STATISTICAL COMMISSION and
ECONOMIC COMMISSION FOR EUROPE

COMMISSION OF THE
EUROPEAN COMMUNITIES

CONFERENCE OF EUROPEAN STATISTICIANS

EUROSTAT

**Joint UNECE/Eurostat Work Session
on Statistical Metadata**
(6 - 8 March 2002, Luxembourg)

Working Paper No. 12
English only

Topic (i): Infrastructure issues for statistical metadata

DEVELOPING A COMMON UNDERSTANDING
OF STANDARD METADATA COMPONENTS:
A STATISTICAL GLOSSARY

Submitted by OECD-Eurostat[1]

Contributed paper

ABSTRACT

The aims of this paper are to clarify some of the issues surrounding the definition and the use of the term "metadata", and describe how the development of a statistical glossary containing definitions of metadata elements describing all aspects of the statistical production cycle, and the use of those definitions by metadata authors in different organisations, could further the cause of data and metadata exchange.

Given the diverse requirements of different national and international organisations, reaching agreement on a common model for metadata collection and dissemination has proven to be an almost impossible task. Agreement on a common highly aggregated model might even hide a serious misunderstanding of the basic metadata components, if a clear definition is not provided. It is probably more realistic to develop more open standards that everyone can implement according to their own requirements. Agreement between organisations on a common set of metadata elements and definitions of those elements covering the entire statistical production cycle would provide the flexibility for each organisation to manipulate these elements to derive a variety of specific dissemination outputs.

One of the objectives of the SDMX Group is ".... to focus on more efficient processes for exchange and sharing of data and metadata within the current scope of our collective activities." This entails the development of a set of metadata content standard definitions that will enable the participating institutions to take maximum advantage of exchange protocols such as GESMES/CB and e-standards such as XML. The focus of co-operation between international organisations in this area is therefore not the development of common metadata templates, prompt points, content models, etc - perhaps impossible - but instead on the identification of discrete and unambiguous metadata elements (and a related glossary) that provide sufficient information about what is required to enable the various authors of metadata in national agencies and international organisations to provide consistent metadata. Such work is essential if metadata provided by different organisations is to be used for comparing national statistical compilation practices.

A generic glossary of standard metadata elements should take into account the need of interoperability between different national, international or supranational statistical institutes. The focus of the glossary

---

[1] Prepared by Denis Ward (OECD) and Marco Pellegrino (Eurostat)

being developed by the authors is on statistical content and on the semantics of statistical metadata elements. The glossary being developed by the authors in conjunction with the SDMX project is a subset of a more comprehensive glossary being developed jointly by OECD and Eurostat that contains a large number of definitions relating to statistical terminology and concepts as well as definitions of variables commonly used in economic and social statistics.

Both the SDMX glossary and the larger, more comprehensive glossary being developed by the OECD and Eurostat draw heavily on the content of other glossaries that use definitions from existing international statistical standards. In fact, the authors emphasise the need to utilise existing standards rather than developing a new set of definitions. Examples of such existing glossaries include: United Nations "Glossary of Classification Terms"; UNECE "Terminology on Statistical Metadata" and Eurostat's CODED Glossary.

I.   INTRODUCTION

1.   The aims of this paper are to clarify some of the issues surrounding the definition and the use of the term "metadata", and describe how the development of a statistical glossary containing definitions of a comprehensive set of metadata elements, and the preparation of statistical metadata by authors in different national agencies and international organisations, based on a common set of definitions, could further the cause of data and metadata exchange, in particular the exchange of consistent and comparable metadata.

2.   There are many different types of metadata covering the entire statistical production cycle, from data collection through to data dissemination. Unfortunately, there is no universally accepted definition of either metadata or metadata models[2] that outline the different types of metadata. Different national agencies (e.g. the US Bureau of Census, the Australian Bureau Statistics) and international organisations (Eurostat, IMF, OECD) have developed their own data and metadata models. Although there is some commonality between these models, each has a slightly different objective and each describes different metadata elements and stages of the statistical production cycle.

3.   The consequences and drawbacks affecting information systems caused by the absence of universally adopted international guidelines and recommendations relating to metadata content[3] are summarised in the International Standard ISO/IEC 11179 (a multi-part standard concerning data element specification and standardisation, prepared by the Subcommittee ISO/IEC JTC1/SC32)[4]. These consequences include:

??   a lack of mechanisms for enabling global data acquisition and interchange, particularly across application areas;

??   the absence of unique global identifiers for standard data elements ;

??   inadequacy of documentation of data element characteristics to support fully automated sharing of data, including locating, retrieving, and exchanging the data;

??   lack of uniform guidance for identification, development, and description of data elements;

??   the impossibility or difficulty of finding and retrieving a specific standard data element among thousands or millions;

??   no universal means for organizing standard data elements exists;

---

[2] The notion of a "metadata model" is also sometimes referred to as a metadata template, metadata prompt points, etc.

[3] The US Bureau of the Census defines a metadata content standard as "a standard that specifies a set of metadata items and does not specify the physical format of the content, the services to be provided, or the syntax used. (*Survey Design and Statistical Methodology Metadata*, Software Standards Management Branch, Systems Support Division, United States Bureau of the Census, Washington D.C., August 1998, Section 1.1, page 1)

[4] ISO/IEC International Standard IS 11179-3, Information Technology - Specification and standardization of data elements - Part 3: Basic attributes of data elements, 1994.

?? though data is sometimes standardized within an organization, there are few common data standards between organizations;

?? exchange of data among organizations results in a proliferation of customized data interchange representations;

?? data definitions and descriptions not being sufficiently precise to support reuse or multiple users of data;

?? inadequacy of current inventory structures for reducing logical data redundancies ;

?? global implementation of Electronic Data Interchange (EDI) is impeded by a lack of standard data elements; standard data elements are needed for the content of EDI messages.

4.   Given the diverse requirements of different national agencies and international organisations, reaching agreement on a common model for metadata collection and dissemination has proven to be an almost impossible task. Effective agreement is relatively harder to achieve if the exercise is only conducted in terms of broad metadata items, especially where the item used (e.g. "scope of the data") in turn also comprises more detailed metadata elements (e.g. "geographical coverage", "activity coverage", "unit coverage", etc). In this case, agreement on a common highly aggregated model might even hide a serious misunderstanding of the basic metadata components, if a clear definition is not provided.

5.   Given the problems of achieving universal acceptance and implementation of elaborate metadata content models, it is probably more realistic to develop more open standards that everyone can implement according to their own requirements. The central theme of this paper is that it is more appropriate for metadata authors in different organisations to adopt a common set of metadata "elements" that can be used to describe the collection, processing and dissemination of statistical data. Agreement between organisations on a common set of detailed metadata elements would provide the flexibility for each organisation to manipulate these elements to derive a variety of specific standalone metadata dissemination outputs or frameworks (such as the SDDS – refer Section (d) below) according to its own needs. An agreed-to list of basic metadata elements would provide the common underlying structure. Depending on the need of each organisation responsible for producing the information, a part of the metadata could then be organised in a more "active" way, for instance, in a proper textbase associated with the database itself.

6.   The enhancement of data and metadata exchange therefore requires development on two fronts: a) on "content" issues, relating to a common understanding between different organisations of the meaning of "chunks" of textual and other information that describes data (i.e. definitions, how it is compiled, processed and accessed), and b) on IT related issues.

7.   As mentioned above, one of the problems in this field is the multitude of definitions currently available for the term "metadata" and related concepts. Attachment 1 provides examples of some that are currently in use. There are no doubt others. Clarification of these concepts requires an understanding of a number of related issues, these being:
?? a broad definition of the term "metadata";
?? a clear understanding of the "types of metadata" for different metadata systems;
?? the concept of a "metadata dissemination typology"(i.e. the format for metadata dissemination used by each organisation); and
?? the identification (and definition) of standard metadata elements.

II.   DEFINITION OF METADATA

8.   Numerous definitions of metadata have been presented in publications/papers prepared by the UN, UNECE, Eurostat, OECD, NSOs, etc. Whilst the precise wording of each definition differ, by and large, they tend to agree on the central aspect of metadata being "information about data". From there differences between the various definitions become more acute.

9.   In the context of a broad definition of metadata it *might* be useful to also include two additional aspects:

?? The fact that metadata comprises a range of different subjects – definitional information about the content of data; the statistical data production cycle processes involving the collection, processing and dissemination of data; information on data quality and measures of performance; and systems related information that describe the IT context, how data can be accessed, etc. These are what could, for want of simplicity, be referred to simply as "types of metadata"; and

?? the dimension of how metadata may be used – active/passive metadata. This dimension is included in some of the existing definitions, and in papers presented at various work sessions such as METIS. A commonly used distinction is that "active" metadata is physically integrated or linked with the information system containing the data, or it can be used to determine the actions of automated processes, whereas "passive" (or descriptive, or stand-alone) metadata is any prepared documentation.

10.   A broad definition of metadata might then be:

> Metadata is information which is needed for the production and usage of statistical data. It provides information on definitional content; on the processes for the collection, processing, storage and dissemination of data; measures of quality; and information of IT and related issues.

11.   One could also add that metadata can be actively linked to data and to IT systems, or simply prepared as stand-alone descriptive files, although that consideration concerns the use rather than the definition of metadata.

III.   METADATA TYPES AND SYSTEMS

12.   Agreement on a broad definition of metadata is only the first step. What is also required, especially in the context of the objectives of the SDMX project, is a more common understanding of the different types of metadata needed by different metadata systems.

13.   A metadata system is one that uses and produces different metadata types by undertaking the following processes:
?? metadata collection;
?? metadata processing;
?? metadata storage; and
?? metadata dissemination (footnotes, table headings, explanatory notes, etc.).

14.   Each phase of the statistical life-cycle may have different requirements. For example, metadata for supporting collection and processing addresses data producers' needs and will be much more detailed than dissemination metadata, the main goals of which are: a) producing summary information for external users, and; b) delivering standard metadata to support analyses of comparability with statistics produced by other institutes. Only a part of the available metadata would normally be disseminated at this stage, though there will also be instances where data users also require access to detailed metadata.

15.   A number of metadata type schemas have been developed in recent years by several national statistical institutes and international organisations based on their own business needs and requirements. Examples include: US Bureau of Census' SDMS (Survey Design and Statistical Methodology); Statistics Canada's IMDB (Integrated MetaData Base); Australian Bureau of Statistics' Data Management Project; OECD's *Main Economic Indicators*; Eurostat's Metadata Dissemination Typology; and the IMF's Special Data Dissemination Standard. More information on the precise meaning of some of these metadata type schemas is also provided in Attachment 1. Much of the effort in some of these schemas appears to have focussed on IT issues rather than to an explanation of the data.

16.   Each schema contains unique features, though there are also a number of common features, especially when describing the more detailed metadata elements. It is not the intention of the authors of this paper to present yet another schema, or even to identify a general common metadata model. The result could be too hard to manage for any producer and too complex for most users. Perhaps complexity

is a reason why elaborate and sophisticated models have had difficulty being implemented, or once implemented, being maintained. As mentioned above (in para. 5), the focus of this paper is the identification of a set of common metadata elements in the form of a glossary. This is explored below in more detail in Section (e).

