Statistical Commission
Thirty-sixth session
1-4 March 2005
Item 7(d) of the provisional agenda
Presentation of statistical data and metadata

Background document
Available in English only

DATA AND METADATA PRESENTATION AND REPORTING HANDBOOK
DRAFT VERSION FOR COMMENT BY THE STATISTICAL COMMISSION

Prepared by the Organisation for Economic Co-operation and Development

DATA AND METADATA PRESENTATION AND REPORTING HANDBOOK

DRAFT VERSION FOR COMMENT BY THE COMMISSION

BACKGROUND DOCUMENT

As per Provisional agenda item 7(d) of the agenda, the attached draft version of the *Data and Metadata Presentation and Reporting Handbook* is presented to the Commission as a background document.

This version incorporates extensive comments and suggested revisions received from international organisations including Eurostat, the IMF and the World Bank as well as from numerous national agencies (e.g. national statistical offices and central banks). As can be seen in the draft, there are several areas indicated where further work is required. These are highlighted by text in square brackets [……].

Issues where Commission comment and guidance are sought

As outlined in Part V of the information note submitted to the Commission, comments and guidance are sought with regard to:

I. The scope of issues currently included in the draft Handbook. Are any relevant / key data and metadata presentation and reporting issues not covered for:

- social and demographic statistics;
- developing countries;
- annual statistics;
- economic statistics?

II. Any relevant sources / references at the national or international levels that are not currently used? These include relevant examples of recommended practice on data presentation and reporting?

Comments and suggestions can also be forwarded directly to Denis Ward at the OECD (denis.ward@oecd.org).

OECD
Paris

1 February 2005

UNSC 2005
Prov. agenda item 7(d)
Background document

DATA AND METADATA REPORTING AND PRESENTATION HANDBOOK

DRAFT VERSION

31 January 2005

Foreword

[To be prepared]

Contents

5.1 Introduction
5.2 Need for metadata content standards
5.3 Recommended practices for the reporting and dissemination of metadata

5.3.1 Where metadata should be disseminated
5.3.2 Access to metadata
5.3.3 Adoption of a set of common metadata items
5.3.4 Adoption of a common set of terminology for metadata preparation
5.3.5 Unambiguous presentation of similar but not identical series

6. Guidelines on key reporting practices for

6.1 Data revision
6.2 Presentation of series breaks
6.3 Linking time series
6.4 Sampling and non-sampling errors
6.5 Re-basing indices
6.6 Citation
6.7 Administrative data

7. References

ANNEXES

1. Summary of data and metadata reporting and presentation recommendations presented in this Handbook
2. Glossary of data and metadata reporting and presentation terms

1. INTRODUCTION AND OVERVIEW

1.1 Need for data and metadata reporting and presentation standards

The need for the dissemination of international guidelines and recommendations in one comprehensive document for both the reporting and presentation of data and metadata by national agencies to international organisations by statistical agencies in both developed and developing countries has been recognised for some time. Data encompasses the whole range of statistics including economic, social and population, and both short-term and structural statistics together with any associated metadata. With the demand for closer integration and co-ordination of the work of the various international organisations, the need for agreement and adoption of a common set of data reporting and presentation practices has become even more pressing. For this demand to be met, agreement on a basic set of practices and guidelines is required not only between international organisations but also by national agencies, the initial source of almost all data and metadata entering the international environment.

Greater economic integration between the major trading blocks around the globe has led to user demand for greater comparability of data between countries. There is also parallel demand for greater data comparability over time for the same time series within the one country. The major causes of real differences in data have been identified (Eurostat 2002a, p. 35) as:

- the use of different variable definitions, concepts, units and classifications; and
- differences in data collection and processing (transformation) practices.

However, even where differences from these factors are minimal, data disseminated by different countries, etc. may actually look different because of the use of different data reporting and presentation practices. Similarly, data for identical statistical domains disseminated by various international organisations, although coming from the same national source, may also appear different due to variation in presentation practices and data transformation processes used at the international level.

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 reporting and presentation of data and metadata. These concern the need to improve data quality and to minimise the reporting burden in the provision of data and metadata to international organisations. These imperatives are discussed in more detail in Section 2.1 below, together with current international initiatives designed address them (Section 2.2). The relationships between these initiatives, both to each other and to the reporting and presentation standards outlined in this Handbook are highlighted in Section 2.3.

Finally, the wider range of options now available for the dissemination of statistics further strengthens the need for the development and adoption of data and metadata reporting and presentation standards by national agencies and international organisations. In the past, most users accessed their statistical needs solely via paper publications and although such publications are still an important means of dissemination, access via electronic media, in particular, through the internet, is now very common. Users frequently determine the availability of the statistics they need in electronic databases through use of search engines, etc. Such searches often involve interrogations across a number of databases prepared by different statistical agencies and the efficiency of this process is enhanced significantly if the statistics are presented in a common format using standardised terminology. These issue are taken up in more detail below in the discussion on terminology (Section 1.5) and key international initiatives (Section 2.2).

The ultimate objectives of the implementation of the recommendations presented in this Handbook are to ensure that all types of users – the general public, informed users and analytical specialists – can access their statistical needs efficiently, and that methodologies used in their compilation are sufficiently transparent for users to assess the relevance of the statistics accessed to their needs.

1.2 Scope of reporting standards referred to in this Handbook

As mentioned above, there is a need for the implementation by international organisations and statistical agencies in both developed and developing countries of a set of consistent data and metadata reporting and presentation standards. The standards outlined in this Handbook are relevant for the whole range of statistics comprising economic, social and population statistics, and are an essential prerequisite for data clarity and transparency and for statistical comparisons between countries around the world.

Data and metadata reporting and presentation covers a broad range of issues and topics, and before proceeding further it is worthwhile outlining the fairly limited scope of the data reporting and presentation issues covered in this Handbook. The Handbook makes the distinction between data presentation standards involving table layout, font, type faces, readability, graphical presentation, etc, used by both international organisations and national agencies, and standards for the reporting of data by national agencies to international organisations through submission of data files by national agencies or through international agency access to national agency databases and the use of web query tools, etc.

Irrespective of how data are exchanged, the boundary between "presentation" and "reporting" issues is often very fuzzy and although the main focus of this Handbook are standards for data reporting and exchange between organisations it necessarily touches on some issues that could be regarded as presentation.

This Handbook provides a number of specific recommendations covering the two broad dimensions in which all statistical data may be reported, namely:

- Types of data – refers either to original data (sometimes referred to as "absolute figures" or "absolute data") or original data that has been transformed for presentation as indices, growth rates, ratios, rates, percentages, etc. Original data may be either stock series which are measures at a point in time, or flow series which comprise measures during periods of time. Original data may be presented either in:

  – terms of physical units (tonnes, cubic metres, gig joules, etc); or

  – in value terms expressed at current or constant prices.

  The dissemination of original data is common for statistics published at annual or less frequent intervals, though such data are also disseminated for many short-term indicators such as monthly or quarterly data on motor vehicle registrations, construction permits, etc. 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 and social phenomena. As will be shown in Section 4.4 below, there are a number of different types of growth rates.

  The Handbook provides recommendations for the reporting and presentation of both quantitative and qualitative statistics. Information collected in qualitative surveys request respondents to assign qualities, rather than quantities, to the variables of interest (OECD 2003f, p. 7). In many instances the recommendations in the current Handbook are relevant to both types of statistics, but in some areas recommendations are specific to qualitative data.

More and more data producers are also expected to provide users with access to micro-level data, in particular, to record-level sample survey datasets. This is particularly the situation for users of social and population statistics derived from household surveys. The proper documentation and dissemination for such data is also of paramount importance and raises issues of confidentiality and others specific to this type of data. Data and metadata standards have been developed for the documentation of micro-level datasets under the Data Documentation Initiative (or DDI standard)[1]. Although, the DDI is briefly described below in Section 2.2.3, the presentation and reporting of micro-data are however outside the scope of this Handbook.

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

The focus of this Handbook is the reporting and presentation of statistical data in tabular form. It is also common practice to present key aggregates disseminated in paper publications, press releases, etc, in graphically. Such forms of data presentation are also outside the scope of this Handbook, though a number of national statistical agencies have prepared manuals outlining recommended practice in this area. [Insert reference to key manuals].

Section 5 of this Handbook emphasises the need for national agencies and international organisations to prepare and disseminate adequate metadata describing concepts, collection and processing processes as well as data reporting and presentation practices. Such metadata must also be readily accessible and understood by users with different degrees of statistical expertise.

The metadata presentation and dissemination guidelines discussed in Section 5 draw heavily on current and previous work of the a number of international initiatives, in particular, that of the UNECE – Eurostat – OECD Work Session on Statistical Metadata (METIS). This Section focuses on recommendations on where and how metadata should be disseminated. Other Sections of this Handbook touch on specific metadata issues such as describing methodology used in seasonal adjustment (Section 4.6) and the provision of information on data revisions (Section 6.1), series breaks (Section 6.2) and sampling and non-sampling errors (Section 6.4).

Ideally, metadata should be expressed by different organisations in different countries around the globe on the basis of a common terminology. The importance of the use of common terminology is discussed below in Section 1.5. The Glossary at the end of this Handbook contains a comprehensive set of definitions for concepts related to data and metadata reporting and presentation.

In addition, there are 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 data. These include different:

- data revision presentation practices;
- reporting practices for the presentation of series breaks;
- 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.

[1] Published in the *Handbook on the Preservation, Documentation and Dissemination of Surveys and Censuses Data and Metadata*

This Handbook outlines key recommendations and/or good practice with regard to each of these. Such practices are consistent with sound governance in statistics encompassed, either implicitly or explicitly, in the UN *Fundamental Principles of Official Statistics* (UNSC 1994) and the UN *Handbook on the Operation and Organisation of a Statistical Agency* (UNSD 2001).

The range of reporting practices outlined initially in this version of the Handbook could be expanded in subsequent versions to include others, in particular, those of specific interest to users in developing countries. Such good practices are also outlined in 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 dissemination[2].

Another dimension relevant to the development data reporting and presentation guidelines and recommendations is the data dissemination medium or format (on-line database, internet web pages, CD-ROM, paper publication or press release) used to disseminate data. The reality is that different forms of data reporting and presentation are more suited to different means of data dissemination.

For quick reference, the data and metadata reporting and presentation recommendations provided in the following Sections of this Handbook are also summarised in Annex 1.

1.3 National and international publishing manuals and guidelines

Almost all agencies at both national and international levels have publishing manuals or guidelines containing technical guidelines that touch on a very wide range of issues such as corporate standards concerning data presentation, layout of tables, graphs, maps, fonts and type faces, etc., for use by authors involved in the preparation of both paper and electronic statistical dissemination. In the main, such manuals tend to focus on corporate policy on the publication preparation process and data and metadata presentation to help ensure that disseminated output comply with a set of professional standards and that the range of output disseminated by the organisation have a common look and feel. However, they also frequently deal with some of the reporting issues discussed in this Handbook such as revision, citation, presentation of sampling and non-sampling errors, etc. Ideally, key elements of international guidelines on data and metadata reporting included in this Handbook should also be imbedded in publishing manuals at both the national and international levels and/or be linked to it.

Although some of the guidelines and practices included in publishing manuals prepared by both national agencies and international organisations may be relevant only in the context of a specific institution or country, the desirable objective of increasing the transparency of internal publication guidelines and processes to external view would be achieved if these manuals were placed on the Internet. At the moment, only a very small number of such manuals are accessible to external readers via the web.

[Insert example of national publication manual placed on Internet – outline content]

[2] In the main, the good practice described in ROSCs relevant to the data reporting practices described in this Handbook are restricted to data revision and the availability of appropriate metadata.

1.3.1 Recommended practices for publication manuals

There are two 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.

  The content of such a manual will of course be determined by corporate priorities and resources available for implementation, though it should at a minimum provide clear guidance on issues relating to: the presentation of statistics in tables (including practices for titles, use of footnotes, etc) and graphs; the use and presentation of growth rates and seasonally adjusted data; citation; and recommended practices for data revision, presentation of series breaks, provision of information on sampling and non-sampling errors and the presentation of administrative data. Where appropriate, links and references to the recommendations in this Handbook could also be provided.

- 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.

1.4 Guidelines and recommendations for the reporting and presentation of statistics in existing international statistical standards

The reporting and presentation recommendations and guidelines presented in this Handbook draw on the extensive range of existing international statistical standards dealing with a broad range of economic, social and population statistical domains (such as labour force, national accounts, education, health) that have been developed by international organisations in co-operation with national agencies ((UNSD 2002a) and (Eurostat 2003b)). The focus of these statistical domain standards are primarily conceptual and encompass definitional issues, classifications, coverage, etc, though some standards also outline recommended practice for the collection of data. However, in the main, international statistical standards are largely silent, or give only brief mention to a limited range of reporting and presentation issues and practices, though there are exceptions. For example:

- The European Commission's 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. 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).

- The IMF Special Data Dissemination Standards (SDDS) (IMF 2003b) and the Fund's Data Quality Assessment Framework (DQAF) (IMF 2003c) both touch on a number of key presentation issues such as the provision of metadata to enhance interpretability, coherence over time and the adoption of good

practice for data revision. The guidelines and recommendations on data revision presented in Section 6.1 of this Handbook draw extensively and directly from text prepared by the Fund in the context of the SDDS and IMF Reports on the Observance of Standards and Codes (ROSC).

- The work of METIS on the presentation of statistical metadata on the internet and terminology on statistical metadata published by the UN in 2000 (UNSC and UNECE 2000 and 2000a).

- Work undertaken by the Committee for the Coordination of Statistical Activities (CCSA) on rebasing and citation prepared by UNCTAD and UNESCO respectively in 2003, and presented below in Sections 6.5 and 6.6.

United Nations Statistical Divisions Methodological Publications in Statistics website

The Methodological Publications in Statistics website was created by the United Nations Statistical Division (UNSD) at the request of the United Nations Statistical Commission in 1999. The website would contain statistical standards (including classifications) across all statistical domains that had been developed by all international organisations in addition to those under development and planned. The website database would be updated annually. All international organisations were requested to provide the appropriate information.

International organisations contributing to the database include: Commonwealth of Independent States, Economic Commission for Europe, Eurostat, Food and Agriculture Organisation of the United Nations, IMF, International Labour Organisation, OECD, UNESCO, United Nations City Groups, Universal Postal Union, UNSD, World Bank, World Health Organisation, World Tourism Organisation

The following information is provided in the database for each of the methodological guidelines and recommended guidelines listed: name of lead organisation; other organisations involved in the development of the standard (if any); formal title of the standard; brief description of the contents of the standard; year published; previous or subsequent version(s) of the standard. In some instances a hyperlink is provided to a site containing a complete electronic version of the standard.

The database is available at http://unstats.un.org/unsd/progwork/pwabout.asp

This Handbook provides a single comprehensive reference source for presentation and reporting standards drawn from the above and other existing international statistical guidelines and recommendations. It also provides a focus for the evolution of new standards on topics / issues not adequately covered in those standards at the moment. The recommendations and practices included this Handbook could either be imbedded in future international statistical standards for specific statistical domains, or be linked to it.

Over the last 10 to 15 years more emphasis has been given to the linkages between international standards that have been developed for different statistical domains in order to ensure conceptual and terminological consistency, etc. Examples of such linkages include the interface between: 1993 System of National Accounts (SNA 93) and the Balance of Payments Manual, Version 5 (BPM5); SNA 93 (Chapter XVII, population and labour inputs) and the relevant International Conference of Labour Statistician (ICLS) Resolutions on employment; and the recently published manuals on consumer price indices and producer price indices. This process will continue as guidelines for emerging statistical domains (such as

information and communication technologies) are developed or existing standards (such as the 1962 ICLS Resolution on working time measurement) are revised to ensure their relevance to current circumstances. This Handbook could be used as a resource to ensure consistency between international standards in the area of data and metadata presentation and reporting.

Because such practices may differ according to the dissemination medium used, where necessary, guidelines are provided for different 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 key statistics in press releases.

1.4.1 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 1.5.1 below. Such use may take the form of either the direct incorporation of text within the future guidelines or by reference / link to the recommendations in this Handbook, etc.

New data and metadata presentation and reporting issues will emerge over time, or existing guidelines will need to be modified to meet the needs of new standards that will be required for various statistical domains in future. Change will also be brought about by the use of new technologies, etc., for the dissemination of statistics to users. It is therefore intended emerging issues, etc., be incorporated in a web version of this Handbook.

1.5 Importance of use of common terminology

A factor complicating the development of international reporting and presentation standards has been the use of different terminology used by different countries, or in the same country for different series, for example, the various forms of growth rates. The use of inconsistent labels for the same concept frequently leads to misunderstanding by users and the risk of inappropriate use of statistics. 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 and outlining national practices with respect to data collection, manipulation, etc. Such inconstancies severely limit the use of much existing metadata in comparing national concepts and practices in the collection, transformation and dissemination of statistics.

1.5.1 Glossary systems

Users of statistics often comment on the different terminologies used to describe the same thing in different publications and databases, and instances where the same database contains different definitions for the same concept. There are two ways of looking at this: the first is to say that because so many institutions use slightly different definitions or terms to describe the same phenomenon, it is acceptable to use different terms interchangeably. The second approach is to say that terminology should be consistent across institutions so that the question of ambiguity does not arise. This Handbook emphasises the second approach.

There are also differences in the impact of ambiguous and unclear terminology on different users with varying degrees of statistical knowledge and expertise. While the general public may not understand what is meant by say, "sampling error", informed users will probably have a reasonable understanding of the

term. Even if they do not have such an understanding, glossaries such as the *OECD Glossary of Statistical Terms* (OECD 2002a) contain definitions that can explain terms which are of interest only to the more informed user of statistical data.

In recent years the importance of national agencies and international organisations adopting common definitions for variables, concepts, etc, based on international statistical standards has received greater recognition as a precondition for the compilation and dissemination of comparable statistics. As mentioned above, there is also a similar need to adopt common terminology in the preparation of metadata that can be used to compare national practices and concepts. To promote the use of common terminology and an understanding of concepts and definitions many national and international organisations have compiled glossaries that provide definitions of key concepts and statistical domain specific variables, etc. However, in the main these glossaries have been domain (or issue) specific and as a result there are many instances of different and inconsistent definitions being disseminated in multiple glossaries containing the same concepts within the one organisation and country.

To help overcome this problem, a number of international organisations have developed extensive glossary databases containing definitions of key concepts and variables covering a wide range of statistical concepts, etc. The *OECD Glossary of Statistical Terms* is one example of such a glossary database, though others have been developed by Eurostat (called CODED) (Eurostat 2003a) and the United Nations Statistical Division (UNSD 2002b). Similarly, an extensive array of concepts and definitions primarily dealing with population statistics have been assembled by the United Nations Development Group for the indicators to be used for monitoring the Millennium Development Goals (refer UNDG 2003)[3]. The definitions included in each of these glossary databases have by and large been drawn from existing international statistical guidelines and recommendations and by and large cover economic, social and population statistics.

The process of compiling these extensive glossary databases has entailed the confrontation of inconsistent definitions covering the same concept that previously resided in different glossaries with narrower domain coverage. The extensive glossary databases developed by the OECD, Eurostat and UNSD also facilitate the preparation of sub-glossaries covering more specific statistical domains with consistent definitional content. In the OECD context, the *OECD Glossary of Statistical Terms* can be used to derive a number of sub-glossaries, including the:

- Metadata Common Vocabulary (MCV) being developed in the context of the Statistical Data and Metadata Exchange (SDMX) initiative (refer Section 2.2.1 below);
- System of National Accounts (SNA) Glossary;
- Data and Metadata Reporting and Presentation Glossary provided in Annex 2 of this Handbook. Many of the definitions in this Glossary are also to be found in the MCV.

Some national statistical agencies and international organisations are taking their use of corporate glossary databases a step further by integrating them into their corporate data and metadata environments that are currently being developed as a tool for reinforcing the use of standard concepts and definitions in data collection questionnaires and disseminated output.

[3] Refer http://www.developmentgoals.org/ and Section 4.5 below for further information on the indicators for the Millennium Development Goals.

1.5.2 Data reporting and presentation terminology

The adoption of common terminology by national agencies and international organisations is also an important element of the data reporting and presentation recommendations embodied in the following Sections of this Handbook. Where appropriate, recommended definitions and terms are provided (in "boxes") in some Sections to ensure a common understanding of the concepts and issues described. Issues of terminology are particularly important in the discussion of growth rates (Section 3.4) and in guidelines for the reporting of different forms of data (Section 4).

A comprehensive consolidated Glossary of key data presentation terms is provided in Annex 2 below. This Glossary includes the source of each definition and in some instances further context information. The definitions in this Glossary are also available in the *OECD Glossary of Statistical Terms* referred to above.

1.5.3 Recommended practices to ensure the use of consistent terminology and definitions

The issue of the use of consistent terminology and definitions in relation to the preparation of metadata is taken up in more detail in Section 5 of this Handbook. The recommended practices outlined below are designed to stem the current common practice of different author areas within national and international organisations developing their own, often inconsistent, sets of terms and related definitions. Such inconsistencies are often inappropriate, though even where appropriate, there is perhaps still the need to provide users with information about the context / reasons for any differences.

Agencies should therefore:

- Establish a process 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 in various thematic glossaries which in turn are applied in questionnaires and disseminated output.

  There are a range of options available to achieve this objective. These include the creation of a corporate glossary 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. An alternative could be to use existing glossary databases that have been developed by international organisations such as the OECD, Eurostat and UNSD or domain specific glossaries developed by the IMF, etc. The definitions in these glossaries may require modification or elaboration to meet individual national circumstances.

  Irrespective of the glossary tool adopted, there is still the need for management within an organisation to ensure that appropriate practices involving the use consistent terminology are adopted across the organisation.

- Ensure that any glossary developed by an agency contains the following minimum structure 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 certain definition is used within one statistical domain or geographical context. Finally, the glossary should also be translated into languages used by the institution in question.

- Provide appropriate cross references 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 terms and definitions more readily available.

- Avoid attaching precisely the same label or title for different definitions (for further discussion refer Section 5.3.5 below).

2. LINKS TO OTHER INTERNATIONAL INITIATIVES

As mentioned in Section 1.1 above, there are two broad imperatives relevant at both the national and international levels that drive the need for the articulation of a comprehensive set of standards for the reporting of data and metadata. These concern the need to improve data quality and to minimise the data reporting burden in the provision of data and metadata to international organisations. Several international initiatives currently underway that impact on these imperatives are described in Section 2.2 below. This is followed (in Section 2.3) by brief notes highlighting the relationships both between the initiatives described and the guidelines and recommendations on data and metadata reporting and presentation outlined in this Handbook.

2.1 Imperatives driving need for articulation of data and metadata reporting standards

2.1.1 Improved data interpretability and coherence

The first imperative relates to the need to improve the quality of statistics presented to users at both the national and international levels, particularly with respect to interpretability and coherence (within datasets, across data-sets, over time and between countries). These are important dimensions of quality imbedded in one form or other in all quality frameworks that have been developed by national agencies and international organisations (such as by Eurostat (Eurostat 2003d), the IMF (IMF 2003c), OECD (OECD 2003e), Statistics Canada (Statistics Canada 2002), etc)[4]. However, beyond stating the case for improvement, such frameworks seldom go into detail about how these quality dimensions would be implemented in the context of data and metadata presentation and reporting, the exception being the treatment of revisions in the IMF quality framework.

Interpretability reflects the ease with which the user may understand and appropriately use and analyse statistical information. The adequacy of concept definitions, target populations, variables and terminology that underlies 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 and which is both well structured (readable) and readily accessible.

With respect to coherence, users are often confronted by the problems outlined in Section 1.1 above when comparing statistics compiled over time within the one agency, by agencies in different countries and by different international organizations. This Handbook focuses on different practices in the reporting of statistics and presents the main reporting practices outlined in Section 1.2 above 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.

2.1.2 Minimisation of reporting burden

The second imperative driving the need for the development of data and metadata presentation and reporting standards refers to the need to minimise the reporting burden of national agencies in their provision of data and metadata to international organisations. Emphasis here is on the development of

[4] An analysis undertaken by the IMF (IMF 2005) compared the high level metadata concepts that are common across the various quality framework initiatives by the IMF, Eurostat, OECD, Statistics Norway, Statistics Sweden, and Statistics Canada. Although not specifically mentioning the impact of good data presentation practice on clarity and interpretability, the Fund's view was that such initiatives could be used to help structure the analyses and assessment of reference metadata content.

more efficient practices and processes for such reporting. In this context, discussions at recent international forums (such as the 2002 CES (OECD/IMF 2002) and the 2003 meeting of the OECD High Level Group for Statistics[5] (OECD 2003a)) 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. Data sharing refers to the sharing of both input files (common questionnaires, etc) and the output of national and international organisations to help ensure the consistency of published statistics. Such a model envisages the extraction of common data requirements by international organisations from data (input or output files) located on national agency websites. This model is particularly beneficial for either statistical series that are collected regularly (i.e. daily, weekly, monthly, quarterly) or where the requirements of international organisations collectively are relatively stable over long periods of time (e.g. for annual national accounts).

International organisations have also introduced a wide variety of other processes to minimise the reporting burden of national agencies. These include the despatch of common questionnaires (for example by Eurostat and the OECD for research and development (R&D) statistics) and the use of regional organisations' data (for example, Eurostat and the European Central Bank) in lieu of direct collection by all international organisations from national sources. Such sharing of data and metadata often requires the sharing of classifications and the adoption of common statistical units.

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 and other processes for minimising reporting burden outlined above. Prerequisites for implementation involve not only the resolution of a number of information technology and communications (ITC) issues but also agreement between national agencies and international organisations on a number of data "content" issues including:

- Identification of a set of common data requirements for key statistical aggregates. A brief outline is provided below in Section 22 below 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.

- Adoption of common or at least consistent classifications such as International Standard Industrial Classification (ISIC), Harmonised System (HS), International Standard Classification of Occupations (ISCO) and the International Standard Classification of Education (ISCED). These and other international classifications are readily available on Eurostat's Classification Server (RAMON) (Eurostat 2003a) or on the United Nations Statistical Division website (UNSD 2004).

- Agreement on key data reporting and 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 different international organisations.

The need for further international work on data presentation was also highlighted in a presentation by Eugenio Domingo Solans (Solans 2003), a member of the Governing Council and Executive Board of the European Central Bank at the 54th Session of the International Statistical Institute in Berlin in 2003. Solans acknowledged the impact of modern website technology in improving timely access to official statistics. However, he also emphasised the need for ITC developments to be accompanied by the development and implementation of standards for data presentation, citing specific issues such as growth rates, revision policy, provision of quality indicators, etc.

The development of the required guidelines in these and other areas is the responsibility of international organisations in co-operation with national agencies. Obviously, the implementation of the data sharing

[5] The forerunner of the OECD Committee on Statistics created in May 2004.

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 reporting 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.

2.2 Key international initiatives

The two key initiatives currently underway at the international level that will impact on the imperatives driving the need for the articulation of data and metadata reporting standards described above are the Statistical Data and Metadata Exchange (SDMX) project and the Workshop on Statistical Metadata (METIS). Both these initiatives would benefit either directly or indirectly from the development and adoption (by international organisations and national statistical agencies around the world) of a common set of data reporting practices. Furthermore, as stated in Section 2.1.2 above, the data reporting standards described in this Handbook are key elements of the implementation of the data sharing model, the aims of which are to:

- avoid duplication and enhance efficiency in the transfer of data between systems whilst at the same time reducing the reporting burden of national agencies; and

- ensure the consistency of data disseminated by different international organisations.

International initiatives on metadata standards also seek to enhance the interpretability of data presented to users.

2.2.1 Statistical Data and Metadata Exchange (SDMX) project

The Statistical Data and Metadata Exchange (SDMX) project[6] is a consortium of seven international organizations (BIS, ECB, Eurostat, IMF, OECD, UNSD, World Bank) working to foster standards for the exchange and sharing of data and metadata. 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 the work of others in the field of statistical information. This would be achieved by taking advantage of existing and emerging:

- exchange protocols, such as GESMES/CB which were implemented by central banks for exchanging time series;

- data dissemination formats, such as that implicit in the IMF's Dissemination Standards Bulletin Board (DSBB); and

- e-standards, such as eXtensible Markup Language (XML).

The new standards developed by SDMX seek to take advantage of new web-based technologies and the expertise of those working on the business requirements and ITC support for the collection, compilation, and dissemination of statistical information.

