



Guidelines for the Customization of the Global Data Structure Definition for Sustainable Development Goals Indicators

Document version: 1.6

Date: Oct 2021

1. Purpose of this document

This document is an annex to the **Guidelines for the Global Data Structure Definition for Sustainable Development Goals Indicators**, provided separately. The purpose of this document is to provide recommendations on how to customize the global SDG DSD for dissemination purposes. The customized DSD may also assist with implementing databases for national SDG indicators.

The 2030 Agenda for Sustainable Development underscores the national ownership of the SDGs and calls for “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.” Consequently, while the global SDG DSD sets a standard for global reporting, the Member States must be able to customize the DSD to suit its own national circumstances, in particular extend it with national indicators and disaggregation. This document outlines recommendations and guidance on how to customize the DSD.

2. SDMX reporting and dissemination for SDG indicators

2.1. Global SDG reporting and dissemination

The first official version of the Global DSD for SDG indicators was released in June 2019 and is available at the Global Registry (<http://registry.sdmx.org>) while guidelines and other supporting documentation are available at <https://unstats.un.org/sdgs/iaeg-sdgs/sdmx-working-group/>

The global SDG DSD has 32 concepts, including 16 dimensions, 15 attributes, and the primary measure. These include 2 “wildcard” (multi-purpose) dimensions, [COMPOSITE BREAKDOWN](#) and [CUST BREAKDOWN](#), which can be used to support multiple breakdowns. The multi-purpose dimensions play a key role in the customization of the DSD.

The global SDG DSD was created on the basis of the SDG global database, of which UNSD is the custodian. The global DSD is intended to be used for formal data transmission (reporting) as follows:

- Between the SDG custodian agencies and UNSD;
- Between countries and the custodian agencies, reflecting established dataflows;
- Between countries and regional agencies.

The countries may also wish to transmit their global indicators to UNSD, using the unmodified global SDG DSD, for the purposes of comparison between the global and national datasets. The countries may delegate this role to regional agencies.

In addition, the SDG DSD is used by UNSD for dissemination from the global SDG database through an SDMX API.

2.2. National SDG reporting and dissemination

The global DSD only covers global SDG indicators and disaggregation. In most countries, national-specific indicators and breakdowns are defined that are not present in the global DSD. To support these for national dissemination, a country needs to extend the DSD with its national-specific indicators, codes, and breakdowns.

It is envisaged that a country may establish two separate dataflows for SDG indicators:

- Global dataflow, which uses the global DSD and can be used for the transmission of nationally-sourced global SDG indicators to the custodian agencies, UNSD, regional agencies, and other recipients.
- National dataflow, which uses the nationally-customized global DSD and can be used for the dissemination of the full set of disaggregated national SDG indicators, including nationally-sourced global indicators and national-specific indicators. The national DSD should be compatible with the SDG DSD conceptual model, i.e. use the same dimensions and mandatory attributes.

The country may also choose to establish additional dataflows optimized for dissemination, using DSDs with reduced number of concepts, which are therefore incompatible with the SDG DSD conceptual model. In some situations, this approach may be very useful since the global DSD is optimized for data exchange and may be cumbersome to use for other purposes such as visualization.

Reporters are encouraged to ensure, however, that a national dataflow is established, as described above, that follows the SDG conceptual model. This maintains a high level of interoperability and simplifies the consumption of data for the global user.

3. Customization of the global SDG DSD

3.1. The objectives of customization

A country may wish to extend the global SDG DSD when one or more of the following applies:

1. National indicator framework includes indicators that are not part of the global DSD, or series for global indicators that are not part of the global DSD.
2. Breakdowns of indicators reported by the country use codes that are not part of the global DSD.
3. The indicators are disaggregated using breakdowns that are not part of the global DSD.
4. Additional information is transmitted that is not expressed in the concepts of the global DSD.

For guidance on (1), see [Adding indicators](#).