17.   There does seem however clear benefit in drawing a distinction between metadata that describes data content (definitional) from that which describes processes for producing data (procedural), and metadata measuring performance and quality. Also, for definitional metadata, a distinction is often made between "administrative" metadata and information more connected with the explanation of the data (definitions and methodology).

IV.   METADATA DISSEMINATION TYPOLOGY

18.   Having defined metadata and discussed metadata types and metadata systems (though avoiding the formulation of yet another metadata type schema), it is now necessary to define the scope of our work in more detail, through the notion of the "metadata dissemination typology". This simply describes a format for the dissemination of metadata by organisations. The term "typology" is used synonymously in this context with the term "format". The SDDS is perhaps the most well known example of such a metadata dissemination typology/format, although there are many such formats around – look at any NSO website or metadata concept publication. The range of metadata elements in each typology may be comprehensive or narrow depending on the needs and objectives of the agency responsible for its preparation and dissemination. The main distinction between a metadata dissemination format and a metadata model described in earlier sections of this paper is that the latter also attempts to define precisely each metadata element. The distinction can be further illustrated through consideration of the SDDS.

19.   The SDDS does what it is designed to do quite well, though it does not attempt to be a fully comprehensive framework covering the whole range of metadata types and metadata elements, nor does it attempt to cover all statistical subjects[5]. The SDDS guidelines for each methodological element (accounting conventions, nature of basic data, compilation practices, etc) of each statistical subject are very useful, however, they are mostly aimed at helping *national* providers. They are of more limited use for describing supranational aggregations (such as those provided by Eurostat) or when making methodological comparisons between countries, primarily because different countries have provided different interpretations of what is required in the text describing each methodological element within the SDDS.

20.   For these reasons, the SDDS dissemination format is only a partial solution for the metadata needs of the OECD, Eurostat and NSOs, whose range of statistical subject responsibilities is invariably broader. The needs of these organisations would be met through the development and adoption of a comprehensive metadata glossary containing definitions of a number of metadata elements, each being defined unambiguously. The development of such an open standard and its adoption by metadata authors in different organisations would then permit organisations to incorporate these elements into their own specific metadata dissemination typology/format without loss of the ability to compare the content of common metadata elements between countries.

V.   IDENTIFICATION OF STANDARD METADATA ELEMENTS: A STATISTICAL GLOSSARY

21.   One of the objectives of the Statistical Data and Metadata eXchange (SDMX) Group, as outlined in the Common Statement by Participating Institutions, is "…. to focus on more efficient processes for exchange and sharing of data and metadata within the current scope of our collective activities." This entails the development of a set of metadata content standard definitions that will enable the participating institutions to take maximum advantage of exchange protocols such as GESMES/CB and e-standards such as XML. Essentially, this entails ensuring that the "chunks" of information about data provided by country A describe the same attribute of the data as information provided by country B for the same

---

[5] Statistical subjects include topics such as national accounts, balance of payments, prices, unemployment, etc., that are included in the scope of the SDDS, and other topics such as retail, that are not.

concept, definition, process, quality standard, etc. This can only be achieved through participating organisations reaching agreement on how those "chunks" of information, or metadata elements about data are defined. Such work is essential if metadata provided by different national and international organisations is to be used for comparing national statistical compilation practices.

22.   An issue here is just how broad or narrow these metadata elements should be. It would probably be quicker to reach agreement on broad terms (e.g. coverage, timeliness, units, etc), though this would not advance the cause of consistency between organisations very far, especially if they were used in defining XML tags. It would be easier, though not essential, to define narrower, more discrete metadata elements in the context of some form of hierarchy comprising the different metadata types covering the entire statistical production cycle, for example, the Eurostat, OECD or ABS schemas presented in Attachment 1. A very *embryonic* example of how such a hierarchy and the component metadata elements might look is presented in Attachment 2.

23.   The bottom detailed level of the hierarchy would comprise what Eurostat refers to as "atomic" items or metadata elements, i.e. those that cannot be defined more unambiguously. The question of the level of detail of such a framework still needs further thought. However, given the range of statistical subjects/topics and the number of distinct elements/aspects of the statistical production cycle, there is need for such a hierarchy to be comprehensive and linked to a glossary containing an actual definition of each metadata element. Ideally, these definitions would be drawn from existing international statistical standards and be accompanied by detailed reference/source details.

24.   The focus of co-operation between international organisations in this area is therefore <u>not</u> the development of common metadata templates, prompt points, content models, etc - perhaps impossible - but instead on the identification of discrete and unambiguous metadata elements (and a related glossary) that provide sufficient information about what is required to enable the various authors of metadata to provide consistent metadata. The elements identified should be comprehensive and capable of being applied to all fields (subjects) of statistics and types of metadata for each of these fields. Once developed, these elements could then be used by different organisations in whatever metadata dissemination framework or model they choose.

25.   Such a generic glossary of standard metadata elements should take into account the need of interoperability between different national, international or supranational statistical institutes. Some of those elements are already considered within the existing models for dissemination (SDDS) or for the exchange of information (such as GESMES). The experience of GESMES/CB has shown that a list of information elements must be unambiguous and generic enough to be implemented without having to change the whole information system accordingly.

26.   A complication in the development of such a glossary is the existence of many other glossaries at the international and national levels. Examples include: United Nations "Glossary of Classification Terms"; UNECE "Terminology on Statistical Metadata" and Eurostat's CODED Glossary. The content of these glossaries overlap as many of the definitions they contain have been drawn from the same international statistical recommendations and guidelines. Furthermore, there are numerous instances of definitional inconsistency between the various glossaries. Also, some glossaries focus on particular subjects/issues: UNECE terminology (metadata), SDMX (standard statistical elements for exchange), subject-matter glossaries (CODED).

27.   The OECD and Eurostat are currently developing a joint glossary that brings together in the one readily accessible location (i.e. on the internet) the contents of existing glossaries. The new glossary will provide a comprehensive set of definitions of variables commonly used in economic and social statistics, and statistical concepts and terminology derived from existing international statistical standards development by the United Nations, ILO, IMF, OECD and Eurostat. This will provide a resource for use by people in national and international organisations seeking definitions for inclusion in questionnaires and publications.

28.   The focus of the glossary being developed by the authors in the context of the SDMX project, which is presented in Attachment 3, is on statistical content and on the semantics of statistical metadata elements. The SDMX glossary contains a subset of the terms included in the more comprehensive glossary being developed jointly by OECD and Eurostat referred to in the previous paragraph. The

SDMX glossary has also drawn heavily on the content of other existing glossaries, as well as on the work undertaken by ISO/IEC JTC1 workgroup for the specification and management of data elements in the framework of ISO/IEC 11179.

29.   The final issue concerns the nature of the processes for developing international statistical standards. Over the last two or three decades an extensive range of statistical guidelines and recommendations have been prepared by international organisations working with national statistical institutes and other agencies responsible for their compilation and dissemination. The aims of the standards are to promote the development of internationally comparable statistics and the use of practice in the compilation of statistics.   A list of current international guidelines and recommendations is maintained by UNSD on their website *Methodological Publications in Statistics* (http://esa.un.org/unsd/progwork).[6] These include standards relating to metadata and IT.

30.   The development of international standards on metadata and IT has also been carried out through the activities of the International Standards Organisation (ISO). There appears to be considerable overlap in the development of standards in these areas by international organisations and the ISO processes and therefore scope for more co-ordinated activity by these two processes for development of international standards in data and metadata exchange.

---

[6] Another list of existing international classifications is provided by Eurostat's Classifications Server, RAMON, at http://europa.eu.int/comm/eurostat/ramon

Attachment 1

CURRENT DEFINITIONS OF METADATA

a. *Terminology on Statistical Metadata*, Conference of European Statisticians Statistical Standards and Studies, No. 53, UNECE, Geneva 2000)

Metadata	Metadata provides information on data – and about processes of producing and using data.
Metadata	Metadata comprises data and other documentation that describes objects in a formalised way.
Metadata layer	A metadata layer is a layer in the reference model for standardisation in statistics used to denote the set of attributes related to statistical metainformation.
Statistical metadata	Statistical metadata is metadata describing statistical data.
Statistical metadata system	A statistical metadata system is a data processing system that uses, stores and produces statistical metadata.

b. *Guidelines for the Modelling of Statistical Data and Metadata*, UNECE and UNSC, Conference of European Statisticians Methodological Material, 1995

Statistical metadata	Statistical metadata are data which are needed for proper production and usage of statistical data. They describe statistical data and – to some extent – processes and tools involved in the production and usage of statistical data. Expressed briefly, statistical metadata are data about statistical data.
Statistical metainformation system	A statistical metainformation system is a system which uses and produces statistical metadata, informing about statistical data, and which fulfils its tasks by means of functions like "statistical metadata collection", "statistical metadata processing", "statistical metadata storage", and "statistical metadata dissemination".
Active or passive metainformation system	A metainformation system may be *active* or *passive*. An active metainformation system is physically integrated with the information system containing the data that the metadata in the metainformation system informs about. A passive metainformation system contains only references to data, not the data themselves.

c. Eurostat Task Force on Metadata, December 2000-May 2001. Identified three broad groups of metadata for use within Eurostat's Reference Environment.

