



OECD SHORT-TERM ECONOMIC STATISTICS WORKING PARTY  
(STESWP)

**Release of the *OECD Data and Metadata Reporting and Presentation Handbook*.  
Where to from here?**

Paper prepared by Denis Ward  
Statistics Directorate, OECD

Submitted to the Working Party under item 5 of the draft agenda

Meeting:  
25-27 June 2007

Franqueville Room  
OECD Headquarters, La Muette, Paris  
Starting at 09.30 a.m. on the first day

## EXECUTIVE SUMMARY

This paper outlines a series of options for consideration by STESWP delegates covering future work on data and metadata reporting and presentation following the recent release of the OECD *Data and Metadata Reporting and Presentation Handbook*. The paper outlines a number of specific issues for written comment by delegates prior to the Working Party meeting in June 2007. These will also be discussed at the meeting

The five issues for written comment are outlined in Section D of the paper. These comprise:

- (i) Should the Handbook recommendations be used as a basis for comparison of current national and international practice in the reporting and presentation of short-term economic indicators (refer para. 16)?
- (ii) Do you endorse the approach outlined in paras. 18-19 of this paper identifying 11 main themes from the extensive Handbook recommendations?
- (iii) Do you endorse the top-down approach for reporting on the national application of these themes from a corporate perspective as well as the extent of their application across the many short-term economic indicator domains (refer paras. 20-21)?
- (iv) Do you have any comments on the draft issues identified for each theme, and priorities outlined in columns (3) and (4) of the table following para. 19 of this paper?
- (v) Finally, do you endorse the processes for the provision of the required summaries and national agencies and international organizations and their subsequent use outlined in paras. 22 and 23 of this paper?

## A. INTRODUCTION

1. The primary objective of this paper is to present the Short-term Economic Statistics Working Party (STESWP) with a number of practical options for discussion at the June 2007 Working Party meeting for future work on issues relating to data and metadata reporting and presentation of short-term economic statistics. This follows the recent publication of the OECD *Data and Metadata Reporting and Presentation Handbook*. The options for future work are outlined below in Section D.

2. The OECD Handbook was developed under the auspices of STESWP over the period 2003-2006 and was formally endorsed by the heads of national statistical agencies in the OECD Committee on Statistics (CSTAT) in the second half of 2006. Draft versions of the Handbook were also presented to the United Nations Statistical Commission (UNSC) in 2005.

3. The internet version of the Handbook was published on the OECD website in late 2006 and is downloadable in its entirety free of charge<sup>1</sup> in pdf format. Paper editions of the Handbook in English and French were released by the OECD in June 2007, and it has also been translated into Japanese by the Japanese Statistics Bureau to facilitate its use by various agencies in the national administration.

---

<sup>1</sup> Available at <http://www.oecd.org/dataoecd/46/17/37671574.pdf>

4. The Handbook was the result of very extensive consultation with STESWP national delegates, initially from the task force established in 2002<sup>2</sup> to formulate recommendations on issues related to data presentation and seasonal adjustment, and subsequently from individual countries for the numerous drafts / manuscripts of the Handbook. Other international / supranational organizations (Eurostat, ECB, IMF and the World Bank) also provided extensive and sometimes very detailed comments and suggestions that were also incorporated into the final version of the Handbook.

## **B. BRIEF OVERVIEW OF HANDBOOK AIMS AND CONTENTS<sup>3</sup>**

5. The main stated aim of the OECD *Data and Metadata reporting and Presentation Handbook* (refer para. 1) was to provide a single comprehensive reference set of international guidelines and recommendations for the reporting and presentation of statistical data and metadata, covering both economic and social statistics. The Handbook was designed for use by statistical agencies in both developed and developing countries and by international organizations.

6. The need for such standards flows from greater economic integration between the major trading blocs around the world resulting in further user demand for improved comparability of statistics between countries. The major causes of differences in statistics arising from the use of different concepts, units and classifications as well as differences arising from varying collection and processing (transformation practices). Even where differences stemming from these aspects are minimal, statistics disseminated by different countries and international organizations may actually appear different because of the application of dissimilar presentation and data transformation processes. An example of the impact of the latter is highlighted in the discussion on the development of guidelines on seasonal adjustment that will be covered in item 8 of the 2007 STESWP agenda.

7. The benefits of providing such international standards in a single reference source were that it would among other things, assist in:

- promoting the consistency of methodologies through the standardisation of format, terminology and dissemination of data and metadata used by international and national agencies;
- promoting the comparability of statistical data;
- promoting the use of recommended practice in the presentation of statistics and improve the transparency of such practices to users;

---

<sup>2</sup> The Presentation and Seasonal adjustment Task Force was established at the inaugural meeting of STESWP in 2002 to formulate recommendations on standard presentation terminology; and the presentation of growth rates, revisions, sampling errors, seasonally adjusted data, etc. The various papers prepared by the Task Force that were presented at the STESWP meeting in 2003 were later incorporated into the OECD *Data and Metadata Reporting and Presentation Handbook*.

<sup>3</sup> For a more detailed outline of the contents of the Handbook refer Section 1.2, pp. 13-16

- minimising the reporting burden of providing data and metadata to international organisations; and
- identifying areas where additional guidelines for the reporting and presentation of data and metadata need to be prepared in the future.

8. All of the issues covered in the Handbook outlined in Section B above are accompanied by a series of guidelines and recommended practices that have either been drawn from existing guidelines or that have been formulated by STESWP.

9. The Handbook covers a very broad range of data reporting and presentation issues that are relevant to both economic and social statistics, structural and short-term indicators. It provides a number of specific recommendations covering different types of data (e.g. original data or original data that has been transformed for presentation as indices, growth rates, ratios, rates, percentages, etc) as well as the different forms of time series data (e.g. raw series, working day adjusted, seasonally adjusted, trend-cycle estimates, etc.).

