Guidelines for the Global Metadata Structure Definition for Sustainable Development Goals Indicators

MSD Version: 1.0

Guidelines version: 1.2

Date: Nov 2023

Contents

I. Introduction

In September 2015, United Nations Member States adopted the 2030 Agenda for Sustainable Development and tasked the United Nations Statistical Commission as a functional commission of the UN Economic and Social Council to develop the global indicator framework. "Data which is high-quality, accessible, timely, reliable and disaggregated by income, sex, age, race, ethnicity, migration status, disability and geographic location and other characteristics relevant in national contexts" is called for (A/RES/70/1).

In March 2015 at its forty-sixth session, the United Nations Statistical Commission created an Inter-agency and Expert Group on SDG Indicators (IAEG-SDGs), which is composed of representatives from a regionally-balanced group of Member States and includes regional and international agencies as well as other key stakeholders, such as civil society, academia and the private sector, as observers. The IAEG-SDGs was tasked with providing a proposal for a global indicator framework (and associated global and universal indicators) for the follow up and review of the 2030 Agenda to be considered by the Statistical Commission at its forty-seventh session in March 2016. At the forty-seventh session of the Commission, the Global indicator framework was agreed upon by the Member States.

To ensure timely and efficient collection, validation, and dissemination of SDG indicators, a data exchange format needs to be agreed upon and used by SDG data providers. This enables the automation of data exchange while simplifying and improving data validation and dissemination.

Statistical Data and Metadata Exchange (SDMX) is a standard sponsored by seven international organizations (BIS, ECB, Eurostat, OECD, IMF, UN, and WB). It was endorsed by the United Nations Statistical Commission in 2008 as a preferred standard for data exchange, and was approved as an ISO standard (ISO/IS 17369:2013). It has been successfully used for data exchange and dissemination in areas such as Macro-Economic Statistics, International Merchandise Trade, and others including MDG indicators.

To facilitate the development of SDMX-based data and metadata exchange formats for SDG Indicators, IAEG-SDGs approved the Terms of Reference of the Working Group on SDMX for SDGs Indicators ("SDMX-SDGs Working Group"). The overall objective of the Working Group is to design a solution for the exchange and dissemination of SDG Indicators based on SDMX standard.

In June 2019, the SDMX-SDGs Working Group published the first official global Data Structure Definition for SDG indicators ("SDG DSD"). The SDG DSD is available at the SDMX-SDGs Working Group page as well as the SDMX Global Registry. The DSD has been regularly updated since, and the global SDG Database published at the UNSD SDMX API. The DSD is actively used for data exchange and dissemination of both global and national SDG datasets.

Subsequent to the release of the SDG DSD, the SDMX-SDGs Working Group embarked on developing a Metadata Structure Definition for the SDG indicators. Following pilot SDG reference metadata exchange with the SDG Custodian Agencies and countries in 2020-2021, and further development of the MSD and SDMX API, the official SDG Metadata Structure Definition was published in March 2022.

II. SDG Metadata Structure Definition

The SDG Metadata Structure Definition (SDG MSD) is based on the SDG metadata template approved by the IAEG-SDGs. The SDG MSD has two principal components:

1. Metadata Concept Scheme (MCS). The MCS is a list of metadata concepts (or "metadata attributes") that describe aspects related to methodology, quality, and others, of the SDG indicators.

2. Metadata Structure, which in turn consists of

    a. Metadata Target. Metadata Target describes the "attachment" of a reference metadata set, i.e. objects or data to which the metadata relates. In the case of the SDGs, the metadata target is a partial key based on the SDG Data Structure Definition.

    b. Report Structure, which defines the structure of a hierarchical metadata report using concepts defined by the MCS.

The components of the MSD are described below.

III. Metadata Concepts

This section describes concepts in the SDG Metadata Concept Scheme as well as how they are organized in the SDG Metadata Report. The SDG Metadata Concepts were aligned, to the extent possible, with the Global Metadata Concept Scheme and SDMX Glossary. See Standardising Reference Metadata Reporting in SDMX and SDMX Glossary for further information.

The SDG Metadata Report has a hierarchical structure. Metadata should be reported at the most detailed level available. When detailed metadata is not available, metadata can be reported at a higher, coarse, level.

The concepts are arranged in 9 sections, numbered 0 to 8. Each section contains one or more concepts related to a specific topic. When there are more than one concepts, these are arranged hierarchically.

In a metadata set, values of the metadata attributes should be reported as HTML wrapped in CDATA sections.

The metadata concepts, their ID's, and arrangement are given below.