Administrative metadata	Administrative items are important for accessing, using or reusing the data or for conducting further analyses on metadata. Broad items: contacts, access conditions, copyright, release policy, dissemination formats, prices, last update, data sources, references (to publications and databases), symbols and abbreviations, on-line help.
Basic descriptive metadata	This group allows the identification of the statistical observation. It consists of the list of dimensions (descriptive attributes) associated with data dictionaries: - List of dimensions used to identify and describe the data syntax and structure. This covers definitions of variables, breakdown dimensions, time and geographical coverage, measures of unit or statistical elaboration, etc. - Contents of dimensions (dictionary files).
Semantic metadata	This includes explanatory notes of various types, in the form of stand-alone text files, or footnotes/flags attached to the data. This is a typically human-oriented group of metadata, which explains the meaning (semantics) of data. Broad items: introductory texts, table of contents, definitions, classification used, periodicity, scope of the data, nature of basic data, data production process, compilation practices, quality assessment (relevance, accuracy, timeliness, accessibility and clarity, comparability, coherence, completeness).

d. Australian Bureau of Statistics, Statistical Integration Through Metadata Management, M. Colledge,

Metadata	Metadata simply means data about data., and refers to the definitions, descriptions of procedures, system parameters, and operational results which characterise and summarise statistical programmes. Metadata may be passive (descriptive), i.e. the form of documentation which is used by agency staff, or may be active (prescriptive), i.e. determining the actions of automated surveys processes.
Definitional metadata	Definitional metadata relate to statistical units and populations, topics, classifications, data item definitions, questions and question modules, and statistical terms.
Procedural metadata	Procedural metadata relate to the procedures by which data are collected and processed.
Operational metadata	Operational metadata arise from and summarise the results of implementing the procedures. They include measures of respondent burden, response rates, edit failure rates, costs, and other quality and performance indicators.
Systems metadata	Systems metadata are active metadata used to drive automated operations, including, for example, file layouts and access paths.
Dataset metadata	Dataset metadata comprise the minimal systems metadata used to describe, access and update datasets.

e. OECD Main Economic Indicators Metadata – List of Metadata Items to Describe the Compilation of Statistics – available at http://www.oecd.org/std/mastnew.htm

Source	source agencies; main national publication sources; breakdown available; methodological references; length of time series; periodicity; unit of measurement
Concepts and coverage	definition; coverage (reference period; geographical coverage; classification coverage; sectoral description; statistical population; particular exclusions)
Standards	standard systems/framework; standard classifications/nomenclatures; international comparability; departures from international standards
Data collection	reporting units; reporting method; survey items (description of questionnaire; master list; survey description; non-response rate); administrative sources (description; update procedures; quality assessment; access and control, for statistical usage; statistical techniques for data processing); multiple sources items (items for use of multiple sources; standard combination framework; standard techniques for combining multiple sources); reporting date in relation to the event measured; time lapse between event and processing
Data manipulation	aggregations/grossing up (aggregation method; grossing up method; weights for aggregation); seasonal and other adjustments; other manipulations
Data quality and timeliness	sampling errors and their corrections; other errors and their corrections; missing data in time series; breaks in time series; preliminary estimates; revision policy; corroborating evidence; timeliness; release dates
Data transmission	transmission medium; transmission identification (including format; transmission date; transmission contact person
Data storage and manipulation by the OECD	database storage (type of data; country; data periodicity; subject/topic; series label; series identifier; unit of measurement); database management (person responsible for updating; access rules); data manipulation (links with other data; seasonal adjustment; rebasing; aggregation; harmonisation); system metadata (physical location of data; data dictionary); metadata items to be derived from the system (storage format; date of last update; links and formulae)
Output preparation and delivery by the OECD	Publication description (title; number of significant digits and decimals; list of metadata items shown in publication); data enquiries (statistical expert for information on dataset; list of publications that cover this dataset)

Attachment 2

<div align="center">

CONTENT METADATA PRESENTATION

</div>

Level 1 – Broad metadata types

1. Administrative/sources

2. Concepts and Coverage/Definitional/Semantic

3. Standards

4. Data collection, manipulation/accounting conventions, etc

5 Quality and performance metadata

1. Administrative/sources

Source agencies	Primary sources (data, metadata)
	Secondary sources (data, metadata)
Reference documents	Publications (hardcopy, electronic)
	Databases (internal, external)
	Internet (internet addresses)
	Methodological references – title of source (summary metadata, detailed metadata, information on major changes) – See also Part 5 (Accessibility and clarity)
Contact	Role/function of contact (e.g. data enquiries, methodological information/assistance, etc)
	Person name
	Position title
	Organisation title
	Organisation address
	Organisation postal address
	Telephone number
	Fax number
	E-mail address
Release information	Periodicity (frequency of data in published source)
	Series available (table of contents)
	Breakdowns available for indicator(s)
	Length of time series in available data
	Missing data in time series
	Description of breaks in time series
	Timeliness (transmission delay, publication delay)
	Prices
	Access conditions (copyright, subscription policy, pricing policy, access by public, internal access before release)
	Revision policy
	Available dissemination media (e.g. paper publication, CD-ROM, diskette, internet) and formats – see also Part 5 (Accessibility and clarity)
	Languages available
	Units of measurement in which data are published)
	Release calendar (listing future release dates)
	Date of last data update
	Date of last metadata update
	Official comments (press release, institutional commentaries)
Institutional framework	Legal framework
Symbols	List of flags (codes added to data)
	List of special values (codes replacing the data)

2. Concepts and Coverage/Definitional/Semantic

Definition	Definition of the indicator
Coverage	Reference period – period to which the indicator relates Geographical coverage Classification coverage, i.e. items in a national/international classification, e.g. activity, occupation, sector. institution Statistical population – reference population covered by the statistics, e.g. units Specific exclusions
Units	Unit definitions Departures from international units definitions

3. Standards

Standard systems/framework	Recognised international/national standards for compiling the data, e.g. SNA
Classification	Classifications used – national/international, name, version number
Departures from international standards	Specific departures from international standards – see also Part 5 (Comparability)

4. Data collection, manipulation/accounting conventions, etc

Data source	Administrative source Household survey Enterprise/establishment survey
Reporting units	Units which report information, e.g. enterprise, establishment, household, individuals
Valuation principals	Time of recording (e.g. price observations) Types of prices – e.g. basic, producers, sellers Valuation principles
Collection	Frequency of collection (e.g. weekly, monthly, quarterly, annually, ad hoc) Collection method - Self enumeration (postal/mail, drop-off mailback/drop-off pickup, electronic form), personal interview (face to face, CAPI), telephone (telephone interview, CATI) Type of enumeration – complete enumeration (census), sample Communication with data providers Editing and correction of data at collection time

	Non-response rate Non-response action
Frames	Type – area, list (business register) Source of information – births, deaths, amendment/changes to existing entries Frame maintenance procedures Frame maintenance issues – missing units, out of scope units, duplications, deaths, nils Dimensions
Sampling	Sample frame Sampling unit Sample size Sample type – non-probability sample (quota sampling, convenience or haphazard sampling, judgment or purposive sampling), probability sample Stratification Sample selection methodologies Weights used Frequency of weight update
Administrative source (issues)	Description – name, primary purpose, legal status, agency responsible for management, list of administrative information available Update procedures Quality assessment – existence of quality and coherence checks by source agency, adequacy of administrative source for statistical use (of definitions, concepts, coverage), quality of coding Access and control for statistical usage – access conditions by source agency, timeliness
Survey development	User groups Identification of user needs Testing/pilot surveys
Processing (data)	Data entry – EDI, scanning, OCR Coding Editing systems – input, output Types of edits - validation, missing data, logical, consistency (reconciliation), range Imputation – non-response (item, complete) Disclosure control - direct disclosure, inadvertent disclosure, residual disclosure, disclosure from external information Processing system Site of processing Reconciliation
Manipulation	Aggregation method – how data are aggregated from elementary cell levels Grossing up – how estimates of the population are made from the sample (number-raised estimation, ratio estimation) Weights used for aggregation Seasonal adjustment – system used (e.g. TRAMO/SEATS, X-11, X-12 ARIMA) Derived items

	Estimation of sampling error
	Index compilation – item selection, base year, type of weights, period of current weights, frequency of weight revision, aggregation structure, compilation formula
	Adjustments for quality differences
Presentation of data	Levels
	Indices
	Quarterly (monthly) data expressed at annual (or quarterly) levels
	Period average
	End-of-period
	Annual growth rates
	Growth over previous period
	Same period previous year
	Annualised growth rates
	Year to date data

5. Quality and performance metadata

Relevance	Summary description of:
	- Users
	- User's needs
	- Relevance of statistical concepts (statistical measure, target population, reference period)
Accuracy	Sampling error (standard error, variance, relative standard error, confidence interval)
	Non-sampling error (coverage errors, measurement errors, processing errors, non-response errors, model assumption errors)
Timeliness and punctuality	With respect to reference period
	With respect to date of publication by member country
	With respect to date of publication in the database
	With respect to date of last update of domain in database
	With respect to date on which user consults the database
Accessibility and clarity	Information on the dissemination process
	Accessibility of information on data (metadata)
Comparability	Comparability over time
	Geographical comparability (international comparability, departures from international statistical standards)
	Comparability between domains
Coherence	Coherence between provisional and final statistics
	Coherence of annual and short-term statistics

	Coherence of statistics in the same socio-economic domain Comparison of statistics with national accounts Consistency over time Spatial consistency Consistency with related series Continuity Additivity
Completeness	Summary description

Attachment 3

SDMX[*] GLOSSARY OF STATISTICAL TERMS

(DRAFT)

Date of last update: 21 September 2001

---

[*] SDMX = Statistical Data and Metadata eXchange.

Accessibility	*Accessibility* reflects the availability of information from the holdings of the agency, also taking into account the suitability of the form in which the information is available, the media of dissemination, the availability of metadata, and whether the user has reasonable opportunity to know it is available and how to access it. The affordability of that information to users in relation to its value to them is also an aspect of this characteristic. (*Statistics Canada Quality Guidelines*, 3rd edition, October 1998, page 5)
Account	An *account* is a tool which records, for a given aspect of economic life, (a) the uses and resources or (b) the changes in assets and the changes in liabilities and/or (c) the stock of assets and liabilities existing at a certain time; the transactions accounts include a balancing item which is used to equate the two sides of the accounts (e.g. resources and uses) and which is a meaningful measure of economic performance in itself. (SNA 2.85 and 2.87) Eurostat – as above
Accounting conventions	See IMF, Guide to the Data Dissemination Standards - Module 1: The Special Data Dissemination Standard, May 1996.
Accuracy	*Accuracy* of data or statistical information is the degree to which those data correctly estimate or describe the quantities or characteristics that the statistical activity was designed to measure. 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 error, or the potential significance of error, introduced through individual major sources of error, e.g. coverage, sampling, non-response, response, processing and dissemination. (*Statistics Canada Quality Guidelines*, 3rd edition, October 1998, page 4)
Active metainformation system	An *active metainformation system* is physically integrated with the information system containing the data that the metadata in the metainformation system informs about. (Guidelines for the Modelling of Statistical Data and Metadata, UNECE and UNSC, Conference of European Statisticians Methodological Material, 1995 – page 1) *See also Passive metainformation system.*
Activity – ISIC Rev. 3	An *activity* is a process, i.e. the combination of actions that result in a certain set of products. An activity can be said to take place when resources such as equipment, labour, manufacturing techniques or products are combined, leading to specific goods or services. Thus, an activity is characterised by an input of resources, a production process and an output of products. (ISIC Rev. 3, para. 29) See also *principal activity, secondary activity, ancillary activity*.
Activity – NACE	An *activity* is said to take place when resources such as equipment, labour, manufacturing techniques, information networks or products are combined, leading to the creation of specific goods or services. An activity is characterised by an input of products (goods or services), a production process and an output of products. In practice the majority of units carry on activities of a mixed character. One can distinguish between three types of economic activity: - Principal activity: The principal activity is identified by the topdown method as the activity which contributes most to the total value added of the entity under consideration. The principal activity so identified does not necessarily account for 50% or more of the entity's total value added. - Secondary activity: A secondary activity is any other activity of the entity that produces goods or services. - Ancillary activity: Principal and secondary activities are generally carried out with the support of a number of "ancillary activities", such as accounting, transportation, storage, purchasing, sales promotion, repair and maintenance etc. Thus ancillary activities are those that exist solely to support the main productive activities of an entity by providing non-durable goods or services for the use of that entity. (NACE Rev. 1, pages 14-15 and Council Regulation (EEC), No. 696/93, Section IV B1 and B4 of 15.03.1993 on the statistical units for the observation and analysis of the production system in the Community and Eurostat)
Activity classification	The main purpose of an *activity classification* is to classify productive economic activities. The main aim is to provide a set of activity categories that can be utilised when dissecting statistics according to such activities. *ISIC* is the United Nations International Standard Industrial Classification of All Economic Activities. The third revision of ISIC is used in the 1993 SNA. (ISIC Rev. 3, para. 16) See also *ISIC*
Aggregation/disaggregation	*Aggregation* is the combination of related categories, usually within a common branch of a hierarchy, to provide information at a broader level to that at which detailed observations are taken. *Disaggregation* is the breakdown of observations, usually within a common branch of a hierarchy, to a more detailed level to that at which detailed observations are taken. (*United Nations Glossary of Classification Terms*; prepared by the Expert Group on International Economic and Social Classifications; available at: www.un.org/Depts/unsd/class/glossary_short.htm)
Analysis	See *Data analysis*.
Analytical unit – SNA	For more refined analysis of the production process, use is made of an *analytical unit* of production: this unit, which is not always observable, is the unit of homogeneous production, defined as covering no secondary activities. (SNA 2.48) See also *Statistical units and Observation units*.
Analytical unit – ISIC Rev. 3	*Analytical units* are created by statisticians, often by splitting or combining observation units with the help of estimations and imputations in order to compile more detailed and more homogeneous statistics than is possible using data on observation units. (ISIC Rev. 3, para. 63)

