

PROPOSAL FOR THE DEVELOPMENT OF INTERNATIONAL GUIDELINES FOR THE
PRESENTATION OF STATISTICAL DATA AND METADATA¹

Report prepared by Denis Ward, Statistics Directorate, OECD

1. At the 36th session of the Administrative Committee on Co-ordination (ACC) Subcommittee on Statistical Activities held in New York on 17-19 September 2002 the Committee agreed to propose in 2004 in the multi-year programme of work of the Statistical Commission, the item "Guidelines on the presentation of statistical data, especially on the internet". The OECD offered to undertake preliminary work on this topic including the preparation of a draft document for discussion and input by the Committee for the Coordination of Statistical Activities (CCSA) at their meeting in September 2003. A draft document will be submitted at the end of August 2003 as a room document for this meeting. A preliminary outline of the draft guideline document is attached to this paper.

2. Over the last 12 months the OECD undertook preliminary work on data and metadata presentation standards, prompted by the need to reduce the reporting burden of national providers of data to the Organisation and to improve data quality. This work identified the need to broaden the scope of the topic proposed at last years ACC meeting, both with regard to the range of presentation issues covered and to remove the sole focus on the internet. Further work on this topic and the preparation of a comprehensive set of international guidelines will require the active involvement of several other members of the CCSA as well as a number of national agencies currently active in this area. To broaden the number of participating organisations and further facilitate the widening of the scope of the project, a revised version of this paper (together with a preliminary outline of a presentation guidelines document) could form the basis of a submission for discussion at the 2004 Statistical Commission highlighting the political need for the articulation of a comprehensive set of international presentation guidelines. The UNSC submission for next year could also contain timelines for the development of a guidelines document, aiming perhaps at the preparation of a preliminary version for the 2005 Commission meeting.

3. This brief CCSA paper:

- provides background information, emphasising the need at both the national and international levels for guidelines on data presentation;
- outlines the possible scope of any draft guidelines;

¹ Paper prepared by Denis Ward, Statistics Directorate, OECD (denis.ward@oecd.org)

- provides links to other related initiatives currently underway at the international level; and
- indicates several areas where input from CCSA members is particularly sought, either at or following the September 2003 meeting.

4. It should be emphasised that the draft guidelines to be presented to the CCSA is very much an early working document and should be regarded as such. The content represents the first set of output from one of the task forces of the OECD's Short-term Economic Statistics Expert Group (STESEG) which at this stage is still coming to grips with the issue of data presentation (Finn 2003), and are attempting to garner existing source information on issues related to the topic from both national and international agencies. Task force work will continue until the early part of 2004. The main purpose of the draft document to be presented for consideration by the CCSA is to help define the boundaries of the guidelines and identify issues that should be included. The current emphasis of the draft document is short-term economic statistics, though many of the presentation issues raised are also relevant for both structural statistics and social statistics. As further mentioned below (in Part (e) of this paper), input/suggestions from CCSA members would be particularly welcome on useful references, and identification of particular issues specifically relevant for structural and social statistics not currently covered.

5. Finally, the draft document will attempt to cover presentation issues across the whole range of data dissemination media (on-line databases, webpages, other electronic, paper publications and press releases). Initial work undertaken to date has shown that in a number of areas there are a range of standards that are relevant/appropriate for different dissemination media and that it would be useful to identify specific areas/issues of difference. Whilst in the end this could prove to be too ambitious and may need to be scaled back, it was thought beneficial not to focus initially on any one dissemination medium such as the internet.

a. Background on the need for clearer articulation of data presentation standards

6. 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 minimise the data reporting burden in the provision of data and metadata to international organisations.

Improved data interpretability and coherence

7. 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 datasets, across data-sets, over time and between countries). These are important dimensions of quality that are already imbedded in one form or other in all quality frameworks that have been developed by national agencies and international organisations (such as by Eurostat, the IMF, OECD, 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 metadata presentation.

8. 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 through the presentation of metadata which is appropriate to the needs of a range of different users and uses of the data which is both well structured and readily accessible.

9. 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 organisations:

- conceptual differences arising from the use of different variable definitions, units and classifications;
- operational differences flowing out of differences in data collection and processing practices countries; and
- different practices in the presentation of data. Such differences include: the type of presentation (absolute figures, indices, growth rates), form of the data (raw, seasonally adjusted, trend-cycle); revision practices; presentation of sampling errors; base years; methodological transparency afforded through ready access to appropriate statistical metadata.