0. Indicator information (SDG_INDICATOR_INFO)

Description of the Goal, Target, Indicator and data series - if applicable. The date/period when this metadata file is last updated should be provided. The linkages with any other Goals and Targets under the SDG framework, if any, should be described. The name of international organisation(s) responsible for monitoring this indicator should be provided.

0.a. Goal (SDG_GOAL)

SDG Goal number and name.

0.b. Target (SDG_TARGET)

SDG Target number and name.

0.c. Indicator (SDG_INDICATOR)

SDG Indicator number and name.

0.d. Series (SDG_SERIES_DESCR)

This section should contain codes and descriptions of all series to which the metadata set applies. While metadata sets are attached to each series individually in the database, and series code to which current metadata set is attached can be looked up in the target, this section should contain all series codes to which the metadata set is attached.

0.e. Metadata update (META_LAST_UPDATE)

The date when this metadata report is last updated.

By convention, this attribute should be reported as a date in the format YYYY-MM-DD, e.g. "2022-05-12" for May 12, 2022. Any other format may create issues during metadata exchange

0.f. Related Indicators (SDG_RELATED_INDICATORS)

Linkages with any other Goals and Targets.

0.g. International organisations(s) responsible for global monitoring (SDG_CUSTODIAN_AGENCIES)

Global reporting: International organizations (departments/offices) responsible for monitoring this indicator at the global level. Country reporting: This concept has no national counterpart.

1. Data reporter (CONTACT)

Information of the contact persons and their organization responsible for the reporting of the indicator or time series specified.

1.a. Organisation (CONTACT_ORGANISATION)

Agency responsible for reporting of the indicator or time series specified below.

1.b. Contact person(s) (CONTACT_NAME)

Name(s) of the contact points for the data or metadata.

1.c. Contact Organisation Unit (ORGANISATION_UNIT)

Organisation unit information of the contact points for the data or metadata.

1.d. Contact Person Function (CONTACT_FUNCT)

Functional title(s) of the contact points for the data or metadata.

1.e. Contact Phone (CONTACT_PHONE)

Phone number(s) of the contact points for the data or metadata.

1.f. Contact Mail (CONTACT_MAIL)

Mailing address(es) of the contact points for the data or metadata.

1.g. Contact emails (CONTACT_EMAIL)

E-mail address(es) of the contact points for the data or metadata.

2. Definition, concepts and classifications (IND_DEF_CON_CLASS)

Precise definition of the indicator including references to standards and classifications, preferably relying on international agreed definitions. The indicator definition should be unambiguous and be expressed in universally applicable terms. It must clearly express the unit of measurement (proportion, dollars, number of people, etc.). Precise definition of all different concepts and terms associated with the indicator, also including reference to any associated classifications.

2.a. Definition and Concepts (STAT_CONC_DEF)

Precise definition of the indicator preferably relying on internationally agreed definitions. The indicator definition should be unambiguous and be expressed in universally applicable terms. Precise definition of all different concepts and terms associated with the indicator, also including reference to any associated classifications.

2.b. Unit of Measure (UNIT_MEASURE)

Description of the unit of measurement (proportion, dollars, number of people, etc.)

2.c. Classifications (CLASS_SYSTEM)

Describe references to both national and international standards and classification being used. [Information to be provided where applicable.]

3. Data source type and data collection method (SRC_TYPE_COLL_METHOD)

Description of data sources, data collection methods, and related information such as descriptions of all data collection methods, data collection and release calendar, data providers and compilers, as well as their institutional mandates.

3.a. Data sources (SOURCE_TYPE)

Description of all actual and recommended sources of data. This description should include, when applicable, any changes of the data source over time, details of denominator (if from a different source) and any other relevant information related to the origin of the source or indicator. Similar details should be given for administrative sources.

3.b. Data collection method (COLL_METHOD)

Description of all methods used for data collection. This description should include, when applicable, the sample frame used, the questions used to collect the data, the type of interview, the

dates/duration of fieldwork, the sample size and the response rate. Some additional information on questionnaire design and testing, interviewer training, methods used to monitor non-response etc. should be provided here. Questionnaires used should be annexed (if very long: via hyperlink).

3.c. Data collection calendar (FREQ_COLL)

Dates when source collection is next planned.

3.d. Data release calendar (REL_CAL_POLICY)

Expected dates of release of new data for this indicator, including the year (or, ideally, the quarter/month when the next data point associated with the indicator will become available).

3.e. Data providers (DATA_SOURCE)

Identification of national and/or international data provider(s), specifying the organization(s) responsible for producing the data.

3.f. Data compilers (COMPILING_ORG)

Organization(s) responsible for compilation of on this indicator either at national or global level.