[6] More detailed information about SDMX and projects currently underway under the umbrella of SDMX are available on the SDMX website (BIS, ECB, Eurostat, IMF, OECD, UN 2003)

Several initiatives conducted under the SDMX umbrella in several institutions are currently underway or have been completed. These comprise:

- a Pilot Project to develop a unified infrastructure of SDMX standards and freely available tools for data exchange and web dissemination and launch a web-based reference implementation;
- a Practical Case Study to investigate and demonstrate ways in which emerging web standards can help simply the collection, compilation and dissemination of statistical information;
- a Batch Data Exchange project to advance and maintain the Generic Statistical Message for Time Series (GESMES/TS) - Version 3.0 issued in 2003;
- the Metadata Repositories project to develop a standard approach to the creation and use of metadata repositories for access, analysis and reuse of statistical metadata; and
- the creation of a common vocabulary for statistical metadata.

Draft versions of these projects are already available for public comment on the SDMX website (at www.sdmx.org), together with Version 1.0 of the SDMX Standards-setting Documents that were posted on the site in September 2004. Following the receipt of further input from users, subsequent versions of these documents will be prepared by the third quarter of 2005.

The last project listed above entails the development of a glossary (the Metadata Common Vocabulary (MCV)) (OECD and Eurostat 2003) as a tool to help ensure the consistency of metadata prepared by authors at the national and international levels, with respect to both the content and the range of methodological issues covered by the metadata. The MCV is designed to include the range of metadata terms and items 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 used in the metadata dissemination models that have been are being developed by international organisations such as Eurostat, European Central Bank, IMF and the OECD. The MCV is also intended to be relevant for metadata models developed by national agencies.

The MCV is an example of a metadata glossary whose content has been jointly developed by several international organisations which is regularly updated by two agencies (Eurostat and the OECD) in order to promote the use of common terminology which is a prerequisite for the efficient exchange of information about statistical data. The MCV is referred to again in Section 5.3.4 below in the context of guidelines for the reporting and dissemination of metadata.

In addition, there are a number of other projects currently being undertaken at the international level which, though not directly being undertaken in the context of the SDMX initiative, either use SDMX standards or are designed to be consistent with and complement the main objectives of SDMX. The focus of these projects is the development of common data requirements by international organisations, the primary purposes of which are to share input files and reduce the reporting load of national agencies, and the dissemination of consistent data at the international level.

The most notable example of common data requirements is the questionnaire used by the Eurostat, IMF, OECD, UNSD and the World Bank for the collection of annual national accounts data[7]. The questionnaire comprises a very detailed common 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.

There is still considerable scope for international organisations to reach agreement on common lists of variables for other structural and short-term statistics domains that require common on-going collection. For short-term economic indicators, the OECD, Eurostat and others have independently developed lists of variables 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 the monthly Main Economic Indicators (MEI) database and the list of variables stated in the European Commission Short-term Statistics Regulation which specifies both the reference period and the form of data to be transmitted to Eurostat by EU member states. [Identify co-ordination agreements for social and population statistics]

A 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[8] and approved by the Statistical Programme Committee (SPC) in September 2002. The list, which will be refined over time, also includes target release dates and other quality objectives. Eurostat compiles and releases PEEIs based on member state contributions on a common dissemination platform accessible via the Euroindicators site covering both EU/Euro area and national indicators compiled according to EU standards. The SDMX Open Data Interchange (SODI) project described below entails the development of mechanisms to improve both the timeliness and accessibility of PEEIs.

There is a need to identify 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 reported in the context of a data sharing model. The OECD is also currently investigating the expanded use of Eurostat short-term economic statistics for European Union member states.

National Accounts World Wide Exchange (NAWWE) project

The 2002 meeting of OECD National Accounts Experts (OECD 2002c) proposed an experiment to test the implementation of the data sharing model for annual national accounts between national agencies and the OECD. The idea behind the NAWWE project is to implement a model in which data are not necessarily transferred across organisations but, rather, published on the web in a form that users could extract by simply using country and variable references. The OECD has set up a demonstration model of this kind of data extraction for two Member countries (Australia and Canada). This pilot study starts from the Excel tables already produced by national agencies for transmitting annual national accounts data to international

[7] Others include: co-ordination arrangements between the OECD and UNSD on merchandise trade statistics; OECD use of structural economic statistics from Eurostat; joint Eurostat – OECD questionnaire for foreign direct investment statistics; IMF use of OECD industrial production indices disseminated for OECD Member countries, etc.

[8] There are equivalent national lists of principal economic indicators for (primarily) short-term statistics, for example the US Principal Federal Economic Indicators (Fedstats 2003)

organisations but the idea is that the necessary XML files should be created directly from the national statistical organisation's on-line databases.

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 for data quality to be maximised for the international statistical community since the data they would use would be those officially disseminated and not specially compiled for and transmitted to international organisations.

SDMX Open Data Interchange (SODI)

The SODI project has been designed to test a new approach for collecting and disseminating short-term statistics from European Union member states, aimed particularly at making such statistics:

- available more quickly at the international level, ideally, at the same time as they are available at the national level; and
- more accessible to users in formats which facilitate further use.

The agencies participating in the project are Eurostat and the national statistical institutes of France, Germany, Netherlands, Sweden and the United Kingdom. The SODI project emerged from concerns expressed in 2003-2004 by the Friends of the Chair Group of the EU Statistical Programme Committee (SPC) about the quality and timeliness of the PEEIs used in the management of the Economic and Monetary Union (EMU) in the EU. The project is centred on the concept of data published at the national level by EU member states being simultaneously made available through a common dissemination environment at the EU level.

More specifically, the SODI project will investigate various approaches, based on SDMX standards and tools, for streamlining the collection of PEEIs. Amongst other things, the project will address:

- the data model and metadata standard to be used;
- the interface to be used;
- possible implementation of the "pull" system of data dissemination[9];
- legal questions relating to the reporting obligations of EU member states;
- questions of validation and adjustment by member states and by Eurostat;
- integration of the SODI concept into the data life cycle for EU statistics.

In its implementation the SODI concept requires a significantly higher degree of harmonisation of statistical dissemination by member states than current data transmission methods, particularly in relation to simultaneous publication and dissemination at the national and international levels, transmission of revisions and the alignment of seasonal adjustment procedures. SODI further demonstrates the need for the

[9] Whereby receiving organisations themselves extract the required statistics and their associated metadata from a provider organisation's website as distinct from a "push" process involving the providing organisation forwarding a file(s) containing the required information to one or more receiving organisations.

formulation of data and metadata standards at the international and their adoption at both the national and international levels.

The initial SODI feasibility study is based on two indicators from the PEEI, quarterly GDP and the monthly industrial production index. If the results of the feasibility study (planned for June 2005) are positive, extension to other indicators from the PEEI is envisaged.

2.2.2 UNECE-Eurostat-OECD Workshop on Statistical Metadata (METIS)

The METIS forum on statistical metadata organised jointly by UNECE, Eurostat and the OECD has met about every 18 months since 1995. The major objectives of the forum, attended by delegates from around ten international organisations and 30 countries, are to exchange experiences on the development of corporate statistical metadata environments and to develop international standards in this area. Discussion at METIS covers a combination of ITC and metadata "content" issues, the latter comprising terminology, metadata models and corporate metadata governance.

The February 2004 meeting of METIS noted the existing wide variation in metadata practices in both international organisations and national agencies and that the evolution of metadata content standards has not kept pace with infrastructure developments, especially web-based technologies. One way of addressing this variation was to ensure that existing and future metadata standards were presented in the context of a framework. One of the issues raised which highlighted the need for such a framework was the apparent fragmented and seemingly unconnected nature of metadata standards developed over the last few years by various initiatives and agencies including METIS, METANET, the Neufchatel Group, Eurostat, IMF, ISO, OECD, etc.

Following the February 2004 meeting METIS undertook to draft and publish a framework that would provide links and context to current and previous metadata standards initiatives together with comparisons of selected examples of current best practice at the national and international levels (UNECE, Eurostat and OECD 2004, para. 7). The framework and links to recommended practices would be developed along a number of key themes such as:

- processes for metadata collection;
- terminology;
- metadata and data interchange entailing identification of common models between international organisations to which national agencies could map;
- migration strategy from existing fragmented metadata environments;
- metadata dissemination and its relation to dissemination of statistics;
- metadata governance and corporate management issues;
- incorporation of usability concerns in metadata management.

These themes touch on some aspects of the guidelines for the reporting and dissemination of metadata outlined in Section 5 below.

2.2.3 Data Documentation Initiative for Micro-data

[Text for 2.2.3 to be provided by World Bank]

[Identify other relevant international initiatives, particularly for socio-demographic statistics]

2.3 Relationship between international initiatives and data reporting and presentation

There are clear relationships both between each of the international initiatives outlined and the reporting and presentation recommendations outlined in this Handbook. For example, the SDMX initiative shows HOW data and metadata can be exchanged more efficiently, either through the implementation of the data sharing model with data being located on national databases (being pilot tested by the NAWWE project) or through dissemination of national data to a common data platform at the international level (being pilot tested by the SODI project). If successful, these approaches could be applied to other statistical domains.

The various initiatives undertaken by international organisations to identify a set of common data needs and requirements essentially describe WHAT data is to be exchanged both between themselves and from national agencies.

The work of METIS seeks to develop international standards, identify good practice, etc, to improve the INTERPRETABILITY of the data to be exchanged.

Finally, the reporting standards, etc, outlined in this Handbook aim at maximising both the INTERPRETABILITY of data exchanged and the CONSISTENCY of data disseminated by both national agencies and international organisations. The need for such consistency is particularly relevant in an environment where users have ready on-line access to databases maintained at both the national and international levels, often with overlapping content.

3. GUIDELINES FOR THE REPORTING OF DIFFERENT TYPES OF DATA

3.1 Introduction

The presentation and reporting of the appropriate type of data, together with the different form of data, is essential to ensure correct use and analyses of the economic, social or population statistics in question. It is the responsibility of the statistician to select the type and form of data appropriate to both the data on hand and the audience of users. Furthermore, some forms of presentation are perhaps more suited to specific methods of data dissemination. For example, the method appropriate for a comprehensive on-line database may not be at all suitable for a press release containing only a limited number of key statistical aggregates. The form and type of data selected should therefore be presented appropriately, if necessary with accompanying metadata, in order to minimise the possibility of misrepresentation by users.

The main types of data are original data (some times referred to as "absolute figures", absolute data", or "raw data"), indices, growth rates, rates, ratios and percentages. Other types of data in common use are composite indicators, survey balances and spatial data[10] where variances of regional indicators are calculated. Discussion in this Section of the Handbook is restricted to original data and the main forms of transformation of such data, namely indices, growth rates, rates and ratios. Section 4 below provides guidelines for the presentation of different forms of data. The main aim of the data transformations listed above is to facilitate the drawing of main features and generalisations out of a mass of original data. However, as will be shown below, it is necessary to use the most appropriate form of transformation (e.g. growth rate, rate, ratio) and to provide adequate information about its calculation or derivation to minimise the risk of misleading users.

3.2 Original data

Original data in the context of this Handbook refers to statistical information that has not undergone any transformation process such as the compilation of indices, growth rates or ratios. As mentioned in the Introduction to this Handbook, original data may be either stock series which are measures at a point in time (for example, a money supply stock series which refers to money supply on the last working day of the reference period) or flow series which comprise measures during periods of time (for example, passenger car registrations where an estimate for a reference month is the sum of daily car registrations). Indices, growth rates and ratios, etc, are further transformations of original data where the precise treatment differs depending on whether such data are stock or flow series.

Original data may be expressed in terms of either physical units (tonnes, cubic metres, gig joules) or value at current or constant prices. Raw or data initially collected at source by direct survey or derived from administrative sources are often expressed initially in terms of original data. The compilation of indices, growth rates, etc, are viewed as further transformation to enhance understanding and appropriate analyses.

The main strength of original data (UNEP 2002) is that they provide users with a sense of scale or relative magnitude about the phenomenon being measured and its contribution to an overall effect. The provision of original data enable users to:

- consistently track data;

[10] Refer Glossary in Section 8 of this Handbook for definitions of these three types of data.

- aggregate data to derive a total measure; and
- undertake transformations other than those originally presented.

In many instances original data will therefore be the most appropriate type of statistical information. On the other hand, the normalisation or standardisation of statistical data through the compilation and presentation of ratios, indices and growth rates, facilitate analysis of change over time, comparisons of change between similar or dissimilar original data or processes and information on the efficiency, intensity or quality of a value or achievement. Furthermore, in situations where data are known to have significant non-sampling errors (refer Section 6.4) the presentation of original data could be misleading and it may be more appropriate to provide results in the form of percentages. For example, in situations where there is known under enumeration of total business units in a population it could be better to present the number of units in different size classes as a percentage instead of absolute figures. Similarly, key measures in economic and social statistics may also be expressed in the form of percentages of total GDP.

The focus of this Section of the Handbook is not how original data should be presented in tables or graphs, which are generally dealt with in publication manuals (referred to in Section 1.3 above), but with the decision on whether or not such data should be presented in lieu of, or in addition to, other types of data in different forms of dissemination media.

3.2.1 Recommended practices 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 original data where they exist for a particular series either directly in the publication or through the provision of references or hyperlinks.

- In situations where absolute figures 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 in, for example, a press release to facilitate interpretation and understanding. Ideally, original data should be presented in addition to the transformed series. However, in some dissemination media this may not be possible 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.

3.3 Indices

3.3.1 Introduction

Indices are indicators of relative change which show the extent of change in prices and/or quantities of a given commodity or commodity group in relative terms from a base period to the comparison period (Statistics Finland 2003). As this definition states, the extent of change over time in quantities or prices may reflect those of an individual commodity or a collection or group of different commodities that have been aggregated or combined in some way. Analysts monitoring economic activity are more often interested in the latter and the most commonly used indexes for short-term economic monitoring are aggregate indices such as industrial production, retail volume, consumer price, etc. Such indices are typically weighted averages of either price or quantity ratios.

The Statistics Finland Quality Guidelines disseminated in English on their website states that the focus of the compilation of key indices for short-term economic analysis is the aggregation of individual series, the weights used, the maintenance of those weights and procedures used to take account of changes in the commodity composition and changes in quality. A detailed exposition of these issues is beyond the scope of this Handbook[11] whose focus is the presentation of appropriate information about the index to enable users to assess its relevance to their particular requirement(s).

3.3.2 Recommended practices for indices

The Statistics Canada Policy on Informing Users of Data Quality and Methodology (Statistics Canada 2000, p. 11) 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 5 below emphasise the need to structure metadata appropriately for users with differing degrees of expertise and need.

3.4 Growth rates

3.4.1 Introduction

Growth rates (or rates of change) express the change in value or magnitude of a time series between two or more different periods of time. For this reason growth rates are often a feature of press releases, analyses and media reports concerning economic activities (such as movements in prices, output or unemployment, etc.) and in social phenomena such as population growth (or decline). Both national agencies and

[11] The Statistics Finland Quality Guidelines provides a small list of sources for further information on index theory in the Bibliography to their guidelines on indicators and indices (Statistics Finland 2003).

international organizations use a wide variety of different forms of growth rates in their disseminated output.

The main issues for growth rates from the perspective of data reporting and presentation entail the use of inconsistent terminology by different agencies to label the different types of rates, and the appropriate use of a specific form of growth rate in different circumstances. Although there may not be an absolutely "correct" form of growth rate, there are instances where some forms are more appropriate than others. Most forms of growth rate can be useful depending on the specific needs of analysts.

Recent work by the OECD Short-term Economic Statistics Expert Group (STESEG) task force on data presentation and seasonal adjustment sought to identify best and appropriate practice for the use of growth rates for different short-term economic statistical indicators or analyses. Issues examined in the context of growth rate presentation include (Eurostat 2002a, p. 135):

- the form of the data, i.e. whether or not the data are working day adjusted, seasonally adjusted or trend-cycle;

- whether such rates should be based on a given period or should the data be smoothed through, for example, the use of a moving growth rate of the latest three months over the three months before;

- the time horizon that should be referred to. For example should comparisons be made in relation to the previous month ((t/t-1), the previous quarter (t/t-4) or for the same month of the previous year (t/t-12)?

The OECD task force also formulated recommendations on terminology for growth rates used in economic statistics.

As outlined in the *Growth rate terminology* information box presented below, growth rates are frequently expressed in terms of a rate of change over the previous year, quarter, month etc., for which the terms "month-on-previous month", "quarter-on-previous quarter" and "annual" growth rates are applied. For some phenomena, for example, human population growth, each of these growth rates may be further differentiated by the nature or pattern of the incremental difference between each successive period, and whether or not the absolute values of these increments over a given period are themselves constant or changing. The most commonly used patterns of growth are: arithmetic, geometric and exponential (Rowland 2003). These growth rates are important elements of population statistics where attempts are made to derive a mathematical model that accurately reflects the rate of growth (or decrease) in human populations. The main attributes of these patterns of growth described by Rowland (2003) are summarised below.

- Arithmetic growth occurs when the annual (or quarterly, monthly) growth rates or change commonly used in economic, social and population statistics increase (or decrease) by a constant number in each succeeding time period being analysed. Arithmetic change results in a linear trend, following a straight line rather than a curve when graphed. Although not commonly used for describing actual human population change, arithmetic growth rates are nevertheless commonly used in demographic and other analyses, e.g. the presentation of *average annual increases*.

- Geometric growth on the other hand refers to the situation where successive changes in a population differ by a constant ratio (as distinct from a constant amount for arithmetic change). In geometric growth incremental changes in population become larger, as increases (or decreases) are said to be "self-reinforcing". The compilation of monthly, quarterly, or annual interest payments using a compound interest formula is an example of the use of geometric change.

- Exponential growth refers to the situation where growth compounds continuously at every instant of time. Because compounding takes place at intervals much longer than an instant, Rowland (2003, p. 51) refers to geometric growth as being a "special case" of exponential growth. When graphed, an exponential graph shows a smooth curve without steps between increments because the change is continuous.

Measures of human population growth based on geometric and exponential growth are commonly used, each yielding similar results where either the time interval in short or the rate is within the typical range for national populations (Rowland 2003, p. 61). Variations of both the geometric and growth rate formulae are derived for calculating end-of-period populations and initial populations.

As will be shown in Section 4, the form of data (seasonally adjusted, trend-cycle, etc) and the time horizon are closely linked. For example, if a growth rate of a month in relation to the same month of the previous year is calculated then it would be inappropriate to use seasonally adjusted or trend-cycle data as the purpose of seasonal adjustment is to facilitate comparisons of different months in the same year.

3.4.2 Recommended terminology for growth rates

As mentioned above, there is considerable inconsistent use of labels attached to the various forms of growth rates by different agencies, particularly with respect to "annual growth rates", "annualised growth rates", "year-on-year change". In order to minimise the risk of misunderstanding a set of standard growth rate terminology is presented below. These relate to growth rates in common use for the presentation of economic, social and population statistics. The definitions are also presented in the Glossary in Annex 2.

Growth rate terminology		
Concept	**Definition** (M_t denotes the value of a monthly time series in month t and Q the value of a quarterly time series in quarter t.)	**Context**
Year-on-year (YoY) growth rates	Year-on-year growth rates are rates of change expressed over the corresponding period (month or quarter in relation to the frequency of the data) of the previous year. Such rates are expressed as $(M_t/M_{t-12})-1$ or $(Q_t/Q_{t-4})-1$.	Also often referred to as Year-over-year growth rates, Year-to-year growth rate, Rate of change from the previous year, or 12-month rate of change.
Year-on-year changes	Year-on-year changes are changes in levels expressed over the corresponding period (month or quarter in relation to the frequency of the data) of the previous year. Such changes are expressed as M_t-M_{t-12} or Q_t-Q_{t-4}.	Also often referred to as Year-over-year changes, or Year-to-year changes.
Month-on-previous-month growth rates	Month-on-previous-month growth rates are rates of change expressed with respect to the previous month. Such rates are expressed as $(M_t/M_{t-1})-1$.	Also often referred to as Month-to-month (Period-to-period) growth rates, Month-over-month growth rates, 1-month rate of change, or Rate of change on the previous month. For some phenomena, month-on-previous month growth rates may be further differentiated by the nature or pattern of the incremental difference between each successive period and whether or not the absolute values of these increments over a given period are themselves constant or changing. The most commonly used patterns of growth are: arithmetic, geometric and exponential – refer below
Month-on-previous month changes	Month-on-previous-month changes are changes in levels expressed with respect to the previous month. Such	Also often referred to as Month-to-month (Period-to-period) changes, or Month-over-month changes.

Growth rate terminology		
Concept	**Definition** (M_t denotes the value of a monthly time series in month t and Q the value of a quarterly time series in quarter t.)	**Context**
	rates are expressed as M_t-M_{t-1}.	
Quarter-on-previous-quarter growth rates	Quarter-on-previous-quarter growth rates are rates of change expressed with respect to the previous quarter. Such rates are expressed as $(Q/Q_{-1}) - 1$.	Also often referred to as Quarter-to-quarter (Period-to-period) growth rates, Quarter-over-quarter growth rates, 1-quarter growth rates, or Rate of change on the previous quarter. For some phenomena, quarter-on-previous quarter growth rates may be further differentiated by the nature or pattern of the incremental difference between each successive period and whether or not the absolute values of these increments over a given period are themselves constant or changing. The most commonly used patterns of growth are: arithmetic, geometric and exponential – refer below
Quarter-on-previous quarter changes Annual growth rates	Quarter-on-previous-quarter changes are changes in levels expressed with respect to the previous quarter. Such rates are expressed as $Q - Q_{-1}$. Annual growth rates are annual rates of change expressed over the previous year. Such rates are expressed as $(Y_t/Y_{t-1}) - 1$.	Also often referred to as Quarter-to-quarter (Period-to-period) changes, or Quarter-over-quarter changes. For some phenomena, annual growth rates may be further differentiated by the nature or pattern of the incremental difference between each successive period and whether or not the absolute values of these increments over a given period are themselves constant or changing. The most commonly used patterns of growth are: arithmetic, geometric and exponential – refer below
Annual changes	Annual changes refer to annual changes in levels expressed over the previous year. Such changes are expressed as $Y_t - Y_{t-1}$.	
Year-to-date totals	Year-to-date totals are data expressed in cumulative terms from the beginning of the year.	Sometimes referred to as cumulative totals or cumulative data.
Annualised growth rate (Annualised rate of change)	Annualised growth rates (Annualised rate of change) show the value that would be registered if the quarter-on-previous quarter or month-on-previous month rate of change were maintained for a full year. Such rates are expressed as $((Q_t/Q_{t-1})^4)-1$ for quarterly data and $((M_t/M_{t-1})^{12})-1$ for monthly data. Annualised growth rates facilitate comparison of data for different time periods (e.g. years and quarters) but may suffer from high volatility in a series.	In addition to the compounded form of the Annualised growth rate presented here, the term "Annualised growth rate" is sometimes used to describe the quarterly or monthly growth rate multiplied by four or twelve. Multiplying the quarterly or monthly growth rate by four or twelve is more appropriately referred to as "Linear approximation of the annualised growth rate". Some agencies use the expression "1-month rate of change, annualised", etc, for such rates.
Linear approximation of the annualised growth rate	Linear approximation of the annualised growth rate is a quick calculation of the annualised growth rate that show the levels that would be measured for a quarter or month were maintained for a full year. Quarterly rates of change are multiplied by 4 and monthly rates of change are multiplied by 12. Such rates are expressed as $4^* (Q - Q_{t-1}/Q_{t-1})$ for quarterly data or $12^* (M_t - M_{t-1}/M_{t-1})$ for monthly data. Such rates facilitate comparison of data for different time periods (e.g. years and quarters).	This approximation is valid if the growth rate is small, but the linear approximation of the annualised growth rate should not be used in official releases. The compound form of the Annualised growth rate should be preferred.

Growth rate terminology		
Concept	**Definition** (M_t denotes the value of a monthly time series in month t and Q_t the value of a quarterly time series in quarter t.)	**Context**
Annualised semi-annual growth rate	Annualised semi-annual growth rates show the value that would be registered if the rate of change measured with reference to two quarters or six months behind were maintained for a full year. Such rates are expressed as $((Q_t/Q_{t-2})^2)-1$, $((M_t/M_{t-6})^2)-1$. Annualised semi-annual growth rates facilitate comparison of data for different time periods (e.g. years and quarters).	
Arithmetic growth	Arithmetic growth refers to the situation where a population increases by a constant number of persons (or other objects) in each period being analysed. In general terms an arithmetic growth rate may be expressed as: $\left(\frac{Pn - Po}{n}\right) \div Po \times 100$ Where P_o = population at the start; P_n = population at the end; n = number of intervals between P_o and P_n (Rowland 2003)	Arithmetic growth rates may take the form of annual growth rates, quarter-on-previous quarter growth rates or month-on-previous month growth rates – see above.
Geometric growth rate	Geometric growth refers to the situation where successive changes in a population differ by a constant ratio (as distinct from a constant amount for arithmetic change). In general terms a geometric growth rate may be expressed as $\sqrt[n]{\frac{Pn}{Po}} - 1$ Where P_o = population at the start; P_n = population at the end; n = number of intervals between P_o and P_n (Rowland 2003)	Geometric growth rates may take the form of annual growth rates, quarter-on-previous quarter growth rates or month-on-previous month growth rates – see above. The geometric growth rate is applicable to compound growth over discrete periods, such as the payment and reinvestment of interest or dividends. Although continuous growth, as modelled by the exponential growth rate, may be more realistic, most economic phenomena are measured only at intervals, in which case the compound growth model is appropriate. Like the exponential growth rate, the geometric growth rate does not take account intermediate values of the series. (World Bank 2004)
Exponential growth rate	Exponential growth refers to the situation where growth compounds continuously at every instant of time. Because compounding takes place at intervals much longer than an instant, a geometric growth as being a "special case" of exponential growth. In general terms an exponential growth rate may be expressed as $\frac{In\left(\frac{Pn}{Po}\right)}{n}$ Where P_o = population at the start; P_n = population at the end; n = number of intervals between P_o and P_n; In =	Exponential growth rates may take the form of annual growth rates, quarter-on-previous quarter growth rates or month-on-previous month growth rates – see above. The exponential growth rate is the growth rate between two points in time for certain population indicators, notably labour force and population. This growth rate is based on a model of continuous exponential growth between two points in time. It does not take into account the intermediate values of the series. Nor does it correspond to the annual rate of change measured at a one-year interval, which is given by:

Growth rate terminology		
Concept	**Definition** (M_t denotes the value of a monthly time series in month t and Q the value of a quarterly time series in quarter t.)	**Context**
	natural logarithm (Rowland 2003)	$(P_n - P_{n-1})/P_{n-1}$ (World Bank 2004)
Least-squares growth rate	The least-squares growth rate, r, is estimated by fitting a linear regression trend line to the logarithmic annual values of the variable in the relevant period. The regression equation takes the form: $\ln X_t = a + bt$ Which is equivalent to the logarithmic transformation of the compound growth equation, $X_t = X_o(1 + r)^t$ In this equation X is the variable, t is time, and $a = \ln X_o$ and $b = \ln(1 + r)$ are parameters to be estimated. If b^* is the least-squares estimate of b, then the average annual growth rate , r, is obtained as $[\exp(b^*) - 1]$ and is multiplied by 100 for expression as a percentage. The calculated growth rate is an average rate that is representative of the available observations over the entire period. It does not necessarily match the actual growth rate between any two periods. (World Bank 2004)	Least-squares growth rates are used whenever there is sufficiently long time series to permit a reliable calculation. No growth rates are calculated if more than half the observations in a period are missing.

3.4.3 Recommended practices for the reporting and presentation of growth rates

The OECD task force undertook a comprehensive review of the issues involved in the presentation of the different types of growth rates for short-term indicators in the dissemination of indicators in press releases, in tables posted on the website of statistical institutions, etc. Any particular rate of change can be useful depending on the needs of a specific analyst. The focus of task force investigations was the identification of the most suitable way(s) of presenting economic indicators to the general public, in order to prevent misunderstandings in their reading of economic events. The recommendations prepared by the task force for the presentation of growth rates are presented below in Section 4.6 in the context of the recommendations of different forms of data. The types of growth rates covered in these recommendations comprise:

- rate of change with respect to previous period;
- the use of annualised growth rates; and
- rates of change with respect to the same period of the previous year.

