19 May 1998 concerning short-term statistics. http://forum.europa.eu.int/irc/dsis/bmethods/info/data/new/1165-98en.pdf, accessed on 29 July 2003.

Eurostat (2002). Methodology of short-term business statistics: interpretation and guidelines. http://forum.europa.eu.int/irc/dsis/bmethods/info/data/new/embs/MM_Eurostat%202002-0507%20vo.2.pdf, accessed on 29 July 2003.

– (2003a). Eurostat concept and definitions database (CODED). http://forum.europa.eu.int/irc/dsis/coded/info/data/coded/en.htm, accessed on 29 July 2003.

– (2003b). Eurostat Classification Server (RAMON). www.oecd.org/findDocument/0,2350,en_2649_37423_1_1_1_1_37423,00.html accessed on 29 July 2003.

Finn, Brian (2003). Some proposals for standard terminologies relating to data presentation. Discussion paper presented at a meeting of the OECD Short-term Economic Statistics Expert Group, Paris, 26-27 June 2003. http://www.oecd.org/dataoecd/12/60/2789033.pdf, accessed on 7 August 2003.

IMF (2003a). Reports on the Observance of Standards and Codes. www.imf.org/external/np/rosc/rosc.asp, accessed on 29 July 2003.

– (2003b). Dissemination Standards Bulletin Board. http://dsbb.imf.org/Applications/web/dsbbhome/, accessed on 1 August 2003.

OECD (2002a). Glossary of Statistical Terms. http://cs3-hq.oecd.org/scripts/stats/glossary/index.htm, accessed on 29 July 2003.

– (2002b). Overview of OECD glossary of statistical terms. Discussion paper presented at an OECD national accounts expert meeting, Paris, October 2002. http://www.oecd.org/dataoecd/57/34/1946320.doc, accessed on 7 August 2003.

– (2002c). National accounts world wide exchange. Discussion paper presented at an OECD national accounts expert meeting, Paris, October 2002. http://www.oecd.org/dataoecd/10/62/1953780.doc, accessed on 7 August 2003.

– (2003). The new OECD statistical information system. Discussion paper presented at a meeting of the OECD High-Level Group on Statistics, Geneva, 13 June 2003. www.oecd.org/dataoecd/27/57/2955756.pdf, accessed on 29 July 2003.

OECD/IMF (2002). Progress report on new developments in data and meta-data collection for international organizations. Discussion paper presented at the Conference of European Statisticians Paris, 10-12 June 2002. http://www.unece.org/stats/documents/ces/2002/8.e.pdf, accessed on 7 August 2003.

OECD/Eurostat (2003). Meta-data common vocabulary project. www.sdmx.org/General/Projects.htm, accessed on 7 August 2003.

United Nations Statistics Division (2002a). Methodological publications in statistics. http://unstats.un.org/unsd/progwork/, accessed on 29 July 2003.

– (2002b). Definitions for a United Nations common database. http://unstats.un.org/unsd/cdb/cdb_help/cdb_quick_start.asp, accessed on 29 July 2003.

# Annex

## Possible content of proposed manual on data and meta-data presentation

Foreword

Contents

1. Introduction and overview

1.1 Need for presentation standards
1.2 Scope of presentation issues referred to in this Manual

2. Links to other initiatives

2.1 Statistical data and meta-data exchange (SDMX) project
2.2 Development of common lists of variables at the international level
2.3 National accounts world wide exchange (NAWWE) project

3. Data presentation terminology

3.1 Problem terms

4. Guidelines for the presentation of different types of data

4.1 Absolute figures
4.2 Indices
4.3 Growth rates

5. Guidelines for the presentation of different forms of data

5.1 Raw data
5.2 Working day adjustment
5.3 Seasonally adjusted data
5.4 Trend cycle

6. Guidelines on key presentation practices

6.1 Data revision
6.2 Presentation of series breaks
6.3 Sampling errors
6.4 Use of common practices in rebasing indices
6.5 Presentation of related but not identical series/variables
6.6 Citation practices
6.7 Practices for maintaining classifications including the historical records of classifications
6.8 Availability and presentation of metadata (including explicit templates for meta-data presentations)

7. Glossary
8. Bibliography