10. Section 6 of the Handbook emphasizes the need for national agencies and international organisations to prepare and disseminate adequate metadata describing concepts, collection and processing processes and data reporting and presentation practices. This Section of the Handbook provides extensive metadata presentation and dissemination guidelines that draw heavily on current and previous work of a number of international initiatives, in particular, that of the UNECE – Eurostat – OECD Work Session on Statistical Metadata (METIS). Section 6 focuses on recommendations on where and how metadata should be disseminated. Other Sections of the Handbook also touch on specific metadata issues such as: those required for micro-data; describing methodology used in seasonal adjustment and the provision of information on data revisions; series breaks; and on sampling and non-sampling errors.

11. The Handbook also emphasizes importance of the adoption of a common set of terminology that describe some of the issues outlined above and presents definitions throughout the body of the text, many of which are also provided in a Glossary at the end of the Handbook.

12. Finally, the Handbook provides recommendations on a small number of key data reporting practices that also have a significant impact on data interpretability, and where different approaches currently used by national statistical agencies and international agencies complicate comparisons of national statistics. These comprise:

- data revision presentation practices;
- reporting practices for the presentation of breaks in time series;
- practices for the reporting of sampling and non-sampling errors;
- base years in the presentation of indices;
- data and metadata citation practices;
- practices for the reporting of administrative data.

13. The Handbook recommendations for these issues are consistent with sound governance in statistics encompassed, either implicitly or explicitly, in the UN *Fundamental Principles of Official Statistics* and the UN *Handbook on the Operation and Organisation of a Statistical Agency*.

## C. FUTURE CHANGES TO THE CONTENT OF THE HANDBOOK

14. It has always been the intention throughout the development of the Handbook that its' content would not remain static and that it would evolve over time to incorporate new issues as they emerge or where existing guidelines would be modified to meet the needs of new standards that will be required in future. Such changes could arise through the use of new technologies and platforms for the dissemination of statistics to users. Examples of such platforms will be presented in item 4 of the 2007 STESWP meeting agenda<sup>4</sup> where recent developments and innovations will be demonstrated in the presentation of short-term indicators and the analytical tools / text aimed at broadening the user community. Proposals received so far for possible new topics for inclusion in the Handbook comprise:

- linking time series – refer STESWP 2007 agenda item 7 (Implementation of revised classifications / backcasting);
- graphical presentation of statistics;
- problems related to data confidentiality;
- presentation of chained indices and chain linking.

15. These and other possible future topics would be incorporated into the internet version of the Handbook. It is also intended to set up a more dynamic web version of the Handbook on the OECD website that would feature additional live links to specific examples of recommended presentation practices at the national and international levels. Resource constraints have prevented this till now.

## D. WHERE TO FROM HERE?

16. Now that the Handbook has been published the question now arises as to what would STESWP like to do with the various recommendations?

- *Option 1* would be to take a passive approach and do nothing. The Handbook would primarily serve as a reference document for possible adoption of the its recommendations by national and international statistical agencies alike.
- *Option 2* would be to take a more active approach and use the recommendations as a basis for comparison of current national and international practice in the reporting and presentation of short-term economic indicators. This approach is perhaps more in line with the mandate of the Working Party to improve the quality of short-term economic statistics which it “would monitor by regular studies [of the] quality and comparability of data in all domains of short-term economic statistics [and] encourage the methodological transparency, and timely availability of statistical series ....”<sup>5</sup>.

### How could this be done?

---

<sup>4</sup> Agenda item 4: Recent developments / innovations in the presentation and dissemination of short-term economic indicators

<sup>5</sup> OECD Short-term Economic Statistics Working Party mandate, page 433 of the *OECD Directory of Bodies: 2007*

17. If STESWP were to accept Option 2 as a basis for future work questions arise as to:

- which of the extensive range of recommendations in the Handbook should be given a higher priority for the review of current national / international practices; and
- which short-term economic statistics domains would one compare?

***Priorities for recommendations***

18. The recommendations included in the Handbook are summarized in Section 1.3.3. These are also provided in the Annex to this paper. In the preparation of the Handbook the issues covered were deemed of sufficient importance to Working Party members to warrant their inclusion, though obviously, not all of the issues covered are of equal importance.

19. Having said this the question now arises as to which would have greater impact on data comparability and transparency? Rather than determining such priorities on the basis of each discrete recommendation (remembering that there are 14 pages of them) it would be more useful to synthesize the recommendations into a small number of broad themes as attempted in the following table where 11 topics have been identified (in col. (2)). If STESWP believes that work on all themes would be too onerous a task over the next 12 months then relative priorities would need to be assigned. A draft set of priorities is provided in col. (4) for Working Party consideration.

**Broad Handbook themes, draft outline of issues to be considered and possible priority for future work to be considered by STESW delegates**

