**UNITED NATIONS STATISTICAL COMMISSION and
ECONOMIC COMMISSION FOR EUROPE
CONFERENCE OF EUROPEAN STATISTICIANS**

**EUROPEAN COMMISSION
STATISTICAL OFFICE OF THE
EUROPEAN COMMUNITIES (EUROSTAT)**

**ORGANISATION FOR ECONOMIC COOPERATION
AND DEVELOPMENT (OECD)
STATISTICS DIRECTORATE**

Joint UNECE/EUROSTAT/OECD Work Session on Statistical Metadata (METIS 2008)
Luxembourg, 9 – 11 April 2008

TERMS RELATED TO OFFICIAL STATISTICS

Proposals and comments by the UNECE for the 2008 public review
of the Metadata Common Vocabulary

INTRODUCTION

1.   The UNECE is since 1992 the custodian of the Fundamental Principles of
Official Statistics. The purpose of this note is to provide an input to the development
of Metadata Common Vocabulary (MCV) focusing on some terms relating to the core
of the official statistics.

2.   Owing to the importance of statistical metadata in managing statistical
activities and the role that metadata have for comparability of data released by
different producers of official statistics, as well as different countries, it is very
important to use commonly agreed terms and concepts. In this respect the MCV
exceeds the original scope of the Statistical Data and Metadata Exchange imitative
(SDMX) that provided a standardization framework used for development of the
MCV. Therefore, the authors of this note appreciate that the MCV, together with
other content oriented guidelines, was put up for a public review.

3.   The following sections comprise our comments and proposals to the MCV.

NATIONAL AND INTERNATIONAL DEFINITIONS

4.   Authors of the MCV made a significant progress towards identifying common
definitions of individual terms, but there are still examples[1] showing that this is not an
easy task. It is not easy to find common grounds between different international
organizations. It may be even more difficult to find a common definition applicable
to national and international statistical institutions.

5.   In justified cases, it may be good to provide "national" and "international"
definitions:
- When the term depends on legislative or substantive context that differs by
  nature between the national and international levels, for example: official
  statistics, data collection, etc.

---

[1] See for example the definition of quality

- When the term is related only to the national or international context, and it is not applicable to the other one. For example:
    o Statistical registers (population, business, etc.) are generally maintained by national statistical offices, and access to administrative records for microdata collection is typical for national level;
    o Data sharing is the foundational idea behind the SDMX and refers to sharing of data among international organizations and decrease of burden on national statistical offices.

6.   In majority of cases the national and international definitions will be identical.

STATISTICAL SURVEY

7.   The MCV defines the term survey in a broad sense, covering statistical activities that use questionnaire or similar data collection exclusively for statistical purposes, as well as data stored in administrative registers and records that were originally collected for non-statistical purposes.

8.   In reality some countries have the same broad understanding or 'survey' as the MCV, but other countries use 'survey' only for the process of primary data collection from respondents to gather unit-level information for exclusively statistical purposes. This issue was discussed at the July 2007 Workshop on Statistical Metadata in Vienna, and the participants proposed the following:

> *"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. Included are primary sources, such as censuses and sample surveys, and other sources, such as the collection of data from administrative records. Derived statistical activities are also included."*

9.   In contrary to the above, and taking into account further discussion within the METIS Steering Group and our own views, we propose the following:
- Define the survey in a narrower sense, that is:
    o *"A systematic process of gathering unit-level information for exclusively statistical purposes by questioning responding units about themselves. The whole data flow from the respondent unit to the statistical producer is controlled by the statistical producer Included are:*
        > *(i)   a census, which attempts to collect data from all members of a population;*
        > *(ii)   a sample survey, in which data are collected from a (usually random) sample of population members";*
- Define the term 'data collection' in a way that it both survey data collection as well as collection of data from administrative sources:
    o *A systematic process of gathering data carried out by a producer of official statistics.*
- Introduce a term 'statistical activity':
    o *"Activity oriented towards a collection, storage, transformation and distribution of statistical information".*

10.   The latter also corresponds to the spirit of the Cross-Domain Concept by the SDMX. A partial discrepancy between the Cross-Domain Concepts and the MCV is a

mirror of the diversity of views, and can be explained by the fact that all documents are still in a dynamic development.

CLASSIFICATION OF METADATA

11.   Everyone intuitively understands that there are different kinds of metadata. Each metadata experts agrees that the structural metadata appearing in a stub and header of a statistical table are not all we need, and there are other very important types of metadata. However, when it comes to describing those other metadata classes, the views differ.

12.   Participants at the 2007 Workshop set this as an important question for the 2008 METIS work session. Working Paper No. 7 prepared by Bo Sundgren will be discussed at METIS 2008. The abstract available at the time of preparing this note suggested to look into metadata classes from four different aspects:
- Who needs the metadata, and for which purposes (e.g. exploratory and explanatory);
- What the metadata inform about: metadata objects (attachment objects) and metadata variables;
- How the metadata are structured and formalised (or not);
- Where the metadata come from (source processes) and where they go (use processes).

13.   In our opinion whatever classifications are agreed, we hope that the classification adopted will be simple. Its main purpose is to have a common language (terminology) for describing metadata and metadata systems.