Term	Target definition
Analytical unit – Eurostat	*Analytical units* represent real or artificially constructed units, for which statistics are compiled. Analytical units can correspond therefore for example to enterprises, local units, kind-of-activity units (KAU), local kind-of-activity units (local KAU) as well as to units of homogeneous production (UHP) and local units of homogeneous production (local UHP). (Eurostat)
Attribute	An *attribute* is a characteristic of an object or entity. (ISO/IEC 11179, Part 1, Framework for the specification and standardization of data elements, 1998) An *attribute* is a statistical concept providing qualitative information about a time series or an observation (e.g. concepts such as units, magnitude, currency of denomination, titles and methodological comments could be used as attributes in the context of an agreed data exchange). (GESMES/CB Brief glossary at http://www.ecb.int/stats/gesmes/gesmes.htm) See also *Basic attributes*.
Attribute value	An *attribute value* is a representation of an instance of an attribute. (ISO/IEC 11179, Part 1, Framework for the specification and standardization of data elements, 1998)
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 year but it may be as short as one day or as long as the average of a group of years.). (*A Dictionary of Statistical Terms*, Fourth Edition, Kendall, Sir Maurice G., and William R. Buckland for the International Statistical Institute, 1982, London, Longman Group, page 14)
Base period – SNA	The *base period* is the period that provides the weights for an index. (SNA 16.16)
Basic attributes	*Basic attributes* are attributes that are frequently needed to specify a data element. Categories of basic attributes: *Identifying* are attributes that are applicable for the identification of a data element. *Definitional* are attributes that describe the semantic aspects of a data element. These attributes may be derived by inheritance from characteristics of data element concepts, objects or entities. *Relational* are attributes that describe associations among data elements and/or associations between data elements and classification schemes, data element concepts, objects, entities. *Representational* are attributes that describe representational aspects of a data element. *Administrative* are attributes that describe management and control aspects of a data element. (ISO/IEC 11179, Part 3, Basic attributes of data elements, 1994)
Census	A *census* is a survey conducted on the full set of observation objects belonging to a given population or universe. (*Terminology on Statistical Metadata*, Conference of European Statisticians Statistical Standards and Studies, No. 53, UNECE, Geneva 2000)
Central product classification (CPC) – SNA	The *central product classification (CPC)* is a classification based on the physical characteristics of goods or on the nature of the services rendered; each type of good or service distinguished in the CPC is defined in such a way that it is normally produced by only one activity as defined in ISIC. Each type of good or service distinguished in the CPC is defined in such a way that it is normally produced by only one activity as defined in the International Standard Industrial Classification of all economic activities (ISIC Rev. 3). Conversely, each activity of the ISIC is defined in such a way that it normally produces only one type of product as defined in the CPC (where each type of product may have a number of individual products coded under it). So far as is practically possible, an attempt is made to establish a one-to-one correspondence between the two classifications, each category of the CPC being accompanied by a reference to the ISIC class in which the good or service is mainly produced. However, such a one-to-one correspondence is not always possible. In practice, therefore, the output of an industry, no matter how narrowly defined, will tend to include more than a single product. (SNA 5.44) The Classification of Products by Activity (CPA) is the equivalent classification at European Union level. Detailed information on CPC Rev. 1 is available at: http://esa.un.org/unsd/cr/registry/regcst.asp?Cl=3&Lg=1
Central product classification (CPC) – UN	The *central product classification (CPC)* is a classification based on the physical characteristics of goods or on the nature of the services rendered. Each type of good or service distinguished in the CPC is defined in such a way that it is normally produced by only one activity as defined in ISIC. The CPC covers products that are an output of economic activities, including transportable goods, non-transportable goods and services. The classification structure comprises: - Sections – one digit code; - Divisions – two-digit code; - Groups – three-digit code; - Classes – four-digit code; - Subclasses – five-digit code The current version (Version 1.0) was last revised in 1997 and is expected to be updated by 2002. (*Central Product Classification* (CPC). Version 1.0. United Nations, New York, 1998, Series M, No. 77, Ver. 1.0) Detailed information on CPC Rev. 1 is available at: http://esa.un.org/unsd/cr/registry/regcst.asp?Cl=3&Lg=1

Term	Target definition
Certified data element	A *certified data element* is a recorded data element that has met the quality requirements specified in ISO 11179. (ISO/IEC 11179, Part 1, Framework for the specification and standardization of data elements, 1998)
Chain indices	*Chain indices* are obtained by linking price (or volume) indices for consecutive periods; the short-term movements which are linked are calculated using weighting patterns appropriate to the periods concerned. (SNA 16.41)
Chained index weighting	*Chained index weighting* is an alternative way of weighting together the subaggregates that form GDP. The key difference to the fixed-weight aggregation, used in most countries, is that the prices are continuously updated and that "substitution bias" is avoided and that measures are independent of the choice of base year. (*The OECD Economic Outlook: Sources and Methods.* Available at www.oecd.org/eco/sources-and-methods)
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. The terms "classification" and "nomenclature" are often used interchangeably, despite the definition of a "nomenclature" being narrower than that of a "classification". The structure of classification can be either hierarchical or flat. Hierarchical classifications range from the broadest level (e.g. division) to the detailed level (e.g. class). Flat classifications (e.g. sex classification) are not hierarchical. The characteristics of a good classification are as follows: - the categories are exhaustive and mutually exclusive (i.e. each member of a population can only be allocated to one category without duplication or omission); - the classification is comparable to other related (national or international) standard classifications; - the categories are stable, i.e. they are not changed too frequently, or without proper review, justification and documentation; - the categories are well described with a title in a standard format and backed up by explanatory notes, coding indexes, coders and correspondence tables to related classifications (including earlier versions of the same classification); - the categories are well balanced within the limits set by the principles for the classification (i.e. not too many or too few categories). This is usually established by applying significance criteria (e.g. size limits on variables such as employment, turnover, etc.); - the categories reflect realities of the field (e.g. the society or economy) to which they relate (e.g. in an industry classification, the categories should reflect the total picture of industrial activities of the country); and - the classification is backed up by the availability of instructions, manuals, coding indexes, handbooks and training. (*United Nations Glossary of Classification Terms.* Prepared by the Expert Group on International Economic and Social Classifications. Available at: www.un.org/Depts/unsd/class/glossary_short.htm)
Classification (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. Available at: www.un.org/Depts/unsd/class/glossary_short.htm) See also *Classification*.
Classification scheme	A *classification scheme* is an arrangement or division of objects into groups based on characteristics which the objects have in common, e.g. origin, composition, structure, application, function, etc. (ISO/IEC 11179, Part 1, Framework for the specification and standardization of data elements, 1998)
Classification scheme item	A *classification scheme item* is the discrete components of content in a classification scheme. These may be the nodes of a taxonomy/ontology, the terms of a thesaurus, etc. (ISO/IEC 11179, Part 1, Framework for the specification and standardization of data elements, 1998)
Classification unit	The *classification unit* is the basic unit to be classified in the classification (e.g. in an activity classification this would be the establishment or enterprise, in an occupational classification it will be the job). (*United Nations Glossary of Classification Terms.* Prepared by the Expert Group on International Economic and Social Classifications. Available at: www.un.org/Depts/unsd/class/glossary_short.htm)
Classified component	A *classified component* is an administered component of a data element that may be classified in one or more classification schemes. These components include the object class, property, representation class, data element concept, value domain, and data element. (ISO/IEC 11179, Part 1, Framework for the specification and standardization of data elements, 1998)
Code list	A *Code list* is a predefined list from which some statistical concepts (coded concepts) take their values. (GESMES/CB Brief glossary at http://www.ecb.int/stats/gesmes/gesmes.htm)
Coherence	*Coherence* of data and information reflects the degree to which the data and information from a single statistical program, and data brought together across data sets or statistical programs, are logically connected and complete. Fully coherent data are logically consistent – internally, over time, and across products and programs. Where applicable, the concepts and target populations used or presented are logically distinguishable from similar, but not identical, concepts and target populations of other statistical programs, or from commonly used notions or terminology (*Statistics Canada Quality Guidelines*, 3rd edition, October 1998, page 5).