| Handbook Section No. | Summary of Handbook recommendation  | Issue(s) to be considered   | Possible priority for future STESWP work* |
|----------------------|---|---|---|
| (1)                  | (2)   | (3)   | (4)                                       |
| 2.2.1                | 1. Preparation of national / international publishing manuals and their publication on the internet   | <ul style="list-style-type: none"> <li>Does the organization have a single document providing guidance to internal authors on publication standards?</li> <li>Is this document made available to the public? If not, could it be?</li> </ul>  | L   |
| 3.5                  | 2. Recommended practices to ensure the use of consistent terminology and definitions across the organisation  | <ul style="list-style-type: none"> <li>Does the organization have a corporate strategy to ensure the adoption of consistent terminology across the different statistical domains within the agency?</li> <li>What corporate tools are used to standardize terminology across the organization, e.g. corporate glossaries, thesauri?</li> </ul>  | M   |
| 4                    | 3. Guidelines for the reporting of different types of statistical data – original data, indices, growth rates, ratios, proportions, percentages and rates.                                | <ul style="list-style-type: none"> <li>Does the organization mandate how the different forms of growth rate are presented in various publications, e.g. through the use of standard terminology and the consistent use of an unambiguous set of expressions for use in analytical text to describe annual or infra-annual changes?</li> </ul>   | H   |
| 5.6                  | 4. Recommended practices for the reporting of different forms of time series data – forms of time series to be presented – working day adjusted, seasonally adjusted, trend-cycle series. | <ul style="list-style-type: none"> <li>Does the organization have a standard set of practices to be applied across statistical domains for the transformation of original series, e.g. use of standard adjustment software (X-13, TRAMO, etc), consistent application of various options available within software used?</li> <li>Does the organization mandate the type and extent of information provided to users describing the transformation of the original series?</li> </ul> | H   |
| 6.3                  | 5. Recommendations for providing access to metadata, adoption of common metadata terminology and the use of a common metadata items   | <ul style="list-style-type: none"> <li>Does the organization have a corporate policy or set of strategies mandating the need for statistics disseminated to be accompanied by metadata?</li> <li>What corporate tools are used to facilitate the provision and dissemination of metadata, e.g. standard set of metadata items / terminology for use across all statistical domains, corporate metadata repository?</li> </ul>   | H   |
| 7.1.3                | 6. Recommended transparent practices for data revision  | <ul style="list-style-type: none"> <li>Does the organization have a corporate policy on data revision that outlines issues such as the provision of a revisions timetable for each domain, backcasting, outline of reasons for revision, information to users on the size of a likely revision, revisions analysis?</li> </ul>  | H   |

| Handbook Section No. | Summary of Handbook recommendation   | Issue(s) to be considered  | Possible priority for future STESWP work* |
|----------------------|--|--|---|
| (1)                  | (2)  | (3)  | (4)                                       |
| 7.2.3                | 7. Recommended practices on the presentation of series breaks                                    | <ul style="list-style-type: none"> <li>Does the organization have a set of standard practices on the clear identification of series breaks in tables and graphs, provision of an outline of the reasons for series breaks, and backcasting?</li> </ul>   | M   |
| 7.3                  | 8. Recommended practices for the presentation of information on sampling and non-sampling errors | <p><i>Sampling errors</i></p> <ul style="list-style-type: none"> <li>Does the organization mandate how sampling errors are reported in publications produced across the agency, method used for their calculation and their significance for data use?</li> </ul> <p><i>Non-sampling errors</i></p> <ul style="list-style-type: none"> <li>Does the organization mandate how the different sources of non-sampling errors are reported in publications, the provision of quantitative measures of such errors where possible, and their impact / limitations they impose on data use?</li> </ul> | M   |
| 7.4.5                | 9. Recommended practices for rebasing  | <ul style="list-style-type: none"> <li>Does the organization have a corporate policy outlining the frequency for rebasing series, the classification level at which index numbers are rebased, and description of the approach used (direct/indirect, weights used)?</li> </ul>  | L   |
| 7.5.4                | 10. Recommended practices for citation   | <ul style="list-style-type: none"> <li>Does the organization have a clear set of guidelines for use by authors across the organization on citation practices to be used for both datasets and references / text in other disseminated outputs?</li> </ul>  | L   |
| 7.6.2                | 11. Recommended practices for the presentation and reporting of administrative data              | <ul style="list-style-type: none"> <li>Does the organization have a set of corporate practices outlining key statistical issues when administrative data are used such as name of source agency, description of original purpose data were collected by source agency, outline of strengths and weaknesses of the data, departures from international standards, etc?</li> </ul>   | M   |

\* H: high; M: medium; L: low



### ***Which short-term economic statistics domains?***

20. One of the issues discussed back in 2002 when STESWP was established was whether or not it was feasible for one Working Party to cover the very wide range of indicators embodied by the notion of “short-term economic statistics”. For example, compendium publications such as the OECD’s *Main Economic Indicators* cover indicators including national accounts, production, composite leading indicators, construction, retail trade, prices, employment / unemployment, interest rates, share prices, exchange rates, merchandise trade, balance of payments. With the exception of work on the development of the index of services production manual most STESWP work to date has been of a cross-domain nature covering broad issues such as data presentation, seasonal adjustment, revisions, etc.

21. One way of proceeding would be to review all short-term economic indicators on a domain by domain basis. Given the extensive number of such domains, past OECD experience<sup>6</sup> has shown that this approach could be a very resource and time consuming process. An alternative approach would be to take a more macro top-down approach and to identify a small number of key issues for each theme that each country and international organization would report on from a corporate perspective, such as :

- the existence of a formal (or written) corporate policy for each priority topic; and also
- the extent of its application / implementation across the various short-term indicator domains, together with any other salient information.

A draft set of possible key issues is provided in col. (3) in the above table.

### ***Who would this and by when?***

22. STESWP is asked to consider the possibility of each country and international organization (including the OECD) compiling and forwarding the summary information required for each (perhaps only high) priority theme outlined in the above table to the Secretariat by the end of December 2007.

### ***Would the information be published?***

---

<sup>6</sup> The OECD faced a similar problem several years ago when it began the publication of a series of MEI Methodological Analysis publications that went beyond the mere presentation of descriptive metadata that used to be published in the MEI *Sources and Definitions* paper publications and which is now continuously updated and disseminated on-line at <http://stats.oecd.org/mei/>. Because of the extent of indicators published in MEI it was decided to prepare separate publications for each indicator domain and these were prepared for industry, retail, construction, CPI, PPI and wage related statistics. Each publication (which are available at [http://www.oecd.org/document/4/0,2340,en\\_2649\\_33715\\_2367940\\_1\\_1\\_1\\_1,00.html](http://www.oecd.org/document/4/0,2340,en_2649_33715_2367940_1_1_1_1,00.html)) contains comparisons of then current national practice on key issues (such as main concept, classifications used, coverage, data source, etc) against the relevant international standard in a series of summary comparison tables. The main issue with respect to future STESWP work involving the Handbook recommendations is that the preparation of the Methodological Analysis publication even for the limited number of domains covered was a very resource consuming exercise.