The difference between (2) and (3) is that in the former case, the breakdown is present in the global DSD but lacks some specific entries used by the country: for example, an age group used in the country is not present on the age code list. In the latter case, the breakdown itself is absent from the global DSD, e.g. national indicators are disaggregated by specific ethnic groups while a breakdown by ethnicity is not in the global DSD. For (2), see [Extending code lists](#), and for (3), [Adding breakdowns](#).

As regards use case (4) above, customizing the SDG Concept Scheme and conceptual model breaks compatibility and should be used with caution. For further information, see [Modifying the data model](#).

When customizing the global DSD, ownership and versioning of affected artefacts should be properly maintained (see below).

3.2. Ownership of the artefacts

While customizing the DSD, it is crucial that proper ownership of the artefacts is maintained. Improper use of ownership or versioning can lead to serious interoperability and compatibility issues, and can be misleading to the user of SDG indicators.

A reporter that chooses to customize the DSD must take ownership – i.e. assign itself as the maintenance agency – of the DSD and any code lists it modifies. A 3-step process should be followed:

- 1) The reporter makes itself the maintenance agency of any code lists it modifies.
- 2) The reporter makes itself the maintenance agency of the DSD.
- 3) The reporter updates references to the modified code lists in the DSD, as appropriate.

Subsequently, the reporter should follow the [Guidelines on the Versioning of SDMX Artefacts](#) to properly version any artefacts it has modified.

Please note that customized versions of the global DSD cannot be used for reporting to the SDG Lab.

Note: [adding breakdowns using the CUST BREAKDOWN dimension](#) does not require a modification of the DSD or transfer of ownership. See below for further detail.

3.3. Adding indicators

SDG indicators are supported through the SERIES dimension and the associated CL_SERIES code list. As explained in the Guidelines for the Global Data Structure Definition for Sustainable Development Goals Indicators, the reason the term Series is used is caused by the fact that a single official SDG indicator may be composed of multiple series or sub-indicators. This concept was also used in the MDG Data Structure Definition.

Global SDG indicators

The CL_SERIES code list is maintained on the basis of series present in the global SDG database, and expanded to the extent possible with series for indicators that are not yet available in the global database. It is possible that a reporter may have data for an official SDG indicator series for which there is no match in the global series code list. In that case, an email should be sent to sdmx@un.org explaining in full which indicator/series is missing from the global code list. Official SDG series could then be added at the next release of the global SDG DSD.

National SDG indicators

Most countries' national SDG frameworks include indicators that are not part of the global indicator framework. To disseminate those indicators, codes for national series need to be added to the CL_SERIES code list.

The following format is recommended for national indicator codes¹:

`_L_[ISO Alpha2 Country Code]_[2-char Goal][2-char Target][2-char Indicator]_[2-char Series]`

¹ Draft version of these Guidelines recommended using the **N_** prefix for national series code, and early adopters may be following this approach.

This allows for identification of national indicators and the country of their origin. For example, some national indicator codes for Kyrgyzstan:

_L_KG_060402_1	Freshwater withdrawal
_L_KG_070301_1	Energy intensity of GDP by industry
_L_KG_070301_2A	Power intensity of GDP by industry

As explained above, an indicator may include several series. The last portion of the indicator code allows for the inclusion of multiple series within the same indicator. Note that series may also be used for [disaggregating an indicator](#).

Note that in the SDG DSD, unit of measure is an attribute. If there are several units of measure used with the same indicator, a separate series should be created for each unit of measure.

3.4. Extending global cross-domain code lists

Often, while a breakdown is included as a dimension in the global DSD, its code list does not include all entries required by the country. For example, an age group used nationally may not be present in the global code list. In such cases, codes need to be added to the global code list.

Generally, when adding codes to a global code list, the codes should follow the pattern used in the code list (also see below notes on specific [cross-domain code lists](#)). For example, when extending the Unit of Measure code list (CL_UNIT), codes have been created as follows:

CL_UNIT_MEASURE Code	CL_UNIT_MEASURE Name
Y	Years
USD_PER_HA	USD per Hectare

A hierarchical relationship using the **parentCode** attribute should be set within the local extension codes as applicable.

It is recommended to send an email to sdmx@un.org to notify UNSD of codes being added, for their possible inclusion in subsequent releases of the global DSD.

