

UNSD-DFID PROJECT ON SDG MONITORING

MODULE 4

Using common terminology to
describe your data



Module structure

- **What is common terminology**
- **Example of glossaries/tools for common terminology**
- **Terminology management**
- **Exercise**



What is common terminology?



What is common terminology?

- It means everyone understands the same meaning when a term is used.
- Often people use different terminologies to describe the same thing, or they use the same terminology to describe different things. This leads to ambiguity.
- Example: Significant
- **Precise terms need to be used when describing your data (creating metadata) to avoid confusion**

Can you think of any other examples?



Why should we use common terminology?

- Inconsistent terminology complicates the reporting and presentation of standards
- These inconsistencies can occur by country (and international agencies); in agencies within the same country; and even parts of the same organization.
- Inconsistent labels for the same concept frequently leads to misunderstanding by users and the risk of inappropriate use of statistics.

With your partner, think of a benefit of using common terminology





Terminology management



Metadata management systems

- Many countries link and combine all their metadata into one system.
- The goal is to harmonize use of different metadata across the NSS and within the NSO – An example: unemployment rate
- Another goal is to have **one repository** for all metadata
- SDG indicators should ideally be part of the broader metadata management system



The role of terminology management

- Terminology management is an essential element of a **metadata management system**. It includes the following components:
 - **Formulation of a common set of terminology**, names and descriptions for standards metadata elements. It entails development of metadata terminology across all processes in the statistical life-cycle.
 - Development of a set of unified concepts **requires inputs from various parts of a statistical organizations** and entities of the NSS.
 - Formulation of **processes and procedures for on-going maintenance and updating** and rules.
 - Adapt international or develop own **tools** such as systems for the storage and retrieval of the metadata

Which of these elements are present in your NSS?



Recommended practices for terminology management

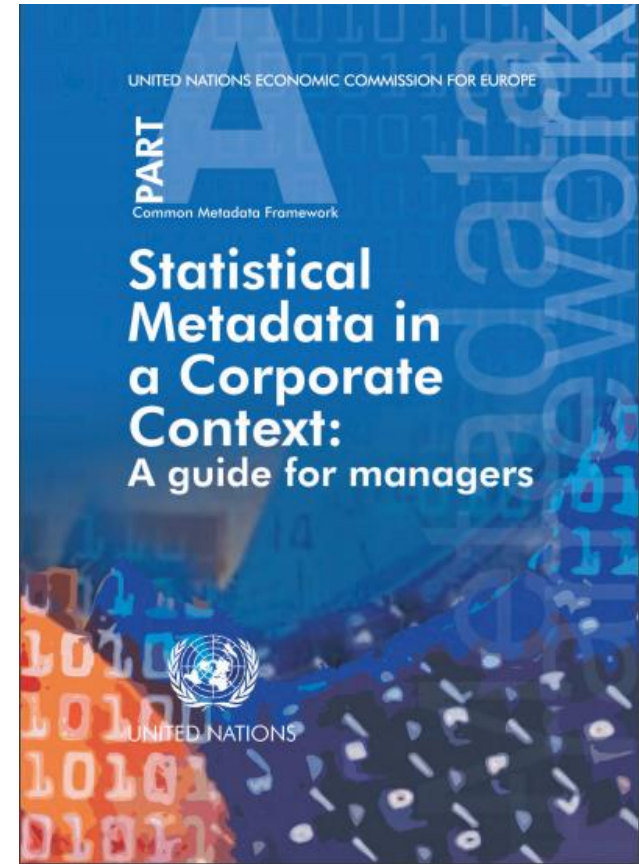
- Establish a terminology management strategy and associated structures. **Use common terminology in various questionnaires and disseminated outputs.**
- Where possible, definitions applied at the national levels should be **aligned** with international definitions, modified as appropriate to meet national circumstances.
- Provide cross references or links
- Adopt a set of common metadata items

Recommended practices for terminology management (cont.)

- **Glossaries should contain the following** to facilitate their interoperability:
 - concept label,
 - definition,
 - detailed source information about where the definition was derived;
 - related terms, and
 - context field providing additional information..
- Glossary should be **translated** to various languages used in the country and by the institution in question.