23. The Secretariat would then assemble the information provided into a single document which would be forwarded to STESWP members for final comment by written procedure by the end of March 2008. It would then be referred to CSTAT, together with an outline of broad issues for consideration and possible discussion at the Committee's June 2008 meeting. The final document would then be published by the OECD.

## **Summary of recommendations presented in the *OECD Data and Metadata Reporting and Presentation Handbook***

### **2.1.2 Recommended practices to ensure consistency in presentation and reporting practices included in future international statistical guidelines and recommendations**

It is recommended that the authors of future international statistical guidelines and recommendations for various statistical domains take a more modular approach in the preparation of those standards by using, as required, existing international recommendations in key areas such as the reporting and presentation practices outlined in this Handbook and terminology and definitions presented international glossaries such as the *Metadata Common Vocabulary (MCV)* described in Section 6.3.4. Such use may take the form of either the direct incorporation of text within future guidelines or by reference / link to the recommendations in this Handbook, etc.

#### **2.2.1 National and international publishing manuals and guidelines**

Recommended practices for publication manuals:

- In the interests ensuring consistency in the presentation of data disseminated across different paper publications and electronic media within an organisation, it is recommended that each statistical agency compile a publishing manual or set of guidelines providing guidance to author areas within the organisation.
- It is also recommended that statistical agencies place such publication manuals and guidelines in the public domain by locating them on their websites. This would give greater transparency and external scrutiny to internal practices and provide a means of disseminating recommended practices to organisations in other parts of the national statistical system.

### **3.5 Recommended practices to ensure the use of consistent terminology and definitions**

The recommended practices outlined below are primarily designed to minimise the current common practice of different author areas within national and international organisations developing their own, often inconsistent, sets of terms and related definitions. Even where differences in terminology are appropriate there is still a need to provide users with information about the context / reasons for such differences.

Agencies should therefore:

- Establish a terminology management strategy and associated structures appropriate to their needs, requirements and resource capacity to reduce the use of inconsistent terminology and related concepts by the different author areas in the organisation which in turn are applied in various questionnaires and disseminated output.

There are a range of options available to achieve this objective. These include the creation of a corporate thesaurus which can be readily accessed by different parts of the organisation and in which existing inconsistent definitions for the same term are confronted and differences either eliminated or explained. Where possible, definitions applied at the national level

should align with international definitions, modified as appropriate to provide further elaboration and / or to meet national circumstances.

- Irrespective of the tool(s) adopted (glossaries, thesauri), there is still the need for senior management within an organisation to ensure that appropriate practices and principles involving the use consistent terminology are developed and adopted across the organisation.
- Ensure that any thesaurus developed by an agency contains the minimum structures outlined in existing international guidelines such as those formulated by the ISO, OMG and W3C.

Corporate glossaries should contain the following in order to facilitate their interoperability: concept label; definition; detailed source information about where the definition was derived; related terms; and context field providing additional information or highlighting how a definition is used within one statistical domain or geographical context. Finally, the glossary should also be translated into various languages used by the institution in question.

- Provide appropriate cross references or links in domain specific glossaries, on Internet or intranet sites to general glossary databases that have been developed either at the national or international levels, the purpose being to make existing standard terms and definitions more readily available.
- Avoid attaching precisely the same label or title for different definitions (for further discussion refer Section 6.3.5).

#### **4. Guidelines for the reporting of different types of statistical data**

##### ***4.2.1 For original data***

- The availability of original data affords maximum flexibility for users interested in undertaking further analysis or transformation beyond how it is presented in the data source on hand. Therefore, users should have access to at least some of the key aggregates of original data where they exist for a particular series, either directly in the publication or through the provision of references or hyperlinks. The provisions under which access to detailed original, confidentialised data is given would be governed by the organisation's dissemination policy.
- In situations where original data are known to have significant non-sampling errors, appropriate metadata should be provided to facilitate appropriate use of the data. In such instances consideration should also be given to the presentation and use of other types of data such as percentages in any analyses of main features.
- Original data are frequently transformed into growth rates or indices, for example, in a press release to facilitate interpretation and understanding. Ideally, original data should be presented in addition to the transformed series. However, this may not be possible in some dissemination media due to space considerations and in this situation it would be sufficient to provide a clear indication of their availability and how they could be accessed.

##### ***4.3.2 For indices***

The Statistics Canada Policy on Informing Users of Data Quality and Methodology (Statistics Canada 2000, Section E.3) states that the provision of an adequate description of characteristics and methodologies specific to indices is as important to users as quality assessments of the data. Statistics Canada therefore recommends the provision of the following metadata:

- precise definitions of the underlying economic concepts the indices are intended to measure. Specific mention should be given to any limitations in the use or application of the index, citing the example of deflation of macro-economic aggregates; and
- descriptions of the methodologies used in the compilation of the index, with particular reference to the:
  - index calculation methods entailing the choice of index formula (e.g. Laspeyres, Paasche, Fisher) and the strategy for constructing the index series (i.e. as either fixed base or chain indices);
  - weighting system used, weight revision practices and frequency of weight revision;
  - computation at various aggregation levels;
  - selection of base year;
  - frequency of re-basing;
  - procedures for linking indices;
  - treatment of changes in the composition of commodities in the market as well as changes in quality.

The methodologies applied should be compared with underlying index concepts and the impact of departures described.

Finally, as much of the above information is of specific interest to specialised users, consideration should be given to having differing levels of information targeted to different kinds of users. The guidelines for the reporting and dissemination of metadata provided in Section 6 emphasize the desirability of structuring metadata appropriately for users with differing degrees of expertise and need.

#### ***4.4.3 For growth rates***

The two key recommendations with respect to growth rate terminology entail the need for statistical agencies to:

- minimise the risk of user misunderstanding of the growth rates being applied through the consistent use of growth rate terminology and the provision of definitions such as those provided in Section 4.4.2 in accompanying metadata; and
- develop a consistent (and unambiguous) set of expressions for use in analytical text to describe changes in annual and infra-annual growth rates and to apply those expressions consistently both within the same press releases and in other press releases disseminated by the same organisation.