Term	Target definition
Comments	*Comments* refer to remarks on the data element. (ISO/IEC 11179, Part 1, Framework for the specification and standardization of data elements, 1998)
Comparability	The comparability component aims at measuring the impact of differences in applied statistical concepts and definitions when statistics are compared between geographical areas, non-geographical domains, or reference periods (*Assessment of Quality in Statistics*, Eurostat, April 2000). See also *Quality - Eurostat*
Compilation practices	See IMF, Guide to the Data Dissemination Standards - Module 1: The Special Data Dissemination Standard, May 1996.
Concept	A unit of thought constituted through abstractions on the basis of characteristics common to a set of objects (ISO 1087). *Concepts* are abstract summaries, general notions, knowledge, etc., of a whole set of behaviours, attitudes or characteristics which are seen as having something in common. Concepts are used to assist in presenting/conveying precise meaning, categorising, interpreting, structuring and making sense of phenomena (such as classifications). (*United Nations Glossary of Classification Terms.* Prepared by the Expert Group on International Economic and Social Classifications. Available at: www.un.org/Depts/unsd/class/glossary_short.htm)
Contact	A *contact* is an instance of a role of an individual or an organization (or organization part or organization person) to whom an information item(s), a material object(s) and/or person(s) can be sent to or from in a specified context. (ISO/IEC FCD 11179-3, Registry Metamodel, Final Committee Draft 2001)
Context	The *context* is a designation or description of the application environment or discipline in which a name is applied or from which it originates. (ISO/IEC 11179, Part 1, Framework for the specification and standardization of data elements, 1998)
Context description language	A *context description language* is the identifier of the language used in the context description. (ISO/IEC FCD 11179-3, Registry Metamodel, Final Committee Draft 2001)
Country identifier	A *country identifier* is an identifier further specifying the geographic region associated with the language. Note: Use the three digit numeric codes from ISO 3166, Part 1, with extensions, if required. (ISO/IEC FCD 11179-3, Registry Metamodel, Final Committee Draft 2001)
Coverage	*Coverage* specifies the population from which observations for a particular topic can be drawn. An understanding of coverage is required to facilitate the comparison of data. Coverage issues are often explained through the use of tables showing linkages (e.g. part or full correspondence); and can also be used to explain the ratio of coverage. The rules and conventions of coverage are largely determined by concept definitions, scope rules, information requirements and, in the case of statistical collections and classifications, collection and counting units and the collection methodology. (*United Nations Glossary of Classification Terms.* Prepared by the Expert Group on International Economic and Social Classifications. Available at: www.un.org/Depts/unsd/class/glossary_short.htm)
Coverage errors	*Coverage errors* arise from failure to cover adequately all components of the population being studied. Incomplete sampling frames often result in coverage errors. (*Handbook of Household Surveys, Revised Edition*, Studies in Methods, Series F, No. 31, United Nations, New York, 1984, para. 8.3)
Coverage ratio	The *coverage ratio* measures the extent to which observations designated as primary to a particular category are undertaken by units primarily involved with the observations related to that category. In industry statistics, the coverage ration is the output of goods and services characteristic of a particular industry in proportion to the total output of the same goods and services by the economy as a whole. (*United Nations Glossary of Classification Terms*. Prepared by the Expert Group on International Economic and Social Classifications. Available at: www.un.org/Depts/unsd/class/glossary_short.htm)
Data - UN	*Data* is the physical representation of information in a manner suitable for communication, interpretation, or processing by human beings or by automatic means. (*Terminology on Statistical Metadata*, Conference of European Statisticians Statistical Standards and Studies, No. 53, UNECE, Geneva 2000)
Data - ISO	*Data* is a representation of facts, concepts, or instructions in a formalized manner, suitable for communication, interpretation, or processing by humans or by automatic means. (ISO/IEC 11179, Part 1, Framework for the specification and standardization of data elements, 1998)
Data analysis	*Data analysis* is the process of transforming raw data into usable information, often presented in the form of a published analytical article, in order to add value to the statistical output. (*Statistics Canada Quality Guidelines*, 3rd edition, October 1998, page 62)
Data collection	*Data collection* is an activity of the survey life cycle for gathering data from respondents and recording it for further processing. (*Terminology on Statistical Metadata*, Conference of European Statisticians Statistical Standards and Studies, No. 53, UNECE, Geneva 2000)

Term	Target definition
Data collection	Data collection is any process whose purpose is to acquire or assist in the acquisition of data. Collection is achieved by requesting and obtaining pertinent data from individuals or organisations via an appropriate vehicle. (Statistics Canada Quality Guidelines, 3rd edition, October 1998, page 31)
Data dictionary - UN	A *data dictionary* is a description of a set of data elements usually associated with a single data set. (*Terminology on Statistical Metadata*, Conference of European Statisticians Statistical Standards and Studies, No. 53, UNECE, Geneva 2000)
Data dictionary - ISO	A *data dictionary* is a database used for data that refers to the use and structure of other data; that is, a database for the storage of metadata [ANSI X3.172-1990]. (ISO/IEC 11179, Part 1, Framework for the specification and standardization of data elements, 1998) See also *Data element dictionary*
Data dissemination	*Data dissemination* is an activity in the survey life cycle to distribute or transmit statistical data to its users. (*Terminology on Statistical Metadata*, Conference of European Statisticians Statistical Standards and Studies, No. 53, UNECE, Geneva 2000)
Data element - UN	A *data element* is the smallest identifiable unit of data within a certain context for which the definition, identification, permissible values and other information is specified by means of a set of attributes. (*Terminology on Statistical Metadata*, Conference of European Statisticians Statistical Standards and Studies, No. 53, UNECE, Geneva 2000)
Data element - ISO	A *data element* is a unit of data for which the definition, identification, representation, and permissible values are specified by means of a set of attributes. (ISO/IEC 11179, Part 1, Framework for the specification and standardization of data elements, 1998)
Data element concept	A *data element concept* is a concept that can be represented in the form of a data element, described independently of any particular representation. (ISO/IEC 11179, Part 1, Framework for the specification and standardization of data elements, 1998)
Data element dictionary	A *data element dictionary* is an information resource that lists and defines all relevant data elements. (ISO/IEC 11179, Part 1, Framework for the specification and standardization of data elements, 1998) See also *Register*
Data element facet	A *data element facet* is any aspect of a data element that is subject to classification. This includes object class, property, representation, and data element concept. (ISO/IEC 11179, Part 1, Framework for the specification and standardization of data elements, 1998)
Data element name	A *data element name* is a single or multi-word designation used as the primary means of identification of data elements for humans. (ISO/IEC 11179, Part 1, Framework for the specification and standardization of data elements, 1998)
Data element registry	A *data element registry* is an information resource kept by a registration authority that describes the meaning and representational form of data elements, including registration identifiers, definitions, names, value domains, metadata and administrative attributes, etc. (ISO/IEC 11179, Part 1, Framework for the specification and standardization of data elements, 1998) See also *Register*
Data element value	A *data element value* is a value out of a set of permissible values pertaining to a data element. (ISO/IEC 11179, Part 1, Framework for the specification and standardization of data elements, 1998) See also *Data value*
Data exchange	See *Information interchange*
Data identifier	A *data identifier* is an identifier of a data element (a string of characters or other graphic symbols) assigned by a registration authority. (ISO/IEC 11179, Part 1, Framework for the specification and standardization of data elements, 1998)
Data item	A *data item* is one occurrence of a data element. (ISO/IEC 11179, Part 1, Framework for the specification and standardization of data elements, 1998)
Data model	A *data model* is a description of the organisation of data in a manner that reflects an information structure (ISO/IEC 11179, Part 1, Framework for the specification and standardization of data elements, 1998)
Data processing	The operation performed on data in order to derive new information according to a given set of rules
Data quality	See *Quality*
Data steward	A *data steward* is a person or organization delegated the responsibility for managing a specific set of data resources. (ISO/IEC 11179, Part 1, Framework for the specification and standardization of data elements, 1998)

Term	Target definition
Datatype	*Datatype* refers to the format used for the collection of letters, digits, and/or symbols, to depict values of a data element, determined by the operations that may be performed on the data element. (ISO/IEC 11179, Part 1, Framework for the specification and standardization of data elements, 1998)
Datatype of data element values	The *datatype of data element values* refers to a set of distinct values for representing the data element value. (ISO/IEC 11179, Part 1, Framework for the specification and standardization of data elements, 1998)
Data value	A *data value* is an element of a value domain. (ISO/IEC 11179, Part 1, Framework for the specification and standardization of data elements, 1998)
Date	A *date* is a time reference. (*GESMES/CB Data Model*, Release 2, July 2000, ECB, BIS, Eurostat et al)
Definition - UN	A *definition* is a statement of the precise meaning of something. In classifications this refers to the explanation of the concepts encompassed in category description and often includes specific examples of what is and is not included in particular categories. (*United Nations Glossary of Classification Terms*, prepared by the Expert Group on International Economic and Social Classifications, available at: www.un.org/Depts/unsd/class/glossary_short.htm)
Definition - ISO	A *definition* is a word or phrase expressing the essential nature of a person or thing or class of persons or things: an answer to the question "what is x?" or "what is an x?"; a statement of the meaning of a word or word group [Webster's Third New International Dictionary of the English Language Unabridged, 1986]. Statement that expresses the essential nature of a data element and permits its differentiation from all other data elements. (ISO/IEC 11179, Part 1, Framework for the specification and standardization of data elements, 1998)
Dimension	*"Dimension"* is a statistical concept used to refer to and identify a time series (e.g. a concept indicating a certain economic activity or a concept identifying a geographical reference area). (GESMES/CB Brief glossary at http://www.ecb.int/stats/gesmes/gesmes.htm)
Dissemination media	*Dissemination* is the release to users of information obtained through a statistical activity. Various release media are possible, for example: electronic format including the internet, CD-ROM; a paper publication; microfiche; a microdata file available to authorised users or for public use; a telephone or facsimile response to a special request; or a public speech, presentation or television or radio interview. (*Statistics Canada Quality Guidelines*, 3rd edition, October 1998, page 59)
Documentation	*Documentation* is descriptive text used to define or describe an object, design, specification, instructions or procedure. (*Terminology on Statistical Metadata*, Conference of European Statisticians Statistical Standards and Studies, No. 53, UNECE, Geneva 2000)
Domain	A domain is the set of possible data values of an attribute. [ISO/IEC 2382]. (ISO/IEC 11179, Part 1, Framework for the specification and standardization of data elements, 1998) See also *Value domain*
Domain of study	A *domain of study* is a major segment of the population for which separate statistics are needed. In the course of tabulation, data may actually be provided for many population segments; however, a domain of study would be a segment identified in the overall statistical plan as one for which a certain level of detail and certain data reliability were required. The domains of study chosen may co-incide with the strata adopted for stratified sampling or may cut across them. A domain could consist of a geographical area such as a region or major population centre. It could also comprise a specified population category, such as a major national or ethnic group. The number of domains has an important bearing on the size and distribution of the sample. (*Handbook of Household Surveys, Revised Edition*, Studies in Methods, Series F, No. 31, United Nations, New York, 1984, paras. 4.6, 4.7)
EDIFACT	*EDIFACT* stands for "Electronic data interchange for administration, commerce and transport". (GESMES/CB Brief glossary at http://www.ecb.int/stats/gesmes/gesmes.htm)
Editing (data)	Data *editing* is the application of checks that identify missing, invalid or inconsistent entries or that point to data records that are potentially in error. (*Statistics Canada Quality Guidelines*, 3rd edition, October 1998, page 35)
Entity	An *entity* is any concrete or abstract thing of interest, including associations among things. [ISO/IEC 2382]. (ISO/IEC 11179, Part 1, Framework for the specification and standardization of data elements, 1998) See also *Object class*
Enumerated domain	A *enumerated domain* is a value domain that is specified by a list of all permissible values. (ISO/IEC 11179, Part 1, Framework for the specification and standardization of data elements, 1998)
Estimation	*Estimation* is a process that consists of assigning values to unknown population parametres by using information from a data set. (*Statistics Canada Quality Guidelines*, 3rd edition, October 1998, page 41)
Form of representation	The *form of representation* refers to the name or description of the form of representation for the data element. e.g. 'quantitative value', 'code', 'text', 'icon'. (ISO/IEC 11179, Part 1, Framework for the specification and standardization of data elements, 1998) See also *Representation term*