[Identify World Bank and other international organisation preferences for different patterns of growth rate.]

3.5 Ratios, proportions, percentages and rates

3.5.1 Introduction

A ratio is a single number that expresses the relative size of two other numbers. They are used widely in social and population statistics[12] for analyzing the composition of a set of events and are usually calculated for subgroups of a population. Examples of such ratios include: sex ratio; ratio of average female wage to male wage; primary school enrolment ratio; house price to income ratio; etc. A proportion is a special type of ratio in which the denominator includes the numerator. A percentage is a special type of proportion where the ratio is multiplied by a constant, 100, so that the ratio is expressed per 100 (Palmore and Gardner 1994).

Ratios, and percentages are used for analyzing the composition of a set of variables or events of a population. On the other hand, a rate refers to the occurrence of events over a given interval in time and are commonly used to analyze the dynamics of change in social and population statistics. Rates are measures that reflect the frequency of an event relative to the population that may experience that event (McManus 2004). Rates may be presented either as a proportion or percentage of a total population (e.g. adult literacy rate as a percentage of the adult population aged 15 years and over) or where the total population is expressed as unity (e.g. number of births per 1 000 population). Examples of such rates include: unemployment rate; birth rate; death rate; net migration rate; total fertility rate; adult literacy rate; infant mortality rate; net reproduction rate.

The relationship between ratios, proportions, percentages, rates, etc, is summarised below in a typology taken directly from Palmore and Gardner 1994.

Terminology for ratios, proportions, percentages and rates

Concept	Definition	Context
Ratio	A ratio is a number that expresses the relative size of two other numbers. The result of dividing a number X by another number Y is the ratio of X to Y, i.e.: $\frac{X}{Y}$ = ratio of X to Y	
Proportion	A proportion is a special type of ratio in which the denominator includes the numerator. An example is the proportion of deaths that occurred to males which would be deaths to males divided by deaths to males plus deaths to females (i.e. the total population).	More common usage of the term "proportion" is in the context of a portion or part in its relation to the whole (OED 1989).
Percentage	A percentage is a special type of proportion where the ratio is multiplied by a constant, 100, so that the ratio is expressed per 100.	
Percentage change	A percentage change on the other hand refers to a change from one period to another expressed as a percentage of its value in the first of the two periods.	

[12] The term "demography" refers to the study of the size, distribution, structure and growth (or decline) of populations. In order to incorporate other characteristics of the population such as ethnic, health, social and economic characteristics, the term "population statistics" has been used in this Handbook in lieu of the term "demographic statistics". (Siddiqui 1999)

Terminology for ratios, proportions, percentages and rates		
Concept	**Definition**	**Context**
Rate	A rate refers to the occurrence of events over a specific interval in time. Similarly, a rate refers to the measure of the frequency of some phenomenon of interest given by: Rate = number of events in a specified period / average population during the period (Everitt 1989)	Caution must be used with the term "rate" as it is sometimes applied to ordinary percentage changes such as a "literacy rate" which is the percentage of a population that is literate. Different constants (commonly 100, 1 000, 100 000) are used in the presentation of different rates (e.g. crude death rates and crude birth rates are usually expressed per 1 000). Palmore and Gardner (1994) recommend that when calculating a rate to proceed without the use of a constant until the final answer is derived and then use the constant to express the rate per 100, 1 000 or whatever is the usual constant for that type of rate. [Identify any conventions in this area.]
Probability	A probability is similar to a rate, the difference being that the denominator comprises all those objects in a given population at the beginning of the period of observation. For example: If 10 people die in one year out of a population of 1 000 at the start of a year, the probability of dying during that year was 10/1 000, or 0.01000.	On the other hand the denominator of a rate is frequently the average or mid-period population exposed to the event in question.

Source: Palmore and Gardner 1994, Ch. 2

Ratios, proportions, percentages and rates are frequently used to compare events both over time and between countries and are widely used in statistics published by United Nations agencies. For example, the extensive set of indicators for monitoring the United Nation's Millennium Development Goals (refer UNDG 2003) contains many of these types of data transformations. A key source of component series used for the derivation of ratios, rates, etc, commonly used in population statistics are registration data for vital events such as births and deaths and other administrative sources for information on migration, health, education, etc. In countries where civil registration of vital events is incomplete, especially in rural areas, key demographic information are obtained from on-going population surveys and periodic population censuses.

Data sources for component series used in the derivation of rates and ratios

There is often a serious gap between quantitative information essential for estimating demographic rates and ratios (e.g. mortality and fertility) in countries and the amount and quality of the data actually available. Countries are making serious efforts to establish registration of births and deaths, while collecting vital statistics from surveys in the interim.

The basic, or direct method, for estimating rates and ratios from surveys is through a retrospective enquiry about birth histories of women of reproductive age and a series of detailed questions. The major problem with this method is the accuracy of reported dates and numbers, especially among older women in the sample. Indirect methods for estimating fertility and mortality were developed to counter this problem. It is based on questions regarding children ever born and children surviving, asked of all women of reproductive age. Methods described in Brass 1975 and UNSC 1983 are applied to these data to estimate period rates 5 or 10 years preceding the survey. Estimates of infant and child mortality and life

expectancy at birth require application of model life tables.

The indirect "sisterhood" method for estimating maternal mortality is based on four questions asked of all adults in the household canvasses (how many sisters, how many alive and dead, and deaths during pregnancy, childbirth or six months post-partum). The approach is analogous to the well established Brass technique for estimating child mortality based on children ever born and surviving.

In order to facilitate the comparison of rates, ratios, percentages, etc, for two or more different geographic areas (regions, countries, etc) it is necessary to "standardise" data to remove the impact of differences in the composition of extraneous variables. Palmore and Gardner (1994, p. 20) refers to the process of removing or controlling the effects of these other variables that could "confound" the comparison. These authors cite the example of a comparison of crude death rates of two populations if they had exactly the same age distribution but each retained their own age-specific death rates. The comparison would hold constant or control the impact of different age distributions in the two populations so that any differences in death rates would result from actual differences in specific mortality rates, rather than differences in the age variable.

Rates, ratios or percentages compiled for population statistics could need to control for differences in a range of extraneous variables used in the composition of rates, etc., for two or more different populations. These include differences in urban or rural residence, occupations, income distributions, gender, marital status. The control of similar differences is relevant in other fields of statistics such as economic statistics. Examples of the standardisation processes are given in Palmore and Gardner (1994, p. 21).

3.5.2 Recommended practices for the presentation of 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 the OECD's Glossary of Statistical Terms (OECD 2002a); Eurostat's CODED (Eurostat 2003a), 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 above in Section 3.5.1, 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.

4. GUIDELINES FOR THE REPORTING OF DIFFERENT FORMS OF DATA

4.1 Introduction

The different forms of time series data discussed in this Handbook are: raw (or original data); working day adjusted; seasonally adjusted; and trend-cycle economic, social and population statistics. There is a wealth of references about the compilation of the different forms of data, and a detailed explanation of methodological issues arising out the different approaches to, for example, seasonal adjustment method, is outside the terms of reference of this Handbook. The focus of this Section is therefore restricted to the reporting and presentation of the different forms of data and the brief introduction below is merely intended as a lead in to reporting and presentation issues.

The main reason for compiling high frequency series such as monthly or quarterly indicators is to form a time series which may be used to monitor level changes and the volatility of those series over time. Time series are of interest to analysts because they are useful for the identification of the position of the economy (or specific sectors) over the business cycle and, more specifically, with respect to turning points in those cycles. An original time series (also known as "raw data") may be decomposed into three basic components:

- Trend-cycle: the underlying path or general direction reflected in the data over the longer term, i.e. the combined long-term (trend) and medium-to-long-term (cycle) movements in the original series.

- Seasonal variations: include seasonal and other systematic effects. Seasonal effects are reasonably stable in terms of annual timing, direction and magnitude. The causes of such effects are natural factors, administrative or legal measures and social traditions. Other effects on time series may be due to variations in the number of working days or trading days in a period, or events that occur at regular intervals such as pay days for large groups of employees, pension payments, etc. Both the seasonal and other effects represent persistent, predictable calendar-related effects.

- Irregular variations: are effects that are very often unpredictable in terms of timing, impact and duration. These may be the result of sampling and non-sampling errors (refer Section 6.4 below), unseasonable weather changes, natural disasters, strikes and socio-economic changes.

It should be emphasised that seasonally adjusted and trend-cycle estimates represent an analytical massaging of the original time series. Furthermore, there is to some extent an amount of subjectivity in the choice between the various options available for the estimation of trend-cycle and seasonally adjusted series. Both seasonally adjusted and trend-cycle estimates complement the original data and can never replace the original series. The non-seasonally adjusted data shows the actual changes that have taken place (subject to the impact of sampling and non-sampling errors) and the seasonally adjusted and trend-cycle estimates represent an analytical elaboration of the data showing the underlying developments (IMF 2001, para. 8.12). The various packages available for seasonal adjustment or trend-cycle analysis will not remove any underlying deficiencies that may be inherent in the basic data[13]. Such elaborations should therefore not be built into the original data compilation process but should be undertaken after the original data has been compiled.

[13] Such as deficiencies in coverage, classifications and definitions, collection practices, compilation practices.

In their *Quarterly National Accounts Manual – Concepts, Data Sources, and Compilation*, the IMF further states that the estimation of seasonally adjusted data exclusively represents a loss of seasonal information to the user. Furthermore, there is no unique solution on how to conduct seasonal adjustment and that seasonally adjusted data are also subject to revisions as future data becomes available even where the original series are not revised. Finally, the IMF states that although errors in source data may be more readily detected in seasonally adjusted series, the source of the error and their correction may be easier through working with unadjusted data (IMF 2001, para. 8.12).

Finally, there is continuing debate among statisticians on which is the most appropriate form for the presentation of a time series to users – original , seasonally adjusted or trend-cycle. The outcome of the discussion is that there is generally no absolute ideal, and as will be discussed below , the final choice depends on the media for the dissemination of data and the main focus or intent of the series. For the former, dissemination of detailed data via an on-line database would imply the availability of original series which affords maximum flexibility to users, whereas dissemination of more aggregated and headline series in a press release would involve the presentation of seasonally adjusted or trend-cycle series, perhaps in addition to original series .

The choice between the presentation of trend-cycle and seasonally adjusted estimates depends on whether the intention of the agency disseminating the data is to place more attention on the underlying movement of the time series or on movements and incident variation. If the former , trend-cycle estimates for year-on-year data would be the preferred form. If the focus is on period-to-period changes in the most recent data then seasonally adjusted estimates would be preferred. Here again, there are no absolutes and to a large degree the underlying "dissemination philosophy" of the agency is an important influence in the final choice.

4.2 Terminology

The following terminology for terms relating to different forms of data are provided in order to ensure a common understanding of the concepts described in subsequent Sections of this Handbook. The terminology presented flows out of recent work of the OECD Short-term Economic Statistics Expert Group (STESEG) task force on data presentation and seasonal adjustment.

Key terminology relating to time series analysis		
Concept	**Definition**	**Context**
Time series	A time series is a set of time-ordered observations on a quantitative characteristic of an individual or collective phenomenon taken at successive, in most cases equidistant, periods of time. .	A time series (TS) can be decomposed into unobservable components. In the most complete case, these components are the trend (T), the cyclical (C), the seasonal (S) and the irregular (I) components. The four components of the time series may each be independent of all the others, in which case the behaviour of the time series is simply the sum of `the components which are additively related (i.e. TS = T+C+S+I). However, most analysts believe that it is unlikely that the time series components are perfectly independent of each other, and are therefore more likely to be multiplicatively related (i.e. TS=T*C*S*I).
Trend component of a time series	The trend is the component of a time series that represents fluctuations of low frequency in a time series,	The trend is normally referred to as the long-term movement in a cyclical context (i.e. the trend

Key terminology relating to time series analysis		
Concept	**Definition**	**Context**
	the high and medium frequency fluctuations having been filtered out. This component can be viewed as those variations with a period longer than a chosen threshold (usually 8 years is considered as the maximum length of the business cycle).	fluctuations have a longer period than the maximum duration of the business cycle). In practice, statistical agencies do not estimate the trend but rather focus on the trend-cycle component (see Trend-cycle).
Trend-cycle	The trend-cycle is the component that represents fluctuations of low frequency in a time series, the high frequency fluctuations having been filtered out. This component can be viewed as those variations with a period longer than a chosen threshold (usually 1½ years is considered as the minimum length of the business cycle).	In practice, statistical agencies estimate trend-cycle by estimating and removing the seasonal and irregular components from the original non-adjusted data
Cyclical component of a time series	The cyclical component of a time series refers to (regular or periodic) fluctuations around the trend, excluding the irregular component, revealing a succession of phases of expansion and contraction. The cyclical component can be viewed as those fluctuations in a time series which are longer than a given threshold, e.g. 1½ years, but shorter than those attributed to the trend.	
Seasonal component of a time series	The seasonal component is that part of the variations in a time series representing intra-year fluctuations that are more or less stable year after year with respect to timing, direction and magnitude.	The seasonal component is also referred to as the seasonality of a time series. The seasonal component reflect "normal" variations that recur every year to the same extent, e.g. weather fluctuations that are representative of the season, length of months, Christmas effect. The seasonal component may also includes calendar related systematic effects that are not regular in their annual timing or are caused by variations in the calendar from year to year.
Calendar effects component of a time series	The calendar effects component is that part of the seasonal component which represents calendar variations in a time series, such as trading days / working days, moving holidays and other calendar effects (such as leap year). The effects of the normal length of a month are assigned to the seasonal component.	The calendar component is often slightly moving and may disturb the stability of the seasonal component.
Irregular component of a time series	The irregular component of a time series is the residual time series after the trend-cycle and the seasonal components, as well as calendar effects, have been removed. It corresponds to the high frequency fluctuations of the series.	The irregular component results from short term fluctuations in a series which are not systematic and in some instances not predictable e.g. uncharacteristic weather patterns. Some irregular effects can however be expected in advance, e.g. changes in value added tax. In a highly irregular series, these fluctuations can dominate movements, which will mask the trend and seasonality. (ABS)
Seasonal adjustment	Seasonal adjustment is a statistical technique to remove the effects of seasonal and calendar influences operating on a series. Seasonal effects usually reflect the influence of the seasons themselves either directly or through institutional factors or social conventions.	Other types of calendar variation occur as a result of influences such as the number of days in the calendar period, the accounting or recording practices adopted or the incidence of moving holidays (such as Easter and Chinese New Year).

Key terminology relating to time series analysis		
Concept	**Definition**	**Context**
Seasonally adjusted time series	Seasonally adjusted time series are series that have been adjusted for seasonal variations, including trading-day (working-day) effects and other regular calendar variations if present.	
Seasonally adjusted component of a time series	The seasonally adjusted component is the result of the extraction of the seasonal component and the calendar effects component from a time series. If neither seasonal nor calendar influences are present in the original data, the seasonally series is given by the original data. For series with no identifiable seasonal variations but with identifiable calendar variations, the seasonally adjusted series is given by the calendar adjusted series.	Trading / working day corrections are alternative ways to normalise the time series.
Calendar adjustment	Calendar adjustment refers to the correction for calendar variations other than seasonal components. Such calendar adjustments include working day adjustments or the incidence of moving holidays (such as Easter and Chinese New Year).	The terms "calendar adjustment" and "working day adjustment" (also known as "trading day adjustment") are often used interchangeably. However, the main difference between the two terms is that working day adjustment is merely one type of calendar adjustment. Another type of calendar adjustment is a "shopping day" adjustment for consumer expenditure or retail trade series. Each variable has its own specific calendar adjustment. The length of month effect is assigned to the seasonal component because it happens year after year in the same period with the exception of leap year effects.
Working / trading day adjustment	Working day or trading adjustments refer to the correction for differences in the number of working or trading days in a given month or quarter which differ from year to year which will impact upon the level of activity in that month or quarter for flow series or the sort / type of day for stock series. In most countries working day adjustment and trading day adjustment are used as synonyms.	The number of working or trading days in a given month or quarter can vary significantly for each statistical domain (e.g. production, merchandise trade) because of differing institutional arrangements, trade specific holidays, etc. Some countries also include bridging effects in working day adjustments. These result from people taking holidays, for example, on Mondays and Fridays when an official public holiday occurs on Tuesdays and Thursdays respectively. The type of working / trading day adjustment carried out needs to be tailored to the cultural and institutional environment operating within individual countries. In the United States, for example, working day adjustment classifies the days of the week into workdays (Monday through to Friday) and non-workdays (Sat. and Sun.) and thus the seasonal adjustment estimates two factors – workday and non-workday. Trading day adjustment (as performed in the US at least) allows for a different effect for each day of the week and computes seven factors – a Monday factor, a Tuesday factor and so on. Essentially, trading day adjustment is a more fine tuned seasonal adjustment method to account for calendar variation. In other countries there may be very distinct differences between workdays and non-workdays in the amount of business conducted, but not so

Key terminology relating to time series analysis		
Concept	**Definition**	**Context**
		much difference between Mondays, Tuesdays, through the Fridays. The US on the other hand, has distinctly different patterns of activity each day of the week and methods used in that country try to account for that.
Moving holidays	Moving holidays are holidays which occur each year, but where the exact timing shifts. Examples of moving holidays include Easter and Chinese New Year. Easter generally falls in April but it can also fall in late March.	Also known as variable holidays
Moving average	A moving average is a method for smoothing time series by averaging (with or without weights) a fixed number of consecutive terms. The averaging "moves" over time, in that each data point of the series is sequentially included in the averaging, while the oldest data point in the span of the average is removed. In general, the longer the span of the average, the smoother is the resulting series.	Moving averages are used to smooth fluctuations in time series or to identify time series components, such as the trend, the cycle, the seasonal, etc. A moving average replaces each value of a time series by a (weighted) average of p preceding values, the given value, and f following values of a series. If $p = f$ the moving average is said to be centered. The moving average is said to be symmetric if it is centered, and if for each $k = 1, 2, \ldots, p = f$, the weight of the k-th preceding value is equal to the weight of the k-th following one. The moving average is not defined for the first p and the last f time series values. In order to compute the moving average for those values, the series must be backcasted and forecasted.

4.3 Working day adjusted data

Many monthly time series contain variations which result from calendar-related systematic effects that are not regular in annual timing and are caused by variations in the calendar from year to year. Such calendar variations include (IMF 2001, para. 8.7) the:

- trading day effects which result from variations from year to year in the number of working days or trading days and the week day composition of a particular month or quarter relative to the "standard" for that particular month or quarter[14]. Such differences arise from factors such as the number of Saturdays and Sundays in a month[15]; and

- the effects of events that occur at regular intervals but not necessarily at the same time each year such as official holidays and regional official holidays, in particular, moving holidays such as Easter, Chinese New Year and Ramadan. Other similar events include paydays for large groups of employees, pension payments, etc.

[14] Stock series should not experience trading day effects as such series only measure the level of activity at a certain point in term and therefore are not affected by how many trading days there are in a given period of time (ABS 2003).

[15] Note that the length of month effect is assigned to the seasonal component – refer to "Calendar adjustment" in Key time series terminology above.

Variations due to these factors can have a significant impact on a time series and may obscure important movements in the series and should be adjusted for. However, working day and trading day effects are part of the overall seasonal variation in a time series and any adjustment of a series for them should be regarded as an integral part of the seasonal adjustment process and not as a separate process (IMF 2001, para. 8.30). However, in terms of timing, working day adjustment is normally undertaken prior to the seasonal adjustment of a series.

4.4 Seasonally adjusted data

Most infra-annual time series show intra-year variations which recur regularly every year, possibly slowly evolving, though normally stable with respect to timing, direction and magnitude. In order to gain insight into the current developments measured by a seasonally varying time series, it is necessary to correct it for these regularly repeating intra-year variations. For example, if a monthly time series decreases every July because of holidays, there is little to be gained by noting that it decreased once more last July as it always does. Analysts are interested in whether the last decrease itself was larger or smaller than usual; furthermore, the removal of regularly repeating seasonal effects from a time series also facilitates geographic and sectoral comparisons.

Another issue concerning analyses of time series subject to seasonal influences is that of obtaining statistically meaningful comparisons of different periods within the same year; for example, comparing January sales levels to say that of September. Seasonal variations, irrespective of their causes, are a reflection of the fact that each period has its own basis of comparison across the years that differ from those of the other periods. Hence, direct comparisons of periods of the same year for seasonal time series is not meaningful in say a press release where the focus of the discussion for many agencies is on the most recent data. Hence the need for seasonal adjustment.

There are numerous seasonal adjustment methods which vary in sophistication. At one end is the simple year-on-year change (or percentage change), and at the more sophisticated end are techniques such as X-12 RegARIMA (Findley et al. 1998) and Tramo-Seats (Gomez and Maravall 1996), BV4, SABLE, DAINTIES, STAMP, etc. For the majority of time series, only the more sophisticated techniques can produce series adequately seasonally adjusted for most purposes. However, because seasonal adjustment can only be achieved through a model of the seasonal behaviour of the series to be adjusted, and since each of the best seasonal adjustment methods encapsulates a broad but restricted family of seasonal models, there is not a unique method that would be applicable to all series. Again, the choice of the most appropriate seasonal adjustment software is outside the scope of this Handbook.

Almost every national statistical agency, central bank and international organization compile or at least disseminate seasonally adjusted data for many of their infra-annual time series, though practices vary with regards to how such series are presented to users and the amount and content of metadata describing the treatment of the series. The work of the OECD Short-term Economic Statistics Expert Group (STESEG) task force on data presentation and seasonal adjustment focused on the issue of how to report seasonally adjusted data, and which information on the seasonal adjustment method and what kind of specification details should be disseminated or made available to users. The recommendations around these issues in Section 4.6 below are also discussed in the content of how the different types of data (particularly, growth rates) are presented to users.

The framework of task force work was set by the identification of three types of users: the general public, informed users, and analytical users. The needs of each group are very different and to meet them statistical agencies need to provide differentiated sets of metadata. The information required by each group entails:

- general public – require "basic" metadata on seasonal adjustment, i.e. a layperson's explanation of the adjustment. Such users are generally not interested in more technical information (such as diagnostics of the procedure);

- informed users - need detailed information on how the statistical program performing the seasonal adjustment was carried out, as well as statistics on the validity of the adjustment for specific series. For

this category of users, statistical agencies generally provide at least one comprehensive document per statistical program;

- analytical users - need some of the results of the statistical program to reprocess them for their own use(s). Such users should be supplied with all the relevant parameters needed to replicate the official seasonal adjustment, and to modify it if they wish.

4.5 Trend-cycle data

The trend component of a time series referred to briefly in the introduction to in Section 4.1 above refers to the long term movements of a series. The trend reflects the underlying movement of the series and is typically due to influences such as population growth, price inflation and general economic development. For the purposes of seasonal adjustment both the long-term trend and medium-term cycles are treated as the trend component of a time series. For this reason, the trend component is sometimes referred to as the trend-cycle (ABS 2001, Section 2.1).

The main issues with respect to the reporting and presentation of trend-cycle data centre on whether or not to produce trend-cycle series for certain types of indicators, rather than to which techniques should be used to produce them. The following points are considered relevant:

- Should trend-cycle estimates be released for volatile series, or not at all (i.e. filtering should be left to users)?

- How to deal with revisions derived from filtering techniques (and with the related end-point problem).

- If volatility is the key feature to discriminate among indicators, perhaps a standard should be set in order to define volatility.

National practice varies considerably across countries with regard to the availability of trend-cycle estimates. For example, the Australian Bureau of Statistics, in addition to raw data and seasonally adjusted data, releases trend data for all series where the irregular component has been filtered out of the seasonally adjusted data. The German Federal Statistical Office publishes the trend-cycle component and the irregular component of the industrial production index. The Korean National Statistical Office produces trend-cycle series as reference material for Composite Indices of Business Indicators. In Italy, ISTAT does not release filtered time series for any indicator although in some press releases, three-term moving averages are added to graphs but with no values or comments. Other countries such as France do not publish trend-cycle data at all.

Much of the discussion on trend-cycle analysis focuses on the end-point problem. Significant differences can arise following revisions to provisional data from which trend data are derived. Since the trend-cycle values at the most recent end of the series are usually estimated by extrapolation, the underlying trend of the most recent series should be presented to users with caution. Data can even change from positive to negative values or vice versa. In this regard particular care is required at turning points where it often takes months until the new correct direction of development appears. Therefore, this problem should be made very clear when publishing trend-cycle data at the current end of the time series. [Insert more context with regard to these issues for seasonally adjusted series which also require revision.]

4.6 Recommended practices for the reporting and presentation of different forms of data

This Section commences with a brief discussion and recommendations on the different forms of data (raw, seasonally adjusted, or trend-cycle) that should be presented, followed by discussion and recommendations on some of the more common transformations applied to seasonally adjusted figures (in particular, the different growth rates described in Section 3.4 above), and finally what information should be provided about seasonal adjustment for each of the three categories of users identified above. The discussion and recommendations use terminology and definitions provided in Sections 3.4.2 and 4.2 above.

4.6.1 `Forms of data to be presented

In the main, the presentation of seasonally adjusted data concerns members of the general public. Since the most appropriate seasonal adjustment techniques are very sophisticated statistical transformations it is not reasonable to expect the general public to have the ability to perform these transformations on their own. Therefore, the general public should expect statistical agencies to seasonally adjust as appropriate the main sub-annual indicators that appear in press releases, etc.

A subtlety is the distinction between the adjustment for seasonal variations per se, and the other adjustments such as those for trading-days and moving holidays such as Easter. Within the statistical community, it is generally understood that "seasonally adjusted" includes all these adjustments unless otherwise specified, not just those for seasonal variations. This definition is also the most widely used by statistical agencies. This aspect is reflected in the definition of a seasonally adjusted time series provided (in Section 4.2) above.

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.

The question now arises as to whether or not the original unadjusted (or original) data should be presented together with the seasonally adjusted figures in data disseminated to the general public, e.g. in press releases? Proponents for presenting the two versions generally point out that many users are interested in seeing the original series, whereas those not in favour of presenting both versions assert that there is a possibility for users to be confused about what is the correct information. The OECD task force supported both points of view. It noted however that if centre stage is given to the seasonally adjusted or trend-cycle estimates through their use in any main features text or broad analyses accompanying the statistics, the risk of confusion is greatly reduced.

By the same token, if any of the intermediate components of a seasonally adjusted series is presented (e.g. the series only adjusted for trading-days) in addition to the unadjusted and seasonally adjusted versions, then the risk of confusing the general public is very real. In addition, it is likely that the statistical agency would then have to provide some explanations about the intricacies of seasonal adjustment with the release, increasing the risk of burying the essential information.

As mentioned above in Section 4.1, the issue of the choice between the presentation of seasonally adjusted or trend-cycle estimates to the general public depends to some extent on the intended focus of the

disseminated time series. If the focus is on underlying medium term movements then trend-cycle estimates would be the preferred form, if on incremental changes (including those in the most recent series) then seasonally adjusted series would be preferred. The OECD task force believed that the focus of press releases in most countries is more likely to be changes in most recent data, though it accepted that this view is not universally shared by all national statistical agencies. The following recommendation was formulated in this context.

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 data.

4.6.2 Analytical transformations

The statistical information reported in press releases and similar forms of data dissemination is of necessity limited, and concentrates on the meaning of the results of a statistical program. In order to help the public at large assimilate this information, some simple transformations are generally presented such as the compilation of the various growth rates presented in Section 3.4 above. If these transformations are not provided in disseminated data, users will of necessity compute them. Hence the following discussion and recommendations which deal with the most common of such 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 or period-to-period (e.g. month-on-previous-month, quarter-on-previous-quarter) change in levels.

Month-on-previous-month and quarter-on-previous-quarter growth rates for original series are not very informative unless seasonal effects are negligible. For this reason statistical agencies seldom use them in their releases of indicators affected by seasonal fluctuations. The growth rate on seasonally adjusted series (or for original data where seasonal factors are of no significance), conveys the most recent information contained in a time series (trend-cycle and irregular movements) and is the best way of presenting short-term developments, even if the irregular component is relatively large. To deal with irregular movements that blur the trend-cycle the rate of change based on two or three month's worth of values[16] can be utilised. This practice, which is customary in some countries, is a convenient (and transparent) way of quantifying the short term movements averaging out a reasonable part of the irregular component.

[16] For example, $(X_t + X_{t-1}) / (X_{t-2} + X_{t-3}) * 100 - 100$ or $(X_t + X_{t-1} + X_{t-2}) / (X_{t-3} + X_{t-4} + X_{t-5}) * 100 - 100$.

