United Nations | DESA Statistics Division

SUSTAINABLE DEVELOPMENT GOALS

SDMX Exchange and Dissemination of Sustainable Development Goals: Lessons Learned

Abdulla Gozalov, Harumi Shibata Salazar

United Nations Statistics Division

Abstract

SDMX-based data exchange of Sustainable Development Goals indicators was piloted in 2018 and launched into production following the first official release of the SDG Data Structure Definition in June 2019. Reference metadata exchange was piloted in 2020 and launched in 2021. At this time, data exchange has been established with countries and SDG custodian agencies, with about 40% of the global SDG database is collected through SDMX APIs and an additional 6% through SDMX-enabled Excel templates. This presentation discusses lessons learned in the development and operation of the SDG data and reference metadata exchange, with a focus on data modeling issues, conversion of reference metadata to SDMX, and improvements to the standard and tools.

Background

- Working Group on Statistical Data and Metadata Exchange for SDG Indicators was established by the Inter-Agency and Expert Group on SDG Indicators
  - Mandate: to develop a solution for the exchange and dissemination of Sustainable Development Goal indicator data and metadata

- June 2019: 1st official data structure definition (DSD) for Goal indicators
- DSD has since been regularly updated in synchronization with the Global Sustainable Development Goal Indicators Database (current version 1.15)

Global SDG DSD – 16 Dimensions

1. Frequency of observation (FREQ)
2. Reporting type (REPORTING_TYPE)
3. Series (SERIES)
4. Reference area (REF_AREA)
5. Sex (SEX)
6. Age (AGE)
7. Degree of urbanisation (URBANISATION)
8. Income or wealth quantile (INCOME_WEALTH_QUANTILE)
9. Education level (EDUCATION_LEV)
10. Occupation (OCCUPATION)
11. Custom breakdown (CUST_BREAKDOWN)
12. Composite breakdown (COMPOSITE_BREAKDOWN)
13. Disability status (DISABILITY_STATUS)
14. Economic activity (ACTIVITY)
15. Type of product (PRODUCT)
16. Time period (TIME_PERIOD)

Global SDG DSD – Attributes

Observation-level Attributes

- Observation Status (OBS_STATUS)
- Unit Multiplier (UNIT_MULT)
- Time Period Details (TIME_DETAIL)
- Time Coverage (TIME_COVERAGE)
- Base period (BASE_PER)
- Nature of data points (NATURE)
- Source details (SOURCE_DETAIL)
- Upper Bound (UPPER_BOUND)
- Lower Bound (LOWER_BOUND)
- Observation level footnote (COMMENT_OBS)

Time series-level attributes

- Unit of measure (UNIT_MEASURE)
- GEO_INFO_URL
- GEO_INFO_TYPE
- Time series level footnote (COMMENT_TS)

Global SDG DSD – Mixed Dimensions

- SERIES
  - **SD_XPD_ESED** Proportion of total government spending on essential services, education [1.a.2]
  - **SG_XPD_HLTH** Proportion of total government spending on essential services, health [1.a.2]
  - **SG_XPD_PROT** Proportion of total government spending on essential services, social protection [1.a.2]
- COMPOSITE_BREAKDOWN
  - **FCC_H** Frequency of Chlorophyll-a concentration: High
  - **FCC_M** Frequency of Chlorophyll-a concentration: Moderate
  - **FIS_POSTFIS_CON_INC** Fiscal intervention stage: Postfiscal consumable income
- CUST_BREAKDOWN
  - **C01** Custom code 01
  - **C02** Custom code 02
  - **C03** Custom code 03