Tools and guidance

- The UN Economic Commission for Europe has coordinated the development of tools and guidances for metadata management.
- A few useful websites/tools to consider when establishing a metadata management system:
 - <https://www.unece.org/stats/mos/stand.html>
 - http://www.unece.org/fileadmin/DAM/stats/publications/CMF_PartA.pdf





Examples of international glossaries/tools for common definitions

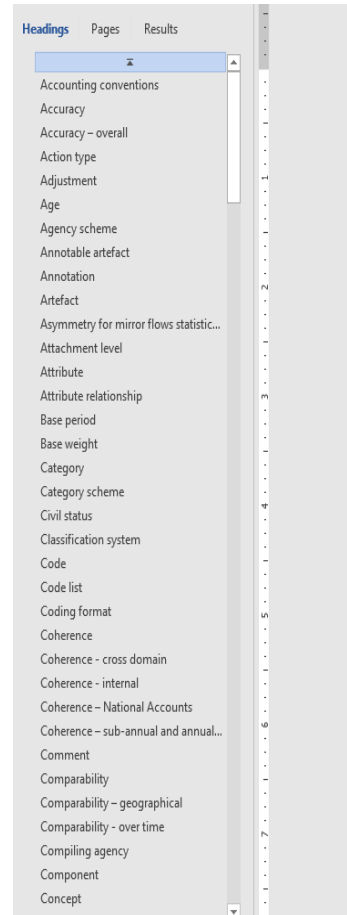


SDMX Glossary

https://sdmx.org/?sdmx_news=new-sdmx-glossary-available

Contains **concepts and related definitions** used in structural and reference metadata of international organisations and national data-producing agencies.

It recommends a **common terminology** that should be used in order to facilitate communication and understanding.



SDMX GUIDELINES

SDMX GLOSSARY

VERSION 1.0

**Please note that Version 1.0
was replaced by Version 2.0
in November 2018**



OECD Glossary of Statistical Terms



GLOSSARY OF STATISTICAL TERMS

STATISTICS PORTAL



- [Glossary Home](#)
- [About](#)
- [Contact Us](#)
- [Downloadable Version](#)
- [Advanced Filter](#)
- [Web Service](#)
- [OECD Statistics](#)

Choose an option to look up statistical terms and definitions

SEARCH FOR A DEFINITION:

ALL WORDS ANY WORDS

SEARCH

ADVANCED FILTER

FOR COMPLEX SEARCH CRITERIA

BROWSE ALPHABETICALLY:

[A](#) [B](#) [C](#) [D](#) [E](#) [F](#) [G](#) [H](#) [I](#) [J](#) [K](#) [L](#) [M](#)

[N](#) [O](#) [P](#) [Q](#) [R](#) [S](#) [T](#) [U](#) [V](#) [W](#) [X](#) [Y](#) [Z](#)



WHO Indicator and Measurement Registry

https://www.who.int/gho/indicator_registry/en/

Global Health Observatory (GHO) data

[Global Health Observatory data](#)

[Data repository](#)

[Reports](#)

[Country statistics](#)

[Map gallery](#)

[Standards](#)

Standards

WHO indicator registry

The WHO Indicator Metadata Registry (IMR) is a central source of metadata of health-related indicators used by WHO and other organizations. It includes indicator definitions, data sources, methods of estimation and other information that allow users to get better understanding of their indicators of interest. It facilitates complete and well-structured indicator metadata, harmonization and management of indicator definitions and code lists, internet access to indicator definitions, and consistency with other statistical domains. It promotes interoperability through the SDMX-HD indicator exchange format and allows incorporation of appropriate international standards such as SDMX MCV (Metadata Common Vocabulary), ISO 11179 (Metadata Registry), DDI (Data Documentation Initiative) and DCMES (Dublin Core).

– [Access the registry](#)

Core health indicators

Standard set of 100 indicators prioritized by the global community to provide concise information on the health situation and trends, including responses at national and global levels.

– [Global Reference List of 100 Core Health Indicators 2015](#)

Contact us

Please [send us](#) your comment c
by e-mail.

Global Health Observatory (GHO) data > Standards



UNICEF briefing notes on SDG indicators

<https://data.unicef.org/resources/sdg-global-indicators-related-to-children/>

 Guidance

Briefing notes on SDG global indicators related to children

April 2018

This is a series of briefing notes for UNICEF regional and country offices on SDG indicators. The first note summarises the development and implementation of the SDG global indicator framework and UNICEF's role in supporting member states to collect, analyse and report on child-related SDG indicators at national and global levels. Briefing notes 2-15 provide detailed information on child-related global SDG indicators for which UNICEF has been identified as custodian, co-custodian, or supporting agency for the purposes of global reporting. Additional briefing notes may be developed in the future covering other global indicators related to children and cross-cutting issues related to SDG monitoring.