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.

The change from previous year (referred henceforth as year-on-year change (YoY)) can be misleading in assessing the cyclical movements of an indicator, due to the compounding of movements over a 12 month span. However, its utilisation is very common in the current practice of users and media. Where necessary, special effects contained in the base period should be highlighted when presenting YoY changes (base effect).

YoY changes should be applied to raw data and to data adjusted for calendar effects if the latter are available. In this way the trading day effects are made clear. Technically, it would not be incorrect to advise against the utilisation of YoY changes on seasonally adjusted data. In particular, when the seasonal component is not deterministic, the rate of change on raw and seasonally adjusted data can be different, conveying conflicting signals, leading the general public and even some informed users to question the validity of the results. However, YoY change calculated on seasonally adjusted series is a very common practice.

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).

In some countries the levels of some seasonally adjusted flow figures are presented at "annual rates", being multiplied by 12 (for monthly series) or 4 (for quarterly series)[17]. This practice is not very common, and is largely restricted to the presentation of sub-annual national accounts estimates by Canada, Japan, Mexico and the United States [18]. The main stated advantage of this practice is that it facilitates the comparison between series of different periodicities (monthly, quarterly and annual). However, given its limited use, it is more of a hindrance when comparing seasonally adjusted figures from different programs within a centralized statistical system such as Statistics Canada, Australia, etc.

Annualising the change of a single month or quarter can result in misleading signals, particularly for series displaying high volatility. Similarly, the notion that level data could be annualised on the basis of seasonally adjusted and calendar adjusted time series which contain only minor irregularities should be treated with caution as it implies that the annualised rates of change should be presented only in those cases where such conditions are considered to hold and to be suppressed in other cases. Changing data presentation like this could call into question the integrity of both the data and the agency producing them. In turn, proposing a minimum length for the period to be annualised (for instance, six months), while

[17] Such series are more appropriately referred to as the "Linear approximation of the annualised growth rate". The compounded form is referred to as the "Annualised growth rate". Refer terminology in Section 5.2 above.

[18] Although Statistics Canada does not present annualised series (apart from monthly-GDP), the Canadian Mortgage and Housing Corporation presents the number of housing starts prepared in annualised form.

correct in principle, seems not very worthwhile in practice, as press releases and other dissemination formats seldom allow for such a kind of data transformation. Furthermore, annualizing semi-annual estimates may also lead to lags in identifying turning points. For these reasons the OECD task force recommended against the use of annualised level changes.

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.

The question now arises as to whether or not period-to-period growth rates should be annualized? In Canada, annualized growth rates are only presented for quarterly estimates of GDP based on the income and expenditure approaches. The monthly growth rates of GDP by industry, the other major sub-annual program that is part of the Canadian System of National Accounts, are not. Also, this practice is not widespread among OECD Member countries, and in countries presenting such rates there are differences in the focus given to them. For example, in the United States the Bureau of Economic Analysis emphasizes the annualized quarterly growth rate of GDP in its press releases[19], whereas in Canada it is the quarterly growth rate itself that is emphasized.

There are two main justifications for the presentation of annualised growth rates. The first is that it provides a forecast for the annual growth rate. The other is that it provides a rate that is interpretable by users on an annual basis similar to, say, the unemployment rate or an interest rate. The first justification for such rates as forecasts is to be rejected as it is only applicable to the first period of the year. The second justification does though reflect a genuine user need. However, annualizing has a very negative aspect in that it exaggerates the volatility of the period-to-period growth rates. It is for this reason that annualizing monthly growth rates is very rarely seen, and is not appropriate.

The task force concluded with the following recommendation.

Recommendation 7

Annualized period-to-period growth rates are not recommended for the presentation of annual or monthly growth rates at annual rates.

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

4.6.3. Information about seasonal adjustment to be provided to users

The general public has an interest in understanding what seasonal adjustment is all about. However, given the sophisticated nature of seasonal adjustment methods, it is not reasonable to expect such users to possess the mathematical and statistical background to understand a technical description of any particular adjustment method.

[19] See http://www.bea.gov/bea/newsrel/gdp103a.pdf.

Accordingly, statistical agencies should provide metadata on seasonal adjustment in the form of a layperson's explanation of the seasonal adjustment process and how seasonally adjusted series should be interpreted. An example of such metadata is that provided by Statistics New Zealand in their user paper, *Seasonal Adjustment Within New Zealand (*Statistics New Zealand 2001*)*.

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.

The task force then formulated the following recommendation for informed users.

Recommendation 9

For the benefit of users requiring information about the validity of the seasonal adjustment method applied, statistical agencies should provide a minimum standard of information that would facilitate an assessment of the reliability of each seasonally adjusted series.

Finally, for analytical users, the task force believed that no additional elements than those provided for informed users and presented in the publication (whatever its format) dedicated to the statistical program need be added. However, for analytical users, the availability of metadata is of paramount importance.

The main elements of this metadata could include the following: a short standardized description of the method used, all the main parameters of the adjustment (e.g. additive versus multiplicative decomposition model), and some of the derived information (e.g. the trading-day weights). The principle to be followed is that the metadata should be of sufficient extent to enable an analytic user to seasonally adjust in a consistent way other series from the same statistical program which may not have been adjusted, or to compare the results obtained from using different options or methods for seasonally adjusting the same series.

[Insert links to any assessment measures, e.g. possible work by Eurostat.]

The task force noted that, to a large extent, the knowledge of which software was used and of the parameters specified for the seasonal adjustment of a particular series is generally sufficient to replicate the process. However, given its limited use this information does not need to be disseminated. Nonetheless, it should be available upon request.

The task force concluded with the following recommendation.

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.

5. GUIDELINES FOR THE REPORTING AND DISSEMINATION OF METADATA

5.1 Introduction

The provision of methodological information or metadata[20] with statistical data outlining concepts, definitions and describing methods used in collection, compilation, transformation and dissemination of statistics is an essential function of all statistical agencies disseminating data at both the national and international levels. The need for such methodological information arises from a desire to lend transparency to economic, social and population statistics so that the typical end-user can make an informed assessment of their usefulness and relevance to his or her purpose. In recognition of this, methodological transparency has been embodied as one of the UN *Fundamental Principles of Official Statistics* (UNSC 1994). The recommended metadata practices provided below in Section 5.3 flesh out this principle.

The provision of metadata is therefore an inescapable role of all statistical agencies in both developed and developing countries and one that requires adequate planning and resources. In recent years even greater emphasis has been given to the importance of ensuring that statistics published by international organisations, national statistical institutes and other agencies are accompanied by adequate metadata. Many statistical agencies have embodied their corporate policy on the provision of metadata in their dissemination standards and author guides[21].

Users of metadata are generally depicted as falling into two broad groups: producers of statistics responsible for designing data collections, actual data collection, processing and evaluation of data, and data dissemination; and end-users of statistics comprising policy analysts, media, academics, students, etc (UNSC and UNECE 2000). The statistical functions of international organisations often fall somewhere in the middle of these broad groups, in that they also perform the role of disseminators of data to internal or external end-users. International organizations also use metadata in evaluations and assessments of data comparability between countries.

In many instances the metadata requirements of users located in national government agencies, the private sector, national statistical agencies and in international agencies frequently overlap. Therefore, determining the amount of metadata to be disseminated by statistical agencies on the basis of the location of users and even the types of uses of metadata may not be all that helpful in that metadata are more often than not presented to an unknown audience of users. The approach used in this Handbook is to differentiate between the amount of metadata detail required for data interpretation by different users and how varying amounts of metadata detail are best presented.

An approach commonly used by national and international agencies entails the presentation of metadata as layers within a pyramid. Using this approach, for any specific statistical domain (e.g. prices, production,

[20] The International Standards Organisation (ISO) definition of metadata is "data that defines and describes other data". A distinction can be made between "structural" metadata – those concepts used in the description, identification and retrieval of statistical data – and "reference" metadata – that describe statistical concepts, methodologies for the collection and generation of data and information on data quality. Reference metadata, sometimes generated, collected or disseminated separately from the data to which they refer can be relevant to all instances of data described: entire collections of data, data sets from a given country, or for a data item concerning one country and one year. For this reason, some overlap may exist between "reference" metadata – which are often disseminated separately from the data they refer – and "structural" metadata used to identify data. Metadata in the context of this Handbook is more akin to "reference" metadata as described above.

[21] An example of such standards are metadata requirements embodied in the UK Office for National Statistics, *National Statistics Code of Practice: Protocol on Data Management, Documentation and Preservation* (ONS 2003, p. 7)

education, health) methodological information describing statistics becomes more detailed as one moves down from the apex of the pyramid. For example:

- at the apex of the pyramid are table headings and footnotes that are generally immediately adjacent to the data and are an integral part of each statistical table. Such metadata are essential for an understanding of the data;

- then there are explanatory notes generally located in the same "publication" and which provide a brief general description of the statistics, definitions, key issues that can impact on the use of the data. In an international context, in the main, explanatory notes do not provide much detail on individual country methodology/practices;

- finally, at the base of the metadata pyramid are detailed methodological information disseminated by national statistical agencies and international organisations in publications and/or on websites. These are potentially the source of the most detailed methodological information available. Some (though not all) statistical agencies publish very detailed concepts, sources and methods for a number of their key annual and infra-annual statistics.

The need for provision of more extensive methodological information, and its accessibility to users through dissemination on the web, is now receiving greater recognition. However, practices in this area vary considerably in the statistical systems in both developed and developing countries with regard to the amount of methodological detail provided on their websites and other dissemination media (even in the national language), frequency of updating, its proximity to the statistics it describes and ease of access by users.

[Insert example of a detailed methodological publication located on the internet for a developed and a developing country]

Metadata disseminated by statistical agencies can therefore be usefully categorised on the basis of both the amount of information provided and its proximity to the statistics it describes. In reality, the typical end-user of statistics seldom requires or uses detailed metadata. Most users of methodological information disseminated by statistical agencies in the context of the dissemination model outlined above, merely access the top layer. If they require more detailed information on specific methodological aspects to determine the relevance of the data to their requirements, they may have to search through succeeding layers where more detailed methodological information is provided. In this context, in addition to the compilation and dissemination of metadata, a key role of the statistician involves the organisation and structuring of metadata that allows users to dig as deeply as necessary into issues of concepts and compilation practices, etc, without being buried in enormous amounts of text.

The primary responsibilities of national and international agencies regarding metadata are therefore to:

- ensure that statistics disseminated are accompanied by appropriate metadata; and to

- provide efficient facilities for the dissemination of appropriate metadata detail to users.

International organizations have a specific additional responsibility to minimise the reporting burden of national agencies supplying both metadata and data.

5.2 Need for metadata content standards

As stated above, there are significant differences between statistical agencies when it comes to the organisation and structure of metadata for statistics which are becoming increasingly accessible via a wider range of dissemination media, in particular, on-line dissemination on the web (in html or databases). The evolution of statistical metadata content standards, i.e. the development of an agreed list of common metadata items and the standardisation of terminology and definitions for these items, has not really kept pace with ITC infrastructure developments. From the perspective of content standards, there are two broad sets of issues:

- accessibility of the metadata. In the context of internet dissemination issues here involve the actual availability of metadata on websites, organisation on the web, provision of search facilities, linkage to data and the financial cost to the user to access the required metadata; and

- significant differences between countries in the actual statistical methodological elements described in metadata for the same statistical domains. In some instances the problem is merely one of terminology where the same term can have different meanings or different terms can have the same meaning. In other cases, the actual metadata is different. From the viewpoint of an international organisation, where there is a frequent need to compare practices used by a number of countries, different metadata content posted on websites or published elsewhere makes any meaningful methodological comparisons a time consuming and costly exercise. The need to compare statistics across countries is by no means restricted to users working in international organisations.

5.3 Recommended practices for the reporting and dissemination of metadata

This Section outlines four key elements of good practice in the compilation and dissemination of metadata which specifically relate: to the need for the compilation of up-to-date by international organisations and national agencies; providing access to metadata; the methodological items (or metadata elements) that should be incorporated in metadata disseminated; and the use of a common set of terminology. Each of these elements of good practice is discussed below. Other Sections of this Handbook outline recommended good practice for ensuring methodological transparency through the provision of metadata in specific areas such as describing methodology used in seasonal adjustment (Section 4.6), the provision of information on data revisions (Section 6.1), series breaks (Section 6.2) and sampling and non-sampling errors (Section 6.4).

5.3.1 The need for metadata

All statistical agencies should:

- compile metadata required for users to understand the strengths and limitations of the statistics it describes;

- keep their metadata up-to-date, incorporating the latest changes in definitions, classifications and methodology, etc;

5.3.2 Access to metadata

National agency and international organisation practices vary significantly with respect to the visibility of metadata they disseminate, in particular, for metadata located on websites. In some instances metadata is easily located by users unfamiliar with the site and in others considerable time and effort is required to navigate through the website to obtain the required information, particularly where metadata for a number of different statistical domains are sought. 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 the internet provides the most accessible medium for obtaining the most up-to-date metadata. It is also good practice for metadata to be structured in such as way as to meet the needs of a range of users with different needs and/or statistical expertise. In this context a layered presentation of metadata is recommended, progressing from summary metadata to more detailed metadata;

- 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;

- structuring the metadata for different statistical domains on the basis of some hierarchic classification. Consideration could be given to the adoption of the UN Administrative Co-ordination Committee's (ACC[22]) Classification of Statistics and Statistical Activities as the international standard for metadata. The classification is available at http://unstats.un.org/unsd/methods/statact/acc-class.htm;

- provision of a local search engine based on free text search;

- good 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[23].

- providing the names of contact persons or email addresses where further information about concepts, definitions and statistical methodologies may be obtained.

5.3.3 Adoption of a set of common metadata items

This issue, together with the adoption of common terminology, is at the heart of current problems and difficulties of comparing methodologies used by different countries in the compilation of the statistics they disseminate. The issue is also associated with the need for international organisations to minimise the metadata reporting burden of national agencies.

[22] Now known as the Committee for the Co-ordination of Statistical Activities (CCSA) which is a body of representatives from all UN and non-UN international organisations involved in statistical activity. The CCSA normally meets once a year.

[23] The World Wide Web Consortium (W3C) document "Cool URIs don't change" (available at http://www.w3.org/Provider/Style/URI) outlines the case for maintaining stable URLs and best practice for designing URLs.

In addition to a perceived lack of co-ordination between international agencies, national agencies faced with the task of providing metadata to different international agencies, often comment on their use of different lists of metadata items used to describe methodologies for the same statistical domains. They also comment on how much easier life would be if different international organizations used either: the same common list of metadata items (or at least a common core set) so that one set of metadata compiled by the national agency would meet the needs of many/all/most international organizations; or at the minimum used a consistent set of individual metadata items that would enable national agencies to map metadata maintained in their corporate metadata facilities to repositories maintained by various international organizations in lieu of direct collection and/or duplicate storage on different databases.

The notion of a minimum core set of metadata required for the correct interpretation of statistics has been discussed at numerous forums on metadata and indeed such a list is included in the UNSC/UNECE guidelines (UNSC/UNECE 2000, p. 5). Similarly, the more comprehensive and hierarchic metamodel for the IMF's SDDS, which is currently being revised, provides another such core list.

The SDMX initiative is currently working on the development of a set of common cross-domain metadata items or categories (referred to as "core statistical concepts") which will become a part of SDMX standards, which if adopted by national and international organizations will enhance the possibility of more efficient metadata interchange. Examples of such concepts are "data source used", "periodicity", "population coverage" and "seasonal adjustments". The core statistical concepts being developed will to a large extent be based on existing template(s) such as those developed by the IMF for the detailed methodological summaries posted on the DSBB. The adoption by both the OECD and Eurostat of metadata items consistent with those used for the DSBB is a step in the direction of the use of consistent if not common metadata items. All three organizations have agreed to adopt the standard set of metadata terminology embodied in the SDMX Metadata Common Vocabulary (see following Section).

5.3.4 Adoption of a common set of terminology for metadata preparation

Considerable resources are often expended by international organizations in verifying text, etc, to ensure that methodological descriptions are as consistent as possible between countries. Not only does the process of metadata verification entail a duplication of effort but it also results in dissemination of different methodological terminology, especially where translation of methodological text into another language is necessary. Ideally, methodological descriptions of the same national statistical collections published by different international organizations should be identical with regards to terminology. 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 organizations which contain definitions derived from those standards. Examples of such glossaries are those maintained by the OECD, Eurostat and UNSD referred to in Section 1.5 above.

The Metadata Common Vocabulary (MCV) developed under the umbrella of the SDMX initiative is specifically aimed at identifying commonly used terms to describe the different types of metadata (SDMX 2003). It is intended to be used by international organizations and national statistical agencies. The MCV contains a core set of metadata items (for both structural and reference metadata) and their related definitions and is designed to improve the standardization of metadata content for the purposes of data exchange and to promote the use of common nomenclatures that can foster international comparability of international data. The current version of the MCV (available on the SDMX website at www.sdmx.org) contains several fields – term title, definition, source, hyperlink to definition source where available, related terms and context.

5.3.5 Unambiguous presentation of similar but not identical series

There are numerous examples where users are confronted by several different versions of seemingly the same series published by different national agencies and international organisations. Examples of such series include:

- Series for individual countries that have been transformed by either national agencies or international organisations to improve their comparability. Examples of such series are: Eurostat's Harmonised Consumer Price Index (HICP) and the OECD's Standardised Unemployment Rates (SURs) that are frequently published alongside similar, but different national series. Similarly, at the national level, there are estimates of employment obtained directly from labour force surveys and labour force series that have been adjusted for SNA labour input purposes.

- Other national series that have been transformed and published by international organisations in some way to improve comparability. In some instances the same national data has been transformed differently by different international organisations resulting in inconsistent output both between themselves and with data published at the national level.

  An example of different transformation processes resulting in different series at the national and international levels is the use of different seasonal adjustment applications by national institutes and the European Commission for business tendency survey data. Other examples are series transformed at the national level specifically to meet the requirements of an international organisation which again results in different data for seemingly the same series being released at the national and international levels.

- The use of different terminology or label for the same concept by different agencies, e.g. "industrial manufacturing" and the heading "intermediate goods".

Recent investigations by the OECD Short-term Economic Statistics Expert Group task force on data presentation and seasonal adjustment sought to formulate a small set of recommendations on practices and processes that would alleviate, in the short term at least, the impact of such inconsistencies by making users more aware of the differences, and the reasons, between similar but non-identical series. In the longer term, the international initiatives to develop standards for the exchange of data and metadata and improved co-ordination between international organizations on the collection of such information from national agencies outlined in Section 2.2. above aim to eliminate most of these differences altogether.

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 organisatons 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 organizations should clearly describe in their metadata specific details of any transformation processes of national data they make to make the series more internationally comparable. Data transformed by international organizations 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 data disseminated by national agencies and international organizations (especially when transformed to an international classification to enhance international comparability) should always be clearly indicated (for instance, NACE Rev. 1, CITI, 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*).

6. GUIDELINES ON KEY REPORTING PRACTICES

As mentioned in the Introduction to this Handbook (in Section 1.2), there are a small number of data reporting practices where different approaches used by national agencies and international organisations can and do complicate the process of comparing data, both between countries and within an individual series over time. This Section provides a brief background on a number of key reporting practices, identifies existing international standards (if any) and outlines recommended practice for implementation by both national agencies and international organisations.

6.1 DATA REVISION

The following discussion on data revision and the need for the formulation by national agencies and international organizations of a comprehensive and transparent revisions policy draws directly and extensively from an IMF Working Paper, Revisions Policy for Official Statistics: A Matter of Governance first presented at the August 2003 International Statistical Institute (ISI) and subsequently revised the following year (IMF 2004). The discussion also uses material from an OECD-ONS workshop on assessing and improving statistical quality revision analysis for national accounts held in Paris on 7-8 October 2004 [insert reference]. Although the IMF working paper was written primarily in the context of economic statistics the terminology, context of revisions and recommended practices are just as relevant for social and population statistics, and irrespective of whether the statistics are published in developed or developing countries.

The basic premise of the IMF working paper is that a sound revisions policy contributes inter alia to good governance in official statistics. It notes that many countries have not yet set out a well-articulated revisions policy. In recent years, however, revisions policy is receiving more emphasis. For example, the *Quarterly National Accounts Manual*, Chapter XI provides a discussion of revisions policy. The Ecofin Council of the European Union, in February 2003, included a section on revisions in its "Code of Best Practices on the Compilation and Reporting of Data in the Context of the Excessive Deficit Procedure." In addition, the IMF's Data Quality Assessment Framework specifies good revision practices.

The need for the development of a set of good practices with regard to data revision, and their application by national agencies is also important at the international level where international organizations such as the IMF, OECD, Eurostat and other regional bodies require consistent data. However, the absence of common policy action in this area among member countries is one of the causes of inconsistency of data at the national and international levels, necessitating constant change due to variations in the timing of revisions.

The purpose of the IMF working paper is to gain widespread acceptance of the importance of a revisions policy, to work towards the development of a comprehensive and internationally accepted set of good practices that would together constitute a sound revisions policy which individual developed and developing countries could fit to their own individual circumstances. The practices outlined below are derived from a discussion of user needs, resource issues, and maintenance of credibility. More specifically, they derive from a selection of examples of good practices in place in various countries that are included in an extensive set of appendices attached to the IMF working paper. These are drawn from national accounts, prices, government finance statistics, monetary statistics, and balance of payments statistics, though again are relevant to the social and population statistical domains.

6.1.1 Typology and terminology

Revisions are defined broadly as any change in a value of a statistic released to the public by an official national statistical agency. The statistic may be a level, such as the value of a flow (for example, GDP) or of a stock (for example, of financial assets), or a change in level, such as the rate of price increase. Revisions can be classified in at least two ways. One way is by the reason for the revision, and another way is by the timing of the revision. It is especially useful to catalogue these in order to establish a common language.

Revisions classified by reason

Revisions may take place for at least eight reasons. In reality, some of the distinctions are blurred because two or more kinds of revisions may be made at the same time. Aside from correction of mistakes, the last item in the list, the reasons tend to break into three groups. The first group is the incorporation of more complete or otherwise better source data, encompassing the first three reasons. The second is routine recalculation, encompassing the next two reasons, and the third is improvements in methodology, encompassing the next two reasons.

- Incorporation of source data with more complete or otherwise better reporting.
- Incorporation of source data that more closely match the concepts.
- Replacement with source data of judgment or of values derived largely by statistical techniques.
- Incorporation of updated seasonal factors.
- Updating of the base period.
- Changes in statistical methods.
- Changes in concepts, definitions, and classifications.
- Correction of errors in source data and computations.

The first reason, incorporation of source data with more complete reporting, cause revisions across a wide spectrum of statistics. At one end of the spectrum, a first report on credit aggregates may be based on the largest financial institutions and then the aggregate is revised when reports from all institutions, including the slower ones that have less sophisticated reporting or are from outside major cities, become available. At the other end of the spectrum, data from monthly samples may be replaced in national accounts components with data from more comprehensive annual samples. Two other reasons for revisions are related. Updating of weights, as for price indexes, brings in information from more recent surveys. Incorporation of audited results, as for budgetary figures and data from financial reports, to replace early results in effect brings in "better" data.

The second reason, the incorporation of source data that more closely match the concept, is most likely to occur in datasets that piece together many data sources in a mosaic that represents a comprehensive picture of some aspect of the economy. The national accounts and balance of payments are prime examples of such datasets. For example, if production is to be measured, source data that represent sales (plus some adjustments) may provide a first estimate and then the estimate is subject to revision as data more closely matching production become available.

In some situations no current data may be available, and a first estimate is based on judgment or statistical techniques. A revision may then occur when data become available. Such situations may arise for quarterly national accounts. The United States uses judgmental extrapolation for the first quarterly estimate for several components, including domestic services and improvements on owner-occupied housing. Subsequently, data become available that can be incorporated.

These first three reasons often appear together, for example, in national accounts and balance of payments. In monetary and government finance statistics, the reasons often boil down to completing institutional coverage and incorporating the outcomes of audited reports.

Incorporation of updated seasonal factors relates closely to the incorporation of additional source data, and some lists of reasons for revisions do not list the two separately. Seasonal factors, such as those that are derived from a moving average of experience or from the most recent year (concurrent seasonal factors), can change as the new experience comes into, and older experience drops out of, the calculations. Some countries rarely revise the consumer price index to bring in new or additional price observations, but do

revise once a year to incorporate updated seasonal factors. For example, the U.S. Bureau of Labor Statistics, with the release of the January index, each year recalculates the seasonal adjustment factors to reflect price movements in the just-completed year. This routine annual recalculation may result in revisions to seasonally adjusted indexes for the previous five years.

Updating of the base year of an index - that is, the year set equal to 100 - is also often a routine reason for revision. This may be carried out as a separate step, but usually it is done when new data underlying the weights for the index are introduced.

Incorporation of changes in statistical methods is sometimes not listed separately because such changes often go hand in hand with changes in source data. However, they can also occur independently. For example, revision studies may reveal that a particular method can be improved or replaced by another to achieve greater accuracy or timeliness. In the last few years, this source of revision has become more prominent as countries moved from fixed-weighted volume and price measures to chain-weighted measures. Changes in concept, definitions, and classifications, often stimulated by the adoption of new international guidelines, are yet another source of revision, for example, when a country moved from following the fourth to the fifth edition of the *Balance of Payments Manual*. Major efforts have been devoted to reaching internationally agreed classifications in recent years. The Classification of the Functions of Government (COFOG) and the Classification of Individual Consumption by Purpose (COICOP) are cases in point. The introduction of new classifications is often undertaken when new concepts and definitions are introduced, but sometimes on its own.

In addition, changes in presentation of statistics should be mentioned. They do not, strictly speaking, fit the definition of revision as a change in a value of a statistic. However, they often take place at the same time as revisions, especially revisions caused by changes in concept, definitions, and classifications. Changes in presentation are also often implemented to respond to the analytical needs of users, for example, the experience of Australia reporting financial derivative asset and liability positions on a gross basis rather than on a net basis.

Finally, revisions occur as errors are corrected. Errors may occur in source data or in processing. For example, reporting institutions may discover after submitting the data that some components are missing or outdated seasonal adjustments may have been inadvertently applied.

A key aspect that distinguishes the first five reasons listed above from the remaining three is the possibility of their being predictable to users in terms of timing. For example, the use of more complete source data as estimates progress from preliminary to final can be undertaken on the basis of a scheduled set of releases. Similarly, the incorporation of updated seasonal factors and base periods are also predictable, regular (albeit with a longer cycle) and known well in advance by the statistical agency with the possibility of their being signaled to users well before implementation. On the other hand, the last three reasons tend to be less predictable though are generally (with the exception of errors) implemented with sufficient lead time to advise users well in advance.

Revisions Classified by Timing

With regard to timing, some revisions are made in the weeks or months shortly after a first release. These are "current revisions" because they affect the current weekly, monthly, or quarterly data. "Annual revisions" are made after data for all the months or quarters of a year become available. Audits are usually undertaken for a calendar or fiscal year's data, although the results may not be available for some time after the close of the year. Both current and annual revisions usually stem from the first four reasons: incorporating source data with more complete reporting; incorporating source data that more closely match concepts; replacing judgment and statistical techniques; and incorporating updated seasonal factors.

Annual revisions often affect several years of data - perhaps three or four years, so an annual estimate may be subject to revision more than once. For example, in the U.S. national accounts, there are three such revisions, as important additional annual source data arrive in each of three years.

Less frequent revisions, often four or more years apart, may be called "comprehensive," "major," "historical," or "benchmark" revisions. Typically they are occasions for major changes in statistical methods and changes in concepts, definitions, and classifications. Often these revisions are carried back, or backcast, for a number of years. Revisions that correct error, of course, have no predictable timing.

Another aspect of timing that impacts on international organizations is the absence synchronization of revisions and their implementation between member states. The receipt of data released by member states at different points in time can have a significant impact on the compilation of zone aggregates or derived statistics (such as purchasing power parities (PPPs)) by international agencies, especially where data for larger economies are involved. The "solution" to this issue involves negotiation and agreement between the international organization and their member states to achieve greater synchronization of both initial releases and revisions, again, especially for the larger economies. Although complete synchronization may not be achievable, the result could be greater coalescing of releases and a narrowing of the band of release dates.