10. This paper and the draft document focuses on the third issue and is an initial 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 organisations and national agencies in their various forms of disseminated output.

Minimisation of reporting burden

11. The second imperative refers to the need to minimise the reporting burden of national agencies in their provision of data and metadata to international organisations. 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 organisations. Such a model envisages the extraction of common data requirements by international organisations from data located on national agency websites.

12. 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 IT issues but also agreement between national agencies and international organisations on a number of data “content” issues including the:

- Identification of a set of common data requirements for key statistics (refer paras. 25-27 below). A brief outline is provided below in Part (d) of this paper on a number of related initiatives designed to further the co-ordinated collection of data and metadata by international organisations from national sources, and either directly or indirectly contribute to the evolution of the data sharing model referred to above.
- Agreement on key data presentation practices that would facilitate both the identification of identical series disseminated by national agencies and international organisations and the dissemination of consistent data, in particular, by international organisations.

13. The development of the required guidelines in these areas is the responsibility of international organisations in co-operation 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 will reside. Data sharing implies a fundamental change in data dissemination with respect to co-ordination between international organisations and the role of national agencies in disseminating data to international organisations through their implementation of data and metadata presentation guidelines that are designed not only to improve the interpretability and coherence of data but also to facilitate dissemination of data, and ultimately minimise their reporting burden.

b. Relevance of existing presentation at national and international levels

14. Where available, the recommendations and guidelines presented in the draft draw on the extensive range of existing international statistical standards that have been developed by international organisations in co-operation with national agencies ((UNSD 2002a) and (Eurostat 2003b)). The focus of these standards are primarily conceptual and also encompass 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 issues². Even the IMF Special Data Dissemination Standards (SDDS) (IMF 2003b) gives only general coverage to presentation standards, and issues 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 harmonisation 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, p. 135). In the light of the absence of existing international standards it may be necessary in some instances to develop new recommendations.

15. Almost all agencies at both the national and international levels have documents containing technical guidelines that touch on organisational 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,

² 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.

and coverage of data presentation issues is often very general. Ideally, key elements of any future international guidelines on data and metadata presentation would be imbedded in such documents and/or be linked to it.

16. The aim of the guidelines outlined in the draft document is therefore 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 para. 18 below. Because such practices may differ according to the dissemination medium used, guidelines will be provided for a range of dissemination media such as on-line databases, data disseminated on websites, in paper publications and other electronic products. Also, because of their increased importance in recent years, consideration is also given to the presentation of data in 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.

c. Possible scope of presentation guidelines

17. 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:

- 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.

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

18. In addition, there are a small number of key data presentation practices that also have a significant impact on data interpretability and where 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:

- data revision
- presentation of series breaks
- sampling errors
- use of common practices in re-basing indices
- presentation of related but not identical series/variables
- citation practices
- availability and presentation of metadata

19. 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 the IMF's Reports on the Observance of Standards and Codes (ROSC) (IMF 2003a) which summarise the extent to which countries observe internationally recognised standards including those related to data dissemination³.

20. 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 metadata text containing definitions, outlining national practices with respect to data collection, manipulation, etc. Such inconsistencies severely limit the use of much existing metadata in comparing national data. To help overcome these problems, a number of international organisations 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 OECD 2002b)) is but one example of such glossaries, though others have been developed by Eurostat (Eurostat 2003a) and UNSD (UNSD 2002b). Also, as will be further discussed in Part (d) below another glossary (the Metadata Common Vocabulary (MCV)) is also being developed in the context of the Statistical Data and Metadata Exchange (SDMX) project.

21. A final key issue related to data presentation standards concerns the need for national agencies and international organisations 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, this metadata should be expressed by different organisations within different countries on the basis of a common terminology. The main terminological problem areas, together with recommended improvements are outlined in the draft document. The Glossary at the back of the document contains a comprehensive set of definitions for concepts related to data presentation.

d. Links to related international initiatives

³ As at July 2003, ROSCs on data dissemination have been undertaken for the following OECD Member countries: Australia, Czech Republic, Hungary, Italy, Korea, Mexico, Sweden, Turkey and the United Kingdom. In the main, the good practice described in ROSCs relevant to the data dissemination issues described in the draft document were restricted to data revision practices.