Term	Target definition
Frame	A *frame* is any list, material or device that delimits, identifies, and allows access to the elements of the target population. (*Statistics Canada Quality Guidelines*, 3rd edition, October 1998, page 16)
GESMES	*GESMES (Generic Statistical Message)* is a United Nations standard (EDIFACT message) allowing partner institutions to exchange statistical multidimentional arrays in a generic but standardised way. It has been designed by Expert Group 6 (Statistics) of the European Board for EDI Standardisation. (GESMES/CB Brief glossary at http://www.ecb.int/stats/gesmes/gesmes.htm) See also *Statistical message.*
GESMES/CB	*GESMES/CB* is a message (a GESMES profile) allowing the exchange of statistical time series, related attributes and structural metadata using a standardised format. (GESMES/CB Brief glossary at http://www.ecb.int/stats/gesmes/gesmes.htm)
Glossary	A collection of terms. (*Terminology on Statistical Metadata* , Conference of European Statisticians Statistical Standards and Studies, No. 53, UNECE, Geneva 2000)
Imputation	*Imputation* is the process used to identify problems of missing, invalid or inconsistent responses identified during editing. This is done by changing some of the resonses or missing values on the record being edited to ensure that a plausible, internally coherent record is created. (*Statistics Canada Quality Guidelines*, 3rd edition, October 1998, page 38)
Identifier	A *identifier* is a language independent unique identifier of a data element within a registration authority. An unambiguous name for an object within a given context. (ISO/IEC 11179, Part 1, Framework for the specification and standardization of data elements, 1998) See also *Data identifier*
Industry – SNA	An *industry* consists of a group of establishments engaged on the same, or similar, kinds of production activity. The classification of productive activities used in the SNA is ISIC (Rev.3). (SNA 5.5 and 5.40)
Industry	*Industry* refers to the activity of the establishment in which the person worked or was employed during the reference period. (*Handbook of Household Surveys, Revised Edition*, Studies in Methods, Series F, No. 31, United Nations, New York, 1984, para. 11.87)
Information	*Information* is knowledge concerning any objects such as facts, events, things, processes or ideas including concepts that within a certain context has a particular meaning (ISO/IEC 2382-1; 1992) (*Terminology on Statistical Metadata* , Conference of European Statisticians Statistical Standards and Studies, No. 53, UNECE, Geneva 2000)
Information (in information processing) - ISO	*Information (in information processing)* refers to knowledge concerning objects, such as facts, events, things, processes, or ideas, including concepts, that within a certain context has a particular meaning. [ISO/IEC 2382] (ISO/IEC 11179, Part 1, Framework for the specification and standardization of data elements, 1998)
Information interchange	*Information interchange* is the process of sending and receiving data in such a manner that the information content or meaning assigned to the data is not altered during the transmission (ISO/IEC 11179, Part 1, Framework for the specification and standardization of data elements, 1998)
Information system	A system which supports decision-making concerning some piece of reality, the object system by giving the decision makers access to information concerning relevant aspects of the object system and its environment. (*Terminology on Statistical Metadata* , Conference of European Statisticians Statistical Standards and Studies, No. 53, UNECE, Geneva 2000) See also *Statistical information system.*
International registration data identifier (IRDI)	The *international registration data identifier* (IRDI) is an internationally unique identifier for a data element. (ISO/IEC 11179, Part 1, Framework for the specification and standardization of data elements, 1998)
Keyword	A *keyword* is one or more significant words used for retrieval of data elements (ISO/IEC 11179, Part 1, Framework for the specification and standardization of data elements, 1998)
Layout of representation	The *layout of representation* is the layout of characters in data element values expressed by a character string representation. (ISO/IEC 11179, Part 1, Framework for the specification and standardization of data elements, 1998)
Lexical	*Lexical* pertains to words or the vocabulary of a language as distinguished from its grammar and construction. (ISO/IEC 11179, Part 1, Framework for the specification and standardization of data elements, 1998)
Maximum size of data element values	The *maximum size of data element values* refers to the maximum number of storage units (of the corresponding datatype) to represent the data element value. (ISO/IEC 11179, Part 1, Framework for the specification and standardization of data elements, 1998)

Term	Target definition
Metadata	*Metadata* is data that defines and describes other data (ISO/IEC 11179, Part 1, Framework for the specification and standardization of data elements, 1998) *Metadata* provides information on data – and about processes of producing and using data. (*Guidelines for the Modelling of Statistical Data and Metadata*, UNECE and UNSC, Conference of European Statisticians Methodological Material, 1995 – page 1). *Metadata* comprises data and other documentation that describes objects in a formalised way. (*Terminology on Statistical Metadata*, Conference of European Statisticians Statistical Standards and Studies, No. 53, UNECE, Geneva 2000) See also *Statistical metadata*.
Metadata registry	A *Metadata registry* is a system for managing structured metadata describing the semantic content of shareable data and metadata. (ISO/IEC FCD 11179-3, Registry Metamodel, Final Committee Draft 2001)
Metadata layer	A *metadata layer* is a layer in the reference model for standardisation in statistics used to denote the set of attributes related to statistical metainformation. (*Terminology on Statistical Metadata*, Conference of European Statisticians Statistical Standards and Studies, No. 53, UNECE, Geneva 2000)
Methodology	A *methodology* is a structured approach to solve a problem. (*Terminology on Statistical Metadata*, Conference of European Statisticians Statistical Standards and Studies, No. 53, UNECE, Geneva 2000)
Minimum size of data element values	The *minimum size of data element values* is the minimum number of storage units (of the corresponding datatype) to represent the data element value. (ISO/IEC 11179, Part 1, Framework for the specification and standardization of data elements, 1998)
Name	The *name* is the primary means of identification of objects and ideas for humans. A single or multi-word designation assigned to a data element. (ISO/IEC 11179, Part 1, Framework for the specification and standardization of data elements, 1998)
Object	An *object* refers to any part of the conceivable or perceivable world. [ISO 1087]. (ISO/IEC 11179, Part 1, Framework for the specification and standardization of data elements, 1998)
Object class	An *object class* refers to a set of objects. A set of ideas, abstractions, or things in the real world that can be identified with explicit boundaries and meaning and whose properties and behavior follow the same rules. (ISO/IEC 11179, Part 1, Framework for the specification and standardization of data elements, 1998)
Object class term	An *object class term* is a component of the name of a data element which represents the object class to which it belongs; e.g. "employee". (ISO/IEC 11179, Part 1, Framework for the specification and standardization of data elements, 1998)
Objectives	*Objectives* are the purposes for which information is required, stated within the context of the program, research problem or hypotheses that gave rise to the need for information. (*Statistics Canada Quality Guidelines*, 3rd edition, October 1998, page 9)
Observation unit	*Observation units* are those entities on which information is received and statistics are compiled. (ISIC Rev. 3, para. 63) See also *Statistical unit* and *Analytical unit.*
Observation unit (Eurostat)	An *Observation* unit represents an identifiable entity, about which data can be obtained. During the collection of data, this is the unit for which data is recorded. It should be noted that this may, or may not be, the same as the reporting unit. *(Eurostat)*
Organization	An *Organization* is a unique framework of authority within which a person or persons act, or are designated to act, towards some purpose. (ISO/IEC FCD 11179-3, Registry Metamodel, Final Committee Draft 2001)
Passive metainformation system	A *passive metainformation system* contains only references to data, not the data themselves. (*Guidelines for the Modelling of Statistical Data and Metadata*, UNECE and UNSC, Conference of European Statisticians Methodological Material, 1995 – page 1. See also *Active metainformation system*.
Period	A *period* is length of time. (*Dictionary of Banking and Finance*, Second Edition, P.H. Collin, 1991, Peter Collin Publishing) The expression period is used to denote the interval or average interval between identifiable points of recurrence. (*A Dictionary of Statistical Terms,* 5th Edition, F.H.C. Marriott, prepared for the International Statistical Institute, Longman Scientific and Technical, 1990)
Periodicity	*Periodicity* refers to the frequency with which the data are compiled. Periodicity specifications for flow data are given in terms of the longest interval to be represented by a single data point; those for stock data are given in the form of the longest interval between compilations. For example, the specification of quarterly periodicity for balance of payments data means that one quarter is the longest interval that may be represented by a single estimate. (*Glossary of Foreign Direct Investment Terms*, OECD, 2001 – unpublished)

Term	Target definition
Permissible data element values	*Permissible data element values* refers to the set of representations of permissible instances of the data element, according to the representation form, layout, datatype, maximum size, and minimum size specified in the corresponding attributes. The set can be specified by name, by reference to a source, by enumeration of the representation of the instances, or by rules for generating instances. (ISO/IEC 11179, Part 1, Framework for the specification and standardization of data elements, 1998)
Population (statistical)	*Population* is the total membership or population or "universe" of a defined class of people, objects or events. There are two types of population, viz, target population and survey population. A target population is the population outlined in the survey objects about which information is to be sought and a survey population is the population from which information can be obtained in the survey. The target population is also known as the scope of the survey and the survey population is also known as the coverage of the survey. For administrative records the corresponding populations are: the "target" population as defined by the relevant legislation and regulations, and the actual "client population". (*United Nations Glossary of Classification Terms.* Prepared by the Expert Group on International Economic and Social Classifications. Available at: www.un.org/Depts/unsd/class/glossary_short.htm)
Property	A *property* is a peculiarity common to all members of an object class. (ISO/IEC 11179, Part 1, Framework for the specification and standardization of data elements, 1998)
Property term	A *property term* is a component of the data element name which expresses a property of an object class. (A component of the name of a data element which expresses the category to which the data element belongs.) (ISO/IEC 11179, Part 1, Framework for the specification and standardization of data elements, 1998)
Reference document	A *Reference document* is a document that provides pertinent details for consultation about a subject. (ISO/IEC FCD 11179-3, Registry Metamodel, Final Committee Draft 2001)
Qualifier	A *qualifier* is a term that helps define and render a concept unique. (ISO/IEC 11179, Part 1, Framework for the specification and standardization of data elements, 1998)
Qualifier term	A *qualifier term* is a word or words which help define and differentiate a name within the database. (ISO/IEC 11179, Part 1, Framework for the specification and standardization of data elements, 1998)
Quality (Eurostat)	*Quality* is defined in the ISO 8402 - 1986 as: 'the totality of features and characteristics of a product or service that bear on its ability to satisfy stated or implied needs'. Eurostat defines the quality of statistics with reference to several criteria. These comprise: - Relevance: an inquiry is relevant if it meets users' needs. The identification of users and their expectations is therefore necessary. - Accuracy: accuracy is defined as the closeness between the estimated value and the (unknown) true value. - Timeliness and punctuality in disseminating results: most users want up-to-date figures which are published frequently and on time at pre-established dates. - Accessibility and clarity of the information: statistical data have most value when they are easily accessible by users, are available in the forms users desire and are adequately documented. - Comparability: statistics for a given characteristic have the greatest usefulness when they enable reliable comparisons of values taken by the characteristic across space and time. The comparability component stresses the comparison of the same statistics between countries in order to evaluate the meaning of aggregated statistics at the European level. - Coherence: when originating from a single source, statistics are coherent in that elementary concepts can be combined reliably in more complex ways. When originating from different sources, and in particular from statistical surveys of different frequencies, statistics are coherent in so far as they are based on common definitions, classifications and methodological standards. - Completeness: domains for which statistics are available should reflect the needs and priorities expressed by the users of the European Statistical System. (*Assessment of Quality in Statistics*, Eurostat, April 2000)
Quality (IMF)	The IMF has developed a "definition" of data *quality* which it uses in the context of its Data Quality Assessment Framework (DQAF). The dimensions quality within the IMF definition are: - integrity; - methodological soundness; - accuracy and reliability; - serviceability; - accessibility. The IMF also states that there are a number of prerequisites for quality. These comprise: - legal and institutional environment; - resources; - quality awareness. Further information on the IMF definition of quality, together with a glossary of terms used within the definition are available at http://dsbb.imf.org/glossary.pdf.