6.1.2 Context of revisions

The context of revisions can be analyzed from three main points of view: user needs, resource issues, and maintenance of credibility.

User Needs

As documented in the data modules of the IMF Reports on the Observance of Standards and Codes (ROSCs), surveys and meetings with users from a wide range of countries confirm their concern about revisions and revisions practices. User needs with respect to revisions fall into the following four categories, the:

- *timeliness* of first release of data and timing of subsequent revisions;
- *accuracy* of first release of data and subsequent revisions;
- *consistency* of data over time;
- *documentation* for the revisions that is provided to users.

Timeliness

Some users, such as policymakers, investors, international organizations, and the media, put strong emphasis on the timeliness of statistics. A key aspect of timeliness is the early release of economic data for them to be relevant to users' particular needs. For a central bank to conduct monetary policy effectively, it will need to analyze data on inflation and growth of monetary aggregates that are as up-to-date as possible. For investors and financial markets to make informed decisions, they also need timely data. For international organizations to monitor adequately economic developments and their funded programs in member countries, it requires the latest data at the earliest possible date.

Another aspect of timeliness that concerns users is that the timing of first release of data and subsequent revisions is predictable and relatively stable from year to year. In addition, the timing of the release may need to be coordinated with preparing important official policy documents, such as government budgets.

Accuracy

While policymakers place a high premium on timely data, they also need a degree of accuracy. The consequences of inaccurate data can be just as serious as late or delayed data, because inaccurate data can cause policymakers to make wrong decisions. Similarly, although investors want timely data on which to base their investment decisions, they do not want to take a decision based on data that are likely to change substantially in the next month or next quarter. Among users, researchers and the academic community place perhaps the highest priority on accuracy, as timely data are less important to them than an accurate and comprehensive time series of data. The needs of researchers and policymakers are related, as researchers use the economic data to test empirically and validate economic theory, which is at the basis of policymakers' decisions.

The importance placed by users on accuracy clearly requires that they be able to judge the accuracy of preliminary data and subsequently revised data. To make informed judgments, revised data must be clearly identified and documentation provided. The documentation should include information on the sources and methods used to prepare data, on changes to be incorporated in upcoming major revisions, and, post-revision, on the sources of the revision. Some indication from statistical agencies of how accurate preliminary or estimated data are would also be useful to researchers, as they may decide not to use these data in their time series if they are outside a certain degree of accuracy.

Consistency

Many users, particularly those engaged in research and forecasting, require consistency of data over time. While they realize that revisions will yield more accurate data, they are concerned that revisions that are frequent or large may disrupt their databases and cause inconsistencies unless the revisions are backcast over a sufficient number of years. As well, users who work with several datasets will be concerned that revisions be carried out in a coordinated way to avoid lengthy periods when one dataset is revised and others are still on the old basis.

Documentation

To lessen the trauma caused by the revisions, users would want clear documentation. Basic documentation should include identifying in statistical publications data that are preliminary (or provisional or estimated) and revised data, explaining the sources of revisions, and explaining breaks in series when consistent series cannot be constructed.

Documentation is particularly important when changes in concepts and definitions are involved because such changes can seriously affect the interpretation of various statistical applications (for example, forecasts) and empirical tests of the validity of economic theory. Meetings and consultations with users arranged by the statistical agency can also be helpful in explaining the reasons for and content of revisions, particularly in advance of the revisions so that users can prepare better to deal with them. It is also important for users that the revised data are as easily accessible as possible, preferably with electronic release of the complete revised time series with explanatory documentation attached.

Resource Issues

Resources affect countries' revisions policies in several ways. On the one hand, there are specific issues of cost effectiveness (that is, is the increased accuracy gained from a revision worth the cost?). On the other hand, there are questions about the basic design of the statistical compilation system itself, which has fundamental implications for the costs of revisions. How effectively a statistical agency addresses the

resource issues that it faces will depend in large part on the quality of its management. Effective management of resources is a critical element of the good governance of statistical agencies.

As described in Section 6.1.1 above, revisions are driven primarily by the arrival of source data. Typically a core set of source data are available for the first estimates that are released to satisfy the timeliness needs of users. Then, as more detailed and comprehensive source data arrive, the first estimates are revised to improve the accuracy of the statistics. In designing the statistical compilation system and defining the surveys and administrative data to be used as source data, it is important to bear in mind the cost implications of alternative designs and definitions. A poor country that tries to implement surveys and administrative data in as much detail and breadth as rich countries may well find that it does not have the means to compile and revise these data in a timely manner. It is not uncommon for poor countries to have compilation systems and sample surveys based on industrial country models that result in piles of collected data sitting unused or never finalized for months and even years. In such countries, an appropriate balance may well require that official statistics rely on less detailed and comprehensive source data that are as representative and timely as the resources of the country will allow.

Statistical agencies of all countries, both rich and poor, must operate within limited budgets and make efforts to ensure the cost effectiveness of their programs, including revisions. Again, it is a matter of balancing, balancing not only timeliness against the accuracy needs of users, but also balancing both timeliness and accuracy needs against the marginal costs of achieving improvements in both areas. Costs are incurred not only by the statistical agencies, but also by the respondents who must take the time and effort to complete the questionnaires and data submissions necessary to comply with data release and revisions policies. A kind of "cost benefit analysis" must be done in order to take realistic and sustainable decisions with respect to the timeliness and frequency of data releases and revisions. It should be conducted in a way that balances needs and costs across different types of data users and different data sets. Unfortunately, no mathematical formula exists to conduct this type of analysis. It must, in effect, be must be accomplished in a less precise way through the difficult process of consultation and coordination among statistical agencies and users, as well as with the political authorities who control the agencies' funding.

In many countries, particularly poor countries, statistical agencies are often seriously under-resourced both in absolute terms and relative to other government agencies. In these circumstances, it will be important that statistical agencies undertake efforts to raise the consciousness of the political authorities to the serious consequences of neglecting to build adequate statistical capacity. International organizations have an important role to play in this arena. With respect to revisions, both statistical agencies and international organizations must impress on the political authorities of countries the critical importance of adequate resources to allow for the timely release and revision of official statistics.

Maintenance of Credibility

Confidence in the figures effectively must be built on confidence in the statistical agency disseminating them. Fundamental to achieving trust in, or credibility of, statistical agencies is *integrity*. Integrity is a central element in the IMF's Data Quality Assessment Framework and is also prominent in the U.N. Fundamental Principles of Official Statistics (UNSC 1994). Providing assurances of integrity involves, at the broadest level, enacting effective statistical legislation and ensuring the professional autonomy of statistical agencies. But establishing a sound revisions policy is also a key element necessary to gain the trust of users.

It is not unusual, even in industrialized countries, for a distrust of government (or the political party in power) to be translated into distrust of official statistics, or at least a healthy degree of skepticism. Revisions can be particularly sensitive if statistical agencies handle them in an unprofessional manner. At

the extreme, users may even suspect the government is intentionally misreporting for its own political or financial motives. For example, investors might suspect the government is intentionally delaying or misreporting data on international reserves to prevent capital flight. Or the media may suspect the government is manipulating statistics to avoid criticism of its policy record. Or an international organization may worry that a government is misreporting to comply with a policy target.

What are the needs of users with respect to revisions and the credibility of official statistics? With respect to the release of first estimates, users need to be able to make informed judgments about the quality of these estimates. How accurate are they? What is the likelihood of further revision, and by how much and in what direction? When will the data be "final"? For the revisions themselves, users need to be informed about the causes of the revisions, as well as have access to complete documentation on methodology and procedures.

Users will also be reassured if they see that revisions take place within the framework of an overall policy and according to a predetermined schedule. If the policy, procedures, and schedule are published, it will be evident that revisions are not ad hoc and for political interests, and that adequate safeguards exist to prevent abuses in this area. Finally, when mistakes are discovered, it is critical that the statistical agency report them to the public as soon as possible and provide satisfactory explanations to reassure users and enable them to distinguish honest mistakes from cases of "misreporting."

6.1.3 Recommended practices for data revision

This Section identifies eight main revisions practices. They are consistent with the general principles of good governance in statistics, such as they appear in the Fundamental Principles of Official Statistics and in the *Handbook on the Operation and Organization of a Statistical Agency* (UNSD 2001). In fact, the revision practices identified can be seen as making explicit the application of these principles about, for example, integrity, responsiveness to users' needs, and professionalism in the context of revisions.

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.

A clear, short summary statement of when to expect revisions and why is readily accessible to users

Most revisions fall under a "revisions cycle." Cycles typically incorporate current (for example, quarterly) and annual revisions as defined in Section 6.1.1 above and less frequent comprehensive or benchmark revisions that usually relate more to the two "improvements" reasons listed in Section 6.1.1.

The UK Office for National Statistics National Statistics Code of Practice: Protocol on Revisions (ONS 2004, p. 8) extends this further by stating that each organization responsible for producing national statistics should publish and maintain a general statement describing its practice on revisions which identifies those outputs subject to scheduled revisions and a separate list of those which deliberately not revised. The Protocol includes all of the recommended good practice for revisions outlined in this Section of the Handbook.

The current revision cycle is relatively stable from year to year

Current and annual revisions are done broadly to incorporate more complete or otherwise better source data. The following practices relate to the timing of current and annual revisions, the:

- revisions are timed to incorporate new source data;
- revision schedule takes into account the timing for preparing important official economic policy documents;
- revision schedule takes into account the timing of revisions in other datasets.

Stability of the revision cycle from year to year is at the heart of good revisions policy. Users place great importance on a revision schedule that is regular. Fortunately, for countries that decide to establish a revisions policy, it is not difficult to ensure that its timing is stable over time. Indeed, it is a logical outcome and one that promotes efficient implementation. The most common basis for stability is the timing of arrival of source data, which then triggers their incorporation into revised data. Occasionally, a balance must be struck between maintaining the stability of the cycle and making unpredictable but important revisions outside the cycle. Coordinating timing with important official economic policy events can also be useful. For example, Italy times the release of national accounts to coincide with the annual presentations to their parliaments on the economic situation. It is also important to coordinate with other macroeconomic sectors to ensure consistency, for example, coordinating revisions of balance of payments statistics with national accounts.

Major conceptual and methodological revisions are usually introduced every four to six years, balancing need for change and users' concerns

Major conceptual and methodological revisions relate mainly to the two "improvements" reasons for revisions outlined in Section 6.1.1 above —to incorporate new statistical methods and new concepts, definitions, and classifications—all superimposed on changes in the structure of the economy. These revisions are typically more far-reaching and complex than current revisions, and can be disruptive and problematic for users if they occur too often or take place in a confusing or unpredictable manner. A reasonable guideline for regular timing would be every four to six years. Timing such as this balances the need to avoid unnecessary disruptions to time series with the need to maintain the quality of statistics in line with international best practices and the changing institutions and structure of the economy.

Although individual countries do not control the timing of major changes in international statistical methodologies (for example, the appearance of *1993 SNA* and the *BPM5*), a four-to-six-year cycle can generally accommodate these changes without undue delays and disruptions. Incidentally, it is also possible and can be helpful to users to coordinate the timing of methodological improvements with the current cycle of revisions timed for the arrival of better source data. Countries do have control, however, over the timing of methodological and classification changes that they undertake to reflect institutional and structural changes in their own economies. These kinds of changes can be accumulated, studied, and prepared for during the four-to-six-year intervals before they are finally published.

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.

Clearly, some revisions are more difficult than others to revise backwards. Among these are data from surveys that have changed, data affected by legal constraints, and data constrained by accounting principles (for example, government finance statistics). Lack of resources also constrains the extent of backward revisions, especially for poor countries. Various second-best approaches are possible, such as the U.S. practice described where GDP series are revised back to the last benchmark (usually five years) and further back for selected series that are particularly important.

Documentation on revisions is readily available to users

Preliminary (or provisional or estimated) data and revised data are identified as such

While this practice may seem obvious, it is not uncommon to find in many countries that preliminary and revised data are not clearly identified. This is especially likely in countries where revisions are not made according to a consistent or clearly stated revisions policy. It also occurs more often for government finance statistics and monetary statistics, where statistical principles may not be as much at the forefront as in national statistical offices. Serious confusion and misunderstandings by users could easily arise from neglect to identify changes in data.

Data revision terminology

Concept	Definition
Data revisions	Data revisions are defined broadly as any change in a value of a statistic released to the public by an official national statistical agency.
Preliminary data / Provisional data	Some statistical agencies use the term "Preliminary data" to describe the first released version of a series and "Provisional data" to describe subsequent versions prior to final amendment. However, the two terms are often used interchangeably though users in general should have no great problem in understanding that data labelled either "preliminary" or "provisional" are subject to revision provided this is clearly highlighted by the agency in the release. Clearly informing the use that the data is subject to revision is more important than the precise term used to describe such data.

The actual notation used by national agencies and international organizations for identifying data in tables or graphs that are either preliminary/provisional or which have been revised varies, though many agencies use "p" or "r" respectively. The main issue is that whichever notation or practice is used it must be unambiguous in meaning and used consistently for a disseminated series and in different publications disseminated by the organisation.

Advance notice is given of major changes in concepts, definitions, and classification and in statistical methods

Users should be alerted in advance of major conceptual and methodological revisions to help them prepare for and understand better the reasons for and nature of the changes. An example is the Australian Bureau of Statistics efforts to prepare users for revised balance of payments statistics according to *BPM*. The Bureau provided a description of the new standard and its benefits in advance, including illustrations of sample draft data tables to begin to acquaint users with the changes. Consultations with key users dealt with the implementation of the new standard, and a number of changes were made in the implementation strategy

and schedule as a result. Various reports and discussion papers published in advance of the revision analyzed and described the effects on Australia's statistics.

The sources of revision are explained when the revised series are released

Breaks in series are documented when consistent series cannot be constructed

Complete and transparent documentation of revisions allows users to understand the sources of revisions and, if needed, adjust their analysis of the data. Perhaps even more importantly, complete documentation serves to promote trust in the credibility and integrity of the data and the institutions responsible for compilation and dissemination. Key parts of the documentation are about the sources of the revisions, including the main flows of source data from the preliminary estimates to the revised data. It is also important that breaks in the series be clearly identified when consistent time series cannot be constructed. Documentation can be available to users in hard copy publications, websites, press releases, and dedicated seminars.

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

It is particularly important for users who make decisions on the basis of preliminary estimates, such as policymakers and investors, to be able to make an informed judgment about the reliability and accuracy of the preliminary, provisional, or estimated data. How much confidence should they have in the first estimates? Accordingly, it is good practice for statistical agencies to conduct periodic analyses of revisions (or "revision studies") and to make them available to users. Today's ITC environment makes such studies less demanding than in the past. The following two good 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 (for example, dispersion and bias);

- the analyses are published for major aggregates to facilitate assessment of the reliability of the preliminary estimates

Measures of the direction and dispersion of revisions are the main topics of most revision studies. With respect to measures of the direction of revisions, if a study shows a systematic bias in the revisions, users can adjust appropriately their interpretation of the preliminary estimates. Alternatively, the discovery of bias by a study may lead to changes in procedures, and these can be announced with the study results. Revision studies can also be used to fine tune the timing revisions within the cycle. Measures of dispersion of the revisions provide users with an indication of the accuracy of the preliminary estimates and enable them to assess the likely size of future revisions.

It is important to report to users not only the statistical analysis carried out in the revision studies, but also the basic data flows from the first estimates through all the revisions. The main conclusions of the studies should be clearly stated. Providing the basic data to users allows them to conduct their own studies of revisions if they wish.

A joint OECD UK Office for National Statistics (ONS) workshop on revisions analysis held in Paris in October 2004 (OECD and ONS 2004) identified marked differences between countries in the degree to which revision analysis is a regular feature of revisions policy, especially in the context of quarterly estimates of GDP. Some countries do not conduct any such analyses, others have conducted them as one-off projects whilst some conduct such analyses on a regular basis.

There was broad consensus at the workshop on the core set of summary indicators that are found in most revision studies, namely:

- average revision $= \sum_i^n (L_i - P_i)/n$ where L is the latest (sometimes: final) estimate, P is the preliminary or first estimate, and n is the number of observations
- average absolute revision $= \sum_i^n |L_i - P_i|/n$
- normalised average absolute revision [insert final version of formula]
- standard deviation of revision $= \left(\frac{1}{n} \sum_i^n (R_i - \overline{R})^2 \right)^{1/2}$ where $R_i = L_i - P_i$ and $\overline{R}_i = (L_i - P_i)/n$

In addition to these numerical summary measures, the workshop recommended the use of graphical presentation to supplement the revision analysis.

In terms of the dissemination of the results of revisions analysis the OECD-ONS workshop recommended that:

- the presentation of summary measures of revisions should have a fairly low profile in short-term publications and press releases. However, it strongly recommended the inclusion of a prominent statement making it clear that the estimates presented are nonetheless subject to revision;

Newfoundland Region Statistical Reports

Introduction

Please read the following information prior to viewing Newfoundland Region data on these pages

The Newfoundland Region of Fisheries & Oceans Canada compiles, interprets, audits and reports information on fishing activities in the Newfoundland Region obtained from a variety of sources within the fishing industry.

View

- Species Quota reports
- Fish Landings and Landed Values reports

Catch, effort, landings and quota data is continually being audited, verified and revised as necessary. The Newfoundland Region links above provide preliminary data which is currently available on DFO information management systems, usually, as of the previous day's date.

Users should be aware that this data is subject to revision at any time.

Source: Statistics Canada, 2004, Statistical Reports – Newfoundland and Labrador Region, available at http://www.nfl.dfo-mpo.gc.ca/publications/reports_rapports.asp?Lang=English

- information in press releases, etc, should be simple and complemented by links to more comprehensive analyses;

- a two-handed approach is useful: concise and unsophisticated information in short-term publications such as press releases; more comprehensive and worked-out analyses in longer term (annual or more) publications.

Finally, the workshop outlined necessary conditions for the availability of data to enable regular revisions analyses. These comprised the:

- systematic archiving of all vintages of data;

- formulation of an archiving strategy when new data are developed, ideally archiving should be organized at the inception of the work. To define documentation and archiving, the ultimate use of this information should be clearly identified;

- inclusion of revision analysis as an integral part of the production process. In this sense such analyses serve a dual purpose: to provide information to users and a tool for the systematic on-going review of the data collection to identify possible causes of revisions with a view to their possible reduction, etc.

Building revisions analysis into the survey processing system: UK experience with the Index of Services

An example of how revisions analysis can be built into the survey production process is provided by the United Kingdom Office for National Statistics (ONS) experimental Index of Services.

The Index was set up by the ONS in 2002 and revisions analysis was used as an on-going measure of the reliability of survey output. Because it was a new collection statisticians at the ONS had the luxury of being able to build the necessary requirements for such analysis into the overall operational survey design. This required around three staff months to undertake , however, staff at the ONS felt that such an investment was worthwhile as

[to be completed]

When a mistake in reporting or processing is made, the revision is made in a transparent and timely manner

Many different types of mistakes occur in official statistics, from simple mathematical and recording errors to misclassifications and mistakes in coverage. The mistakes may be by the statistical agency, or by the reporters of source data. It is critical for the integrity of a country's statistical system that any errors are not only reported to users as soon as possible, but also explained in a way that gives assurance that the mistakes were not politically motivated. Explanations for mistakes are much easier when users are already well informed by complete metadata and related documentation on the compilation procedures and sources and flows of data used by the statistical agency. In such a transparent environment, it is just as likely that users will detect errors as the statistical agency, or will at least quickly understand the source of the error.

6.2 PRESENTATION OF SERIES BREAKS

6.2.1 Introduction

A time series is a set of time-ordered observations of a quantitative characteristic of an individual or collective phenomenon taken at different points of time. Normally, time series are a set of observations of a given variable in sequential order, at constant intervals (months, quarters, years, etc). The continuity of a time series not only implies that the observations are continuous over time but also that the same definitions, classifications, processes, etc, have been applied in the collection and compilation of each observation.

The application of inconsistent definitions and classifications, etc, for each observation over time, in theory, constitutes a series break. However, in reality not all changes to concepts, etc, constitute a break in series that has a significant impact on their use. In reality, statistical agencies responsible for the collection of data frequently apply changes to questionnaires, registers, concepts, to their monthly, quarterly and annual collections, many of which have no appreciable impact on the continuity of the series. Changes to annual collections, in particular, are a fact of life.

Statistical agencies, analysts and government agencies use time series data for economic research and business cycle analysis to interpret current economic events. Statistical agencies require long time series to carry out seasonal adjustment and calendar effect correction.Time series are also fed into models to produce projections and forecasts about future economic conditions. For these reasons users in national agencies and institutions and international organisations attach very high importance to time series continuity. In fact, such continuity within a series is often of greater importance than comparability between countries.

However, the uses of time series statistics outlined above are frequently hampered by series breaks or shortness of the series length. The main causes of series breaks are similar to some of the reasons for revising data described in Section 6.1.1 above, such as:

- changes to the base year which may co-incide with updating of the weighting system which in turn may involve changes in the sample of respondents and the sample of products; and
- the implementation of changes in concepts, definitions and classifications, methodology, sampling, estimation.

To a large extent, these factors derive from within the statistical agency responsible for the initial compilation of the data and are usually intentional (US Bureau of Labour Statistics 1996). However, some changes stem from external influences that may be outside the control of the statistical agency, in particular, where the data are derived from administrative sources. These include changes in laws or administraive procedures, changes in the organisational structure of business through mergers, etc.

6.2.2 Approaches to minimising the impact of time series breaks

National statistical agencies normally attempt to minimise the frequency of series breaks, and when they occur, use a number of approaches to reconstruct series based on the new concepts, classification, etc.

- The most commonly used approach involves the compilation of the series using both the old and new methods, classifications, etc., for a specified period around the time of implementation. The high cost of compling dual series severely restricts their availability and length.

  The availability of dual information enables an objective measure of the impact of change to be assessed and perhaps a concordance between the new and old series at the time of the series break. The concordance “coefficient” so calculated may be used to splice or link the series break. Caution is

required in the application of such coefficients to the historic time series as it is only really applicable over the time dual series were compiled. It may not reflect the economic reality of the entire historical series (US Bureau of Economic Analysis 1993). The difficulty is determining when or how far back the conversion coefficient ceases being accurate.

- Alternatively, agencies may refer back to highly disaggregated data (or even unit record information) and recompile the series based on the new methodology, etc. In practice however this approach is very labour-intensive and may only be possible for key highly aggregated series (OECD 2000). Finally, historical estimates may be made on the basis of a related indicator that exhibits the same or similar changes over time as the series where the series break occurred.

6.2.3 Recommended practices for the presentation and reporting of information about series breaks

Recommended good 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 (US Bureau of Economic Analysis 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 implemention 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 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 ocurred. 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 (togther 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 manipulation, concepts, administrative procedures with regard data from administrative sources;
- an assessment of the magnitude of the effect of the change, where possible, with a quantitative measure.
- reference period of the survey where

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 break in series 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.

6.3 LINKING TIME SERIES

6.3.1 Introduction

6.3.2 Approaches to linking time series

6.3.3 Recommended pra ctices

[to be prepared]

6.4 SAMPLING AND NON-SAMPLING ERRORS

The aim of any statistical agency is to compile and disseminate statistics appropriate to user needs in terms of quality. Quality is commonly defined as "fitness for use" (Statistics Canada 2002, p. 2)[24] and comprises a number of dimensions described in quality frameworks developed over the last few years by both national agencies and international organisations (referred to briefly above in Section 2.1.1). One of the quality dimensions frequently referred to is "accuracy" which is the degree to which the data correctly estimate or describe the quantities or characteristics they are designed to measure. Accuracy refers to the closeness between the values provided and (unknown) true values. Accuracy has many attributes, and in practical terms there is no single aggregate or overall measure of it. Of necessity these attributes are typically measured or described in terms of the error, or the potential significance of error, introduced through individual major sources of error.

An aspect of accuracy is the closeness of the initially released value(s) to the subsequent value(s) of estimates. In light of the policy and media attention given to first estimates, a key point of interest is how close a preliminary value is to subsequent estimates. In this context it useful to consider the sources of revision referred to in Section 6.1.1. in the discussion on data revision. Smaller and fewer revisions is an aim, however, the absence of revisions does not necessarily mean that the data are accurate.

The focus of this Section are recommendations on the presentation of information to users on the several types of errors that originate from processes and methodologies used to collect data. The dissemination of such information to users is an aspect of good metadata reporting referred to above in Section 5 of this Handbook. The most widely used typology of such errors involves the distinction between sampling errors and non-sampling errors.

6.4.1 Sampling errors

Background

Sampling errors occur because not all units in a target population are enumerated in a sample survey. As a result, the information collected on the units in the sample may not perfectly reflect the information that could have been collected had the entire population of units been counted. There are various formulae for estimating the sampling error when probability sampling is applied (Eurostat 2002b).

Survey estimates are subject to two types of errors: bias and variable errors. Bias refers to errors that affect the expected value of the survey estimate, taking it away from the true value of the target parameter. Variable errors on the other hand affect the spread of the survey estimates over potential repetitions of the sample selection process. In the context of sampling errors, bias is minimized through the adoption of adequate sampling procedures, sample size and estimation methods. Therefore the spread is the main aspect of the distribution of the sampling error that needs to be considered. The key parameter describing this spread is the standard error, namely the standard deviation of the sampling error distribution (ENCE 2003) [25].

The relative magnitude of sampling error in relation to non-sampling errors increase as more detailed or disaggregated estimates are compiled. Information on the magnitude of sampling error is an essential

[24] The International Standards Organisation defines quality as "the totality of features and characteristics of a product or service that bear on its ability to satisfy stated or implied needs." (ISO 8402, 1986)

[25] A detailed description of methods, etc, for the compilation of sampling errors is provided in the United Nations publication, Sampling Errors in Household Surveys, United Nations, New York, 1993

factor in determining the level to which data may be meaningfully disaggregated and analysed (UNSD 1993 p. 9).

The provision of information on sampling errors is essential in ensuring the appropriate use of data subject to such error in analyses by users with widely differing levels of statistical expertise. Its availability also facilitates evaluation and improvements to statistical design and procedures (UNSD 1993, p. 9). Particular care is required in the publication of data subject to very high sampling error, and appropriate balance is required in deciding whether such information should be disseminated at all, or presented in such a way as to make limitations in use apparent to users with only minimal knowledge of statistics. There is a need for agencies disseminating data subject to sampling error to prepare a strategy for estimating and publishing such errors. The current practice in many national statistical institutes is in fact to disseminate data subject to high sampling error on the assumption of user responsibility to use the data wisely. In this context the statistical agency does not accept this responsibility though it does have the professional obligation to attach appropriate warnings and caveats about its use (ABS 2004, Section 11:04:01).

Investigations into national practices regarding presentation of sampling error estimates carried out by the OECD Short-term Economic Statistics Expert Group (STESEG) task force on data presentation and seasonal adjustment in 2003 (Graf 2003) showed that much research has been done on the subject but that often no actual estimates of variance or covariance tends to be published for either annual or short-term statistics, particularly for the latter. A review of relevant current national data sources showed wide variation in national practices, with respect to:

- the amount of information provided about standard errors;
- their presentation in technical notes accompanying statistical data; and
- ease of access and proximity to statistical data.

A detailed study of United States agency reporting practices on sampling error published by the US Office of Management and Budget in 2001 (OMB 2001) further highlighted this variation in practice, not only across statistical agencies but also within the same agency for different surveys. The adoption of different practices even for the same survey published in different dissemination media is largely a reflection of the different audiences addressed by each medium with varying degrees of statistical expertise and need for detailed information on sampling error. The recommendations of good presentation practice outlined in the US study were highly dependent on the type of dissemination media, though the study emphasized that irrespective of the medium used it was not a question of whether or not information on sampling error should be provided but rather how essential information is to be conveyed to users.

Forms of presentation of information on sampling error

From it's analysis of practices across US agencies, the OMB study identified two broad methods of presenting information on the precision of estimates (OMB 2001, p.3-4). These entail:

- Reporting a direct estimate of the error through the presentation of an estimate of sampling error for every statistic in a statistical report. Users are therefore provided with both the statistic and its sampling error at the same time. An obvious limitation of this approach is the expansion of the publication. Another reason for not recommending this approach cited by the Australian Bureau of Statistics in its *Publishing Manual* (ABS 2004 p. 500) is that it gives the impression that sampling errors are the only or main source of error and that the interpretation of sampling errors is equally important for all statistics.