22. There are several initiatives currently underway at the international level that would benefit either directly or indirectly from the development and adoption (by international organisations and national agencies) of a common set of data presentation practices. The three projects described in the draft document are brought together under recent initiatives to develop the data sharing model. As stated in para. 12 above, the data presentation standards described in the draft document are a key element of the implementation of this model.

Statistical Data and Metadata Exchange (SDMX) project

23. The Statistical Data and Metadata Exchange (SDMX) project⁴ is a consortium of seven international organizations (BIS, ECB, Eurostat, IMF, OECD, UNSD 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 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 standardisation activities that could allow them 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 another agenda item at this years CCSA meeting.

24. One of the four current SDMX projects is the development of a glossary (the Metadata Common Vocabulary (MCV)) (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 MCV is designed to encompass the range of metadata terms used in the different metadata models that have been developed by national and international agencies. In the context of the SDMX project, particular care is being taken to ensure MCV coverage of terms in the IMF SDDS metadata model, though it is also intended for use in metadata models developed by other international organisations and national agencies.

Development of common lists of variables at the international level

25. The most notable example of a common list of variables in operation at the international level is the is the questionnaire used by the OECD, Eurostat, IMF, World Bank, IMF and UNSD 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 organisations, 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 (NAWWE) project described below.

⁴ More detailed information about SDMX and the four projects currently underway under the umbrella of SDMX are available on the SDMX website, available from www.sdmx.org

26. 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 on-going collection by a number of international organisations, for both structural and short-term indicators. For short-term economic indicators, the 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:

- 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 organisations 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.
- 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 EU member states. A further list containing a subset of priority short-term economic indicator requirements for the European Statistical System (ESS), 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 EU/Euro area and national indicators compiled according to EU standards.

27. 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 organisations. This need was recognised at the June 2003 meeting of the OECD High Level Group on Statistics which called for international organisations 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.

National Accounts World Wide Exchange (NAWWE) project

28. The 2002 meeting of OECD National Accounts experts proposed an experiment to test the implementation of the data sharing model in the national accounts area amongst national agencies and the OECD. The idea behind the NAWWE project (OECD 2002c) 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 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 organisations described above.

29. Another objective of the NAWWE project is to have the data collected by international organisations to be the data officially disseminated by national agencies. The two advantages of

this model are that the burden of reporting to international organisations would be 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.

e. Future work and areas where CCSA input is particularly sought

30. Work on the draft document outlining data and metadata presentation good practice and recommendations will continue following the September 2003 CCSA meeting. This will incorporate CCSA comments and the work of the OECD STESEG task force over the next seven months. As mentioned in the introduction to this paper (in para. 2) a revised version of this paper (together with a preliminary outline of a presentation guidelines document) could form the basis of a submission for discussion at the 2004 Statistical Commission highlighting the political need for the articulation of a comprehensive set of international presentation guidelines. The UNSC submission for next year could also contain timelines for the development of a guidelines document, aiming perhaps at the preparation of a preliminary version for the 2005 Commission meeting.

31. Comments are therefore sought from CCSA members with regard to:

- any of the issues raised above in this paper, particularly with respect to the broadening of the topic and involvement of international and national agencies in future work on international data and metadata presentation guidelines;
- the draft presentation guidelines document to be disseminated to CCSA members as a room document at the end of August 2003;
- any specific data and metadata presentation issues not currently covered either above or in the draft document, particularly those relating to annual statistics and/or social statistics;
- identification of any relevant reference material on presentation issues prepared by either national agencies or international organisations which CCSA members believe would be of use in the preparation of future versions of the draft document;
- obtaining access to any existing organisational publication or database author manuals (if not confidential) that contain guidelines, etc., on any of the presentation issues outlined in this paper.