[DOWNLOAD BRIEFING NOTES](#)

1. NATIONAL AND GLOBAL MONITORING OF CHILD-RELATED SDG INDICATORS
2. CHILD POVERTY
3. NUTRITIONAL STATUS
4. MATERNAL MORTALITY & SKILLED ATTENDANT AT BIRTH
5. CHILD MORTALITY
6. UNIVERSAL HEALTH COVERAGE
7. LEARNING
8. EARLY CHILDHOOD DEVELOPMENT
9. VIOLENCE AGAINST GIRLS AND WOMEN
10. HARMFUL PRACTICES
11. DRINKING WATER
12. SANITATION & HYGIENE
13. CHILD LABOUR
14. ABUSE, EXPLOITATION & VIOLENCE
14. BIRTH REGISTRATION



FAO Portal on SDG indicators

<http://www.fao.org/sustainable-development-goals/indicators/en/>

Key information and statistics

Indicators under FAO custodianship



2.1.1 Hunger

2.1.2 Severity of food insecurity

2.3.1 Productivity of small-scale food producers

2.3.2 Income of small-scale food producer

2.4.1 Agricultural sustainability

2.5.1 Conservation of genetic resources for food and agriculture

2.5.2 Risk status of livestock breeds

2.a.1 Public Investment in agriculture

2.c.1 Food price volatility



5.a.1 Women's ownership of agricultural land

5.a.2 Women's equal rights to land ownership



14.4.1 Fish stocks sustainability

14.6.1 Illegal, unreported unregulated fishing

14.7.1 Value added of sustainable fisheries

14.b.1 Access rights for small-scale fisheries



6.4.1 Water use efficiency

6.4.2 Water stress



15.1.1 Forest area

15.2.1 Sustainable forest management

15.4.2 Mountain Green Cover



IMF Dissemination Standards Bulletin Board (DSSB)



INTERNATIONAL
MONETARY FUND

Search IMF



Send us your Feedback

HOME ABOUT THE IMF RESEARCH COUNTRIES CAPACITY DEVELOPMENT NEWS VIDEOS DATA PUBLICATIONS

DISSEMINATION STANDARDS BULLETIN BOARD (DSBB)

DSBB Home

Country NSDP

SDDS Plus

SDDS

e

2. Methodology

2.1 Concepts and definitions

2.1.1 Concepts and definitions

Annual central government data comprises the operations of the budgetary central government. Data presentation is broadly in line with the analytical framework of the *Government Finance Statistics Manual of 1986 (GFSM 1986)*. Annual data correspond to the sum of monthly and quarterly data that correspond to the Final Accounts for each fiscal year. However, the audited Final Accounts for 2000 through 2003 have not been published.

2.2 Scope

2.2.1 Scope

- **Scope of the data**

The budgetary central government operations cover the national territory and outside the national territory for diplomatic offices located outside the country. The Ministry of Local Governments and Rural Development includes transfers to districts and municipalities. However, own revenue raised by these units (local fees for services) and related expenditure is not included in the budgetary central government operations.

Fiscal data correspond to the operations of the budgetary central government on a cash basis (revenues are recorded when cash is received or checks deposits are made, expenditures are recorded when checks are issued) and are compiled by using administrative and accounting records of the Controller and Accountant General's Department (CAGD), revenue agencies, MOFED and BOG. Data on revenue correspond to deposits in the BOG, expenditure data correspond to checks issued by the CAGD, domestic financing data are compiled by the BOG, and data on foreign grants and loans are compiled by the BOG and the Aid and Debt Management Unit (ADMU) of the MOF.

Data cover the operations of the budgetary central government. The Social Security and National Insurance Trust (SSNIT) that is apparently a central government social security fund is treated as a unit of the public financial corporation subsector. This is a clear deviation from the *GFSM 1986* methodology. Data cover the operations of districts and municipalities budgeted through the Ministry of Local Government and Rural Development. The transfers appropriated to several statutory funds are included in the budget as those of the Educational Trust Fund, Road Fund, Petroleum related Fund, and District Assembly Common Fund.

Transaction coverage appears to be incomplete in data compiled by the CAGD and BOG as some important revenue and expenditure are missed. Examples are a greater part of internally generated revenue...



SUSTAINABLE DEVELOPMENT GOALS

Exercise

What standard glossaries are used for data collection and data description in your organization?

