

10TH EXPERT GROUP MEETING ON

# Statistical Data and Metadata eXchange

JANUARY 25-28, 2021

## **SDG DSD**

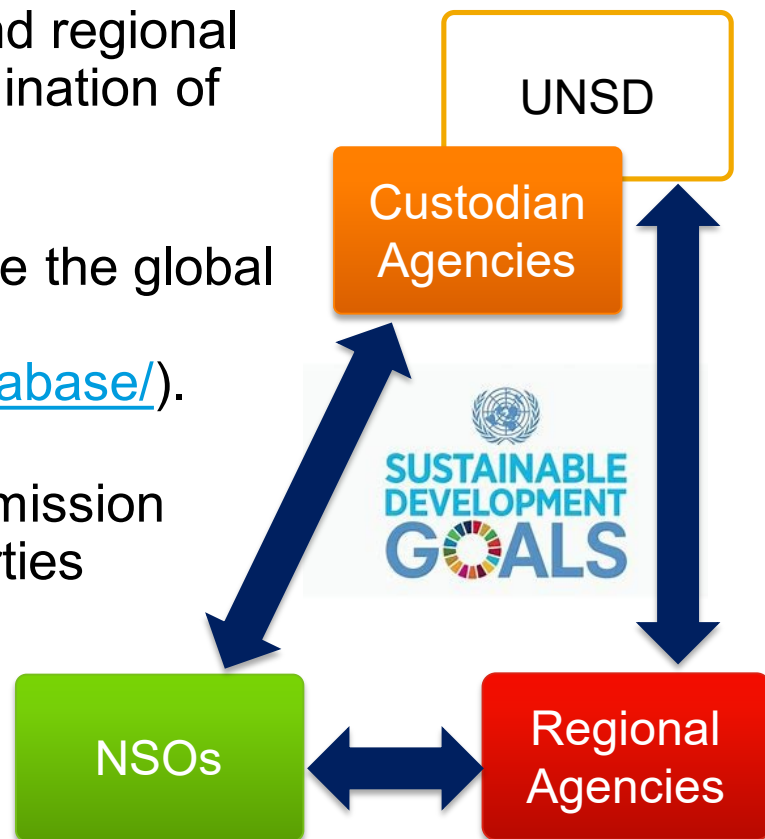
January 26, 2021

Juan Muñoz

Director of IT Planning and Standardization / INEGI

# Purpose of the SDG DSD

- To support the transmission of all SDG indicators and their disaggregation, as approved by the United Nations Statistical Commission
  - Adopted by countries and international and regional organizations for collection and/or dissemination of official SDGs data and metadata.
  - Exchanged information is used to populate the global SDG database, custodied by UNSD (<https://unstats.un.org/sdgs/indicators/database/>).
  - Intended to be used for formal data transmission (reporting) of information between the parties



# Challenges Faced for the SDG DSD Definition

- Integration of information from many countries with different IT development levels
- Continuous changes due to new definitions of indicators required by an evolving framework
- Heterogeneous generation of information due to different capabilities, contexts and interests of the countries
- Compatibility between the Global Dataflow and the National Dataflows (for reporting)

# Design Approach

- One DSD covering all indicators
  - Open to definition of additional DSD if it is needed
  - Four versions (one in 2019 and three in 2020).  
Most recent published on December 24<sup>th</sup>, 2020
- Tested in pilot projects
- Customizations are possible accordingly to agreed guidelines
  - Note: for dissemination purposes
- Continuous integration of feedback by means of the Working Group
  - Participation of international organizations and national statistical offices



<https://unstats.un.org/sdgs/files/SDG-DSD-Customization-Guidelines.pdf>

# SDG DSD as a Global Artifact

The screenshot shows the SDMX Global Registry interface. The main content area displays the 'Data Structure Definition Details' for the 'SDG DSD'. The details include:

- URN: urn:sdmx.org:sdmx:informodel:datadefinition:DataStructure-IAEG-SDGs-SDG(1.3)
- URL: [https://registry.sdmx.org/urn:sdmx.org:sdmx:informodel:datadefinition:DataStructure-IAEG-SDGs-SDG\(1.3\)](https://registry.sdmx.org/urn:sdmx.org:sdmx:informodel:datadefinition:DataStructure-IAEG-SDGs-SDG(1.3))
- Version: 1.3
- State: Public (Final)
- Revisions [2]: 2020-12-24T16:23:40Z (Download Revision, Compare Revision)
- Description: Data Structure Definition for Sustainable Development Goals Indicators
- Annotations: N/A
- Valid From: (empty)
- Valid To: (empty)
- Reference Metadata: (empty)
- Cross References: urn:sdmx.org:sdmx:informodel:codebook:Codebook-IAEG-SDGs-CL\_ACTIVITY(1.1), urn:sdmx.org:sdmx:informodel:codebook:Codebook-IAEG-SDGs-CL\_AGE(1.1)

- Published in the SGR (<https://registry.sdmx.org/>)

- Supported by SDMX Sponsors Committee

- One definition for any organization requesting information about SDG

- Additional documents can be found in the IAEG-SDGs Working Group on SDMX site (<https://unstats.un.org/sdgs/iae-g-sdgs/sdmx-working-group/>)

The screenshot shows the 'SDG Data Structure Definition' page. It includes the following content:

- SDG DSD v1.3 was released on 24 Dec 2020. **NEW**
- The SDG Data Structure Definition can be downloaded using the links below.
- List of links:
  - SDG DSD: This is the DSD in the SDMX-ML format, downloadable from the SDMX Global Registry
  - SDG DSD Matrix: This is an Excel representation of the DSD. As of v1.1, the spreadsheet is implemented with the SDMX Matrix Generator tool, and can be used to customize and generate the DSD in the SDMX-ML format.
  - Guidelines for use of the Global SDG DSD: This document contains additional information about the DSD and guidelines for its usage.
  - Guidelines for the customization of the Global SDG DSD: This document contains information and recommendations on customization of the global SDG DSD for national use.
- Buttons for downloading:
  - SDG DSD Matrix V1.3
  - Global SDG DSD v1.3
  - SDG Dataflows and Content Constraints v1.3
  - Guidelines for the Global DSD for SDGs
  - Guidelines for the customization of the Global DSD
- SDMX API section with a button for 'SDMX-SDGs API Manual'.