Statistics Directorate
OECD
Paris

8 August 2003

BIBLIOGRAPHY

European Commission, 1998, *Council Regulation (EC) No. 1165/98 of 19 May 1998 concerning short-term statistics*, European Commission, Brussels, available from <http://forum.europa.eu.int/irc/dsis/bmethods/info/data/new/1165-98en.pdf> [Accessed 29 July 2003]

Eurostat, 2002, *Methodology of Short-term Business Statistics: Interpretation and Guidelines*, Eurostat, Luxembourg, available from http://forum.europa.eu.int/irc/dsis/bmethods/info/data/new/embs/MM_Eurostat%202002-0507%20vo.2.pdf [Accessed 29 July 2003]

Eurostat, 2003a, Eurostat concept and definitions database (CODED), Eurostat, Luxembourg, available from <http://forum.europa.eu.int/irc/dsis/coded/info/data/coded/en.htm> [Accessed 29 July 2003]

Eurostat, 2003b, Eurostat Classification Server (RAMON), Eurostat, Luxembourg, available from www.oecd.org/findDocument/0,2350,en_2649_37423_11_1_1_37423,00.html [Accessed 29 July 2003]

Finn, Brian, 2003, “Some proposals for standard terminologies relating to data presentation”, discussion paper prepared for the OECD Short-term Economic Statistics Expert Group (STESEG), Paris, 26-27 June 2003, OECD, Paris, available from <http://www.oecd.org/dataoecd/12/60/2789033.pdf> [Accessed 7 August 2003]

IMF, 2003a, *Reports on the Observance of Standards and Codes (ROSCs)*, IMF, Washington DC, available from www.imf.org/external/np/rosc/rosc.asp [Accessed 29 July 2003]

IMF, 2003b, *Dissemination Standards Bulletin Board (DSBB)*, IMF, Washington DC, available from <http://dsbb.imf.org/Applications/web/dsbbhome/> [Accessed 1 August 2003]

OECD, 2002a, *Glossary of Statistical Terms*, OECD, Paris, available from <http://cs3-hq.oecd.org/scripts/stats/glossary/index.htm> [Accessed 29 July 2003]

OECD, 2002b, “Overview of OECD Glossary of Statistical Terms”, discussion paper presented at *OECD National Accounts Expert Meeting, Paris, October 2002*, OECD, available from <http://www.oecd.org/dataoecd/57/34/1946320.doc> [Accessed 7 August 2003]

OECD, 2002c, “National Accounts World Wide Exchange (NAWWE)”, discussion paper presented at *OECD National Accounts Expert Meeting, Paris, October 2002*, OECD, available from <http://www.oecd.org/dataoecd/10/62/1953780.doc> [Accessed 7 August 2003]

OECD/IMF, 2002, “Progress Report on New Developments in Data and Metadata Collection for International Organisations”, discussion paper presented at *Conference of European Statisticians*

(CES), Paris, 10-12 June 2002, UNECE, available from <http://www.unece.org/stats/documents/ces/2002/8.e.pdf> [Accessed 7 August 2003]

OECD, 2003, “The new OECD statistical information system”, Discussion paper at *Meeting of the OECD High Level Group on Statistics, Geneva, 13 June 2003*, OECD, Paris, available from www.oecd.org/dataoecd/27/57/2955756.pdf [Accessed 29 July 2003]

OECD/Eurostat, 2003, “Metadata Common Vocabulary Project”, presented on the SDMX website, IMF, available from www.sdmx.org/General/Projects.htm [Accessed 7 August 2003]

UNSD, 2002a, *Methodological Publications in Statistics*, UNSD, New York, available from <http://unstats.un.org/unsd/progwork/> [Accessed 29 July 2003]

UNSD, 2002b, *Definitions for United Nations Common Database*, UNSD, New York, available from http://unstats.un.org/unsd/cdb/cdb_help/cdb_quick_start.asp [Accessed 29 July 2003]

PRELIMINARY CONTENTS DRAFT DATA AND METADATA PRESENTATION GUIDELINES

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 Metadata 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

3.2 Non-problem areas

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 re-basing indices

6.5 Presentation of related but not identical series/variables

6.6 Citation practices

6.7 Availability and presentation of metadata

7. Glossary

8. Bibliography

ANNEXES

1. OECD main Economic Indicators (MEI) List of Target Indicators

2. Data Requirements Under the Short-term Statistics Regulation (EC) no. 1165/98, 19 May 1998 and Under the Regulation Concerning Labour Cost Index (EC) No. 2001/0166, 23 July 2001

3. European Commission Principal European Economic Indicators (PEEI) Periodicity and delays in Calendar days