Recommendations for the presentation of growth rates in the context of different forms of data used in time series analyses are summarised in Section 4.6.

#### **4.5.2 For ratios, proportions, percentages and rates**

The main issues for the presentation of and ratios centre around the provision of appropriate methodological information (metadata) describing both the actual rates / ratios and, the component series used in their derivation. Precision is required in the provision of information for the user about the time period referred to, the nature of the population being described and the type of occurrence being measured (Palmore and Gardner 1994). More specifically:

- the term “rate”, “percentage”, or “ratio” should be included in the actual label, e.g. maternal mortality rate, crude death rate, etc., to ensure user understanding that the original data has been transformed;
- for rates where the total population is expressed as unity the unit of measurement used in the population should be included in the table heading for the rate, e.g. per 1 000 live births;
- the series labels in the components that make up the rate or ratio should be based where possible on existing international terminology. These are generally outlined in the international guidelines and recommendations for the relevant statistical domain (refer UNSD 2002a) or glossary databases disseminated by international organisations such as the OECD’s *Glossary of Statistical Terms* (OECD 2002a); the Eurostat *Concepts and Definitions Database* (CODED) (Eurostat 2006a), or UNSD’s *Definitions for United Nations Common Database* (UNSD 2002b). Departures from international concepts should be documented in the metadata accompanying the rate / ratio;
- detailed information about the source(s) of the component series used in the derivation of the rate or ratio should be provided. Minimum information comprises: type of data source (administrative, household survey or census, business survey or census), reference period, full official title of the series, full name of the source agency or institution;
- users should have access to the original data used in the derivation of the rate / ratio. This could be included in the body of the publication where the rates or ratios are disseminated (e.g. as annex tables) or through the provision of sufficient reference information or hyperlinks that will enable users to access the original data;
- users should be provided with information on methodologies used in the compilation of the component series used in the derivation of the rate / ratio. The quality (in particular, comparability both over time and between countries) of the rate or ratio is only as good as the quality of the series used in its calculation, and appropriate metadata is therefore essential to enable users to form an understanding of quality and relevance of the rate /ratio for a particular need or purpose.

The final issue involves the need to standardise extraneous variables used in the compilation of ratios, proportions, rates, etc. As mentioned in Section 4.5.1, such standardisation is necessary to enable the comparison of ratios, etc. between countries, regions, etc., and in some instances over time in the same geographic area.

### **5.6 Recommended practices for the reporting of different forms of time series data**

#### ***Forms of time series data to be presented***

*Recommendation 1:* When seasonality is present and can be identified, sub-annual indicators should be made available in seasonally adjusted form. The level of detail of indicators to be

adjusted should be chosen taking into account user demand and cost-effectiveness criteria. The adjustment should be applied appropriately using the method chosen as a standard by the agency. The method used should be explicitly mentioned in metadata accompanying the series.

*Recommendation 2:* When applicable, the focus of press releases (or similar releases to the general public) concerning the main sub-annual indicators should be on their appropriately seasonally adjusted version. Users should also be given access to the original (or raw) series, either in the publication (if space permits) or by reference to it.

Where there is a user demand, the agency may also disseminate intermediate components of the seasonal adjustment process (e.g. series adjusted for calendar effects) and / or trend-cycle estimates but it should be clearly indicated that the focus is on the seasonally adjusted estimate when short-term variation is of interest.

### ***Analytical transformations***

*Recommendation 3:* Press releases presenting seasonally adjusted flow series should at the minimum provide period-to-period growth rates for the latest period and, if space permits, period-to-period (e.g. month-on-previous-month, quarter-on-previous-quarter) change in levels.

*Recommendation 4:* For month-on-previous-month and quarter-on-previous-quarter rates of change, seasonally adjusted data is the best way of presenting information about a time series (trend-cycle and irregular movements) and for presenting short-term developments, even if the irregular component is relatively large. To deal with irregular movements that blur the trend the rate of change bases on two or three months' (or quarters) worth of values can be utilised.

*Recommendation 5:* For rate of change with respect to the same period of previous year the year-on-year changes should be applied to raw data and to data adjusted for calendar effects if the latter are available. Where necessary, special effects contained in the base period should be highlighted when presenting YoY (base effect).

*Recommendation 6:* Because of the risk of providing misleading signals, especially where series display significant volatility, the presentation of annualised level changes is not recommended, especially as the key headline series.

Where annualised changes are used, users should be provided with information regarding the possibility of misleading signals due to series volatility.

*Recommendation 7:* Annualized period-to-period growth rates are not recommended for the presentation of quarterly or monthly growth rates.

Preference should be given to the use of year-on-year growth rates.

### ***Information about seasonal adjustment to be provided to users***

*Recommendation 8:* Statistical agencies should disseminate a non-technical explanation of seasonal adjustment and its interpretation for the benefit of, and aimed at, the general public.

*Recommendation 9:* For the benefit of users requiring information about the appropriateness of the seasonal adjustment method applied, statistical agencies should provide a minimum amount

of information that would enable an assessment of the reliability of each seasonally adjusted series.

*Recommendation 10:* Statistical agencies should maintain metadata on seasonal adjustment of sufficient extent to enable outside users to seasonally adjust in a consistent way other series from the same statistical program that may not have been seasonally adjusted.

### **6.3 Recommended practices for the reporting and dissemination of metadata**

#### **6.3.1 The need for metadata**

All statistical agencies should:

- compile metadata required for users to understand the strengths and limitations of the statistics it describes; and
- keep their metadata up-to-date, incorporating the latest changes in definitions, classifications and methodology, etc.