# Some Facts About the SDG DSD 1.3

- Concept Scheme composed by 33 concepts:
  - **Dimensions (16):** `FREQ`, `REPORTING_TYPE`, `SERIES`, `REF_AREA`, `SEX`, `AGE`, `URBANISATION`, `INCOME_WEALTH_QUANTILE`, `EDUCATION_LEV`, `OCCUPATION`, `CUST_BREAKDOWN`, `COMPOSITE_BREAKDOWN`, `DISABILITY_STATUS`, `ACTIVITY`, `PRODUCT`, `TIME_PERIOD`
  - **Attributes (16; 3 mandatory and 13 conditional):** `OBS_STATUS`, `UNIT_MULT`, `UNIT_MEASURE`, `BASE_PER`, `NATURE`, `TIME_DETAIL`, `COMMENT_OBS`, `TIME_COVERAGE`, `UPPER_BOUND`, `LOWER_BOUND`, `SOURCE_DETAIL`, `COMMENT_TS`, `GEO_INFO_URL`, `GEO_INFO_TYPE`, `CUST_BREAKDOWN_LB`, `DATA_LAST_UPDATE`
  - **Measure (1):** `OBS_VALUE`
- Two defined Dataflows:
  - DF\_SDG\_GLC, Reporting and dissemination dataflow for non-harmonized global SDG indicators reported by countries
  - DF\_SDG\_GLH, Reporting and dissemination dataflow for harmonized global SDG indicators.
- 577 series of indicators defined in `CL_Series`
- 208 breakdown codes defined in `CL_COMP_BREAKDOWN`

# SDMX Matrix Generator

- This Excel tool ([SIS-CC / SDMX Matrix Generator · GitLab](#)) facilitates the customization of the DSD.
- Since version 1.1 of the SDG DSD has been published using this tool

HOW TO COMPLETE THIS WORKSHEET [Open User Guide](#)

This worksheet is used to define a Concept Scheme. It has two sections: The **CONCEPT SCHEME ARTEFACT INFORMATION** which is the "header" information for the Codelist, and the **CONCEPTS** section which is used to define each Concept on a row. Required fields are marked with a \*.

You may click on the **Update DSD matrix...** button automatically add the Concepts and Codelist to the DSD matrix and go to it.

Dimension (D) Attribute (A) Measure (M)	Concept ID	Code List or Uncoded	Type (for Uncoded Concepts)	Min length†	Max length	Concept Role	Concept Name:en	Concept Description:en
D	FREQ	CL_FREQ					Frequency of observation	Time interval at which observations occur over a given time period.
D	REPORTING_TYPE	CL_REPORTING_TYPE					Reporting type	Represents whether the official SDG series is reported by a national, regional, or global level.
D	SERIES	CL_SERIES					SDG Series	SDG Indicator or Series. Uses the GETS convention for encoding.
D	REF_AREA	CL_AREA					Reference area	Country or geographic area to which the measured statistical phenomenon is referred.
D	SEX	CL_SEX					Sex	State of being male or female.
D	AGE	CL_AGE					Age	Length of time that an entity has lived or existed.

# Sustainable Dataflows in ONS

Challenges faced by the ONS to establish sustainable dataflows:

- Fluent (continuous) data feed
  - Inefficient manual work by making Excel transformations
  - (Implementation and) Mapping of dissemination databases
  - Remapping databases to new DSD versions
- Integrating information
  - Implementation of SDMX in Domestic statistical systems
  - Different disaggregation levels for domestic and international purposes
  - Additional indicators and dimensions for internal purposes
  - Not enough skilled resources in the domestic statistical systems
- Duplication of work disseminating SDGs
  - Domestic and international channels (services and sites)
- Dealing with different languages
  - Translations to national languages



# Optimization of the DSD for Dissemination

- SDG DSD designed for reporting, not for dissemination
- Possible issues when used for dissemination:
  - Sparseness of the SDG DSD and complex dimensions complicates data selection.
  - Code labels too long to be displayed on the screen.
  - Captions of multi-purpose dimensions vary depending on the context, not supported by most dissemination software.
- Possible solutions:
  - Hiding dimensions unused in a particular context.
  - Shortening and optimizing code and concept Names for dissemination.
  - Updating Custom Breakdown labels.
  - Use of dissemination software designed or customized to support the specific DSD (Open SDG platform, .Stat system, SDMX-RI, etc.)

# Conclusions

- SDG DSD is the result of a collaborative effort between International Organizations and National Statistical Offices
- Common agreements to solve discordant operational issues must be reached all the time that the flows are operating
- The SDG DSD is operational and ready to be implemented, but there will be new versions as the SDG project continues advancing \*
- Structures, supporting documentation and software tools must be maintained as an integrated set, preserving the work previously made \*
- Translating the SDG SDMX documentation is a must \*
- Regional workshops focused on implementing the SDG DSD and the SDG MSD are needed to develop reporting capacities of the countries \*

# Questions?

[Juan.Munoz@inegi.org.mx](mailto:Juan.Munoz@inegi.org.mx)