Reference area (CL_AREA)

The global SDG reference area code list generally only includes areas at the national level. The countries commonly extend the code list with their subnational areas.

At the second administrative level (province or state level), it is suggested to use the ISO3166 country code followed by subdivision code. Subdivision codes may either follow the [ISO 3166-2](#) standard, which uses up to a 3-character subdivision code, or another coding system used in the country. For example, some of subdivisions for Kyrgyzstan include:

REF_AREA Code	REF_AREA Description
KG05	Batken
KG11	Bishkek (city)
KG08	Chui

Note that the ISO3166-2 standard includes a dash between the country and subdivision code, e.g. US-NY, FR-01. These dashes have been removed by many reporters in data dissemination or exchange projects to improve compatibility, and should not be used in the SDG CL_AREA code list.

ISO 3166-2 subdivision codes for each country can be obtained from national sources, and are widely available on the Web.

Age group (CL_AGE)

The Age code list (CL_AGE) is a cross-domain code list, created in compliance with the [recommendation of the SDMX Statistical Working Group \(SWG\)](#). The global code list captures commonly-used age groups, and should be suitable in most situations, Nevertheless, it is possible that a country may use age groups that are not on the code list. In that case, new age group codes should be created in accordance with the [recommendation](#). For example:

CL_AGE Code	CL_AGE Description
Y40T59	40 to 59 years old
Y20T39	20 to 39 years old

The use of LT and LE (“less than” and “less than or equal”) operators is not recommended. For example, an age group “under 21 years of age” should have a code of Y0T20, instead of using LE or LT. Similarly, use of the GT (“greater than”) operator is not recommended; GE (“greater than or equal”) should be used instead, e.g. Y_GE65.

Economic activity (CL_ACTIVITY)

CL_ACTIVITY is a cross-domain code list. As of August 2019, it includes top-level codes of ISIC rev.4. If a more detailed breakdown is needed, entries from the [cross-domain code list](#) should be used.

An economic activity code list used in SNA93, based on ISIC rev.3, is available at the [SDMX Global Registry](#). If ISIC3 entries need to be included, codes from that code list should be used.

When adding a national variant of the classification to the global code list, it should be prefixed as follows:

L [National Code]

Education level (CL_EDUCATION_LEV)

Currently, the SDG Education Level code list includes top entries of ISCED-11 and ISCED-97, as well as custom codes used in SDG indicators. A more detailed education level code list is used in the global Education DSD and is available at the [SDMX Global Registry](#). If codes need to be added to the SDG Education Level code list, they should be taken, to the extent possible, from the global Education Level code list.

When adding a national variant of the classification to the global code list, it should be prefixed as follows:

L [National Code]

Occupation (CL_OCCUPATION)

The SDG code list for occupation includes entries from ISCO-08, ISCO-88, and ISCO-68, as well as SDG-specific entries. The global cross-domain code list, which includes the full set of ISCO-08 codes, is available at sdmx.org and the [SDMX Global Registry](#). Since the global code list only includes ISCO-08 entries, it uses a different convention than the SDG code list. When extending the SDG occupation code list, the SDG code list convention should be used.

When adding a national variant of the classification to the global code list, it should be prefixed as follows:

L [National Code]

Product or commodity type (CL_PRODUCT)

The CL_PRODUCT code list includes the following types of codes:

- Material flows;
- Product codes that use the Central Product Classification (CPC);
- Custom SDG codes.

These codes are used in various agriculture, environment, trade, and other indicators.

Material flow codes were imported from the CL_MATERIAL code list of the global SEEA DSD. More detailed material flow codes should be added from the same code list. The SEEA DSDs, including the CL_MATERIAL code list, are available at the [SDMX Global Registry](#).

The CPC codes were imported from the CL_COMMODITY code list of the DSD for International Merchandize Trade Statistics DSD (IMTS). Further CPC codes should be added from the same code list, available at the [SDMX Global Registry](#).