- The use of indirect estimates of sampling errors. These in turn involve a number of options, namely:

    - The provision of a procedure to enable the user to compute approximate sampling errors for the estimates disseminated. The two major disadvantages of this approach cited in the US report were that sampling errors would not necessarily be presented beside the data and the fact that the procedure used to compute the approximate sampling errors may not provide values as accurate as direct estimates.

    - The provision of sampling errors for a selected number of key estimates. An advantage of this approach is the provision of direct estimates of sampling errors close to the data, albeit only for a limited range of data. Users therefore need to extrapolate errors for data for which they had not been compiled directly.

    The ABS also recommends the use of some form of notation (e.g. an asterisk) to highlight those statistics subject to very high sampling (or non-sampling) error. Furthermore, the Bureau recommends the presentation of supplementary data and methods for approximating the standards errors of derived statistics such as estimates of change, ratios of different estimates for the same period, changes in ratios over different periods, etc (ABS 2004, p. 501).

Irrespective of the approach used for the presentation of information on sampling error it should be remembered that this information is an adjunct to the substantive results of the survey and should not therefore clutter or obscure these results. In other words it should serve the purpose of highlighting the reliability of the data (UNSD 1993, p. 176).

Recommended practices for the presentation and reporting of information on sampling error

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 emphasises the importance of ensuring that the chosen method is clearly and unambiguously described and presented with accompanying definitions and notation.

[insert examples of recommended practice]

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.

6.4.2 Non-sampling errors

Background

Non-sampling errors are errors in published data which cannot be attributed to sampling fluctuations (ISI 2003). Non-sampling errors may arise from a wide variety of sources including coverage defects in the frame, deficiencies in the collection instrument or questionnaire, problems in the processing system and difficulties in achieving acceptable response to the enquiry.

Information from the draft United Nations Handbook on household surveys in developing and transition countries (UNSD 2003, ch. 11) and the Eurostat standard quality report (Eurostat 2003c, pp. 7-11) has been used in the following typology of non-sampling errors. This is merely intended to illustrate the diversity of such errors. In reality, the distinction between some of the different forms of non-sampling error presented may not be all that distinct, particularly with respect to identifying cause[26].

Non-sampling error terminology	
Non-observation errors	*Coverage errors* – resulting from divergences between the target population and the frame population. Coverage errors comprise under-coverage; over-coverage; multiple inclusion in the frame; incorrect auxiliary information provided in the frame (incorrect activity, size classification, location, etc). *Non-response errors* – resulting from the failure to obtain data for target units in the census or survey. These comprise: unit non-response when no data are obtained for a target unit; and item non-response when data for some, but not all, of the collection variables are obtained for a target unit.
Observation errors	*Measurement errors* – that occur during data collection and result in the recorded values of variables differing from the true value. Such errors arise from: imperfections in the survey instrument (form, questionnaire or measuring device) which leads to the recording of incorrect values; respondents consciously or unconsciously providing incorrect data; interviewer influencing answers given by respondents. *Processing errors* – originating from processes used by the statistical agency following receipt of data from the source. These include: coding, data entry; data editing; imputation errors. *Model assumption errors* – arising from statistical models estimated and used in the estimation phase of a survey. Errors here stem from the selection of the appropriate model, collection of relevant data and estimation of the model's parameters.

The very diversity of non-sampling errors outlined above presents a problem in providing users with information about both their existence in published output, and in particular, quantitative measures of magnitude. For this reason data disseminated by both national agencies and international organisations frequently provide very detailed information on sample errors but only general, broad statements on the existence of non-sampling errors. This may convey a misleading impression that sampling errors are far more important than non-sampling errors.

UN Handbook on household surveys in developing and transition countries raises the issue of a lack of standard methods for estimating parameters for the different components of non-sampling error outlined above. It also refers to the absence of a culture within statistical agencies that recognises the importance of such errors and that they should receive as much attention as sampling errors in the provision of

[26] Although a little dated in places, the United Nations National Household Survey Capability Programme publication, *Non-sampling Errors in Household Surveys: Sources, Assessment and Control*, New York 1982, provides a detailed description of the causes of non-sampling errors, their measurement and recommended processes to minimise them

information to users. Nevertheless, some national statistical agencies and international organisations (in particular, Eurostat 2003c) do provide recommendations on the presentation of non-sampling errors, and in some instances how such measures could be quantified. Although quantitative measures for some non-sampling errors are generally not readily available, there are some that are, e.g. coverage rates, non-response rates.

The absence of readily available quantitative measures for all types of non-sampling error does not however remove the professional obligation on the part of organisations disseminating data to provide at least some information. Depending on the nature of the survey, qualitative information indicating the potential main sources of non-sampling error will assist users in their interpretation and use of data disseminated.

Recommended practices for the presentation and reporting of information on non-sampling error

The focus of the recommended good 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 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. Such information should accompany the data disseminated or at the very minimum 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. The ultimate test will be the provision of 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.

  In this context, the following examples of specific non-sampling error information across the whole range of such errors are merely intended to illustrate the types of information that could be provided. These examples have been drawn directly from Eurostat standard reports for the assessment of quality in statistics (Eurostat 2003c).

Examples of reporting indicators for non-sampling error	
Type of non-sampling error	**Reporting indicator**
Coverage errors	• Information about the frame, reference period, updating actions, quality review actions. • Areas of specific strengths and weaknesses of coverage and a qualitative assessment of possible bias. • Type and size of coverage errors (e.g. coverage rates).
Non-response errors	• Non-response rates. • Imputation methods used (if any) • Indications of the impact of remaining non-response on disseminated data, areas of possible bias, etc. • Indications of the causes(s) of non-response • Information about the actions and incentives to minimise non-response
Measurement errors	• Indications about the causes of measurement errors in the published output and their impact (e.g. areas of imperfection in the collection instrument). • Actions taken to minimise measurement errors in the design and testing phases. • Any quantitative measures (e.g. the mean and variance of measurement error per variable of interest)
Processing errors	• Outline of processes used to minimise processing errors (e.g. staff training, data editing used, use of automated IT C processes, etc). • Indication of remaining errors and their impact on statistics, possible bias, etc.
Model assumption errors	• Models used in the production of the survey's statistics and the assumptions on which they rely. • Evidence about the validity of the assumptions. • Statement about the accuracy of any additional data used in the model's estimation. • Indication about any remaining (unaccounted for) bias, etc, which could affect the statistics.

Source: Eurostat, 2003, "Methodological Documents: Standard Report", discussion item 4.2B, *Assessment of the Quality in Statistics*, 6th meeting, Luxembourg, October 2003

6.5 REBASING INDICES

6.5.1 Introduction

Rebasing refers to the establishment of a new base period (year / month) for an index and the recalculation (rereferencing) of index numbers to reflect the new base. A long time series of fixed base indexes is usually compiled converting the indices released in the previous base (for example 1995 = 100) into indices expressed in the more recent base (for example 2000 = 100) by a simple transformation (called "rebasing") and linking them with the new indices.

As mentioned in Section 6.2 above in the discussion on series breaks, the usefulness of statistics is frequently diminished by breaks in time series, in particular, when fixed base indexes are involved. One of the main causes of series breaks for indices concerns changes to the base year [27]. Ruptures in the time series in many countries are exacerbated by other changes taking place place concurrently with rereferencing such as updating of the weighting system[28] which may also involve changes to the sample of respondents and the sample of products.

Much of the text in this Section is drawn extensively and directly from a paper written by Anna Ciammola (2003) for the 2003 meeting of the OECD Short-term Economic Statistics Expert Group.

6.5.2 Existing international standards

The only existing international statistical standard providing explicit guidelines and recommendations on the rebasing of indices is the European Commission's 1998 Short-term Statistics Regulation (European Commission 1998) which requires that rebasing should take place every five years (Article 11) and within three years from the end of the base year. For example, rebasing to the 2000 base year should be undertaken by the end of 2003 at the latest.

There is however room for interpretation as to whether this requirement means that indices are compiled in January 2004 (which in the case of monthly data may refer to October or November 2003) must be based on 2000 or whether it is the first delivery of data for the first reference period in 2004 (January or the first quarter 2004) that needs to be in the new base year (Eurostat 2002a, Part 8.5.1, Base years and base year changes).

More recent work in this area has been undertaken by the United Nations Committee for the Co-ordination of Statistical Activities (CCSA) at the end of 2003 in an attempt to harmonise the base years of indices disseminated by international organisations (UNCTAD 2003). The indices currently disseminated by international organisations contain a mixture of practices including the publication of base years as reported by member countries for their national series or national series that are rereferenced by the international organisation to a common base year for all member countries. A review of current international organisation practices prepared by UNCTAD for the CCSA showed that for most statistical domains a delay of three to six years is usually needed by international organisations to release their rebased index numbers. The reason for this delay is the availability of data on which index numbers are based, either the series themslves or data used as weights in the calculation of indices. In 2003 the current

[27] As Eurostat's Short-term Statistics Guidelines point out, the term base year is used "to describe the year with respect to which the value of all other reference periods are compared. In a series of index numbers it is the year that takes the value 100". Note that the Handbook on price and values measures in National Accounts (NA) contains a different definition of base year. (Eurostat 2002a, Part 8.5.1, Base years and base year changes). The reference year concept on the other hand refers to the length of time (week, month, quarter or year) for which data are collected. The reference period could also be limited to a specific point in time.

[28] The European Commission's Short-term Statistics Regulation requires that a new weighting system is introduced at least every five years and coordinated with changes in the base years. (European Commission 1998)

base years were reported as being 1990 and 1995 and more rarely, 1980 and 2000. In some cases, the weights used for the base year were the average of three years.

The outcome of the CCSA discussion was the need for international organisations to harmonise their base years and to commence rebasing of their index numbers using 2000 as base year with a view to finalisation of a common base year by the end of 2005.

6.5.3 Compiling rebased (fixed) index numbers

The main criteria for the selection of a suitable base year is that it should be "normal" or "average" and not subject to any major usual circumstance or influence. Common international practice involves the updating of base years every five years with the year being one ending in a "0" or "5".

The criterion normally used by statistical agencies to rebase index numbers is that any conversion into a different reference base period must leave month-to-month and year-to-year percent variations (computed on the rebased indices) identical to those characterising the original series, even though the rebased index level changes. Both the direct and the aggregative approaches described below fulfil this criterion.

Direct approach

In the direct approach the indices at each level of the activity classification, coming from a linear combination of the indices at lower levels, weighted by the original weighting system, are rebased independently. The result of this independent rebasing is that the additivity property of fixed-base indices is lost, which implies that indices at more detailed level of activity cannot be aggregated to produce indices at higher and higher levels by means of the original weighting system.

Indirect or aggregative approach

This approach is computationally more expensive than the direct approach and is performed through the following steps:

- rebasing of the elementary indices (i.e. at the most detailed level of product breakdown);
- updating of the previous weighting systems[29];
- aggregating the rebased elementary indices by means of the updated weights.

The Eurostat Short-term Statistics Guidelines refer to the *Handbook on Price and Volume Measures in National Accounts*, that recommends use of the direct approach and this approach has been adopted by many European national statistical institutes. For instance Italy, Belgium, Austria, Denmark, Ireland used the direct approach after the recent migration to base year 2000, even though the same institutes may have different practices for different indicators.

Rounding policy

Rounding is performed to prepare index numbers for publication and generally data are rounded to the first decimal place. National practices vary considerably with regard to the stage at which index figures are rounded. Some agencies base their calculations on unrounded data that are rounded only at the final stage (as the US BLS does for compiling their producer price index) for dissemination purposes. Others round data at each step, from elementary to the most aggregated indices. The latter practice cannot be replicated for the calculation of retrospective indices, as this may affect month-to-month and year-to-year percentage changes constrained to be equivalent to those already published. Rounding to the third decimal place in the

[29] Updating of the original weights is carried out to give higher (lower) importance to a certain economic sector or product, when the average level of its index, not rebased, over the new year base is greater (smaller) than the average level of the most aggregated index, not rebased, over the same time span.

computation of the rebased indices assures the fulfilment of the above constraint. For dissemination purposes, the rebased data are rounded to a decimal place at the final step. Rebased data have less precision after rounding and the loss of precision due to it is more serious when the rebased index values are smaller than the originally released ones.

6.5.4 Presentation and dissemination of rebased indices

The presentation and dissemination practices of national statistical institutes for index rebasing are very different though the direct approach is the method usually adopted. Influencing factors are the sector breakdown at which index values are released and the impact of a new activity classification and weighting system on the retrospective indices. These issues are discussed below.

Introduction of a new activity classification

If the activity classification system used in the series is unchanged rebasing could conceivably be left to the final users and statistical agencies could then simply maintain the database with the historical weights and indices. However, the introduction of a new activity classification may necessitate the calculation of new index values, especially at the more disaggregated classification levels where there could be significant impact on indices already released (i.e. the general index, the most aggregated level could remain unchanged). In such situations, statistical agencies would need to compute the indices on the previous base using the new activity classification and splicing coefficients (i.e. the averages (in the previous base year) of the retrospective indices over the new base year (at each level of the classification)). Agencies could then:

- provide users with this set of retrospective index values together with the splicing coefficients;
- rebase the indices and provide users with spliced series.

The second solution is widely practised by national statistical agencies as it is more user-friendly than the first approach.

Historical data, rebased historical data, "linking" year and base year

At each rebase, two sets of retrospective indices are available from the first period of the base year onwards (for example, from January 2000 to December 2002, if the indices base 2000 = 100 are presented starting from January 2003). These comprise old indices rebased by means of one of the two approaches described above and the new indices computed according to the new weighting system, sample of respondents, activity classification, etc. This means that statistical agencies may:

- replace the old index values with the new ones over the whole time span (2000-2002 in the example);
- replace the old index values with the new ones starting from the year following the base (the time span 2001-2002);
- maintain the old index rebased until December 2002.

In previous cases, the "link" year, i.e. the year at the beginning of which the old rebased index series are linked with the new index series is, respectively, 2000, 2001 and 2003. For presentation purposes, the importance of the "link" year arises because year-to-year percentage changes are computed comparing the index numbers belonging to different structures. Each of these approaches are common practice for statistical agencies. The first is especially appreciated by time series analysts as it removes structural

breaks, introduced at the beginning of the "link" year, from the end of the series. The third approach leaves index percentage changes already published unchanged.

6.5.5 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 re-base 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 re-base;
- the link year;
- the classification level at which index numbers are re-based 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 the 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 [30].

[30] If a statistical agency adopts the direct approach to rebase the indices, it does not provide the updated weights to the final users and the original weighting system cannot be used to aggregate the rebased indices (the additivity property is lost when the rebasing is carried out).

6.6 CITATION

This Section draws directly and extensively from a paper prepared by the UNESCO Institute for Statistics (UIS) presented at the 2nd session of the Committee for the Coordination of Statistical Activities (CCSA) in Geneva on 8-10 September 2003. The context of the paper was the need for the adoption of good citation practice by international organisations, though the practices cited are also relevant for national agencies. The focus of the paper was citation of datasets, though brief mention in this Section is also given to text citation.

6.6.1 Reasons for citation

Proper citation is an essential element of data and metadata reporting. Citation refers to the process of acknowledging within the organisation's current database or text the document, database or other source from which information has been obtained. A reference on the other hand refers to the detailed description of the actual source from which information has been obtained. A bibliography is a list of references consulted (Caledonian University 2003).

There are many reasons why citation of data is as important as citation of other published sources of material. These include:

- evaluation of the value of datasets is assisted by being able to track usage accurately. The inclusion of a feedback and tracking mechanism as part of a data citation policy is very useful in this regard;
- it facilitates assessment of the reliability of the information on the basis of its provenance and context. Also, additional information provided at the original source may permit the user to go more deeply into the subject and to verify sources and authenticity;
- the importance of giving appropriate credit to the producers of datasets. This is particularly the case in the increasingly competitive academic sector where credit needs to be attached to the production of high quality, well-documented datasets. It is also important in other sectors in an environment where different agencies re-use one another's data;
- it can enable other researchers to locate the exact version of the data used so that they might re-analyze the data to amplify, extend, confirm or refute the author's interpretation of it, all of which is an important part of the scientific process;
- it can enable other researchers to locate current versions of the same dataset or similar datasets from the same source;
- it is important that producers of the data should be able to locate quickly and accurately the exact version of the data supplied so they can answer queries quickly and can also resolve problems with the data.

These reasons underline the fact that effective citation also places an obligation on the data user to subsequently follow common citation best practice. To encourage effective citation, an obligation is therefore also placed upon the data provider to provide the necessary information (metadata) in conjunction with the dataset, e.g. as outlined in Section 6.6.4 below.

There is a more fundamental issue however. Even with all of the problems and challenges mentioned previously, it is possible for users to provide a basic citation for a dataset based on the guidelines that have

been provided in the various citation styles list in Section 6.6.2 below . This basic citation may not address all of the concerns that have been identified as the citation of datasets is not viewed in the same manner as citation of other materials. Unfortunately, the use of appropriate citation practices by both national agencies and international organisations is more the exception than the norm, particularly with respect to providing adequate reference information to statistics accessed from other agencies. The norm in relation to data sourced from databases accessed via websites or other online facilities is merely to cite the name of the organisation and (perhaps) almost generic descriptions of the actual database. There is considerable variation in the amount of citation information provided for data obtained from online sources.

6.6.2 Existing citation standards

There are a number of well developed and effective standards and styles for citation and bibliographic reference of material that been developed over many years and are in widespread use. Each of these styles have been updated so that they contain the necessary guidelines to effectively cite many forms of traditional information sources as well as newer information sources, such as the internet. The most popular of these styles include:

APA	generally used in psychology, education, and other social sciences
MLA	literature, arts, and humanities
AMA	medicine, health, and biological sciences
Turabian	designed or college students to use with all subjects
Chicago	used with all subjects by books, magazines, newspapers, and other publications.
Harvard	commonly used in the United Kingdom academia and in the legal profession.

Two ISO standards have also been created in the area of bibliographic references:

ISO 690	Information and documentation – Bibliographic references – Content, form, and structure.
ISO 690-2	Information and documentation – Bibliographic references – Part 2: Electronic documents or parts thereof.

ISO 690-2 specifies the elements to be included in bibliographic references to electronic documents. It sets out a prescribed order for the elements of the reference and establishes conventions for the transcription and presentation of information derived from the electronic source document. ISO 690-2 is intended for use by authors and editors in the compilation of references to electronic documents for inclusion in a bibliography, and in the formulation of citations within the text corresponding to the entries in that bibliography. It does not apply to full bibliographic descriptions as required by librarians, descriptive and analytic bibliographers, indexers, etc.

Even with all of this work in the area of citation, citation of datasets is still relatively unexplored. The complexity of dataset citation is increasing due to the ease in which data is redistributed and reused so that the original source may be a number of stages back. Data may be transformed accidentally or deliberately at any of these stages. Data may also be delivered embedded in software and will require metadata for

informed understanding. Data may also be very dynamic or provided via a database environment, which could make it difficult in the future to reproduce the state of the data at the time that it was cited. In international organizations, how to credit the data sources within countries for the provision of their data is also an issue.

6.6.3 Data management issues

The two of the reasons for data citation outlined in Section 6.6.1 above imply that:

- an historical copy is being maintained of datasets in an organization; and that the
- exact version of the dataset can be located based upon the information available to the user at the time they initially accessed the dataset.

In a highly dynamic environment where the data is constantly changing, these issues become very complex. The introduction of databases also complicates the matter. How to effectively cite the information in a highly dynamic environment will depend upon whether or not the organization providing the data can recreate the environment at the time of data retrieval. The notion of recording both the date and time of dataset access, as part of the dataset citation may be exactly what is required. Ultimately however, the recommendation of how to cite this dataset effectively will have to be provided by the organization that provides the dataset. Different technical implementations may require different information be included in the citation.

In addition, datasets are copied and redistributed in many forms to meet the needs of the moment. For internet dissemination, a dataset may be placed in an online database environment with interactive access to the data. The same dataset may be used as the foundation for analytical papers or may be placed in reference databases, CD-ROMs, or publications. The end result is that there are multiple uses of the original dataset both internal and external to the organization that created the dataset.

Organizations need to have an effective data management and data archival policy that will keep an historical record of the datasets. The retention period for various datasets will be different depending upon the data and requirements.

6.6.4 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 organizations to:

- Formulate and then place their data citation policy in an obvious position on websites, and also provide the policy in conjunction with any electronic datasets. This citation policy could be as simple as the following statement:

Citation Information

In the event that data from the <organization name> are incorporated into your research or publication, please supply the following acknowledgment within your published work: "These data are distributed by the <organization name> <organization website>"

If possible, please email or send us reprints/citations of papers or oral presentations founded on <organization name> data (see below for email and mailing address). This will help us to stay informed of how our data are being utilized.

There are no restrictions for use of data from the <organization name> unless otherwise expressly stated.

If you have any questions, please contact: <contact information>

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

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

- 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.

Traditionally, a citation only cites the most recent use of a reference even though it may have passed through a number of different organizations since the responsible organization first created it. That is, hypothetically, if UNESCO data was provided to the World Bank who then provided it to another organization, the World Bank would cite UNESCO and the other organization would cite the World Bank thus creating a chain of citations. The rationale is that by following the chain of citations, the original source of the reference and the responsible organization can be found. While this may not be the preferred approach for datasets, it is the most manageable approach. Adhering to this common citation practice would be the recommendation for dataset citations.

The challenge of effectively citing data sources in countries can be addressed by following this common citation practice of citing the most recent source. If the data is simply collected from countries by an

international organization, not modified in any way, and placed into the dataset, then the country should be cited as the source of the data. If the data is collected as part of a survey or statistical activity, which acts upon the data and subsequently generates a dataset, then the documentation of the survey activity should credit the data sources in the countries for providing all of the data. The dataset itself should reference the survey activity as the source of the data since it has gone through a lot more than simply a collection process. However, if a publication is produced by the same organization that has managed the data collection process, then thanking and providing credit to the countries for the original data would be appropriate.

If an organization takes a dataset and modifies it in some way before redistributing the dataset in whole or as part of a publication, the citation for the modified dataset must indicate the source or the original dataset but the citation must also indicate that the data was modified from its original state, together with information about the types of transformation undertaken.

Citation of text

The main recommendation for text citation is the systematic use in all metadata of one of the widely accepted bibliographic reference styles listed in Section 6.6.2 above. The two commonly used systems for presenting references in text for a bibliography are the Harvard system and the Numeric system[31]. It is beyond the scope of the current Manual 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.
-

[Insert other specific issues]

[31] Refer: Dee, Marianne (ed.) (1998); and Caledonian University (2003), respectively.

6.7 ADMINISTRATIVE DATA

6.7.1 Introduction

The publication, *Measuring the Non-Observed Economy: A Handbook*, compiled by the OECD, IMF, ILO and the Statistical Office of the Commonwealth of Independent States in 2002 cites two basic mechanisms used by statistical agencies for the collection of data – accessing data already being collected for administrative purposes and direct survey by an agency. The Handbook then describes the relative strengths and merits of these sources (OECD et al, p. 105-108).

Administrative data are derived from administrative processes established by government in response to legislation and regulation. The Handbook cites administrative sources producing two types of data that can be used for statistical purposes:

- registration data describing the institutional units that are required to register under legislation or regulation. Such information may be used to establish and maintain lists of units used for direct survey;

- transaction data describing the transactions administered under legislation or regulation. This information may be useful to supplement or replace direct survey.

Eurostat's *Use of Administrative Sources for Business Statistics Purposes: Handbook of Good Practices* provides the following comprehensive list of administrative sources that are potentially available in many countries (Eurostat 1999, p. 21):

- value added tax (VAT) data;
- personal income tax data;
- business (including corporate) taxation data;
- social security data;
- business registration and administration records;
- published business accounts;
- records held by central banks;
- records other than VAT held by customs and excise authorities;
- records of non-domestic consumers held by public utilities;
- records held by regulators of public utilities, transport and financial services;
- records held by associations of employers, of employees and of businesses and professions;
- records held by other private sector bodies, e.g. credit-rating agencies.

Such sources are common to both developed and developing countries. In addition to these, many countries also maintain a vast store of administrative data for births, deaths, education, health and unemployment registration purposes.

In recent years almost all national statistical and other national agencies have made greater use of some or most of the types of administrative data cited above, either in lieu of direct survey collection or to supplement data collected from households and business entities.

The catalysts for this increased use vary by country, with the main ones being: advances in information technology making access and manipulation of administrative data more feasible; and restrictions in budgets, etc, prompting the need for alternatives to direct collection for small and medium size units who contribute to a large proportion of activities in key sectors of the economy, e.g. services. However, perhaps

the main reason has been to meet increased pressure from government for statistical agencies to reduce the reporting burden of enterprises and households by making increased use of administrative data as either a substitute for direct collection or to supplement such sources for units with particular characteristics such as size, activity, geographic location etc.

As a result of these incentives, agencies at the national level make use administrative data for a wide range of purposes including (Eurostat 1999, pp. 17, 41):

- as a source of legal units for business registers;
- for coverage improvement;
- as a size indicator for units;
- as a sample stratification variable;
- as a universe for grossing up survey returns;
- as a data source for small and medium sized units excluded from direct survey collection;
- for the production of provisional results to improve timeliness;
- for compiling regional data.

In many instances, national statistical institutes have been vested with legal powers to access administrative data, to promote such access and even influence the quality of such data through mandatory use by administrative agencies of standard statistical classifications, common business identity numbers, etc.

6.7.2 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 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 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[32];

- 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.

[32] A task force of the OECD Short-term Economic Statistics Expert Group (STESEG) was established at the end of 2004 to review and provide recommendations on transformation processes that could expand the use of administrative source data which are available in countries but which for one or more reasons are not used. Information about this task force are available at …..

7. REFERENCES

ABS (Australian Bureau of Statistics), 2001, *Introductory Course on Time Series Analysis*, ABS, Canberra

ABS, 2003, Time Series Analysis: Issues With Seasonal Adjustment, Canberra, available at http://www.abs.gov.au/websitedbs/d3310114.nsf/0/70611eabf58a97acca256ce10018a0d3?OpenDocument [Accessed on 15 November 2004]

ABS, 2004, Australian Bureau of Statistics Publishing Manual, Canberra, unpublished

BIS, ECB, Eurostat, IMF, OECD, UN, 2003, SDMX – Statistical Data and Metadata Exchange, available at www.sdmx.org [Accessed 29 July 2003]

Bloem, A.M., Dippelsman, R.J., Mæhle, N.Ø. (2001) *Quarterly National Accounts Manual. Concepts, Data Sources, and Compilation.* International Monetary Fund. Available at: http://www.imf.org/external/pubs/ft/qna/2000/Textbook/index.htm.

Brass, W., 1975, *Methods for Estimating Fertility and Mortality from Limited and Defective Data*, Chapel Hill, North Carolina: Laboratories for Population Studies

Bureau of Economic Analysis, July 1993, The Impact of Classification Revisions on Time Series, Issues Paper No. 5, Economic Classification Policy Committee, Bureau of Economic Analysis, U.S. Department of Commerce, Washington D.C.

Caledonian University, 2003, Citation Guide, Caledonian University, Glasgow, available at http://www.lib.gcal.ac.uk/training/citguide.htm#Furthinf [Accessed 29 July 2003]

Carson, Carol, 2000, Toward a Framework for Assessing Data Quality, IMF Statistics Department, Washington DC

Ciammola, Anna, 2003, "Presentation of Retrospective Fixed Base Indexes", paper prepared for meeting of the OECD Short-term Economic Statistics Expert Group, 26-27 June 2003, available at http://www.oecd.org/dataoecd/61/54/2959065.pdf [Accessed 26 April 2004]

Dee, Marianne (ed.), 1998, Quote, Unquote: Guide to the Harvard Type of Referencing, Leeds Metropolitan University, Leeds, available at www.lmu.ac.uk/lss/ls/docs/Harvard/uploads/quoteunq.doc [Accessed on 29 July 2003]

ENCE (Escola Nacional de Ciências Estatísticas do IBGE), 2003, "Reporting and compensating for non-sampling errors for surveys in Brazil : current practice and future challenges", Ch. 11 of draft UN publication, *Household Surveys in Developing and Transition Countries: Design, Implementation and Analysis*, United Nations, New York 2003

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, 1999, *Use of Administrative Sources for Business Statistics Purposes: Handbook of Good Practices*, Version 1, Luxembourg, available from ….