Give information on any customization of these glossaries by your organization

UNSD-DFID PROJECT ON SDG MONITORING

MODULE 4.1

Most frequently used terminology



Module structure

- Commonly used terms
 - Indicator
 - Time Series
 - Observation value
 - Frequency
 - Reference area
 - Unit of measure
 - Average, rate, ratio, percentage
 - Source; type of data sources (census, survey, admin data)



Indicator

- A statistical measure of a quantitative characteristic of an individual or collective phenomenon. (Adapted definition)
- First building block of any statistical series, i.e. what to measure?
- Highest level, in the metadata hierarchy
- Issues around indicators relate to whether they actually measure the phenomenon they are expected to measure

Exercise: Give examples of challenges around the concept of indicators in your organization

Time series

- A set of ordered observations on a quantitative characteristic of an individual or collective phenomenon taken at different points of time.
- Middle level, in the metadata attachment hierarchy (i.e. composed of observation values).

Observation value

- The value of a particular variable at a particular period.
- The observation value is the field which holds the data.
- Lowest level, in the metadata attachment hierarchy.

Metadata attachment hierarchy

Indicator	Time series	Observation value
Literacy rate	<ul style="list-style-type: none">Percent, Male, 15 – 24 yr oldPercent, Female, 15 – 24 yr oldPercent, Both sexes, 15 – 24 yr old	<ul style="list-style-type: none">1990,1991,1992,1993,...
Population using improved sanitation facilities	<ul style="list-style-type: none">Percent, Both sexes, All ages, UrbanPercent, Both sexes, All ages, RuralPercent, Both sexes, All ages, Total (National)	<ul style="list-style-type: none">2000,2005,2010,...
Malaria deaths	<ul style="list-style-type: none">per 100,000 population , Both sexes, Under 5yrsper 100,000 population, Both sexes, All ages	<ul style="list-style-type: none">1994,1995,1996,1997,...



Frequency

- The time interval at which observations occur over a given time period.
- If constant time interval between its observations, this determines the frequency.
- "Frequency" - also called "periodicity" - may refer to several stages in the production process.
- Could mean "Frequency - data collection" or "Frequency – data dissemination"

Frequency - Issues with time period

- Often observation value are presented for single calendar years, even when they are derived from multiple year sources.
- This makes comparison difficult when different methodologies are applied to do this.
- Time period should be based on period when the data is collected (the reference date of the data) and not when the report is published or the data finally made available (the production date).
- Need to find a standard approach to how the time period is described consistently, especially for multi-year periods.

Reference area

- This is the country or geographic area to which the measured statistical phenomenon relates.
- Refers to country, geographical or political group of countries or regions within a country.
- Subject to a variety of hierarchies, as countries comprise territorial entities.
- And for which statistical data are produced internationally on a separate and independent basis.

Frequency – challenges with Time period

- Descriptions of data time periods must be based on policies
 - e.g. to describe all waves of a panel survey with one metadata record/ create a record for each wave/year

Exercise: Give examples of challenges faced by your organization when collating or describing data sources over different time periods



Units of measure

- The unit in which the data values are measured.
- Quantity by which something is counted or described, such as kg, mm, simple number counts or index numbers.
- Some types (e.g. currency, weight) are defined in connection with the unit multiplier, provides the scale, power or magnitude for the value of the variable.



Average

- This is calculated by adding up all the data and dividing by the number of pieces of data.
- So if the hourly rate of pay for 5 employees was as follows:
500, 550, 600, 650, 700 dollars
- The average hourly rate of pay per employee:
$$\frac{(500+ 550+600+650+700)}{5} = 600 \text{ dollars}$$
- Affected by unusually high or low values in the dataset, may result in a figure not typical.

Proportions

- A proportion, or share, is one number divided by another
- Percentages, rates and ratios are different types of proportions
- They are also units of measurement

Percentages

$$\frac{3}{4} = \frac{75}{100} = 75\%$$

"per cent"



Numerator has to be part of the denominator

- Percent means “per 100” (per cent) → Computing by hundredths
- When calculating percentages, the numerator needs to be part of the denominator – a percentage shows the relation between a share and the total.

- Example: Percentage of children born at home:

$$= \frac{\text{number of children born at home} \times 100}{\text{total number of births}}$$

- Percentages are sometimes computed for an array of subcategories using the total number as the base denominator. In this case, a percentage distribution can be obtained, and all percentages should add up to 100 percent. This type of percentage can be used as a summary statistic.