#### **6.3.2 Access to metadata**

Key recommendations in this area include:

- ensuring that users have ready access to such metadata through its dissemination via a range of different media – paper publications, CD-ROMs, etc. However, it is important for all metadata to be available to users on the Internet, given that it provides the most accessible medium for obtaining the most up-to-date metadata. It is also recommended practice for metadata to be structured in such a way as to meet the needs of a range of users with different requirements and/or statistical expertise. This does not necessarily entail the physical presentation of different metadata to each group of users with different statistical expertise. However, a layered presentation of metadata is recommended, progressing from summary metadata to more detailed metadata. Each layer should use clear and precise text;
- dissemination of metadata free of charge on the Internet. There is strong support for the notion that metadata describing statistics has a high public good component and should therefore be disseminated free of charge on the Internet even if the actual economic and social statistics they describe and paper publication versions of the metadata are subject to an organisation's price regime;
- active linkage of metadata to the statistical tables and graphs they describe and vice versa;
- the availability of metadata not only in the national language but also, where resources permit, in a common language such as English;
- structuring the metadata for different statistical domains on the basis of some hierarchic classification. Consideration could be given to the adoption of the UNECE *Classification of International Statistical Activities* as the international standard for metadata. The September 2005 version is available at <http://www.unece.org/stats/documents/ece/ces/bur/2005/5.e.pdf>;
- provision of a local search engine based on free text search;

- implementing recommended practice for ensuring either the stability of URLs (Uniform Resource Locators) or providing links between the old and new URLs that will redirect users to the new address. This is a key issue given the importance of links between websites;
- providing the names of contact persons or email addresses where further information about concepts, definitions and statistical methodologies may be obtained. In some organisations the “contact” would be a generic corporate contact point or referral service for all client enquiries.

### ***6.3.3. Adoption of a set of common metadata items***

International agencies should work with national statistical agencies to develop a core set of non-domain specific metadata items (or prompt points) such as those being developed under the Statistical Data and Metadata Exchange (SDMX) initiative – refer para. 175.

### ***6.3.4 Adoption of a common set of terminology for metadata preparation***

Considerable resources are often expended by international organisations in verifying text, etc., to ensure that methodological descriptions are as consistent as possible between countries. A mechanism for achieving this would be the rigorous use of terminology imbedded in the various international statistical guidelines and recommendations. This could be facilitated by the use of glossaries published by international organisations which contain definitions derived from those standards, in particular, the *Metadata Common Vocabulary (MCV)* developed under the umbrella of the SDMX initiative

### ***6.3.5 Unambiguous presentation of similar but not identical statistical data***

Five broad recommendations of good practice in this area comprise (Friez 2003):

- Similar but different series should be given different titles to facilitate clear differentiation by users.
- International organisations that disseminate national data should always be aware of and clearly state in their metadata whether or not the precise series they disseminate that are derived from national sources are also disseminated in the country of origin, or compiled and / or transformed by national agencies specifically to meet the requirements of international organisations.
- International organisations should clearly describe in their metadata, specific details of any transformation of national data they perform to make the series more internationally comparable. Data transformed by international organisations should be clearly indicated as such, particularly, but not only where, published alongside different national series for the same statistical domain. The two sets of series must be clearly differentiated in the mind of the user.
- The precise name of the classification used in statistics disseminated by national agencies and international organisations (especially when transformed to an international classification to enhance international comparability) should always be clearly indicated (for instance, NACE Rev. 1, CITI, Main Industrial Grouping (MIG) or national classification) so that when the

same denomination is used in various classifications such as intermediate goods, the user clearly knows which classification has been used.

- When a field of activity is only partially covered (such as MIG-intermediate goods or MIG-consumer goods in the new orders indicators of the European Commission's Short-term Statistics Regulation), it should be clearly indicated for instance with an asterisk or a footnote (for example, in the Eurostat's short-term statistics new orders series, MIG-non durable goods (1) - (1) *Partial ; does not include NACE 151-155, 158, 159, 16, 19, 22, 364-366*).

### **7.1.3 Recommended practices for data revision**

#### ***7.1.3.1 Consultations with users elicit views about revisions practices***

Preliminary to elaborating a country's revisions policy, it is important to consult the main users of official statistics to identify needs and priorities specific to individual countries. Their views could be sought, for example, about their particular needs for timeliness of data, problems they experience because of revisions, and their priorities about balancing timeliness with accuracy and consistency.

#### ***7.1.3.2 Provision of a clear, short readily accessible summary statements to users of when to expect revisions and why revisions are undertaken***

#### ***7.1.3.3 Where possible establish a stable revisions cycle from year to year.***

#### ***7.1.3.4 Balance stability of a time series against the need to revise series because of the need introduce new methodologies, concepts, classifications, etc.***

#### ***7.1.3.5 Ensure that revisions are carried back several years to give consistent time series***

To maintain the serviceability of data following major revisions, data should be revised back as far as is reasonable based on a balancing of user needs, costs, and availability of source data. The revised time series should be released simultaneously with the revised current data or soon thereafter, preferably in easily accessible electronic format. The revised series should be of sufficient detail and not so aggregated that users are not able to detect the sources of the changes.

#### ***7.1.3.6 Provide appropriate documentation on revisions***

Such documentation should include:

- clear identification of preliminary (or provisional) data and revised data;
- provision of advance notice of major changes in concepts, definitions, and classification and in statistical methods;
- information on the sources of revision are explained when the revised series are released;
- information on breaks in series when consistent series cannot be constructed;
- information on the size of possible future revisions based on past history;

#### ***7.1.3.7 Users are reminded of the size of the likely revisions based on past history***

Users should be provided with information sufficient for them to make an informed judgment about the reliability and accuracy of preliminary or provisional statistics. The following two recommended practices for revision studies have been identified:

- periodic analyses of revisions investigate the sources of revision from earlier estimates and statistical measures of the revisions;
- the analyses are published for major aggregates to facilitate assessment of the reliability of the preliminary estimates.

#### ***7.1.3.8 Transparent and timely reporting of errors and mistakes***

#### **7.2.3 Recommended practices on the presentation of series breaks**

Recommended practice with regards to time series breaks entails:

- The compiling agency taking all possible steps to avoid and minimise changes to questionnaires, definitions and classifications used to collect and compile data. Methodologies should be developed to reduce the frequency of revisions.