Eurostat, 2002a, *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, 2002b, "Standard Quality Report", discussion item 4, *Assessment of the Quality in Statistics*, fifth meeting, Luxembourg, May 2002, Eurostat, Luxembourg

Eurostat, 2003a, Eurostat concept and definitions database (CODED), Eurostat, Luxembourg, available at 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 at http://europa.eu.int/comm/eurostat/ramon/index.cfm?TargetUrl=DSP_PUB_WELC Accessed 10 December 2004]

Eurostat, 2003c, "Methodological Documents: Standard Report", discussion item 4.2B, *Assessment of the Quality in Statistics*, sixth meeting, Luxembourg, October 2003.

Eurostat, 2003d, *Standard Quality Indicators: Producer-Oriented*, discussion item 7, Working Group "Assessment of Quality in Statistics, sixth meeting, Luxembourg, October 2003

Everitt, B.S., 1989, *The Cambridge Dictionary of Statistics, Cambridge University Press*, Cambridge, UK

Findley, D.F., Monsell, B.C., Bell, W.R., Otto, M.C., Chen, B. –C., 1998, "New Capabilities and Methods of the X-12-ARIMA Seasonal Adjustment Program", (With discussion), *Journal of Business and Economic Statistics*, 16, 2, 127-177, available from www.census.gov/srd/www/x12a/a12down_pc.html [Accessed .....]

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]

Friez, Adrien, 2003, "Some proposals about identification of published series", 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/12/61/2789026.pdf [Accessed 26 April 2003]

Gomez, V., Maravall, A., 1996, "Programs TRAMO and SEATS", *Documentode trabajo* n. 9628, Banco de Espana, Madrid

Graf, Monique, 2003, "Sampling error and seasonal adjustment", 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/61/60/2959043.pdf [Accessed 26 April 2003]

IMF, 1995, *Balance of Payments Compilation Manual*, IMF, Washington DC, Chapter XXI. Publication of Balance of Payments Statistics, paras. 1130-1147

IMF, 2001, "Seasonal Adjustment and Estimation of Trend-Cycles", Chapter 8 of *Quarterly National Accounts Manual – Concepts, Data Sources, and Compilation*, IMF, Washington DC, available from http://www.imf.org/external/pubs/ft/qna/2000/Textbook/ch8.pdf [Accessed 26 August 2003]

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

IMF, 2003c, *Introduction to the Data Quality Reference Site*, IMF, Washington DC, available from http://dsbb.imf.org/Applications/web/dqrs/dqrsintroduction/ [Accessed 25 August 2003]

IMF, 2003d, *Fifth Revision of the Fund's Data Standards Initiatives*, IMF, Washington DC, available from http://www.imf.org/external/np/sta/dsbb/2003/eng/061803.pdf [Accessed 25 August 2003]

IMF, 2003e, Data Quality Assessment Framework and Data Quality Program, IMF, Washington DC, available from http://www.imf.org/external/np/sta/dsbb/2003/eng/dqaf.htm#II [Accessed 25 August 2003]

IMF, 2003f, "SDDS Overview", *Special Data Dissemination Standards*, IMF, Washington DC, available from http://dsbb.imf.org/Applications/web/overview/ [Accessed 26 August 2003]

IMF, 2004, Carson, C. S., Khawaja, S., and Morrison, T.K., "Revisions Policy for Official Statistics: A Matter of Governance", IMF, Washington DC, available from ….. [Accessed ……]

IMF, 2005, Common Domain Reference Metadata Concepts: Proposal for Contents Standards, preliminary draft version 0.5 / 21 January 2005 presented at Workshop on SDMX Issues, Paris 24-26 January 2005, IMF, Washington DC

INSEE, 2003, *Informations rapides.* No. 340. November 12, 2003.

ISI (International Statistical Institute), 2003, *The Oxford Dictionary of Statistical Terms*, edited by Yadolah Dodge, Oxford University Press.

Kirchner, Robert, 2003, "Presentation of Growth Rates", discussion paper prepared for the OECD Short-term Economic Statistics Expert Group (STESEG), Paris, 26-27 June 2003, OECD, Paris, available at …. [Accessed on …]

Ladiray, D, Quenneville, B., 2001, *Seasonal Adjustment with the X-11 Method.* Springer-Verlag.
Rhoades, D., Elhawary-Rivet, N. (1983) Measuring the Current Rate of Inflation. *Current Economic Analysis*, **3**(1), xxv-xxxiii.

McManus, R., 2004, "Population and Migration Studies, 2004-2005", Population Tools, Dept. of Geography, St Patricks College, Dublin University, Dublin, Ireland, available at http://www.spd.dcu.ie/rmcmanus/Population%20Course/lectures_to_date.htm

OECD, 2000, "Basic Principles and Practices in Rebasing and Linking National Accounts Series", paper prepared by Derek Blades presented at Joint ADB/ESCAP *Workshop on Rebasing and Linking of National Accounts Series*, 21-24 March 2000, Bangkok

OECD, 2002a, *Glossary of Statistical Terms*, OECD, Paris, available at http://cs3-hq.oecd.org/scripts/stats/glossary/index.htm [Accessed 29 July 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, ILO, CIS Stat, 2002, *Measuring the Non-Observed Economy: A Handbook*, Paris

OECD, 2003a, "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, 2003b, *OECD Main Economic Indicators*, monthly, OECD, Paris

OECD, 2003c, *OECD Quarterly National Accounts*, quarterly, OECD, Paris

OECD, 2003d, "Manuals, Sources and Methods", OECD Statistics Portal, OECD, Paris, available at www.oecd.org/statistics/manuals [Accessed 29 July 2003]

OECD, 2003e, Quality Framework for OECD Statistics, Paris. Available at www.oecd.org/statistics [Accessed 21 April 2004]

OECD, 2003f, *Business Tendency Surveys: A Handbook*, OECD, Paris

OECD, 2004, "The OECD-JRC Handbook on Practices for Developing Composite Indicators", paper presented at a meeting of the OECD Committee on Statistics, 7-8 June 2004, OECD, Paris

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

OECD and 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 and ONS, 2004, Accessing and Improving Statistical Quality – Revisions Analysis for the National Accounts: OECD – ONS Workshop, 7-8 October 2004, OECD, Paris. Available from http://www.oecd.org/document/23/0,2340,en_2825_293564_33729303_1_1_1_1,00.html [Accessed 8 December 2004]

OED (The Oxford English Dictionary), 1989, eds. Simpson, J.A & Weiner, E.S.C., Volume 1, Clarendon Press, Oxford

OMB (United States Office of Management and Budget), 2001, "Measuring and Reporting Sources of Error in Surveys", *Statistical Policy Working Paper 31, July 2001*, Statistical Policy Office, Office of Information and Regulatory Affairs, Office of Management and Budget, Washington DC, available from …. [Accessed …]

Palmore, James A., Gardner, Robert W., 1994, *Measuring mortality, fertility and natural increase: a self-teaching guide to elementary measures*, Rev. ed., East-West Centre, Honolulu, Hawaii

Rowland, D., 2003, *Demographic Methods and Concepts*, Section 1: Population Dynamics, Part 2: Population growth and decline, Oxford University Press, Oxford, UK, available at http://www.oup.co.uk/pdf/bt/rowland/ch02.pdf [accessed 31 January 2005]

Siddiqui, M.K., 1999, Utilization of Statistical Information: the Case of Population Statistics, paper presented at Asia-Pacific OPOIN-IFLA Joint Workshop on Demographic Information, Bangkok, August 1999

Solans, Eugenio Domingo, 2003, "Official Statistics for a Global Economy", presented at the *54th Session of the International Statistical Institute, Berlin, 13-20 August 2003*, European Central Bank, Frankfurt

Statistics Canada, 2000, *Policy on Informing Users of Data Quality and Methodology*, Statistics Canada, Ottawa, available from http://www.statcan.ca/english/concepts/policy-infousers.htm [Accessed 31 August 2003]

Statistics Canada, 2002, *Statistics Canada Quality Guidelines*. 3rd ed., Statistics Canada, Ottawa, available from http://www.statcan.ca/english/freepub/12-586-XIE/12-586-XIE02001.pdf [Accessed 25 August 2003]

Statistics Finland, 2003, "Indicators and Indices", Quality Guidelines, Helsinki, available from www.stat.fi/tk/tt/laatuatilastoissa/lm021400/pe_en.html [Accessed 3 May 2004]

Statistics New Zealand, 2001, Seasonal Adjustment Within Statistics New Zealand, Statistics New Zealand, Wellington, available at http://www.stats.govt.nz/products-and-services/Articles/sea-adj-Apr01.htm [Accessed 15 November 2004]

UNCTAD, 2003, "Harmonisation of Base Years for Index Numbers", prepared for presentation at the 2nd session of the Committee for the Co-ordination of Statistical Activities, Geneva, 8-10 September 2003, UNCTAD,

UNDG (United Nations Development Group), 2003, *Indicators for Monitoring the Millennium Development Goals: Definitions, Rationale, Concepts and Sources*, United Nations, New York

UNECE (United Nations Economic Commission for Europe), 2000, "Report of the September 1999 Work Session on Statistical Metadata", prepared for presentation at the 48th plenary session of the Conference of European Statisticians, Paris 13-15 June 2000, UNECE, Geneva, p. 3 (CES/2000/20).

UNECE, Eurostat and OECD, 2004, Report of the February 2004 Joint UNECE/Eurostat, OECD Work Session on Statistical Metadata, UNECE, Geneva. Available at http://www.unece.org/stats/documents/ces/ac.71/2004/22.add.1.e.pdf [Accessed 8 April 2004]

UNEP (United Nations Environment Programme), 2002, "GRI Indicators", *Global Reporting Indicators: A Common Framework for Sustainability Reporting*, United Nations, New York. Available at http://www.globalreporting.org/guidelines/2002/dannex5.asp [Accessed 7 May 2004]

UNESCO, 2003, "Good Practices in Citation in the Outputs of International Statistical Offices: The Citation of Datasets – Report Prepared by UNESCO", Paris, presented at the 2nd Session, Committee for the Coordination of Statistical Activities, Geneva, 8-10 September 2003. Available at [Accessed]

UNSC (United Nations Statistical Commission), 1994, *Fundamental Principles of Official Statistics*, UNSD, New York, available from http://unstats.un.org/unsd/goodprac/bpabout.asp [Accessed 25 August 2003]

UNSC and UNECE, 2000, *Guidelines for Statistical Metadata on the Internet*, Conference of European Statisticians Statistical Standards and Studies – No. 52, United Nations, Geneva, available at

http://www.oecd.org/findDocument/0,2350,en_2649_37423_1_1_1_1_37423,00.html [Accessed 29 July 2003]

UNSC and UNECE, 2000a, *Terminology on Statistical Metadata*, Conference of European Statisticians Statistical Standards and Studies – No. 53, United Nations, Geneva. Available at http://www.unece.org/stats/documents/statistical_standards_&studies/52.e.pdf [Accessed 16 December 2003]

UNSC and UNECE, 2001, *Best Practice in Designing Websites for Dissemination of Statistics*, UNECE, Geneva, available at www.unece.org/stats/publications/websitebestpractice.pdf [Accessed on 29 July 2003]

UNSD (United Nations Statistical Division), 1964, *Recommendations for the Preparation of Sample Survey Reports (Provisional Issue), Statistical Papers, Series C, No. 1 Rev. 2*, New York

UNSD, 1982, *Non-sampling Errors in Household Surveys: Sources, Assessment and Control*, United Nations, New York

UNSD 1983, *Manual X: Techniques for Demographic Estimation*, Population Studies No. 81, United Nations, New York

UNSD, 1993, *Sampling Errors in Household Surveys*, UN Department for Economic and Social Information and Policy Analysis, Statistical Division, New York. Available at http://unstats.un.org/unsd/pubs/gesgrid.asp?ID=316 [Accessed on 3 November 2004]

UNSD, 2001, *Handbook on the Operation and Organisation of a Statistical Agency*, UNSD, New York, available from ….. [Accessed ….]

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]

UNSD, 2003, *Household Surveys in Developing and Transition Countries: Design, Implementation and Analysis*, New York, available from http://unstats.un.org/unsd/HHsurveys/index.htm [Accessed 22 April 2004]

UNSD, 2004, Economic and Social Statistical Classifications, United Nations, New York, available from http://unstats.un.org/unsd/class/default.htm [Accessed 10 December 2004)

UK Office for National Statistics (ONS), 2003, *National Statistics Code of Practice: Protocol on Data Management, Documentation and Preservation – Draft for consultation*, March 2003, Newport

UK Office for National Statistics (ONS), 2004, *National Statistics Code of Practice: Protocol on Revisions*, Newport

US Bureau of Economic Analysis, July 1993, "The Impact of Classification Revisions on Time Series", *Issues Paper No. 5*, Economic Classification Policy Committee, Bureau of Economic Analysis, U.S. Department of Commerce, Washington D.C.

US Bureau of Labour Statistics, 1996, "The Tension Between Continuity and Relevance of Economic Time Series", paper prepared by Thomas Plewes presented at UNECE/UNSC *Seminar on Official Statistics – Past and Future*, Lisbon, 25-27 September 1996.

US Census Bureau, 2002, *Census Bureau Releases Benchmark Report for Monthly Manufacturers' Shipments, Inventories, and Orders Survey*, US Census Bureau, Washington DC, available at www.census.gov/indicator/www/m3/bench/bench.htm [Accessed 29 July 2003]

U.S. Census Bureau, 2002, *X-12-Arima Reference Manual. Version 0.2.10.* Available at: http://www.census.gov/srd/www/x12a/x12down_pc.html

US Census Bureau, 2003, *Frequently asked questions*, US Census Bureau, Washington DC, available at www.census.gov/mcd/mcdfaqs.html [Accessed 29 July 2003]

US Fedstats, 2003, *Schedule of Release Dates for Principal Federal Economic Indicators for 2003*, US Fedstats, Washington DC, available from http://www.fedstats.gov/policy/pei_calendar2003.pdf [Accessed 26 August 2003]

World Bank, 2004, *2004 World Development Indicators*, World Bank, Washington DC

ANNEXES

Annex 1

SUMMARY OF DATA AND METADATA REPORTING AND PRESENTATION STANDARDS PRESENTED IN THIS HANDBOOK

[to be prepared]

GLOSSARY OF DATA AND METADATA REPORTING AND PRESENTATION TERMS

Absolute figures	See Original data
Annual changes *(draft definition)*	Annual changes refer to annual changes in levels expressed over the previous year. Such changes are expressed as Y_t-Y_{t-1}. Y_t denotes the value of an annual time series in year t. Task force on data and metadata presentation for the OECD Short-term Economic Statistics Expert Group (STESEG), Paris, 2004
Annual growth rates (draft definition)	Annual growth rates are annual rates of change expressed over the previous year. Such rates are expressed as (Y_t/Y_{t-1})-1. Y_t denotes the value of an annual time series in year t. Task force on data and metadata presentation for the OECD Short-term Economic Statistics Expert Group (STESEG), Paris, 2004
Annualised growth rate (Annualised rate of change) *(draft definition)*	Annualised growth rates (Annualised rate of change) show the value that would be registered if the quarter-on-previous quarter or month-on-previous month rate of change were maintained for a full year. Such rates are expressed as $((Q_t/Q_{t-1})^4)$-1 for quarterly data and $((M_t/M_{t-1})^{12})$-1 for monthly data. M_t denotes the value of a monthly time series in month t and Q_t the value of a quarterly time series in quarter t. Annualised growth rates facilitate comparison of data for different time periods (e.g. years and quarters) but may suffer from high volatility in a series. *Context:* In addition to the compounded form of the Annualised growth rate presented here, the term "Annualised growth rate" is sometimes used to describe the quarterly or monthly growth rate multiplied by four or twelve. Multiplying the quarterly or monthly growth rate by four or twelve is more appropriately referred to as "Linear approximation of the annualised growth rate". Some agencies use the expression "1-month rate of change, annualised", etc, for such rates. Task force on data and metadata presentation for the OECD Short-term Economic Statistics Expert Group (STESEG), Paris, 2004
Annualised growth rate, linear approximation	See Linear approximation of the annualised growth rate
Annualised semi-annual growth rate *(draft definition)*	Annualised semi-annual growth rates show the value that would be registered if the rate of change measured with reference to two quarters or six months behind were maintained for a full year. Such rates are expressed as $((Q_t/Q_{t-2})^2)$-1, $((M_t/M_{t6})^2)$-1. M_t denotes the value of a monthly time series in month t and Q_t the value of a quarterly time series in quarter t. Annualised semi-annual growth rates facilitate comparison of data for different time periods (e.g. years and quarters). Task force on data and metadata presentation for the OECD Short-term Economic Statistics Expert Group (STESEG), Paris, 2004

Arithmetic growth rate

Arithmetic growth refers to the situation where a population increases by a constant number of persons (or other objects) in each period being analysed.

In general terms an arithmetic growth rate may be expressed as:

$$\left(\frac{Pn - Po}{n}\right) \div \text{Po} \times 100$$

Where P_o = population at the start; P_n = population at the end; n = number of intervals between P_o and P_n

Rowland, D., 2003, *Demographic Methods and Concepts*, Section 1: Population Dynamics, Part 2: Population growth and decline, Oxford University Press, Oxford, UK

Balances, survey

Balances (also called net balances) are used to summarise answers to multiple-choice questions in business tendency and consumer opinion surveys. "No-change" answers (such as "normal" or "same") are ignored and the balance is obtained by taking the difference between the weighted percentages of respondents giving favourable and unfavourable answers.

OECD, 2003f, *Business Tendency Surveys: A Handbook*, OECD, Paris, Glossary

Context: Survey balances can be positive or negative

Base

A number or magnitude used as a standard of reference. It may occur as a denominator in a ratio or percentage calculation. It may also be the magnitude of a particular time series from which a start is to be made in the calculation of a new relative series – an index number – which will show the observations as they accrue in the future in relation to that of the base period.

A Dictionary of Statistical Terms, 5th edition, prepared for the International Statistical Institute by F.H.C. Marriott. Published for the International Statistical Institute by Longman Scientific and Technical.

Base period

The period of time for which data used as the base of an index number, or other ratio, have been collected. This period is frequently one of a year but it may be as short as one day or as long as the average of a group of years.

In the System of National Accounts (SNA) 1993, par.16.16, the base period is described as the period that provides the weights for an index.

Context: Under the SDDS, this refers to the period when the published index = 100, or the reference period to which the average level and/or constant price series data refer.

The International Statistical Institute, "The Oxford Dictionary of Statistical Terms", edited by Yadolah Dodge, Oxford University Press, 2003.

Base weight

The weights of a weighting system for an index number computed according to the information relating to the base period instead, for example, of the current period.

The International Statistical Institute, "The Oxford Dictionary of Statistical Terms", edited by Yadolah Dodge, Oxford University Press, 2003.

Business cycles

Business cycles are recurrent sequences of alternating phases of expansion and contraction in the level of a time series.

OECD Leading Indicator Website, Glossary, 2001.

Calendar adjustment

(draft definition)

Calendar adjustment refers to the correction for calendar variations other than seasonal components. Such calendar adjustments include working day adjustments, changes in accounting or recording practices adopted or the incidence of moving holidays (such as Easter and Chinese New Year).

Context: The terms "calendar adjustment" and "working day adjustment" (also known as "trading day adjustment") are often used interchangeably. However, the main difference between the two terms is

	that working day adjustment is merely one type of calendar adjustment. Another type of calendar adjustment is a "shopping day" adjustment for consumer expenditure or retail trade series. Each variable has its own specific calendar adjustment. The length of month effect is assigned to the seasonal component because it happens year after year in the same period with the exception of leap year effects. Task force on data and metadata presentation for the OECD Short-term Economic Statistics Expert Group (STESEG), Paris, 2004
Calendar effects component of a time series *(draft definition)*	The calendar effects component is that part of the seasonal component that represents calendar variations in a time series, such as trading days / working days, moving holidays and other calendar effects (such as leap year). The effects of the normal length of a month are assigned to the seasonal component. *Context:* The calendar component is often slightly moving and may disturb the stability of the seasonal component. Task force on data and metadata presentation for the OECD Short-term Economic Statistics Expert Group (STESEG), Paris, 2004
Chain index	An index number in which the value at any given period is related to a base in the previous period, as distinct from one which is related to a fixed base. *Context:* The comparison of non-adjacent periods is usually made by multiplying consecutive values of the index numbers, which, as it were, form a chain from one period to another. In practice chain index numbers are usually formed from weighted average of link-relatives, namely the values of magnitudes for a given period divided by the corresponding values in the previous period. The International Statistical Institute, "The Oxford Dictionary of Statistical Terms", edited by Yadolah Dodge, Oxford University Press, 2003.
Chain linking	Joining together two indices that overlap in one period by rescaling one of them to make its value equal to that of the other in the same period, thus combining them into single time series. More complex methods may be used to link together indices that overlap by more than period. *Context:* Also known as "chaining". PPI Manual – Glossary of Terms, Draft, Producer Price Index Technical Expert Group, November 2002.
Classification	A classification is a set of discrete, exhaustive and mutually exclusive observations which can be assigned to one or more variables to be measured in the collation and/or presentation of data. *Context:* The terms "classification" and "nomenclature" are often used interchangeably, despite the definition of a "nomenclature" being narrower than that of a "classification". "United Nations Glossary of Classification Terms" prepared by the Expert Group on International Economic and Social Classifications; unpublished on paper.
Classifications, standard	Standard classifications are those that follow prescribed rules and are generally recommended and accepted. They aim to ensure that information is classified consistently regardless of the collection, source, point of time, etc. "United Nations Glossary of Classification Terms" prepared by the Expert Group on International Economic and Social Classifications, unpublished on paper. *Context:* In the international context standard classifications include ISIC Rev. 3, ISCO, CPC, etc. Many national statistical systems also have their own versions of standard classifications which in the

	main are consistent with international standard classifications, though modified to meet national circumstances.
Coefficient of variation	The standard deviation of a random variable divided by the mean. The International Statistical Institute, "The Oxford Dictionary of Statistical Terms", edited by Yadolah Dodge, Oxford University Press, 2003. *Context:* The US Bureau of Census alternatively refers to the coefficient of variation as the ratio of the standard error to the value being estimated, usually expressed in terms of a percentage. Also known as the relative standard deviation. (United States Bureau of Census, Glossary of Selected Abbreviations and Acronyms - refer http://eire.census.gov/cgi-bin/ssd/Glossary<BR< a>
Composite indicator	A composite indicator is formed when individual indicators are compiles into a single index, on the basis of an underlying model of the multi-dimensional concept that is being measured. *Context*: A composite indicator measures multi-dimensional concepts (e.g. competitiveness, e-trade or environmental quality) which cannot be captured by a single indicator. Ideally, a composite indicator should be based on a theoretical framework / definition, which allows individual indicators / variables to be selected, combined and weighted in a manner which reflects the dimensions or structure of the phenomena being measured. OECD, 2004, "The OECD-JRC Handbook on Practices for Developing Composite Indicators", paper presented at the OECD Committee on Statistics, 7-8 June 2004, OECD, Paris
Constant prices	Constant prices are obtained by directly factoring changes over time in the values of flows or stocks of goods and services into two components reflecting changes in the prices of the goods and services concerned and changes in their volumes (i.e. changes in "constant price terms"); the term "at constant prices" commonly refers to series which use a fixed-base Laspeyres formula. SNA 1993 paras. 16.2, 16.71.
Cumulative data	See Year-to-date-data
Current prices	[definition to be developed]
Cyclical component of a time series (draft definition)	The cyclical component of a time series refers to (regular or periodic) fluctuations around the trend, excluding the irregular component, revealing a succession of phases of expansion and contraction. The cyclical component can be viewed as those fluctuations in a time series which are longer than a given threshold, e.g. 1½ years, but shorter than those attributed to the trend. Task force on data and metadata presentation for the OECD Short-term Economic Statistics Expert Group (STESEG), Paris, 2004
Decomposition	The act of splitting a time series into its constituent parts by the use of statistical methods. A typical time series is often regarded as composed of four parts: (a) a long-term movement or trend; (b) oscillations of more or less regular period and amplitude about this trend; (c) a seasonal component; (d) a random, or irregular, component. Any particular series need not exhibit all of these but those which are present are presumed to act in an additive fashion, i.e. are superimposed; and the process of determining them separately is one of decomposition. A Dictionary of Statistical Terms, 5th edition, prepared for the International Statistical Institute by F.H.C. Marriott. Published for the International Statistical Institute by Longman Scientific and Technical.
Deflation	The division of the value of some aggregate by a price index - described as a "deflator" - in order to revalue its quantities at the prices of the price reference period or to revalue the aggregate at the

general price level of the price reference period.

PPI Manual – Glossary of Terms, Draft, Producer Price Index Technical Expert Group, November 2002

Dissemination, data

Dissemination is the release to users of information obtained through a statistical activity.

Statistics Canada, "Statistics Canada Quality Guidelines", 3rd edition, October 1998, page 59.

Context: Data dissemination consists of distributing or transmitting statistical data to users. Various release media are possible; for example: electronic format including the internet, CD-ROM, paper publications, files available to authorised users or for public use; fax response to a special request, public speeches, press releases.

Dissemination formats. Under the SDDS, the concept of dissemination formats is divided into two categories: "hardcopy" and "electronic" publications, which detail the reference documents through which users may access the data described in the metadata and, where relevant, detailed components beyond the minimum prescribed.

Double deflation

Double deflation is a method whereby gross value added is measured at constant prices by subtracting intermediate consumption at constant prices from output at constant prices; this method is feasible only for constant price estimates which are additive, such as those calculated using a Laspeyres' formula (either fixed-base or for estimates expressed in the previous year's prices).

SNA 1993 para. 16.5

Exponential growth rate

Exponential growth refers to the situation where growth compounds continuously at every instant of time.

Because compounding takes place at intervals much longer than an instant, a geometric growth as being a "special case" of exponential growth.

In general terms an exponential growth rate may be expressed as

$$\frac{In\left(\frac{Pn}{Po}\right)}{n}$$

Where P_o = population at the start; P_n = population at the end; n = number of intervals between P_o and P_n; In = natural logarithm

Rowland, D., 2003, *Demographic Methods and Concepts*, Section 1: Population Dynamics, Part 2: Population growth and decline, Oxford University Press, Oxford, UK

Filter

Any method of isolating harmonic constituents in a time series; a mathematical analogy of the "filtering" of a ray of light or sound by removing unsystematic effects and bringing out the constituent harmonics.

A Dictionary of Statistical Terms, 5th edition, prepared for the International Statistical Institute by F.H.C. Marriott. Published for the International Statistical Institute by Longman Scientific and Technical.

Flow series / data

Statistical series presented as flow series/data are cumulated during the reference period, for example, passenger car registrations, where the figure for the reference period is the sum of daily registrations.

Organisation for Economic Co-operation and Development (OECD), "Main Economic Indicators", monthly.

Forecasting

"Forecasting" and "prediction" are often used synonymously in the customary sense of assessing the magnitude which a quantity will assume at some future point of time: as distinct from "estimation" which attempts to assess the magnitude of an already existent quantity. For example, the final yield of a crop is "forecast" during the growing period but "estimated" at harvest.

A Dictionary of Statistical Terms, 5th edition, prepared for the International Statistical Institute by F.H.C. Marriott. Published for the International Statistical Institute by Longman Scientific and Technical.

Geometric growth rate

Geometric growth refers to the situation where successive changes in a population differ by a constant ratio (as distinct from a constant amount for arithmetic change).

In general terms a geometric growth rate may be expressed as

$$\sqrt[n]{\frac{Pn}{Po}} - 1$$

Where P_o = population at the start; P_n = population at the end; n = number of intervals between P_o and P_n

Rowland, D., 2003, *Demographic Methods and Concepts*, Section 1: Population Dynamics, Part 2: Population growth and decline, Oxford University Press, Oxford, UK

Growth rates

Growth rates are ratios of total change in a specified time reference period to values at the beginning of the period or at a specified earlier time reference.

Adapted from Multilingual Demographic Dictionary, English Section, United Nations, 1958, Department of Economic and Social Affairs, Population Studies, No. 29.

Indicator, statistical

A statistical indicator is a data element that represents statistical data for a specified time, place, and other characteristics.

Economic Commission for Europe of the United Nations (UNECE), "Terminology on Statistical Metadata", Conference of European Statisticians Statistical Standards and Studies, No. 53, Geneva, 2000.

Index number

An index number is a quantity that shows by its variations the changes over time or space of a magnitude.