Global SDG DSD – Mixed Dimensions

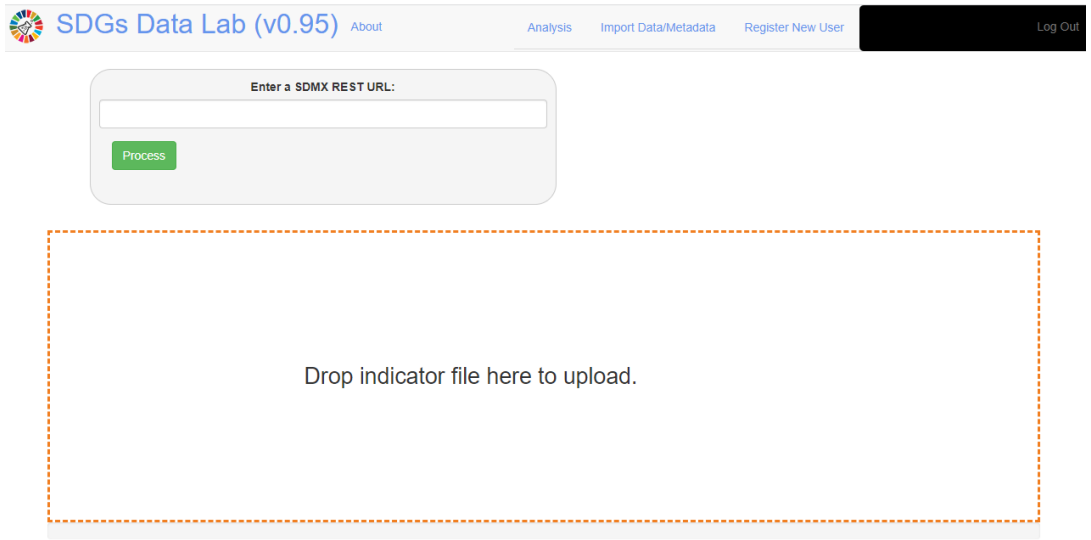
- COMPOSITE_BREAKDOWN
  - Level/Status
  - Name of international institution
  - Type of skill
  - Mode of transportation
  - Name of non-communicable disease
  - Type of speed
  - Migratory status
  - IHR Capacity
  - Policy Domains
  - Policy instruments
  - Type of waste treatment
  - Grounds of discrimination
  - Parliamentary committees
  - Cause of death
  - Substance use disorders
  - Mountain Elevation
  - Deviation Level
  - Frequency of Chlorophyll-a concentration
  - Food Waste Sector
  - Fiscal intervention stage
  - Level of requirement
  - Type of support
  - Report Ordinal
  - Severity of price levels
  - Level_of_government
  - Type of renewable technology
  - Service Attribute
  - Land cover
  - Illicit Financial Flows
  - Nutrient Loading
  - Type of OFDI scheme

Global SDG DSD – Mixed Dimensions

- CUST_BREAKDOWN
  - Cities
  - Counterpart
  - Sampling stations
  - Government name

SDG - SDMX Data transmission

- Data exchange has been established with about 36 countries and six custodian agencies, and the database is published in its entirety through the Statistics Division SDMX application programming interface.



SDGLab accepts:

- SDMX REST URLs
- xml files

Depending on the Reporting type:

- National data stays in the SDGLab database
- For agency data, and SDMX message is sent to our internal SDG data processing system

Current data transmission

	Q1	Q2	Q3	Q4	Total
Number of "unique" series updated or included for the first time					
2022	402	93	52	31	625 series by Q4
2023	501	116	16		662 series by Q3
SDMX transmitted in 2023					
API	156	63	8		171 series (35% of the observations)
SDMX template	32	7	-		35 series (6% of the observations)

Note: series are counted only once even when they might have been released more than once in the indicated quarter.

Challenges

The SDG dataset is highly multi-dimensional and heterogeneous leading to a sparse cube and…

… new breakdowns for the over 500 series are added at a regular basis!!!

