



SDG Data Structure Definition

Workshop on Data and Metadata Sharing
Bangkok, 10-14 December 2018

Abdulla Gozalov, UNSD



SDG Data Structure Definition

- Developed by the Working Group on SDMX for SDG Indicators, which was established by the Interagency Expert Group on SDG Indicators (IAEG-SDGs) in April 2016
- Currently available as a draft
- Official release expected in March 2019

SDG DSD (cont'd)

- Single DSD used for all SDG indicators
 - Support for diverse indicators means not all dimensions are applicable in all cases
 - E.g. AGE is not applicable to indicator “Land area covered by forest”
 - Value **_T** (no breakdown) is used when a dimension or attribute is not applicable.



Dimension: Frequency (FREQ)

- “Indicates rate of recurrence at which observations occur (e.g. monthly, yearly, biannually, etc.).”
- By convention, the SDG DSD currently only supports annual frequency.
- Where the frequency is not annual (e.g. two-year average), detail should be provided in the TIME_DETAIL attribute.



Dimension: REPORTING_TYPE

- Used to distinguish between National, Regional, Global Reporting
- Countries to use value **N** (national reporting)
- Regional organizations to use value **R** (regional reporting)
- Custodian agencies to use value **G** (Global reporting)

Dimension: Series (SERIES)

- Used to represent indicators
- A single indicator can have multiple series
 - Not to be confused with SDMX time series
- E.g. *5.5.1 Proportion of seats held by women in (a) national parliaments and (b) local governments* has 4 series:
 - ***SG_GEN_PARL*** *Proportion of seats held by women in national parliaments*
 - ***SG_GEN_PARLN*** *Number of seats held by women in national parliaments*
 - ***SG_GEN_PARLNT*** *Number of seats in national parliaments*
 - ***SG_GEN_LOCG*** *Proportion of seats held by women in local governments*



Dimension: Age (AGE)

- “Age - or age range - of the individuals the observation refers to.”
- Use **_T** where not applicable

Dimension: Sex (SEX)

- Gender condition: male or female. This dimension applies only if data can be disaggregated by sex.
- Use **_T** where not applicable
- For gender indicators must be set to **F** as applicable
 - E.g. for series *Proportion of seats held by women in national parliaments*



Dimension: Reference Area (REF_AREA)

- Country or geographic area to which the measured statistical phenomenon relates
- It is envisaged that countries will report national-level values but may wish to extend the code list with its sub-national areas for dissemination

Dimension: Urban/Rural location (URBANISATION)

- Has 3 codes
 - **_T** (Total)
 - **_U** (Urban)
 - **_R** (Rural)
- Use **_T** where not applicable



Dimension:

INCOME_WEALTH_QUANTILE

- Used for disaggregating the data by income or wealth quintile of the population
- In the future can be extended to cover decile, percentile, etc
- Use **_T** where not applicable

Dimension: Education Level (EDUCATION_LEV)

- “Highest level of an educational programme the person has successfully completed.”
- Supports top categories of ISCED11 and ISCED97, as well as custom SDG codes
- Use **_T** where not applicable



Dimension: OCCUPATION

- “Job or position held by an individual who performs a set of tasks and duties.”
- Supports top categories of ISCO-08, ISCO-98, ISCO-68
- Use **_T** where not applicable



Dimension: Disability Status (DISABILITY STATUS)

- Used to break down SDG indicators by disability
- At the moment, only used to distinguish between persons with a disability, and persons without a disability
- Use **_T** where not applicable



Dimension: Custom Breakdown (CUST_BREAKDOWN)

- Special dimension introduced to facilitate non-standard breakdowns, primarily in national context
- At the moment empty but in the future will be populated with generic codes (e.g. CODE1, CODE2, etc), to which data providers will assign meaning in their own context



Dimension:

COMPOSITE_BREAKDOWN

- Mixed dimension: represents several merged code lists
 - E.g. International Organizations, Product Type, Material Flows, etc
- Used for breakdowns that are only used in 1 or 2 indicators, in order to avoid creating too many dimensions
- Use **_T** where not applicable



Time Dimension: TIME_PERIOD

- The observation corresponds to a specific point in time ... or a period...”
- The convention for SDGs is to always provide a four-digit year in the TIME_PERIOD concept. Further info must be placed in TIME_DETAIL, and structured period information in TIME_COVERAGE.



Primary Measure: Observation value (OBS_VALUE)

- Used to convey the value of a variable at a period of time
- Should be a floating-point number

Attribute: Unit Multiplier (UNIT_MULT)

- “Exponent in base 10 specified so that multiplying the observation numeric values by $10^{\text{UNIT_MULT}}$ gives a value expressed in the unit of measure”
- If the observation value is in millions, unit multiplier is 6; if in billions, 9, and so on. Where the number is simple units, use 0.
- Mandatory attribute



Attribute: Observation Status (OBS_STATUS)

- “Information on the quality of a value or an unusual or missing value”
 - E.g. can be used to indicate a break in series

Attribute: Time Period Details (TIME_DETAIL)

- “When TIME_PERIOD refers to a date range, this attribute is used to provide metadata on the actual range the observation refers to (e.g. for period ‘2001-2003’ TIME_PERIOD would be 2002 but the actual dates --2001-2003-- would be expressed here).”
- Optional free-text attribute


Attribute: TIME_COVERAGE

- ISO8601 representation of the actual time interval to which the observation refers
- While TIME_PERIOD should always be expressed as a year, and TIME_DETAIL is free-text with additional information, TIME_COVERAGE can optionally be used to provide the exact interval in a structured format



Attribute: Base Period (BASE_PER)

- “Period of time used as the base of an index number, or to which a constant series refers”
- Where a base period applies, it is expected to always be set to a year
- Typically, used for constant prices, as in “2005 USD dollar”
- Optional attribute



Attribute: Unit of Measure (UNIT_MEASURE)

- “Unit in which the data values are expressed”
- It may not be obvious which is the correct unit in some cases. Coding guidelines are available and will be further developed.
- Unit of Measure was a dimension in the MDG DSD but was changed to attribute based on the experience with MDG data exchange



Attribute: Nature of data points (NATURE)

- “Information on the production and dissemination of the data (e.g.: if the figure has been produced and disseminated by the country, estimated by international agencies, etc.)”
- Optional attribute
- Normally set to C (Country Data) in national reporting



Attribute: Source details (SOURCE_DETAIL)

- Provides additional textual information on the data source, e.g. a specific survey that was used to generate the indicator.
- Optional free-text attribute



Attributes: UPPER_BOUND and LOWER_BOUND

- Where the observation value represents a point estimate, can be used to convey the Upper and Lower bounds
 - In MDG DSD, separate series codes had to be created for upper and lower bounds



Attribute: COMMENT_OBS (footnotes)

- “Additional information on specific aspects of each observation, such as how the observation was computed/estimated or details that could affect the comparability of this data point with others in a time series.”
- The concept for footnotes was renamed to COMMENT_OBS, in line with other implementations



Attribute Attachment Level

- Currently, all attributes in the SDG DSD are attached to the observation.
- The official DSD is expected to have both time series and observation level attributes.



SDG DSD: Mappings

- Due to its support for heterogeneous indicators, it's not always obvious which values should be used in some dimensions
- What should be SEX in indicator "Births attended by skilled personnel":
 - Not Applicable? Total? Female?



SDG DSD: Mappings (2)

- Inconsistent mappings lead to duplications and other anomalies
- Coding guidelines are available and will be further developed and enforced through content constraints
- The use of a single code for no breakdown (e.g. for Total and Not Applicable) simplifies the mappings.