Important features in the construction of an index number are its coverage, base period, weighting system and method of averaging observations.

Context: The above definition relates to the usual meaning of the expression “index number”. In full generality, however, the term can also be applied to a series of values which are standardised by being referred to a basic period or area, e.g. if the price of a fixed commodity in a basic year is 40 units and those in succeeding years are 60 and 68 units, the index number of those years would be, on the basis of 100 for the first year, 150 and 170. Such simple cases are, however, usually referred to as “relatives” and the index number is constructed as an average of a number of relatives.

The International Statistical Institute, “The Oxford Dictionary of Statistical Terms”, edited by Yadolah Dodge, Oxford University Press, 2003.

Irregular component of a time series

(draft definition)

The irregular component of a time series is the residual time series after the trend-cycle and the seasonal components, as well as calendar effects, have been removed. It corresponds to the high frequency fluctuations of the series.

Context: The irregular component results from short term fluctuations in a series which are not systematic and in some instances not predictable e.g. uncharacteristic weather patterns. Some irregular effects can however be expected in advance, e.g. changes in value added tax.

In a highly irregular series, these fluctuations can dominate movements, which will mask the trend and seasonality. (ABS)

Task force on data and metadata presentation for the OECD Short-term Economic Statistics Expert Group (STESEG), Paris, 2004

Least-squares growth rate

Level (monthly, quarterly, annual) data

Data expressed as levels are expressed in absolute terms (values, numbers, and units) for a given period (month, quarter, and year).

Organisation for Economic Co-operation and Development (OECD), "Main Economic Indicators", monthly.

Linear approximation of the annualised growth rate

(draft definition)

Linear approximation of the annualised growth rate is a quick calculation of the annualised growth rate that shows the levels that would be measured for a quarter or month were maintained for a full year. Quarterly rates of change are multiplied by 4 and monthly rates of change are multiplied by 12. Such rates are expressed as $4^* (Q_t-Q_{t-1}/Q_{t-1})$ for quarterly data or $12^* (M_t- M_{t-1}/M_{t-1})$ for monthly data.

M_t denotes the value of a monthly time series in month t and Q_t the value of a quarterly time series in quarter t.

Such rates facilitate comparison of data for different time periods (e.g. years and quarters).

Context: This approximation is valid if the growth rate is small, but the linear approximation of the annualised growth rate should not be used in official releases. The compound form of the Annualised growth rate should be preferred.

Task force on data and metadata presentation for the OECD Short-term Economic Statistics Expert Group (STESEG), Paris, 2004

Metadata

Metadata is data that defines and describes other data.

ISO/IEC FCD FCD 11179-1 "Information technology - Metadata registers-Part 1: Framework", May 2003.

Metadata, statistical

Statistical metadata are data about statistical data.

United Nations Statistical Commission and Economic Commission for Europe of the United Nations (UNECE), "Guidelines for the Modelling of Statistical Data and Metadata", Conference of European Statisticians, Methodological material, United Nations, Geneva, 1995.

Context: Statistical metadata provide information on data and about processes of producing and using data. Metadata describe statistical data and - to some extent - processes and tools involved in the production and usage of statistical data (UNECE, "Guidelines for the Modelling of Statistical Data and Metadata", 1995).

Monthly average

By analogy with annual averages and moving averages generally this term ought to refer to the average values of a time series occurring within a month, the resulting figure being representative of that particular month.

Context: In practice the phrase is sometimes used to denote the averaging of monthly values occurring in the same month, e.g. January from year to year, the object being to provide a pattern of seasonal fluctuation. This is objectionable and a better expression would be “seasonal average by months”.

A Dictionary of Statistical Terms, 5th edition, prepared for the International Statistical Institute by F.H.C. Marriott. Published for the International Statistical Institute by Longman Scientific and Technical.

Month-on-previous-month growth rates

(draft definition)

Month-on-previous-month growth rates are rates of change expressed with respect to the previous month. Such rates are expressed as $(M_t/M_{t-1})-1$.

M_t denotes the value of a monthly time series in month t.

Context: Also often referred to as Month-to-month (Period-to-period) growth rates, Month-over-month growth rates, 1-month rate of change, or Rate of change on the previous month.

Task force on data and metadata presentation for the OECD Short-term Economic Statistics Expert

Group (STESEG), Paris, 2004

Month-on-previous month changes

(draft definition)

Month-on-previous-month changes are changes in levels expressed with respect to the previous month. Such rates are expressed as M_t-M_{t-1}.

M_t denotes the value of a monthly time series in month t.

Context: Also often referred to as Month-to-month (Period-to-period) changes, or Month-over-month changes.

Task force on data and metadata presentation for the OECD Short-term Economic Statistics Expert Group (STESEG), Paris, 2004

Moving average

(draft definition)

A moving average is a method for smoothing time series by averaging (with or without weights) a fixed number of consecutive terms. The averaging "moves" over time, in that each data point of the series is sequentially included in the averaging, while the oldest data point in the span of the average is removed. In general, the longer the span of the average, the smoother is the resulting series.

Context: Moving averages are used to smooth fluctuations in time series or to identify time series components, such as the trend, the cycle, the seasonal, etc.

A moving average replaces each value of a time series by a (weighted) average of *p* preceding values, the given value, and *f* following values of a series.

If $p = f$ the moving average is said to be centered.

The moving average is said to be symmetric if it is centered, and if for each $k = 1, 2, \ldots, p = f$, the weight of the *k*-th preceding value is equal to the weight of the *k*-th following one.

The moving average is not defined for the first p and the last f time series values. In order to compute the moving average for those values, the series must be backcasted and forecasted.

Task force on data and metadata presentation for the OECD Short-term Economic Statistics Expert Group (STESEG), Paris, 2004

Moving holidays

(draft definition)

Moving holidays are holidays which occur each year, but where the exact timing shifts. Examples of moving holidays include Easter and Chinese New Year. Easter generally falls in April but it can also fall in late March.

Context: Also known as variable holidays

Task force on data and metadata presentation for the OECD Short-term Economic Statistics Expert Group (STESEG), Paris, 2004

Non-sampling error

An error in sample estimates which cannot be attributed to sampling fluctuations.

Non-sampling errors may arise from many different sources such as defects in the frame, faulty demarcation of sample units, defects in the selection of sample units, mistakes in the collection of data due to personal variations or misunderstanding or bias or negligence or dishonesty on the part of the investigator or of the interviewer, mistakes at the stage of the processing of the data, etc.

The International Statistical Institute, "The Oxford Dictionary of Statistical Terms", edited by Yadolah Dodge, Oxford University Press, 2003.

Original data

Original data in the context of this Handbook refers to statistical information that has not undergone any transformation process such as the compilation of indices, growth rates or ratios. Original data are normally expressed either in physical unit terms (tonnes, cubic metres, etc) or as values (at current or constant prices).

Context: Original data are also commonly referred to as "absolute figures" or "absolute data". Original data could also be described as data received directly from survey or administrative sources.

OECD

Percentage (form)	
Percentage change	The change in an index [or other] series from one period to another expressed as a percentage of its value in the first of the two periods. Producer and International Trade Price Indexes, Glossary of terms – Australian Bureau of Statistics.
Preliminary / Provisional *(draft definition)*	Some statistical agencies use the term "Preliminary data" to describe the first released version of a series and "Provisional data" to describe subsequent versions prior to final amendment. However, the two terms are often used interchangeably though users in general should have no great problem in understanding that data labelled either "preliminary" or "provisional" are subject to revision provided this is clearly highlighted by the agency in the release. Clearly informing the use that the data is subject to revision is more important than the term used to describe such data. OECD, "Some proposals for standard terminology relating to data presentation", paper presented to the OECD Short-term Economic Statistics Expert Group (STESEG), 28-30 June 2004, Paris, 2004
Price index	A price index reflects an average of the proportionate changes in the prices of a specified set of goods and services between two periods of time. SNA 1993 para. 16.14 Usually a price index is assigned a value of 100 in some selected base period and the values of the index for other periods are intended to indicate the average percentage change in prices compared with the base period. (PPI Manual – Glossary of Terms, Draft, Producer Price Index Technical Expert Group, November 2002.
Probability	
Proportion	
Qualitative data	Qualitative data is data describing the attributes or properties that an object possesses. The properties are categorized into classes that may be assigned numeric values. However, there is no significance to the data values themselves, they simply represent attributes of the object concerned. Economic Commission for Europe of the United Nations (UNECE), "Glossary of Terms on Statistical Data Editing", Conference of European Statisticians Methodological material, Geneva, 2000.
Quantitative data	Quantitative data is data expressing a certain quantity, amount or range. Usually, there are measurement units associated with the data, e.g. meters, in the case of the height of a person. It makes sense to set boundary limits to such data, and it is also meaningful to apply arithmetic operations to the data. Economic Commission for Europe of the United Nations (UNECE), "Glossary of Terms on Statistical Data Editing", Conference of European Statisticians Methodological material, Geneva, 2000. *Context:* Strictly, this term, as contrasted with qualitative data, should relate to data in the form of numerical quantities such as measurements or counts. It is sometimes, less exactly, used to describe material in which the variables concerned are quantities, e.g. height, weight, price as distinct from data deriving from qualitative attributes, e.g. sex, nationality or commodity. This usage is to be avoided in favour of such expressions as "data concerning quantitative (qualitative) variables" or "data concerning numerical variables (attributes)". (A Dictionary of Statistical Terms, 5th edition, prepared for the International Statistical Institute by F.H.C. Marriott. Published for the International Statistical Institute by Longman Scientific and Technical).
Quantity index	A measure reflecting the average of the proportionate changes in the quantities of a specified set of goods and services between two periods of time. Usually a quantity index is assigned a value of 100 in some selected base period and the values of the index for other periods are intended to indicate the average percentage change in quantities compared with the base period. PPI Manual – Glossary of Terms, Draft, Producer Price Index Technical Expert Group, November

2002.

Quarter-on-previous-quarter growth rates

(draft definition)

Quarter-on-previous-quarter growth rates are rates of change expressed with respect to the previous quarter. Such rates are expressed as $(Q_t / Q_{t-1}) - 1$.

Q_t denotes the value of a quarterly time series in quarter t.

Context: Also often referred to as Quarter-to-quarter (Period-to-period) growth rates, Quarter-over-quarter growth rates, 1-quarter growth rates, or Rate of change on the previous quarter.

Task force on data and metadata presentation for the OECD Short-term Economic Statistics Expert Group (STESEG), Paris, 2004

Quarter-on-previous quarter changes

(draft definition)

Quarter-on-previous-quarter changes are changes in levels expressed with respect to the previous quarter. Such rates are expressed as $Q_t - Q_{t-1}$.

Q_t denotes the value of a quarterly time series in quarter t.

Context: Also often referred to as Quarter-to-quarter (Period-to-period) changes, or Quarter-over-quarter changes.

Task force on data and metadata presentation for the OECD Short-term Economic Statistics Expert Group (STESEG), Paris, 2004

Rate

A rate is a stated numerical proportion between two sets of things, the second generally expressed as unity.

Oxford University Press, 1995, The Concise Oxford Dictionary of Current English, eds. Fowler, H.W. & Fowler, F.G.

Context: In social and population statistics a rate refers to the occurrence of events over a given interval of time. They are used to study the dynamics of change, e.g. total fertility rate, crude death rate.

Rates of change

See Growth rates

Ratio

A ratio refers to the quantitative relation between two similar magnitudes determined by the number of times one contains the other integrally or fractionally.

Oxford University Press, 1995, The Concise Oxford Dictionary of Current English, eds. Fowler, H.W. & Fowler, F.G.

Context: More specifically for social and population and other statistics, a ratio is a single number that expresses the relative size of two other numbers. Ratios are useful for analyzing the composition of two sets of events.

Rebasing

In the course of time, the pattern of relative prices in the base period tends to become progressively less relevant to the economic situations of later periods to the point at which it becomes unacceptable to continue using them to measure volume measures from one period to the next; it may then be necessary to update the base period, a process which is commonly referred to as “rebasing".

SNA 1993 para. 16.31.

Context: There is some ambiguity due to ambiguity in the concept of the base year. Rebasing may mean changing the index reference-period of an index number series, or it may mean changing the weights in the index. Both the index reference-period and the weight reference-period may be changed at the same time, but not necessarily so.

PPI Manual – Glossary of Terms, Draft, Producer Price Index Technical Expert Group, November 2002

Reference period

In one sense, this is synonymous with base period. It may also refer to the length of time, e.g. week or year, for which data are collected.

The International Statistical Institute, “The Oxford Dictionary of Statistical Terms”, edited by Yadolah Dodge, Oxford University Press, 2003.

Context: Population, statistical units and variables relate to specific times, which may be limited to a reference time point (e.g. a specific day) or a reference period (e.g. a month, calendar year or fiscal year) (Eurostat, "Quality Glossary").

Relative standard error — See Coefficient of variation

Release calendar — A general statement on the schedule of release of data.

International Monetary Fund (IMF), "Guide to the Data Dissemination Standards, Module 1: The Special Data Dissemination Standard", Washington, May 1996.

Context: An advance release calendar provides a general statement on the schedule of release of data, which is publicly disseminated so as to provide prior notice of the precise release dates on which a national statistical agency, other national agency, or international organization undertakes to release specified statistical information to the public. Such information may be provided for statistical releases in the coming week, month, quarter or year.

Advance release calendar information is one of the requirements of the IMF's Special Data Dissemination Standards (SDDS). Such information is disseminated on the Internet on the IMF`s Data Standards Bulletin Board (DSBB) or on national websites.

Revisions, data — Data revisions are defined broadly as any change in a value of a statistic released to the public by an official national statistical agency.

IMF, Carson, Carol C., “Revisions Policy for Official Statistics: A Matter of Governance”, presented at the *2nd Session of the Committee for the Coordination of Statistical Activities, Geneva, 8-10 September 2003*, IMF, 2003, Washington DC

Revision policy — A policy or set of policies, aimed at ensuring the transparency of disseminated data whereby preliminary data are compiled that are later revised when more and better source data become available.

International Monetary Fund (IMF)," Quarterly National Accounts Manual", Washington D.C., 2001.

Context: Providing users with documentation regarding the source data used and the way they are adjusted gives compilers with the possibility to incorporate new and more accurate information into estimates, thus improving their accuracy without introducing breaks in the time series.

Data may also be subject to ad hoc revisions as a result of the introduction of new classifications, compilation frameworks and methodologies which result in the compilation of historical data that replaces previously released data. Whether or not such changes constitute an actual "revision" or the compilation of a "new" series is a matter of judgment on the part of the statistical agency.

Under the requirements of the Special Data Dissemination Standard (SDDS), an organisation's revision policy for specific statistics is disseminated on the Internet on the IMF`s Dissemination Standards Bulletin Board (DSBB).

Sampling error — That part of the difference between a population value and an estimate thereof, derived from a random sample, which is due to the fact that only a sample of values is observed; as distinct from errors due to imperfect selection, bias in response or estimation, errors of observation and recording, etc.

The totality of sampling errors in all possible samples of the same size generates the sampling distribution of the statistic which is being used to estimate the parent value.

The International Statistical Institute, “The Oxford Dictionary of Statistical Terms”, edited by Yadolah Dodge, Oxford University Press, 2003.

Context: Sampling errors arise from the fact that not all units of the targeted population are enumerated, but only a sample of them. Therefore, the information collected on the units in the

	sample may not perfectly reflect the information which could have been collected on the whole population. T he difference is the sampling error (Eurostat, Quality Glossary).
Sampling variance	The variance of a sampling distribution. The word "sampling" can usually be omitted, as being defined by the context or otherwise understood. The sampling variance of a statistic is the square of its standard error. A Dictionary of Statistical Terms, 5th edition, prepared for the International Statistical Institute by F.H.C. Marriott. Published for the International Statistical Institute by Longman Scientific and Technical.
Seasonal adjustment *(draft definition)*	Seasonal adjustment is a statistical technique to remove the effects of seasonal and calendar influences operating on a series. Seasonal effects usually reflect the influence of the seasons themselves either directly or through institutional factors or social conventions. *Context:* Other types of calendar variation occur as a result of influences such as the number of days in the calendar period, the accounting or recording practices adopted or the incidence of moving holidays (such as Easter an d Chinese New Year). (A slightly modified version of the definition from: An Analytical Framework for Price Indexes in Australia: Glossary and References, Australian Bureau of Statistics, Canberra, 1997) Task force on da ta and metadata presentation for the OECD Short -term Economic Statistics Expert Group (STESEG), Paris, 2004
Seasonal adjustment programs	Seasonal adjustment is normally done using off-the- shelf programs—most commonly worldwide by one of the programs in the X-11 family. Other programs in common use include the TRAMO-SEATS package developed by Bank of Spain and promoted by Eurostat and the German BV4 program. *Context:* The original X-11 program was developed in the 1960s by the U.S. Bureau of the Census. It has subsequently been updated and improved through the development of X -11-ARIMA by Statistics Canada and X-12-ARIMA by the U.S. Bureau of the Census, which was released in the second half of the 1990s. The core of X-11-ARIMA and X-12-ARIMA is the sam e basic filtering procedure as in the original X-11. Quarterly National Accounts Manual – Concepts, Data Sources and Compilation, IMF, 2001, para. 8.13.
Seasonal component of a time series (draft definition)	The seasonal component is that part of the variations in a time series representing intra -year fluctuations that are more or less stable year after year with respect to timing, direction and magnitude. *Context:* The seasonal component is also referred to as the seasonality of a time series. The seasonal component reflect "normal" variations that recur every year to the same extent, e.g. weather fluctuations that are representative of the season, length of months, Christmas effect. The seasonal component also includes calendar related systematic effects that are not stable in their annual timing or are caused by variations in the calendar from year to year. Task force on data and metadata presentation for the OECD Short -term Economic Statistics Expert Group (STESEG), Paris, 2004
Seasonal variation (seasonal component) *(draft definition)*	The seasonal variation is that part of the variations in a time series representing intra-year fluctuations that are repeated more or less regularly in the same period year after year. OECD, "Some proposals for standard terminology relating to data presentation", paper presented to the OECD Short-term Economic Statistics Expert Group (STESEG), 28-30 June 2004, Paris, 2004
Seasonally adjusted component of a time	The seasonally adjusted component is the result of the extraction of the seasonal component and the calendar effects component from a time series. If neither seasonal nor calendar influences are present

series	in the raw data, the seasonally series is given by the raw data.
(draft definition)	For series with no identifiable seasonal variations but with identifiable calendar variations, the seasonally adjusted series is given by the calendar adjusted series. *Context:* Trading / working day corrections are alternative ways to normalise the time series. Task force on data and metadata presentation for the OECD Short-term Economic Statistics Expert Group (STESEG), Paris, 2004
Seasonally adjusted time series	Seasonally adjusted time series are series that have been adjusted for seasonal variations, including trading-day (working-day) effects and other regular calendar variations if present.
(draft definition)	Task force on data and metadata presentation for the OECD Short-term Economic Statistics Expert Group (STESEG), Paris, 2004
Smoothing	The process of removing fluctuations in an ordered series so that the result shall be "smooth" in the sense that the first level differences are regular and higher order differences small. Although smoothing can be carried out by freehand methods, it is usual to make use of moving averages or the fitting of curves by least squares procedures. In fact, the concept is closely tied to that of trend fitting. A Dictionary of Statistical Terms, 5th edition, prepared for the International Statistical Institute by F.H.C. Marriott. Published for the International Statistical Institute by Longman Scientific and Technical
Spatial data	Spatial data is any data with a direct or indirect reference to a specific location or geographical area. World Meteorological Organisation, 2004, Report of CBS Expert Team on Integrated Data Management, Fourth Meeting, Geneva, 1-3 September 2004
Special Data Dissemination Standard (SDDS)	The Special Data Dissemination Standard (SDDS) was established by the International Monetary Fund (IMF) to guide members that have, or that might seek, access to international capital markets in the provision of their economic and financial data to the public. Subscription to the SDDS was opened in early April 1996. The SDDS identifies four dimensions of data dissemination: a) The data: coverage, periodicity, and timeliness; b) Access by the public; c) Integrity of the disseminated data; and d) Quality of the disseminated data. The SDDS prescribes that subscribing members provide a summary description of methodology for each data category, including statements of major differences from international guidelines. The term "methodology" is used in the SDDS in a broad sense to cover the aspects of analytical framework, concepts, definitions, classifications, accounting conventions, sources of data, and compilation practices. International Monetary Fund (IMF), "Guide to the Data Dissemination Standards, Module 1: The Special Data Dissemination Standard", Washington, May 1996.
Standard error	The positive square root of the variance of the sampling distribution of a statistic. It includes the precision with which the statistics estimates the relevant parameter as contrasted with the standard deviation that describes the variability of primary observations. The International Statistical Institute, "The Oxford Dictionary of Statistical Terms", edited by Yadolah Dodge, Oxford University Press, 2003.

Stock series / data	Statistical data presented as stock series/data are measured at the end of the reference period, for example, money supply data which can refer to an observation on the last working day of the reference period. Organisation for Economic Co-operation and Development (OECD), "Main Economic Indicators", monthly.
Tendency	The term tendency is used in business tendency surveys where the respondent is asked for a judgment on the direction of changes (e.g. up/down/same). Organisation for Economic Co-operation and Development (OECD), "Main Economic Indicators", monthly.
Time series *(draft definition)*	A time series is a set of time-ordered observations on a quantitative characteristic of an individual or collective phenomenon taken at successive, in most cases equidistant, points of time. *Context:* A time series (TS) can be decomposed into unobservable components. In the most complete case, these components are the trend (T), the cyclical (C), the seasonal (S) and the irregular (I) components. The four components of the time series may each be independent of all the others, in which case the behaviour of the time series is simply the sum of `the components which are additively related (i.e. TS = T+C+S+I). However, most analysts believe that it is unlikely that the time series components are perfectly independent of each other, and are therefore more likely to be multiplicatively related (i.e. TS=T*C*S*I). Task force on data and metadata presentation for the OECD Short-term Economic Statistics Expert Group (STESEG), Paris, 2004.
Time series breaks	Breaks in statistical time series occur when there is a change in the standards for defining and observing a variable over time. Such changes may the result of a single change or the combination of multiple changes at any one point in time of observation of the variable. The specific causes of breaks in a statistical time series include changes in: classifications used, definitions of the variable, coverage, etc. *Context:* Statistical agencies and users of time series data for economic research to analyse and interpret economic and social events and conditions attach very high importance to the continuity and consistency of data over time. However, it should be emphasised that the occurrence of time series break may not necessarily jeopardise the reliability of a time series. Statistical agencies frequently apply a number of techniques to ensure the continuity of a time series. Finally, the impact of a time series break is often a matter of judgement on the part of the user and depends on the use(s) to which the data are put. Statistical Data and Metadata Exchange (SDMX) – BIS, ECB, Eurostat, IBRD, IMF and OECD – Metadata Common Vocabulary, Release 1, December 2003.
Trend component of a time series *(draft definition)*	The trend is the component of a time series that represents fluctuations of low frequency in a time series, the high frequency fluctuations having been filtered out. This component can be viewed as those variations with a period longer than a chosen threshold (usually 1½ years, the average length of the business cycle). *Context:* The trend is normally referred to as the long-term trend in a cyclical context (i.e. the trend is longer than the average duration of the business cycle). In practice, statistical agencies estimate trend by estimating and removing the seasonal and irregular components. This will normally give a time series containing part of the cyclical component and is referred to as the trend-cycle component. Task force on data and metadata presentation for the OECD Short-term Economic Statistics Expert Group (STESEG), Paris, 2004

Trend-cycle (draft definition)	The trend-cycle is the component that represents fluctuations of low frequency in a time series, the high frequency fluctuations having been filtered out. This component can be viewed as those variations with a period longer than a chosen threshold (usually less than 1½ years). *Context:* In practice, statistical agencies estimate trend-cycle by estimating and removing the seasonal and irregular components. Task force on data and metadata presentation for the OECD Short-term Economic Statistics Expert Group (STESEG), Paris, 2004
Trend estimates	Trend estimates are derived from seasonally adjusted estimates via an averaging process which attempts to remove the irregular component of the time series. This allows the underlying direction of a time series to be identified. An Analytical Framework for Price Indexes in Australia: Glossary and References, Australian Bureau of Statistics, Canberra, 1997.
Trend fitting	The general process of representing the trend component of a time series. A trend may be represented by a particular curve form, e.g. the logistic, or by a particular form of the general class of polynomial in time, or by a moving average. A Dictionary of Statistical Terms, 5th edition, prepared for the International Statistical Institute by F.H.C. Marriott. Published for the International Statistical Institute by Longman Scientific and Technical.
Volume index	A volume index is most commonly presented as a weighted average of the proportionate changes in the quantities of a specified set of goods or services between two periods of time; volume indices may also compare the relative levels of activity in different countries (e.g. those calculated using PPPs). A major aim of economic analysis is to develop an understanding of changes taking place in an economy over time. This includes the measurement of short-term growth or decline. To achieve this for key economic value aggregates, such as the value of industrial production or the value of retail turnover, it is necessary to distinguish between changes arising solely from price changes and those arising from other influences such as quantity and quality, which are referred to as changes in "volume". SNA 1993 para. 16.11.
Working / trading day adjustment *(draft definition)*	Working day or trading adjustments refer to the correction for differences in the number and/of sort/type of working or trading days in a given month or quarter which differ from year to year which will impact upon the level of activity in that month. In most countries working day adjustment and trading day adjustment are used as synonyms. *Context:* The number of working or trading days in a given month or quarter can vary significantly for each statistical domain (e.g. production, merchandise trade) because of differing institutional arrangements, trade specific holidays, etc. Some countries also include bridging effects in working day adjustments. These result from people taking holidays, for example, on Mondays and Fridays when an official public holiday occurs on Tuesdays and Thursdays respectively. The type of working / trading day adjustment carried out needs to be tailored to the cultural and institutional environment operating within individual countries. In the United States, for example, working day adjustment classifies the days of the week into workdays (Monday through to Friday) and non-workdays (Sat. and Sun.) and thus the seasonal adjustment estimates two factors – workday and non-workday. Trading day adjustment (as performed in the US at least) allows for a different effect for each day of the week and computes seven factors – a Monday factor, a Tuesday factor and so on. Essentially, trading day adjustment is a more fine tuned seasonal adjustment method to account for calendar variation. In other countries there may be very distinct differences between workdays and non-workdays in the

amount of business conducted, but not so much difference between Mondays, Tuesdays, through the Fridays. The US on the other hand, has distinctly different patterns of activity each day of the week and methods used in that country try to account for that.

For trading day adjustment the Demetra User Manual states that if there are differences in economic activity between all days of the week, six regression variables are introduced to account for these differences. Similarly for working day adjustment if there are no differences in economic activity between the working days (Monday to Friday), but between these and non-working days (Saturday, Sunday), one regression variable is introduced to account for that difference.

Task force on data and metadata presentation for the OECD Short-term Economic Statistics Expert Group (STESEG), Paris, 2004

Year-on-year change

See Year-on-year growth rates

Year-on-year (YoY) growth rates

(draft definition)

Year-on-year growth rates are rates of change expressed over the corresponding period (month or quarter in relation to the frequency of the data) of the previous year. Such rates are expressed as (M_t/M_{t-12}) -1 or (Q_t/Q_{t-4}) -1.

M_t denotes the value of a monthly time series in month t and Q_t the value of a quarterly time series in quarter t.

Context: Also often referred to as Year-over-year growth rates, Year-to-year growth rate, Rate of change from the previous year, or 12-month rate of change.

Task force on data and metadata presentation for the OECD Short-term Economic Statistics Expert Group (STESEG), Paris, 2004

Year-on-year changes

(draft definition)

Year-on-year changes are changes in levels expressed over the corresponding period (month or quarter in relation to the frequency of the data) of the previous year. Such changes are expressed as M_t-M_{t-12} or Q_t-Q_{t-4}.

M_t denotes the value of a monthly time series in month t and Q_t the value of a quarterly time series in quarter t.

Context: Also often referred to as Year-over-year changes, or Year-to-year changes.

Task force on data and metadata presentation for the OECD Short-term Economic Statistics Expert Group (STESEG), Paris, 2004

Year-to-date data

Data expressed in cumulative terms from the beginning of the year; sometimes referred to as cumulative data.

Statistical Data and Metadata Exchange (SDMX) – BIS, ECB, Eurostat, IBRD, IMF and OECD – Metadata Common Vocabulary, Release 1, December 2003