When adding a national variant of the classification to the global code list, it should be prefixed as follows:

L [National Code]

3.5. Aggregated codes

In statistics, **aggregate data** are data combined from several measurements. When data is aggregated, groups of observations are replaced with summary statistics based on those observations. Common aggregate functions (i.e. the summary statistics) include Sum, Average (i.e., arithmetic mean), Count, Maximum, Median, Minimum, etc.

Each of these observations is identified by a set of codes from the codelists representing each dimension. The aggregation operations are typically aimed at defining aggregation hierarchies in one particular dimension. When a set of item codes are aggregated, the result has to be assigned to a new item in the same dimension, i.e. in the same codelist. Each **aggregated item** will be the result of

combining items from the codelist, either consecutive or not. The following set of standard "operators" is proposed²:

- "T" for expressing ranges (e.g. from 3 to 9) with more than 2 codes. Combination of codes using “_” should be used when 2 consecutive codes are combined.
- “_” for the combination of two codes, whether consecutive or not (e.g. 5 and 6; A and F), and also for a list of more than two non-consecutive codes.
- "X" for expressing "except" or "excluding",
- "GT" for "greater than", "LT" for "less than", "GE" for "equal to or greater than" and "LE" for "equal to or less than".

For example, some aggregates are usually defined in ISCO08³ with the following codes (and rules):

Code	Name	Rule
ISCO08_1T3	Managers, professionals, technicians and associate professionals	ISCO08_1+ ISCO08_2+ ISCO08_3
ISCO08_4_5	Clerical support workers, service and sales workers	ISCO08_4+ ISCO08_5
ISCO08_6_7	Skilled agricultural, forestry and fishery workers, craft and related trades workers	ISCO08_6+ ISCO08_7
ISCO08_8_9	Plant and machine operators, and assemblers and elementary occupations	ISCO08_8+ ISCO08_9
ISCO08_0	Armed forces occupations	ISCO08_0

When the objective of aggregating is to generate a new classification variant that can be applied to any version (variant) of the classification with the aim of harmonizing heterogeneous data sources, a similar convention is followed but assigning a new prefix to identify the new variant:

It is a good practice specifying the aggregation rule(s) for each item in an annotation.

Code	Name	Rule
SECTOR_AGR	Agriculture	ISIC4_A ISIC3_A +ISIC3_B ISIC2_1
SECTOR_IND	Industry	ISIC4_B +ISIC4_C+ISIC4_D+ISIC4_E+ISIC4_F ISIC3_C +ISIC3_D+ISIC3_E+ISIC3_F ISIC2_2 +ISIC2_3+ISIC2_4+ISIC2_5
SECTOR_NAG	Non-agriculture	ISIC4_TOTAL-ISIC4_A ISIC3_TOTAL-ISIC3_A-ISIC3_B ISIC2_TOTAL-ISIC2_1
SECTOR_SER	Services	ISIC4_G +ISIC4_H+ISIC4_I+ISIC4_J+ISIC4_K+ISIC4_L+ISIC4_M+ISIC4_N+ISIC4_O+ISIC4_P+ISIC4_Q+ISIC4_R+ISIC4_S+ISIC4_T+ISIC4_U ISIC3_G +ISIC3_H+ISIC3_I+ISIC3_J+ISIC3_K+ISIC3_L+ISIC3_M+ISIC3_N+ISIC3_O+ISIC3_P+ISIC3_Q ISIC2_6 +ISIC2_7+ISIC2_8+ISIC2_9
SECTOR_T	Total	ISIC4_TOTAL ISIC3_TOTAL ISIC2_TOTAL
SECTOR_X	Not classified	ISIC4_X ISIC3_X ISIC2_0

²See “[Guidelines for the creation and Management of SDMX Code Lists](#)”

³ International Standard Classification of Occupations, 2008

3.6. Adding breakdowns

It is possible that a national indicator(s) may be disaggregated using a breakdown that is not present in the global DSD. In that case, the DSD may need to be customized to add the new codes in support of the breakdown.

There are 3 principal methods of adding a breakdown to the global DSD:

- Use the SERIES dimension.
- Use the COMPOSITE_BREAKDOWN dimension.
- Use the CUST_BREAKDOWN dimension.