District	Share of total registration
A	22%
B	11%
C	9%
D	34%
E	24%
Total	100%



Rates and ratios

Ratio

$$\frac{\text{C miles}}{\text{D miles}}$$

Same units

“Completeness ratio”

$$\frac{\text{Registered events}}{\text{Total events}}$$

Rate

$$\frac{\text{X miles}}{\text{Y hour}}$$

Different units

Measure of frequency

$$\frac{\text{Event}}{\text{Time}}$$



Ratio

- A ratio is a single number that expresses the relative size of two other numbers, expressed as a number (i.e. can be > 1)

$$\frac{X}{Y} = \text{ratio of } X \text{ to } Y$$

- They are used widely in SDGs for analysing the composition of a set of events, often calculated for subgroups of a population.
- Examples include: gender parity ratio; ratio of average female wage to male wage; primary school enrolment ratio; etc.

Exercise

- Provide an example of
 - A percentage
 - A rate
 - A ratio

Source

- A specific data set, metadata set, database or metadata repository from where data or metadata are available. (MCV 2009)
- Source is often used as a synonym for "data provider". However, data provider is the organisation or individual from where statistics are obtained.
- The term "source" refers to the origins, characteristics and components of the raw data used for compiling statistical aggregates, e.g. administrative database, census, sample survey.

Source

- Sources can be distinguished according to the following classification:
 - Statistical source - data collected exclusively for statistical purposes
 - Non-statistical source - data not primarily collected for statistical purposes.
 - Mixed source - data from a combination.

Administrative data

- Administrative systems are there to manage a process. For example, records are created as measles immunizations are performed.
- These records can also be used to compute statistical indicators.
- Most administrative systems update data on a 'live' basis and dissemination occurs at regular intervals, but usually with a time delay.
- There may be biases in who is covered in these data sources and there may be inconsistent use of definitions. Such shortcomings should be specified in the metadata.



Census data

- A census is a vehicle for collecting data from every member of a given population.
- Time consuming and expensive, information is usually collected on a limited range of issues.
- Prior to carrying out a census, the whole population must be identified and a set of questions developed.
- Carried out in most countries once every 10 years and the first results are released 1 or 2 years after data collection.

Survey data

- A survey involves identifying and collecting data from a randomly selected subset (sample) of the population under investigation.
- The sample is usually drawn from a national sampling frame developed from the latest census.
- Unlike administrative systems, the primary purpose is to collect data. Cheaper and convenient alternative to the census.
- Despite sampling error, if properly designed and carried out can be less biased and produce better estimates than administrative data or a census.

Household survey

- Household surveys are multi-purpose surveys to collect information from people living in private households.
- The most common household surveys are described below:
 - Multi Indicator Cluster Surveys (MICS)
 - Demographic and Health Surveys (DHS)
 - Living Standard Measurement Surveys (LSMS)
 - Core Welfare Indicator Questionnaires (CWIQ)
 - Household Budget / Income and Expenditure Surveys
 - Labour Force Surveys (LFS)



Key features of different sources

CHARACTERISTIC	ADMIN	SURVEY	CENSUS
<i>INCLUSION CRITERION</i>	All events registered	All sampled units	All units (i.e. 100% coverage of the population)
<i>BIAS</i>	They can be biased if incomplete	Usually there is bias, but if well designed bias is minimal	Theoretically there is no bias, but lack of coverage may lead to it
<i>COST</i>	Low	Medium	High
<i>FREQUENCY</i>	Ongoing	3-5 years	10 years
<i>DISSEMINATION – YEARS AFTER COLLECTION</i>	1-2 years	1 year	1-2 year
<i>POTENTIAL FOR POLICY MAKING AND ADVOCACY</i>	Very good but limited by poor coverage	Good, but not in the short run	Good, but not in the short run



Footnotes

- A note or other text located at the bottom of a page of text, manuscript, book or statistical tabulation that provides comment on or cites a reference for a designated part of the text or table. (MCV 2009)
- Denoted by a number, mark, etc, in the main body of the text.
- Related to but of lesser importance than the larger work in the main body of the text/ table.

Footnotes

- Footnotes could be used to :
 - Explaining a concept
 - specify challenges with comparability
 - specify challenges with accuracy
 - References/relevant links
 - Symbols or abbreviations

All of these are also part of metadata

Footnotes

Contributing family worker population age 15-64, by level of education and sex, Country A, 2010

Contributing Family Worker

	Male	Female
Total	28,899	18,953
No education / not stated	14,352	11,396
	49.7%	60.1%

Contributing family worker: Persons who were working without pay in the business or farm of another household/family member.

Source: Country A, Population and Housing Census, 2010





Thank you for your attention