Where metadata come from (source processes) and where they go (use processes)

14.   One of the possible classifications comes from the conclusion of the 2007 Workshop:
- (i)   Survey Metadata
- (ii)   Definitional Metadata
- (iii)   Methodological Metadata
- (iv)   System Metadata
- (v)   Operational Metadata
- (vi)   Quality metadata

What metadata inform about – objects to which metadata can be attached

15.   The following is our suggestion:
- System as a whole
- Individual producer
- Statistical activity
  - Activity as such
  - Instance
- Microdata for one instance
  - Set
  - Characteristics
  - Unit of group of units
  - Item

- Results from a statistical activity (can be from more than one instance)
  - Set
  - Aggregates/indicators (conceptual metadata)
  - Dimensions (structural metadata)
  - Individual items of a dimension (including time)
  - Individual cells
- Classifications
- Data transfers
  - Regular transfers of similar data sets (metadata flow definitions)
  - Transfer instance / transfer at a specific date (administrative metadata)

QUALITY

16.   Quality information is important for the internal management of statistical activities within statistical offices, as well as for internal and external users of statistics. Therefore, communicating quality is one of the important roles of metadata. The present version of the MCV comprises multiple definitions of quality, based on quality frameworks created by different international organizations.

17.   We recommend to us the ISO 9000/2005 definition:
        The degree to which a set of inherent characteristics fulfils requirements

18.   There can be also identified a common core in different of quality characteristics provided within the MCV. We tried to identify this core, and attempted to unify the definitions. We recommend to introduce a term (or a set of terms) called 'components of quality'. These may be further described through:
- Objects to which quality attributes can be attached; and
- Quality attributes.

Objects to which quality attributes can be attached

- System as a whole
- Individual producer
- Statistical activity
  - Single instance
  - Accumulated instances
- Microdata for one instance
  - Set
  - Single characteristics
  - Specific groups of units
- Results for a statistical activity
  - Single instance
  - Accumulated instances

Quality attributes
- Relevance
- Accuracy
- Timeliness and punctuality
- Accessibility and clarity of information
- Transparency
- (Interpretability)

- Comparability
- Coherence
- (Methodological soundness)
- Integrity / Credibility
- Serviceability

Relationship between objects and the attributes

19.   Individual attributes may be attached to some, but not all of the objects listed above. The following tables attempts to describe the possible combinations:

	Relevance	Accuracy	Timeliness and punctuality	Accessibility	Transparency	Comparability	Coherence	Integrity / Credibility	Serviceability
System as a whole	x				x		x	x	x
Individual producer	x				x		x	x	x
Statistical activity - single instance		x	x		x		x	x	
Statistical activity – accumulated instances	x	x	x		x		x	x	
Microdata set		x	x	x	x	x	x		
Microdata single characteristic		x	x		x				
Microdata – specific group of units		x	x		x				
Results – single instance		x	x	x	x	x			
Results – accumulated instances	x	x	x	x	x	x			

STATISTICAL BUSINESS PROCESS CYCLE

20.   This was the main subject of the 2007 METIS Workshop. The draft Generic Model for the Statistical Business Process Cycle will be discussed at METIS 2008 as Working Paper No. 17. The paper provides a three level model that further explain the nine processes above. The MCV should accommodate the terminological aspect of the finally agreed model that is definitions of processes and subprocesses and related definitions indirectly invoked by the model.

SUGGESTED RE-DEFINED TERMS

21.   We propose amending definitions of the following terms and will provide comments to the SDMX sponsors. Issues related to 'Quality' and 'Survey' are explained above in more detail.
- Accessibility
- Aggregation
- Data capture
- Data editing
- Longitudinal data (Longitudinal microdata set)
- Quality
- Questionnaire
- Survey

ADDITIONAL TERMS

22.   We also suggest inclusion of additions terms. At the same time we know that the list will grow even further as an outcome of the ongoing activities and further development. The terms proposed by us are:

- Access
- Access to microdata
- Administrative record
- Administrative register
- Data from administrative sources
- Data collection from admin. Sources
- Administrative source   Administrative source of statistical data
- Aggregated information
- Anonymised microdata
- Automatic registration and measurement devices
- Basic statistical unit
- Benchmark data
- Break → Time series break; or Break in time series
- Components of quality
- Composite statistical units
- Country
- Data imputation
- Data confidentiality
- Data sharing
- Dissemination
- Edit (editing rule/criteria)
- Global system of official statistics
- Grossing up
- Indicator (statistical indicator)
- International system of official statistics
- Longitudinal Microdata sets
- Macrodata set (Macrodata)
- Matching
- Measurement unit
- Microdata set (Microdata)
- National system of official statistics
- Official statistics
- Owner of administrative records
- Population
- Producer of official statistics
- Register based census
- Release
- Rolling census
- Special aggregation service
- Statistical Classification
- Statistical Nomenclature
- Statistics
- Statistical Office
- Statistical (survey) cycle

- Statistical activity
- Statistical information
- Statistical product
- Statistical programme
- Statistical Register
- Support activities
- Statistical business process cycle (working title)'
- Survey instance
- → Instance
- Transformation
- Unit level data
- User
- Value

* * * * *