While any of these methods can be used to add a breakdown, each is most suitable in certain specific situations. Recommendations on the use of each of these methods are given below.

Adding breakdowns using the SERIES dimension

A breakdown can be implemented through the SERIES dimension by adding a separate series code for each breakdown category for each indicator disaggregated by the breakdown. For example, for the number of air passengers disaggregated by type, the following series were created (Cambodia⁴):

N_KH_090102_01	9.1.2 Air transport, international passengers carried
N_KH_090102_02	9.1.2 Air transport, domestic passengers carried

This approach is best suited in situations when the following conditions apply:

- The breakdown is applicable to a very small number of indicators (1-2);
- The breakdown includes a very small number of entries.

If the breakdown applies to a larger number of indicators, the DSD maintenance burden increases since a separate series code needs to be added for each breakdown entry. This issue is further exacerbated as the number of entries in the breakdown grows. In addition, this approach then complicates the processing and presentation of the data.

However, when these conditions do apply, as in the case of 9.1.2 above, then this approach is often optimal, and can reduce the maintenance burden.

When using this approach, the reporter must take ownership of the CL_SERIES code list and the DSD.

Adding breakdowns using the COMPOSITE_BREAKDOWN dimension

The COMPOSITE_BREAKDOWN dimension, and its CL_COMP_BREAKDOWN code list, are used to represent multiple breakdowns in the global DSD. This code list can be used to add national-specific breakdowns.

The format used in the code list is as follows:

[breakdown abbreviation]_[code]

⁴ As an early adopter, Cambodia used prefix N_ for national series. As described [above](#), prefix _L should be used in new implementations.

Code description should start with a description of breakdown followed by description of entry.

For example, the following codes were added to the CL_COMP_BREAKDOWN code list for the Cambodian national DSD to support a breakdown by the source of light:

SL_SOLAR	Source of light: Solar
SL_PUBLIC	Source of light: Publicly-provided electricity
SL_PRIVATE	Source of light: Privately-generated electricity
SL_OTHER	Source of light: Other
SL_NONE	Source of light: None
SL_KEROSENE_LAMP	Source of light: Kerosene lamp
SL_CANDLE	Source of light: Candle
SL_BATTERY	Source of light: Battery

When using this approach, the reporter must take ownership of the CL_COMP_BREAKDOWN code list and the DSD.

Note: caution should be used when adding codes to the CL_COMP_BREAKDOWN code list, since only one breakdown at a time in this code list can be used to disaggregate an indicator. Any breakdown to be added, therefore, should be checked against all existing breakdown in this code list to ensure a conflict is impossible, or at least improbable. If it does appear likely that a breakdown to be added will conflict with other breakdowns on the code list, another method should be used, e.g. using the SERIES codes list as [above](#).

If a national indicator is disaggregated by two breakdowns not present in the global DSD, one of them may be implemented through the CL_COMP_BREAKDOWN code list, and the other through either the CL_SERIES code list, or the CL_CUST_BREAKDOWN code list.

[Adding breakdowns using the CUST_BREAKDOWN dimension](#)

The CUST_BREAKDOWN dimension primarily serves the purpose of supporting volatile breakdowns, i.e. those whose code lists change often, lack international or even national classifications, or are otherwise difficult to maintain, such as ethnocultural characteristics or others derived from write-in responses on a census or survey form.

For the custom breakdown dimension, its generic code list CL_CUST_BREAKDOWN is used to transmit the codes, while attribute CUST_BREAKDOWN_LB is used to transmit descriptions of the codes directly in the dataset. CUST_BREAKDOWN_LB should start with a description of breakdown followed by description of entry. For example, if disaggregation by the source of light was implemented through custom breakdown (rather than composite breakdown as above), the values of the CUST_BREAKDOWN dimension and CUST_BREAKDOWN_LB attribute in the dataset would look as follows:

CUST_BREAKDOWN code	CUST_BREAKDOWN_LB
C01	Source of light: Solar
C02	Source of light: Publicly-provided electricity
C03	Source of light: Privately-generated electricity
C04	Source of light: Other
C05	Source of light: None

C06	Source of light: Kerosene lamp
C07	Source of light: Candle
C08	Source of light: Battery

Note that implementation of a breakdown through the CUST_BREAKDOWN dimension has no impact on the DSD, i.e. it does not result in any change to the global DSD. The CUST_BREAKDOWN code list is generic and stays intact, while the CUST_BREAKDOWN_LB attribute is set at the time the data is converted to SDMX. This approach may thus reduce the maintenance burden, or in some cases even make it possible to use the global DSD as is in support of national dissemination.