However, there comes a time when the time series may be disrupted even when outdated classifications, concepts and questionnaires are maintained. In such instances a complete break in series may be preferred to series that continue to be collected on the basis of outmoded classifications and concepts that do not approximate reality. There is clearly a tradeoff between costs imposed by breaking a time series on one hand and the benefits from improving the relevance of the time series on the other (BEA 1993).

- Where significant breaks in a time series are unavoidable, users should be given warning well in advance of the implementation of the series break outlining the timing of implementation and a detailed explanation of the reason(s) for the change. “In advance” is taken to mean not just the time of implementation but sufficient time to enable users to implement modifications to their systems, programmes or databases and to seek further clarification if necessary. A common practice adopted by many statistical agencies is to issue a detailed discussion paper many months in advance of the change.
- Actual breaks in the series should be clearly identified in both the statistical table and any accompanying graphs. A variety of methods are commonly used by national agencies and international organisations to highlight in tables that a series break has actually occurred. These include the insertion of a line in the table at the break point, inclusion of a footnote or tabular presentation as an entirely new series. Whichever method is adopted, the main point is that the break is completely clear to users. Consideration will also need to be given to the identification of series breaks (together with appropriate explanatory information) in data disseminated electronically such as via on-line databases, etc.

The following information drawn directly from Eurostat guidelines should also be provided (Eurostat 2003c, p. 16):

- the reference period of the survey where the break occurred;
- whether or not the difference reported is one-off with limited implications for the time series and / or if the reported change led to harmonisation with any standards;

- a precise outline of the difference in concepts and methods of measurement before and after the series break;
- a description of the cause(s) of the difference, e.g. changes in classification, in statistical methodology, statistical population, methods of data transformation, concepts, administrative procedures with regard to statistical data from administrative sources;
- an assessment of the magnitude of the effect of the change, where possible, with a quantitative measure.

Links and references to more detailed information should also be provided.

- Points in line graphs should not be joined across discontinuities in data. The reason for the series break should be explained in a footnote accompanying the graph with appropriate links or references to more detailed explanations of the causes of the breaks.
- When methodological changes are introduced, an attempt should be made to revise the historical series as far back as data and available resources permit. Ideally, such backcasting should extend back 2-3 years to reflect the new methodology, etc.

### **7.3 Presentation of information on sampling and non-sampling errors**

#### **7.3.1 Sampling errors**

In the interests of data transparency, and to help ensure the appropriate use of data, statistics derived from all sample surveys should be accompanied by information on sampling errors. Such information should be provided for all dissemination media – online databases, websites, other electronic products, paper publications and press releases. It is also important for the information to be expressed in non-technical terms capable of being understood by the non-specialist user. The mode of presentation and the amount of detail provided should therefore meet the specific needs of particular categories of users (UNSD 1993, p. 176). The required information comprises the provision of the following information in accompanying or clearly linked technical notes outlining (OMB 2001, p. 3-8):

- Alerting users to the fact that data are derived from a random or non-random sample. If the latter then inference implications should be clearly stated.
- Sampling error should be identified as a source of error which should be explained and interpreted for data users through provision of a brief definition of sampling error. For example, strong warnings about the unreliability of data with high sampling error.
- Sampling errors must be presented in the context of total survey error. In this context users should be made aware of the fact that sampling error is just one, and often not the most significant, component of total error (UNSD 1993, p. 176, 7.1 (1)).
- If statistical tests are used in the report, the significance level at which statistical tests are conducted should be stated explicitly.
- Sampling errors for key estimates should be available to the user either in a table in the publication or linked on the Internet. Some form of notation should also be placed directly beside estimates with very high sampling (or non-sampling) error.

Sampling errors may be presented in one of a number of different forms, for example:

- as absolute values of the standard error (se);
- as relative values, standard error divided by the estimate (rse); or
- in the form of probability or confidence intervals.

The preferred use of either the absolute or relative forms depends on the nature of the estimate and readers are referred to the United Nations publication, *Sampling Errors in Household Surveys* (UNSD 1993, p.178) for a detailed evaluation of the different forms of presentation and several examples of recommended practice. The UN evaluation emphasizes the importance of ensuring that the chosen method is clearly and unambiguously described and presented with accompanying definitions and notation.

In order to ensure consistency in the dissemination of this information across the organisation in all published output subject to sampling error, some statistical agencies mandate a standard set of words to be included in all relevant publications.

Where space considerations preclude the inclusion of detailed information, either references or hyperlinks to more detailed technical reports or user manuals should be provided. Such information should enable specialist users to analyse detailed data or compile new tabulations and would therefore:

- identify the specific method used for calculating the sampling error;
- provide sampling error calculations (tabulations) for different types of estimates (e.g. levels, percents, ratios, movements, means and medians) for a number of variables and disaggregations. The aim is to provide a basis for extrapolation to statistics for which sampling errors have not been computed by the source agency (UNSD 1993, p. 180);
- contain evaluations of the procedures used for estimating sampling errors.

### **7.3.2 Non-sampling errors**

The focus of the recommended practice on non-sampling errors outlined below is not the methods by which national agencies minimise their impact but rather guidelines on the type of information on such errors to be reported with disseminated statistics:

- As for the reporting of information on sampling error for all sample surveys, all statistical output disseminated by national agencies and international organisations should be also accompanied by information on non-sampling errors. Such information should be accessible for statistics disseminated on all types of media – online databases, websites, other electronic products, paper publications and press releases. It is also important for the information to be expressed in non-technical terms capable of being understood by the non-specialist user. Such information should either accompany the data disseminated or be provided in clearly linked technical notes.
- Where possible, quantitative measures of non-sampling error should be provided. However, because of the difficulty in quantifying some non-sampling errors, agencies will need to

disseminate a mixture of quantitative and qualitative information that enables a non-technical user to clearly understand the strengths and limitations of the data. In particular, information on non-sampling errors should clearly convey to the user the fact that such errors, either individually or in total, may have a greater impact on the reliability of the data than sampling error and that the “ready” availability of quantitative measures of sampling error is not necessarily an indication of their relative significance.

