



**United
Nations**

Department of Economic and Social Affairs
Statistics

SDG Data Structure Definition



United Nations

Department of Economic and Social Affairs
Statistics

<p>1 NO POVERTY</p>	<p>2 ZERO HUNGER</p>	<p>3 GOOD HEALTH AND WELL-BEING</p>	<p>4 QUALITY EDUCATION</p>	<p>5 GENDER EQUALITY</p>	<p>6 CLEAN WATER AND SANITATION</p>
<p>7 AFFORDABLE AND CLEAN ENERGY</p>	<p>8 DECENT WORK AND ECONOMIC GROWTH</p>	<p>9 INDUSTRY, INNOVATION AND INFRASTRUCTURE</p>	<p>10 REDUCED INEQUALITIES</p>	<p>11 SUSTAINABLE CITIES AND COMMUNITIES</p>	<p>12 RESPONSIBLE CONSUMPTION AND PRODUCTION</p>
<p>13 CLIMATE ACTION</p>	<p>14 LIFE BELOW WATER</p>	<p>15 LIFE ON LAND</p>	<p>16 PEACE, JUSTICE AND STRONG INSTITUTIONS</p>	<p>17 PARTNERSHIPS FOR THE GOALS</p>	<p>THE GLOBAL GOALS For Sustainable Development</p>

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

- A 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.

Current Version

- SDG DSD Current Version 1.9, released Jul 2022
- Manuals
 - Guidelines for use of the Global SDG DSD
 - Guidelines for the customization of the Global SDG DSD
 - SDMX API Manual

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
- Global code list contains ISO 3166-Alpha 2 (two-letter) and M49 (numerical) country codes, as well as numeric SDG region codes.
- 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 the following codes
 - _T (Total)
 - U (Urban)
 - R (Rural)
 - CITY (City)
 - TSUB (Town and semi-dense area)
- 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
- 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
- Populated with generic codes (e.g. C01, C02,...C999), 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 and content constraints 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
- 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 Global Dataflows

- **DF_SDG_GLH** – Harmonized Global Dataflow. This dataflow is used by the Custodian Agencies to report SDG indicators that are part of the global dataset, regardless of how the data was obtained. This dataflow is also used to disseminate the global dataset at the SDMX API.
- **DF_SDG_GLC** – Country Global Dataflow. This data is used by countries to report data to UNSD, as well as to disseminate national data in compliance with the SDG Global DSD.

SDG Cube Region Content Constraints

- **CN_SDG_GLC**, attached to dataflow **DF_SDG_GLC**
 - Restricts the dimension **REPORTING_TYPE** to code **N** (“National”)
 - Ensures that data from countries always have REPORTING_TYPE=N, i.e. the countries always use correct Reporting Type for national dataset.
- **CN_SDG_GLH**, attached to dataflow **DF_SDG_GLH**
 - Restricts the dimension REPORTING_TYPE to code **G** (“Global”)
 - Ensures that data from custodian agencies always have REPORTING_TYPE=G, i.e. the agencies always use correct Reporting Type for the global dataset.