Term	Target definition
Quality of life	*Quality of life* is the notion of human welfare (well-being) measured by social indicators rather than by "quantitative" measures of income and production. *(Glossary of Environment Statistics*, Studies in Methods, Series F, No. 67, United Nations, New York, 1997)
Quantitative data	*Quantitative data* is data expressing a certain quantity, amount or range of values related to an object. The quantitative data is usually associated with measurement units. (*Terminology on Statistical Metadata*, Conference of European Statisticians Statistical Standards and Studies, No. 53, UNECE, Geneva 2000)
Questionnaire	A *questionnaire* is an identifiable instrument containing questions for gathering data from respondents. (*Terminology on Statistical Metadata*, Conference of European Statisticians Statistical Standards and Studies, No. 53, UNECE, Geneva 2000)
Recorded data element	A *recorded data element* is a submitted data element which contains all mandatory attributes and has been recorded but the contents may not meet the quality requirements specified in other parts of ISO/IEC 11179. (ISO/IEC 11179, Part 1, Framework for the specification and standardization of data elements, 1998)
Reference document	A *reference document* is a document that provides pertinent details for consultation about a subject (ISO/IEC 11179, Part 1, Framework for the specification and standardization of data elements, 1998)
Reference period	In connection with price or volume indices, the *reference period* means the period to which the indices relate; it is typically set equal to 100 and it does not necessarily coincide with the "base" period that provides the weights for the indices. (SNA [16.16])
Register - UN	A written and complete record containing regular entries of items and details on particular set of objects. (*Terminology on Statistical Metadata*, Conference of European Statisticians Statistical Standards and Studies, No. 53, UNECE, Geneva 2000)
Register - ISO	A *register* is a set of files (paper, electronic, or a combination) containing the assigned data elements and the associated information. (ISO/IEC 11179, Part 1, Framework for the specification and standardization of data elements, 1998) See also *Data Element registry*
Registration	*Registration* is the assignment of an unambiguous identifier to a data element in a way that makes the metadata about those data elements available to interested parties. (ISO/IEC 11179, Part 1, Framework for the specification and standardization of data elements, 1998)
Registration applicant	A *registration applicant* is an organization, individual, etc, which requests the assignment of an identifier from a registration authority. (ISO/IEC 11179, Part 1, Framework for the specification and standardization of data elements, 1998)
Registration authority	A *registration authority* is an organization authorized to register data elements or other objects. (ISO/IEC 11179, Part 1, Framework for the specification and standardization of data elements, 1998)
Registration authority identifier	A *registration authority identifier* is an identifier assigned to a registration authority. (ISO/IEC 11179, Part 1, Framework for the specification and standardization of data elements, 1998)
Registration status	*Registration status* is a designation of the position in the registration life-cycle of a data element. (ISO/IEC 11179, Part 1, Framework for the specification and standardization of data elements, 1998)
Related data reference	A *related data reference* is a reference between a data element and any related data. (ISO/IEC 11179, Part 1, Framework for the specification and standardization of data elements, 1998)
Release calendar	A general statement on the schedule of release of data (IMF, Guide to the Data Dissemination Standards - Module 1: The Special Data Dissemination Standard, May 1996).
Relevance	The *relevance* of data or of statistical information is a qualitative assessment of the value contributed by these data. Value is characterised by the degree to which the data or information serve to address the purposes for which they are produced and sought by users. Value is further characterised by the merit of those purposes, in terms of the mandate of the agency, legislated requirements and the opportunity cost to produce data or information. (*Statistics Canada Quality Guidelines*, 3rd edition, October 1998, page 4)
Reporting unit	A *reporting unit* is a unit that supplies the data for a given survey instance. (*Terminology on Statistical Metadata*, Conference of European Statisticians Statistical Standards and Studies, No. 53, UNECE, Geneva 2000)
Reporting unit (Eurostat)	The *reporting unit* is the unit that reports to the survey authority. It reports information for each of the observation units. In certain cases it may correspond to an observation unit. (Eurostat)

Term	Target definition
Response errors [in household surveys]	*Response errors* [in household surveys] may be defined as those arising from the interviewing process. Such errors can result from a number of circumstances, such as the following: - inadequate concepts or questions; - inadequate training; - interviewer failures; - respondent failures. (*Handbook of Household Surveys, Revised Edition*, Studies in Methods, Series F, No. 31, United Nations, New York, 1984, para. 8.6)
Representation	*Representation* is the combination of a value domain, datatype, and, if necessary, a unit of measure or a character set. (ISO/IEC 11179, Part 1, Framework for the specification and standardization of data elements, 1998)
Representation category	A *representation category* is the type of symbol, character, or other designation used to represent a data element. (ISO/IEC 11179, Part 1, Framework for the specification and standardization of data elements, 1998)
Representation term	A *representation term* is a component of a data element name which describes the form of representation of the data element. (ISO/IEC 11179, Part 1, Framework for the specification and standardization of data elements, 1998)
Responsible organisation	A *responsible organization* is the organization or unit within an organization that is responsible for the contents of the mandatory attributes by which the data element is specified. (ISO/IEC 11179, Part 1, Framework for the specification and standardization of data elements, 1998)
Sample	A *sample* is a subset of a frame where elements are selected based on a randomised process with a known probability of selection. (*Terminology on Statistical Metadata*, Conference of European Statisticians Statistical Standards and Studies, No. 53, UNECE, Geneva 2000)
Sample survey	A *sample survey* is a survey which is carried out using a sampling method, i.e. in which a portion only, and not the whole population is surveyed. (*A Dictionary of Statistical Terms*, 5th Edition, F.H.C. Marriott, prepared for the International Statistical Institute, Longman Scientific and Technical, 1990)
Sample survey (2)	A *sample survey* is a survey in which data are collected from a (usually random) sample of population members. (*Statistics Canada Quality Guidelines*, 3rd edition, O(ctober 1998, page 7)
Sampling	*Sampling* is the process of selecting a number of cases from all the cases in a particular group or universe. (*Handbook of Vital Statistics Systems and Methods, Volume 1: Legal, Organisational and Technical Aspects*, United Nations Studies in Methods, Glossary, Series F, No. 35, United Nations, New York 1991)
Sampling error	*Sampling errors* occur when the results of a survey are based on a sample of the population rather than the entire population. (*Statistics Canada Quality Guidelines*, 3rd edition, October 1998, page 52)
Sampling fraction	The *sampling fraction* is the ratio of the sample size to the population size. (*Statistics Canada Quality Guidelines*, 3rd edition, October 1998, page 22)
Sampling frame	A *sampling frame* is a body of information about the population being investigated which is used as the basis for selecting samples and in subsequent estimation procedures. (*Handbook of Household Surveys, Revised Edition*, Studies in Methods, Series F, No. 31, United Nations, New York, 1984, para. 4.8)
Sampling unit	A *sampling unit* is one of the units into which an aggregate is divided for the purpose of sampling, each unit being regarded as individual and indivisible when the selection is made. Such units may be defined on some natural basis, such as a household or a person, or on some arbitrary basis, such as areas defined by co-ordinates on a map. (*Handbook of Vital Statistics Systems and Methods, Volume 1: Legal, Organisational and Technical Aspects*, United Nations Studies in Methods, Glossary, Series F, No. 35, United Nations, New York 1991)
Scope (universe)	The *scope* is the coverage or sphere of what is to be observed. It is the total membership or population of a defined set of people, object or events. (*United Nations Glossary of Classification Terms*. Prepared by the Expert Group on International Economic and Social Classifications. Available at www.un.org/Depts/unsd/class/glossary_short.htm)
Seasonal adjustment	*Seasonal adjustment* is a statistical technique to remove the effects of seasonal calendar influences operating on a series. Seasonal effects usually reflect the influence of the seasons themselves either directly or through production series related to them, or social conventions. Other types of calendar variation occur as a result of influences such as number of days in the calendar period, the accounting or recording practices adopted or the incidence of moving holidays (such as Easter). (*An Analytical Framework for Price Indexes in Australia: Glossary and References*, Australian Bureau of Statistics, Canberra, 1997) See *Trend estimates*
Seasonal adjustment	*Seasonal adjustment* consists of estimating seasonal factors and applying them to a time series to remove seasonal variations. These variations represent the composite effect of climatic and institutional factors that repeat with certain regularity within the year. (*Statistics Canada Quality Guidelines*, 3rd edition, October 1998, page 46)