- With respect to precisely what information on non-sampling errors that should be reported, the ideal recommendation is for national agencies and international organisations to disseminate information on all of the non-observation and observation errors summarised above. The second-best option is the adoption of a more pragmatic approach which entails national agencies using their professional judgement and more detailed knowledge about the data to identify a sub-set of key non-sampling errors that have a significant impact on the reliability of the data in question. The important thing is for these agencies to develop a culture of critical appraisal of their statistical output and for key strengths and weaknesses to be documented and disseminated.

#### **7.4.5 Recommended practices for rebasing**

It is recommended that rebasing be undertaken every five years and within three years from the end of the base year.

Unless the year was “unusual” it is also recommended that the base year selected be one ending with a “0” or “5”.

In order to provide sufficient transparency to users with regards to a rebase it is necessary to ensure that the following metadata accompanies any rebased data, either directly or through the provision of: appropriate references or links:

- the methodological approach adopted for the rebase, in particular, the processes actually undertaken during the rebase, e.g. simple rereferencing, introduction of new weights, etc.;
- the link year;
- the classification level at which index numbers are rebased and disseminated;
- the rounding policy followed in the rebasing, even though rounding should only be carried out at the very last stage for presentation purposes;
- a transition table from the old to the new classification system, if this is introduced;
- the description of any new weighting system and its impact on the aggregation of lower level indices;
- when the direct approach is adopted, a note of caution is useful to alert users that any aggregation of rebased indices needs the updating of the weights of the previous bases

#### **7.5.4 Recommended practices for citation**

##### *Citation of datasets*

If citation of datasets is to be taken seriously, a concerted effort must be made by national agencies and international organisations to:

- Formulate and then place their data citation policy in an obvious position on websites, including the policy for the citation of data disseminated via electronic datasets. Furthermore,

this policy should be accompanied by detailed sample citations to be included in specific web pages for users to copy as required. This makes it easy for users to include the correct information in the citation. The following example of this practice is from Statistics Canada's Census web module *Community Profiles* where the following citation instruction is provided at the bottom of the webpage in printer friendly format:

*How to cite: Statistics Canada. 2002. 2001 Community Profiles  
Released June 27, 2002. Last modified: 2005-11-30.  
Statistics Canada Catalogue no. 93F0053XIE.*

- Secondly, encourage a culture of data citation both inside and outside the organisation wherever data is being used. This awareness can be raised by contacting all known users of an organisation's data, all editors of publications known to use an organisation's data, etc., requesting that they follow the citation policy for the organisation in future publications.

A simple but effective citation style for datasets would be to include the following elements:

- unambiguous name of the dataset;
- author of the dataset;
- agency (or part of the agency) responsible for the dataset;
- date of the dataset (or version number);
- contact details for queries;
- address of the archive or other place of storage or system for accessing data;
- publisher (if this is different from the author, though for many agencies' publications the author and publisher are the same);
- if appropriate, the paragraph, table or page number.

This citation style should be followed for any data that is published internally or externally as well as for the documentation of any datasets that are created or modified. The actual ordering of the elements outlined above, punctuation, use of italics, etc., is a matter of individual (or organisational) choice.

### *Citation of text*

The main recommendation for text citation entails the systematic use in all metadata of one of the widely accepted bibliographic reference styles listed in Section 7.5.2. The two commonly used systems for presenting references in text for a bibliography are the Harvard system and the Numeric system. It is beyond the scope of the current Handbook to outline these systems in any detail beyond outlining a number of specific areas in metadata presentation where such systems should be used. These include the provision of:

- References or source for concept or variable definitions used in all published output, e.g. definitions appearing in explanatory notes, glossaries, etc. At the moment it is almost impossible to identify the primary source of concept and variable definitions published by both national agencies and international organisations. In particular, it is seldom possible for the user to identify whether or not a specific definition: has been taken directly from existing international statistical standards; is a modified version adapted for a specific use (say at the national level); or an entirely new definition.

- Sufficient reference (citation) information to enable the user to readily identify the availability of more detailed information on definitions and concepts, collection methodology, etc.

An example of a clear statement of citation policy at the national level is provided in Statistics Canada's publication, *How to Cite Statistics Canada Products* (Statistics Canada 2006a), which provides examples of recommended citation practices for a wide range of statistics products: publications; data products; census products; microdata products; maps and geospatial products; and E-STAT products.

### **7.6.2 Recommended practices for the presentation and reporting of administrative data**

Because statistics derived from administrative sources will be based on data that were not originally compiled or produced for statistical purposes and frequently by other non-statistical agencies, Statistics Canada in their policy guidelines for informing users on data quality and methodology (Statistics Canada 2000, p. 12) states that it is particularly important for such data to be methodologically transparent to users and stressed the need for such data to be accompanied by the following types of metadata:

- the name of the source agency for the administrative data. If more than one agency or ministry provides the services and collates data on these (e.g. health or education services provided by several agencies in some countries) specific information should be provided as to whether or not the data are from all agencies, or only from the main agency or ministry;
- a precise description of the purposes for which the statistical data were originally compiled and collected by the administrative agency;
- an outline of the strengths and weaknesses of the data in terms of the statistical application of the data. Particular attention should be given to the impact of issues relating to coverage and possible coverage bias, differences in concepts from international statistical guidelines and recommendations, in particular, the use of non-standard classifications and the use of unit concepts that differ from statistical units concepts;
- a description of processing or transformation (if any) undertaken by the statistical agency following receipt of the administrative data. Such processing may attempt to reduce or minimise inherent weaknesses in the original data;
- descriptions of the reliability of the data, including adherence to international norms and standards and caveats / limitations on the statistical use(s) of the data, e.g. for social indicator generation.