SDG Series Content Constraints

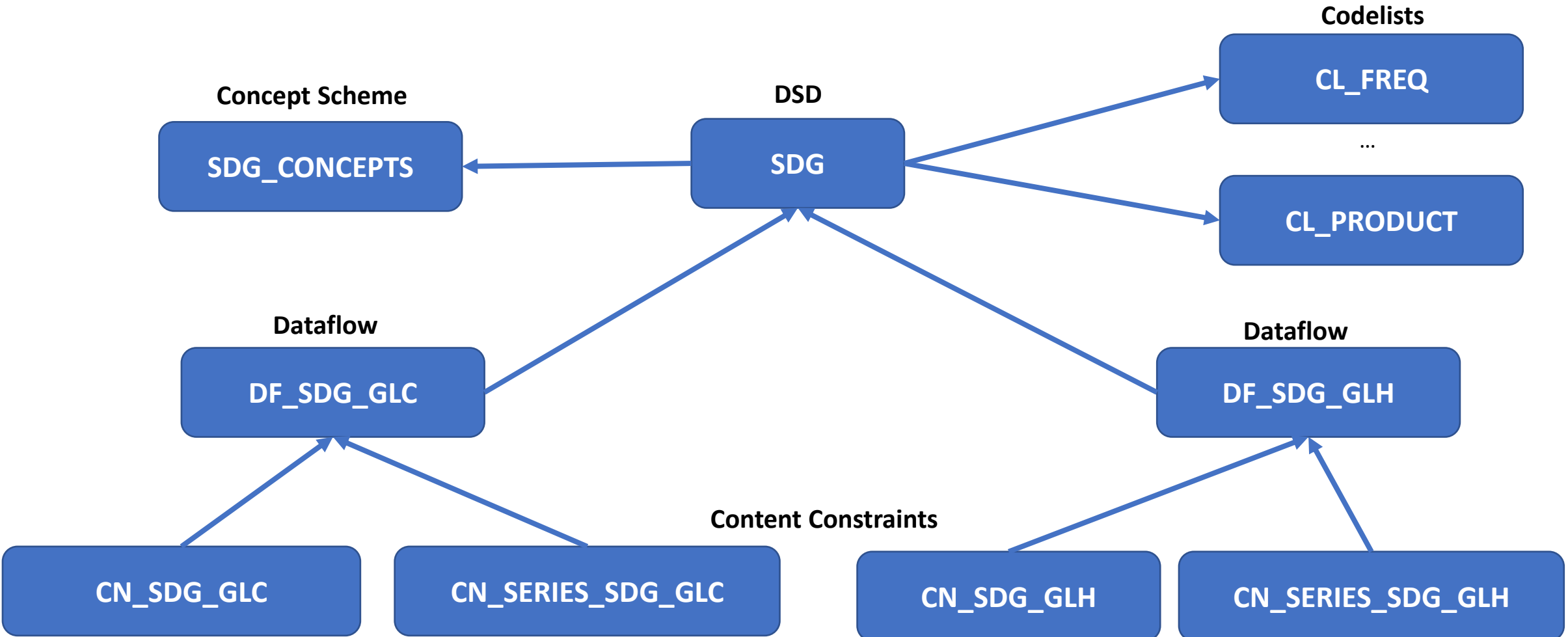
- **CN_SERIES_SDG_GLC**, attached to dataflow **DF_SDG_GLC**
- **CN_SERIES_SDG_GLH**, attached to dataflow **DF_SDG_GLH**
- Although separate, the constraints are identical in terms of content
 - For practical reasons and to make them future-proof
- Provide all valid combinations of SDG dimension codes.
- Can be downloaded from the [SDMX Global Registry](#) or [SDMX-SDG page](#).
- An [Excel matrix](#) representing the series content constraints can also be downloaded from the SDMX-SDG page.

SDG Content Constraints Matrix

- Informal representation of SDG series content constraints in CSV/Excel
- Can be used to determine how to correctly map an SDG series

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	SERIES	Name	UNIT_MEASURE	UNIT_MULT	SEX	AGE	URBANISATION	COMPOSITE_BREAKDOWN	EDUCATION	DISABILITY_STATUS	OCCUPATION	INCOME_WEALTH	PRODUCT	ACTIVITY
83	SH_STA_MORT	Maternal mortality ratio	PER_100000_LIVE	ALL	F	ALL	ALL	_T;MS_MIGRANT;MS_NOM	ALL	ALL	ALL	ALL	_T	_T
84	SH_STA_BRTC	Proportion of births attended by skilled health personnel	PERCENT		0	ALL	_T		_T	_T	_T	ALL	_T	_T
85	SH_DYN_IMRTN	Infant deaths (number)	NUMBER	ALL	ALL	Y0	ALL	_T	_T	_T	_T	ALL	_T	_T
86	SH_DYN_MORT	Under-five mortality rate	PER_1000_LIVE	ALL	ALL	Y0T4	ALL	_T	_T	_T	_T	ALL	_T	_T
87	SH_DYN_IMRT	Infant mortality rate	PER_1000_LIVE	ALL	ALL	Y0	ALL	_T	_T	_T	_T	ALL	_T	_T
88	SH_DYN_MORTN	Under-five deaths (number)	NUMBER	ALL	ALL	Y0T4	ALL	_T	_T	_T	_T	ALL	_T	_T
89	SH_DYN_NMRTN	Neonatal deaths (number)	NUMBER	ALL	ALL	M0	ALL	_T	_T	_T	_T	ALL	_T	_T
90	SH_DYN_NMRT	Neonatal mortality rate	PER_1000_LIVE	ALL	ALL	M0	ALL	_T	_T	_T	_T	ALL	_T	_T
91	SH_HIV_INCD	Number of new HIV infections	PER_1000_UNINF	ALL	ALL	ALL	ALL	_T;MS_MIGRANT;MS_NOM	ALL	ALL	ALL	ALL	_T	_T

Diagram of SDG artefacts





**United
Nations**

DESA
Statistics Division



Thank you.