




## The SDG Data Structure Definition



# SDG Data Structure Definition

- Developed by the Working Group on SDMX for SDG Indicators, established by IAEG-SDGs in April 2016
- First version officially released on 14 June 2019

 [SDG DSD Matrix Version 1.0](#)

 [Global SDG DSD v1.0](#)

 [Guidelines for the Global DSD for SDGs](#)

<https://unstats.un.org/sdgs/iaeg-sdgs/sdmx-working-group/>

# SDG Data Structure Definition

- One single DSD is 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 an dimension is not applicable.

## Dimension: Frequency (FREQ)

- “Indicates rate of recurrence at which observations occur (e.g. monthly, yearly, biannually, etc.).”
- By convention, SDGs 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 “sub-indicators”
  - A single indicator can have multiple series
  - Not to be confused with SDMX time series (each series can have multiple time series, i.e., multiple disaggregation with observations organized over time)
- Example: Indicator 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: 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: 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: Age (AGE)

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

## 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: Economic Activity (ACTIVITY)

- “High-level grouping of economic activities based on the types of goods and services produced.”
- Consists of top-level ISIC categories.
- Use **\_T** where not applicable.

# Dimension: Product Type (PRODUCT)

- Product or commodity code
- Combines SDG-specific entries from several classifications including CPC, Material Flows, and non-standard
- 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
- Used in conjunction with attribute CUST\_BREAKDOWN\_LB, which transmits description of the custom code.
- Use **\_T** where not applicable

## Dimension: COMPOSITE\_BREAKDOWN

- Mixed dimension: represents several merged code lists
  - E.g. by International Organizations, Hazard Type 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: 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
- Mandatory observation-level attribute

## 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 observation-level 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.
- Mandatory time series-level attribute

## 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 observation-level 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
- Optional observation-level attribute.

## 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 observation-level attribute.

# Attribute: Nature of data points (NATURE)

- Information on the production and dissemination of the data
- Expresses whether a data point has been produced and disseminated by the country, estimated by international agencies, etc.
- Normally set to C (Country Data) in national reporting
- Optional observation-level attribute

## **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 observation-level 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
- Optional observation-level attributes

# Attributes: Footnotes (COMMENT\_OBS and COMMENT\_TS)

- “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.”
- Attribute COMMENT\_OBS is used for observation-level footnotes, and COMMENT\_TS for time series-level footnotes. Both are optional.

## Attribute: GEO\_INFO\_URL

- Provides web address of a geoinformation file. Used in conjunction with attribute GEO\_INFO\_TYPE.
- Optional time series-level attribute.

## Attribute: GEO\_INFO\_TYPE

- Specifies type of geoinformation file provided in attribute GEO\_INFO\_URL.
- Optional time series-level attribute.



## 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

- Inconsistent mappings lead to duplications and other anomalies
- Coding guidelines will be 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.



**United  
Nations**

DESA  
Statistics Division



**Thank you.**