Term	Target definition
Seasonal adjustment (Eurostat)	To examine the extent to which a change in the original short -term data (gross or unadjusted data) indicates the "real" development of the business cycle, the influence of seasonal fluctuations must be considered. The normal decomposition is the decompositi on into trend, cycle, seasonal variation and irregular fluctuations. (Adapted from Methodology of Short-term statistics, Eurostat 1997 p. 31)
Seasonal adjustment (ECO)	Series are adjusted for seasonal variations and in some cases for calendar working days variations. When available, the seasonally adjusted data are taken directly from national statistical sources; otherwise, the method used for de-seasonalisation is the standard X -11 ARIMA which was developed by the US Bureau of Census and incorporates general smoothing techniques and spectral analyses. (Further details may be found in Technical Paper No. 15 of the Bureau of the Census.). Where appropriate, series are also corrected for calendar variations (e.g. work-days per month) and constrained for annual coherency. *(The OECD Economic Outlook: Sources and Methods.* Available at www.oecd.org/eco/sources-and-methods)
Semantics	*Semantics* is the branch of linguistic science which deals with the meaning of words ((ISO/IEC 11179, Part 1, Framework for the specification and standardization of data elements, 1998) [from *Webster*].
Separator	A *separator* is a symbol or space enclosing or separating components within a name; a delimiter. (ISO/IEC 11179, Part 1, Framework for the specification and standardization of data elements, 1998)
Shareable data	*Shareable data* is data that has precise identifiers, meaning, structures, and values. (ISO/IEC FCD 11179-3, Registry Metamodel, Final Committee Draft 2001)
Standardised data element	A *standardized data element* is a certified data element within the data element registry that is preferred for use. (ISO/IEC 11179, Part 1, Framework for the specification and standardization of data elements, 1998)
Statistical concept	A *statistical concept* is a statistical characteristic of a time series or an observation. (GESMES/CB Brief glossary at http://www.ecb.int/stats/gesmes/gesmes.htm)
Statistical data collection	*Statistical data collection* is the operation of statistical data processing aimed at gathering of statistical data and producing the input object data of a statistical survey. (*Terminology on Statistical Metadata* , Conference of European Statisticians Statistical Standards and Studies, No. 53, UNECE, Geneva 2000)
Statistical data editing	*Statistical data editing* is the operation of detecting and correcting errors in statistical data. (*Terminology on Statistical Metadata* , Conference of European Statisticians Statistical Standards and Studies, No. 53, UNECE, Geneva 2000)
Statistical indicator	A data element that represents statistical data for a specified time, place, and other characteristics. (*Terminology on Statistical Metadata* , Conference of European Statisticians Statistical Standards and Studies, No. 53, UNECE, Geneva 2000)
Statistical information system	The information system oriented towards the collection, storage, transformation and distribution of statistical information. (*Terminology on Statistical Metadata* , Conference of European Statisticians Statistical Standards and Studies, No. 53, UNECE, Geneva 2000)
Statistical message	A *Statistical message* is a message carrying statistical data. (*Terminology on Statistical Metadata* , Conference of European Statisticians Statistical Standards and Studies, No. 53, UNECE, Geneva 2000) A *Statistical message* is a predefined and agreed way of representing syntactically sets of statistical data, attributes and structural metadata which need to be exchanged between partners. (GESMES/CB Brief glossary at http://www.ecb.int/stats/gesmes/gesmes.htm)
Statistical metadata	*Statistical metadata* is metadata describing statistical data. (*Terminology on Statistical Metadata*, Conference of European Statisticians Statistical Standards and Studies, No. 53, UNECE, Geneva 2000). *Statistical metadata* are data which are needed for proper production and usage of statistical data. They describe statistical data and – to some extent – processes and tools involved in the production and usage of statistical data. Expressed briefly, statistical metadata are data about statistical data. (*Guidelines for the Modelling of Statistical Data and Metadata* , UNECE and UNSC, Conference of European Statisticians Methodological Material, 1995 – page 2). See also *Metadata.*
Statistical metadata system	A *statistical metadata system* is a data processing system that uses, stores and produces statistical metadata. (*Terminology on Statistical Metadata* , Conference of European Statisticians Statistical Standards and Studies, No. 53, UNECE, Geneva 2000) See also *Statistical metadata* and *Metadata.*
Statistical metainformation system	A *statistical metainformation system* is a system which uses and produces statistical metadata, informing about statistical data, and which fulfils its tasks by means of functions like "statistical metadata collection", "statistical metadata processing", "statistical metadata storage", and "statistical metadata dissemination". A statistical metainformation system may be either active or passive. (*Guidelines for the Modelling of Statistical Data and Metadata* , UNECE and UNSC, Conference of European Statisticians Methodological Material, 1995 – page 2.)
Statistical production	The activity that is carried out within statistical information system and aimed at producing of statistics. (*Terminology on Statistical Metadata* , Conference of European Statisticians Statistical Standards and Studies, No. 53, UNECE, Geneva 2000)

Term	Target definition
Statistical units - ISIC	*Statistical units* are the entities for which information is sought and for which statistics are ultimately compiled. These units can, in turn, be divided into observation units and analytical units. The statistical units in ISIC Rev. 3 comprise the: - enterprise; - enterprise group; - kind-of-activity unit (KAU); - local unit; - establishment; - homogeneous unit of production (ISIC Rev. 3, paras. 63, 76)
Stratification	*Stratification* consists of dividing the population into subsets (called strata) before the selection of a sample within each of these subsets. (*Statistics Canada Quality Guidelines*, 3rd edition, October 1998, page 20)
Structural metadata	*Structural metadata* are the concepts, structure definitions and code lists defined by a centre institution for the exchange of statistical information with its partners and other organisations. (GESMES/CB Brief glossary at http://www.ecb.int/stats/gesmes/gesmes.htm)
Structure definition	Structure definition: in order to exchange statistical information, a central institution has to agree with its partners which statistical concepts are necessary for identifying the series (and for use as dimensions) and which statistical concepts to be used as attributes. These form the structure definition. (GESMES/CB Brief glossary at http://www.ecb.int/stats/gesmes/gesmes.htm)
Structure set	A *structure set* is a method of placing objects in context, revealing relationships to other objects. Examples include entity-relationship models, taxonomies, and ontologies. (ISO/IEC 11179, Part 1, Framework for the specification and standardization of data elements, 1998)
Submitting organisation	The *submitting organization* is the organization or unit within an organization that has submitted the data element for addition, change, cancellation, or withdrawal in the data element registry. (ISO/IEC 11179, Part 1, Framework for the specification and standardization of data elements, 1998)
Survey	A *survey* is an investigation about the characteristics of a given population by means of collecting data from a sample of that population and estimating their characteristics through the systematic use of statistical methodology. (*Terminology on Statistical Metadata*, Conference of European Statisticians Statistical Standards and Studies, No. 53, UNECE, Geneva 2000)
Survey	The term *survey* covers any activity that collects or acquires statistical data. Included are censuses, sample surveys, the collection of data from administrative records and derived statiscal activities. (*Statistics Canada Quality Guidelines*, 3rd edition, October 1998, page 7)
Survey design	*Survey design* covers the definition of all aspects of a survey from the establishment of a need for data to the production of final outputs (the microdata file, statistical series, and analysis). (*Statistics Canada Quality Guidelines*, 3rd edition, October 1998, page 7)
Survey types (for collecting demographic information)	For the purpose of collecting demographic information, in general there are three types of surveys which have been used in many countries. These are: - single-round surveys; - multi-round surveys; and - dual record systems. (*Handbook of Household Surveys, Revised Edition*, Studies in Methods, Series F, No. 31, United Nations, New York, 1984, para. 9.80)
Synonymous name	A *synonymous name* is a single or multi-word designation that differs from the given name, but represents the same data element concept. (ISO/IEC 11179, Part 1, Framework for the specification and standardization of data elements, 1998)
Syntax	*Syntax* is the structure of expressions in a language, and the rules governing the structure of a language. The relationships among characters or groups of characters, independent of their meanings or the manner of their interpretation and use. (ISO/IEC 11179, Part 1, Framework for the specification and standardization of data elements, 1998)
System of National Accounts (SNA)	The *System of National Accounts* (*SNA*) consists of a coherent, consistent and integrated set of macroeconomic accounts, balance sheets and tables based on a set of internationally agreed concepts, definitions, classifications and accounting rules. (SNA 1.1) The *System of National Accounts* 1993 (SNA) has been prepared under the joint responsibility of the United Nations, the International Monetary Fund, the Commission of the European Communities, the OECD and the World Bank. The SNA is a comprehensive, consistent and flexible set of macroeconomic accounts intended to meet the needs of government and private-sector analysts, policy makers and decision takers. It is designed for use in countries with market economies, whatever their stage of economic development, and also in countries in transition to market economies. The system has been welcomed and unanimously approved by the Statistical Commission of the United Nations. should be used as the international standard for the compilation of national accounts statistics in order to promote the integration of economic and related statistics. (SNA 1993, Foreword)

Term	Target definition
Target population	The *target population* is the set of elemnts about which information is wanted and estimates are required, although practical considerations may dictate that some units are excluded. (*Statistics Canada Quality Guidelines*, 3rd edition, October 1998, page 16)
Taxonomy	*Taxonomy* refers to classification according to presumed natural relationships among types and their subtypes. (ISO/IEC 11179, Part 1, Framework for the specification and standardization of data elements, 1998)
Term - UN	A *term* is a word or phrase used to designate a concept (*Terminology on Statistical Metadata*, Conference of European Statisticians Statistical Standards and Studies, No. 53, UNECE, Geneva 2000)
Term - ISO	A *term* is a designation of a defined concept in a special language by a linguistic expression. [ISO/IEC 1087] (ISO/IEC 11179, Part 1, Framework for the specification and standardization of data elements, 1998)
Terminology	*Terminology* is a set of terms (*Terminology on Statistical Metadata*, Conference of European Statisticians Statistical Standards and Studies, No. 53, UNECE, Geneva 2000)
Thesaurus	A *thesaurus* is a controlled vocabulary arranged in a given order in which relationships among terms are displayed and identified. (ISO/IEC 11179, Part 1, Framework for the specification and standardizat ion of data elements, 1998)
Timeliness	*Timeliness* of information reflects the length of time between its availability and the event or phenomenon it describes, but considered in the context of the time period that permits the information to be of value and still acted upon. It is typically involved in a trade-off with reliability. (*Statistics Canada Quality Guidelines*, 3rd edition, October 1998, page 5)
Time of recording	The *time of recording* for transactions and, hence, for holdings is governed by the principle of accrual accounting. For financial claims and liabilities, changes of ownership are considered to have taken place at (or be proxied by) the time that the parties to the transactions record them in their books or accounts. If no precise date can be fixed, the reporter may use the date on which the creditor received payment or the date on which some other financial claim was satisfied. For direct investment income data, dividends should be recorded as the date they are payable and income on debt as it is accrued. *(Glossary of Foreign Direct Investment Terms*, OECD, 2001 – not published)
Time series	A *time series* is a time-ordered vector of observations. (*GESMES/CB Data Model*, Release 2, July 2000, ECB, BIS, Eurostat et al)
Trend	In time series analysis, a given time series can be decomposed into: - a cyclical component - a trend component - a seasonal component - an irregular component The method of *trend* estimation adopted by the OECD in the compilation of the composite leading indicator is a modified version of the phase-average trend (PAT) method developed by the United States NBER. (*OECD Leading Indicator Website*, Glossary, 2001. Available at http://www.oecd.org/std/cli/cliglo ss.htm)
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) See also *Seasonal adjustment*
Type of relationship	*Type of relationship* is an expression that characterizes the relationship between the data element and related data. (ISO/IEC 11179, Part 1, Framework for the specification and standardization of data elements, 1998)
Unit of measure	A *Unit of measure* is a system of measurement. (ISO/IEC FCD 11179-3, Registry Metamodel, Final Committee Draft 2001)
Unit of measure precision	*Unit of measure precision* is the degree of specificity for a unit of measure (ISO/IEC FCD 11179-3, Registry Metamodel, Final Committee Draft 2001)
Units, statistical	See *Statistical units*
Value domain	A *value domain* is a set of permissible values. (ISO/IEC 11179, Part 1, Framework for the specification and standardization of data elements, 1998)
Version	A *version* is the identification of an issue of a data element specification in a series of evolving data element specifications within a registration authority. (ISO/IEC 11179, Part 1, Framework for the specification and standardization of data elements, 1998)
Version identifier	A *version identifier* is an identifier assigned to a version under which a data element registration is submitted or updated. (ISO/IEC 11179, Part 1, Framework for the specification and standardization of data elements, 1998)