3.g. Institutional Mandate (INST_MANDATE)

Description of the set of rules or other formal set of instructions assigning responsibility as well as the authority to an organisation for the collection, processing, and dissemination of statistics for this indicator.

4. Other methodological considerations (OTHER_METHOD)

Other methodological considerations. These include rationale behind the indicator, comments and limitations, computation methodology at the country, regional, and global level, treatment of missing values if applicable, and information related to quality assurance.

4.a. Rationale (RATIONALE)

Description of the purpose and rationale behind the indicator, as well as examples and guidance on its correct interpretation and meaning.

4.b. Comment and limitations (REC_USE_LIM)

Comments on the feasibility, suitability, relevance and limitations of the indicator. Also includes data comparability issues, presence of wide confidence intervals (such as for maternal mortality ratios); provides further details on additional non-official indicators commonly used together with the indicator.

4.c. Method of computation (DATA_COMP)

Explanation of how the indicator is calculated, including mathematical formulas and descriptive information of computations made on the source data to produce the indicator (including adjustments and weighting). This explanation should also highlight cases in which mixed sources are used or where the calculation has changed over time (i.e., discontinuities in the series).

4.d. Validation (DATA_VALIDATION)

Description of process of monitoring the results of data compilation and ensuring the quality of the statistical results, including consultation process with countries on the national data submitted to

the SDGs Indicators Database. Descriptions and links to all relevant reference materials should be provided.

4.e. Adjustments (ADJUSTMENT)

Global reporting: Description of any adjustments with respect to use of standard classifications and harmonization of breakdowns for age group and other dimensions, or adjustments made for compliance with specific international or national definitions. National reporting: This concept is typically not applicable for national reporting.

4.f. Treatment of missing values (i) at country level and (ii) at regional level. (IMPUTATION)

Global reporting: (National level) Description of the methodology employed for producing estimates for the indicator when country data are not available, including any mathematical formulas and description of additional variables used as input into the estimation process. (Regional level) Description of how missing values for individual countries or areas are imputed or otherwise estimated by international agencies to derive regional or global aggregates of the indicator. National reporting: This concept is not applicable for national reporting.

4.g. Regional aggregations (REG_AGG)

Global reporting: Description of the methodology, including any mathematical formulas, used for the calculation of the regional/global aggregates from the country values. Description of the weighting structure used for aggregating country indicator values to regional and global levels. Additional methodological details on how the data from countries or areas is assembled by custodian international agencies to provide regional and global aggregates. This is distinct from the method of computation, which looks at how the indicator is compiled at a national level. National reporting: This concept is not applicable for national reporting.

4.h. Methods and guidance available to countries for the compilation of the data at the national level (DOC_METHOD)

Global reporting: Description of methodology used by countries for the compilation of data at national level and the relevant international recommendations and guidelines available to countries. Descriptions and links to all relevant reference materials should be provided. National reporting: For national reporting a country may refer to the globally available metadata and explain how it is being used.

4.i. Quality management (QUALITY_MGMNT)

Description of systems and frameworks in place within an organisation to manage the quality of statistical products and processes.

4.j. Quality assurance (QUALITY_ASSURE)

Description of practices and guidelines focusing on quality in general and dealing with quality of statistical programmes at your agency, including measures for ensuring the efficient use of resources.

4.k. Quality assessment (QUALITY_ASSMNT)

Description of overall evaluation of fulfilling quality requirements, based on standard quality criteria.

5. Data availability and disaggregation (COVERAGE)

Global reporting: Indicate for how many countries the data for this indicator are already currently available on a regular basis. Data availability by regional breakdowns and time periods can also be descibed here. Describe the specification of the dimensions and levels used for disaggregation of the indicator (e.g., income, sex, age group, geographic location, disability status, etc.). National reporting: Data availability by sub-national breakdowns and time periods can be descibed here. Describe the specification of the dimensions and levels used for disaggregation of the indicator (e.g., income, sex, age group, geographic location, disability status, etc.).

6. Comparability / Deviation from international standards (COMPARABILITY)

Explanation on the differences between country produced and internationally estimated data on this indicator, highlighting and summarising the main sources of differences.

7. References and Documentation (OTHER_DOC)

Descriptions and links to all relevant reference materials related to this indicator.

8. Translations (TRANSLATIONS)

Additional information on translation of reference metadata set.

IV. Metadata Target

SDMX metadata target describes the reference metadata set's attachment, i.e. the artefact or data that the metadata applies to. Since in this case the reference metadata describe SDG indicators, the target in a metadata set must reference the SDG DSD and dataflows.

The metadata target includes two components: KeyDescriptorValuesTarget and IdentifiableObjectTarget.