When using this approach, no modification to the global DSD or transfer of artefact ownership is required.

3.7. Modifying the data model

Modifying the SDG data model includes adding to or removing concepts from the SDG Concept Scheme, adding or removing a concept from the conceptual model (e.g. removing an attribute or dimension), or changing the role of a concept (dimension to attribute or vice versa).

While using a reduced or increased number of concepts may be desirable in certain situations (such as optimizing the DSD for visualization or implementing well-established national disaggregation), these actions break compatibility with the global DSD. Reporters are encouraged to ensure that at least one dataflow exists that follows the SDG conceptual model, which facilitates interoperability and is certainly of benefit to international or national users. Concepts used in the SDG DSD should be sufficient for the expression of SDG-related data for dissemination purposes.

Adding breakdowns

Breakdowns unavailable in the global DSD can be [added as described above](#), and should not require additional dimensions. Dimensions should only be added when compatibility and interoperability are not a consideration.

Adding textual metadata

Adding textual metadata should not require a modification to the data model. As explained in the **Guidelines for the Global Data Structure Definition for Sustainable Development Goals Indicators**, the SDG DSD includes 4 free-text attributes that can be used for that purpose:

- **SOURCE_DETAIL** can be used to provide additional textual information on the data source, e.g. a specific survey that was used to generate the indicator.
- **TIME_DETAIL** can be used to provide additional textual information when the time period of the indicator differs from the calendar year, or for additional information on the base period.
- **COMMENT_OBS** and **COMMENT_TS** are used for footnotes, or comments, at the observation and time series level, respectively, and can be used to provide any additional relevant textual information.

Adding other conditional attributes

Adding conditional attributes to the DSD usually does not break software built for the consumption of data that conforms with the global SDG conceptual model. It still leads to increased maintenance burden, however, as well as the risk that the consumer may overlook the additional attributes and information therein. It should be considered whether existing concepts could be used to convey the requisite information.

3.8. Summary of scenarios and impact on interoperability

The table below lists the various approaches to implementing custom disaggregation as described above, and their impact on interoperability.

Action	When to use	Advantages	Disadvantages
Using CUST_BREAKDOWN	For volatile or sensitive breakdowns	Full compatibility with the global DSD; no maintenance burden	No persistence of code lists, difficult to use. Should not be used to extend breakdowns already in the DSD.
Adding disaggregation to CL_SERIES	For breakdowns that have a very small number of codes and are applicable to a very small number of indicators	No change to the conceptual model; lowest maintenance burden when used appropriately	Can lead to greatly increased maintenance burden when the number of applicable indicators or codes is greater than 1-2.
Adding disaggregation to CL_COMP_BREAKDOWN	For national breakdowns that are not part of the global DSD	No change to the conceptual model	Risk of collision between breakdowns, since only 1 can be used at a time. Increased maintenance burden.
Extending other global code lists	For breakdowns that are part of the global DSD	No change to the conceptual model	Increased maintenance burden. More difficult to synchronize global and national DSDs.
Adding conditional attributes	For well-established national concepts that are difficult to implement within the SDG conceptual model	Can make it easier to reuse existing concepts or code lists	Can make it more difficult for the global user to consume the information being added. Increased maintenance burden.

<ul style="list-style-type: none">• Adding or removing dimensions• Adding mandatory attributes• Removing existing attributes	When compatibility is not a consideration, e.g. to create a DSD optimized for visualization	Derived DSDs can be easier to use for certain purposes such as visualization	Breaks compatibility with the global DSD and makes it substantially more difficult for the global user to consume.
--	---	--	--