- **KeyDescriptorValuesTarget**. This indicates a partial key.

- **IdentifiableObjectTarget**. This indicates that the reference metadata set is linked to a dataflow.

a. KeyDescriptorValuesTarget

In SDG metadata exchange, the partial key is based on a subset of dimensions of the SDG Data Structure Definition. Together, dimension values define the national or global series that the reference metadata set describes. The 3 SDG dimensions used in the SDG metadata exchange are as follows:

- **REPORTING_TYPE**. This dimension is represented by the code list **CL_REPORTING_TYPE** in the SDG DSD, and specifies whether the data is global (reported by the custodian agency, regardless of the origin) or national (reported by a national government). At this time, either code **G** for the global dataset, or code **N** for country-provided data, can be used.

- **SERIES**. This dimension is represented by the code list **CL_SDG_SERIES** in the SDG DSD. Code from the code list must be used to reference the series.

- **REF_AREA**. This dimension is represented by the code list **CL_AREA** in the SDG DSD, and contains reference area codes in the M49 and ISO-3166 alpha2 format. Global metadata sets should normally use code 1 ("World"). National metadata must use reference code of the reporting country. M49 codes are used at the UNSD data and metadata APIs.

For further information on these dimensions, please see the Guidelines for the SDG DSD.

b. IdentifiableObjectTarget

In SDG metadata exchange, this component describes the dataflow to which the reference metadata applies. It should, therefore, reference one of the following two dataflows:

- **DF_SDG_GLC** (SDG Country Global Dataflow). This dataflow is used for reporting country data reported by national statistical authorities. If this dataflow is referenced in a metadata set, REPORTING_TYPE="N" must be used in the KeyDescriptorValueTarget.
- **DF_SDG_GLH** (SDG Harmonized Global Dataflow). This dataflow is used for reporting global, harmonized data reported by SDG custodian agencies. If this dataflow is referenced in a metadata set, REPORTING_TYPE="G" must be used in the KeyDescriptorValueTarget.

For further information on these dimensions, please see the Guidelines for the SDG DSD.

V. Example metadata set encoding

This example shows metadata set for indicator 1.1.1, series "Employed population below international poverty line". The target as well as part of the metadata report.

Metadata Target	
KeyDescriptorValuesTarget	
REPORTING_TYPE	**G** - Global
SERIES	**SI_POV_EMP1** – Employed population below international poverty line [1.1.1])
REF_AREA	**1** (World)
IdentifiableObjectTarget	
Dataflow	**IAEG-SDGs:DF_SDG_GLH(1.7)** – SDG Harmonized Global Dataflow, v1.7
Metadata Report	
SDG_GOAL	Goal 1: End poverty in all its forms everywhere

SDG_TARGET	Target 1.1: By 2030, eradicate extreme poverty for all people everywhere, currently measured as people living on less than $1.25 a day
SDG_INDICATOR	Indicator 1.1.1: Proportion of the population living below the international poverty line by sex, age, employment status and geographic location (urban/rural)
SDG_SERIES_DESCR	SI_POV_EMP1 Employed population below international poverty line [1.1.1]
META_LAST_UPDATE	2022-03-31
SDG_RELATED_INDICATORS	1.3.1, 8.2.1, 8.3.1, 8.5.1, 8.5.2, 10.4.1
SDG_CUSTODIAN_AGENCIES	International Labour Organization (ILO)
CONTACT_ORGANISATION	International Labour Organization (ILO)
STAT_CONC_DEF	**Definition:** The proportion of the employed population below the international poverty line of US$1.90 per day, also referred to as the working poverty rate, is defined as the share of employed persons living in households with per-capita consumption or income that is below the international poverty line of US$1.90. **Concepts:** Employment: All persons of working age who, during a short reference period (one week), were engaged in any activity to produce goods or provide services for pay or profit Poverty Line: Threshold below which individuals in the reference population are considered poor and above which they are considered non-poor. The threshold is generally defined as the per-capita monetary requirements an individual needs to afford the purchase of a basic bundle of goods and services. For the purpose of this indicator, an absolute international poverty line of US$1.90 per day is used Household in poverty: Households are defined as poor if their income or consumption expenditure is below the poverty line taking into account the number of household members and composition (e.g., number of adults and children). Working poor: Employed persons living in households that are classified as poor, that is, that have income or consumption levels below the poverty line used for measurement.

VI. Links and contact information

Further information and materials can be found at the [official SDMX-SDG page](#) hosted by the United Nations Statistics Division.

Please contact **sdmx@un.org** with enquiries regarding the SDG Data Structure Definition and related SDMX artefacts.