However, adding breakdowns under Composite_Breakdown is increasingly difficult…

It contains already so many that there is always risk that a new breakdown will potentially conflict with an existing one (e.g. for social indicators - migratory status is under composite breakdown)

SDG DSD has to be adapted to the continuous changes due to the new definitions of indicators or revisions in the framework

Challenges

SDG DSD was designed for reporting, not for dissemination

Possible issues when used for dissemination:

- Sparseness of the SDG DSD and complex dimensions complicates data selection
  - Mixed dimensions a particular challenge
- Code labels too long to be displayed on the screen

Unit of measure

- In the current SDG DSD, UNIT_MEASURE is a mandatory attribute of the series that provide the unit in which data values are expressed
  - Attributes provide additional information about data, but are not used to identify a series or observation
  - Unit of measure was made an attribute based on experience with the MDG data exchange

- With hindsight, UNIT_MEASURE should have been made a Dimension, enabling a breakdown by unit of measure
  - Many pairs of number-rate series would have been easier to handle

NATURE and OBS_STATUS

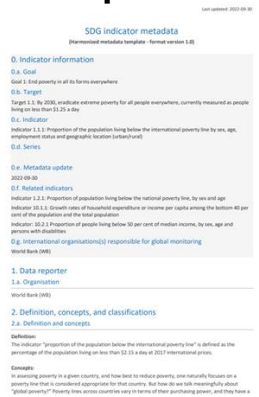
- NATURE: Information on the production and dissemination of the data (e.g.: if the figure has been produced by the country, estimated by an international agency, etc.)
  - Legacy attribute dating back to the MDG data exchange in early 2000s, widely used by reporters and consumers of the global SDG dataset.
- OBS_STATUS: Information on the quality of a value or an unusual or missing value
  - Widely used cross-domain concept and code list.
- Significant overlap between the two attributes and potential for contradiction
- Guidance to be developed explaining the relationship between and proper use of the two attributes; however, there are important questions on the use of the OBS_STATUS values in addition to the enormous amount of additional information to be collected from agencies.

Background SDG reference metadata

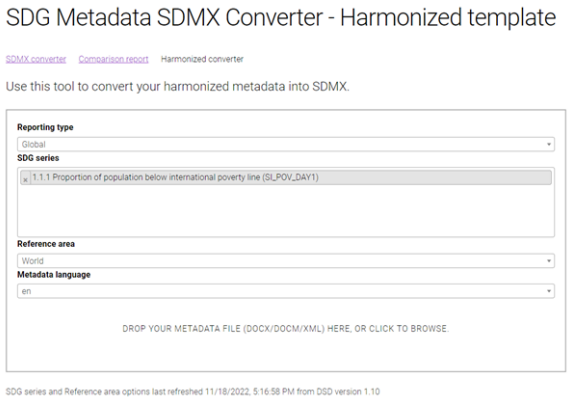
- The first official metadata structure definition for Goal indicators was published in February 2022.
- Reference metadata exchange is facilitated by a set of tools developed by UNSD in cooperation with the World Bank.
  - Retrieves metadata text from a Word template and converts to SDMX-ML
- The reference metadata set for the global Goal indicators is now published through an SDMX application programming interface.
- The reference metadata API provides machine-readable reference metadata
  - Facilitates the computer-assisted translation of Sustainable Development Goal reference metadata into a number of languages, as part of a World Bank project.
  - Used by reference metadata dissemination site, enabling the user to quickly make and download a selection of metadata

Process to publish metadata at the API

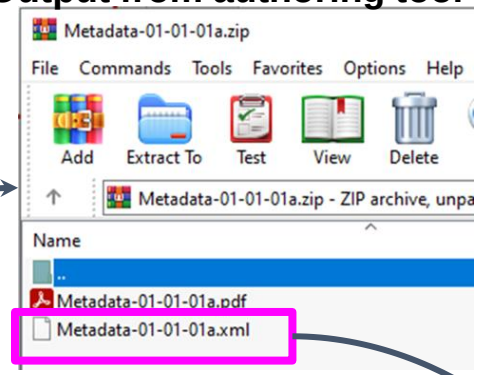
Metadata ready to be published



Authoring tool

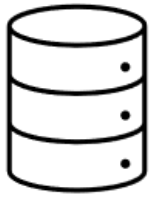


Output from authoring tool



Upload metadata into SDGLab database

SDMX Metadata API



SDGLab database

Challenges in the conversion of reference metadata

- Long and complex files
  - We currently have 258 metadata files

- Technical language used, often with graphics and equations
- Reference metadata are routinely updated
- Content of the reference metadata is maintained by custodian agencies
  - Each one has a style (= different formatting)

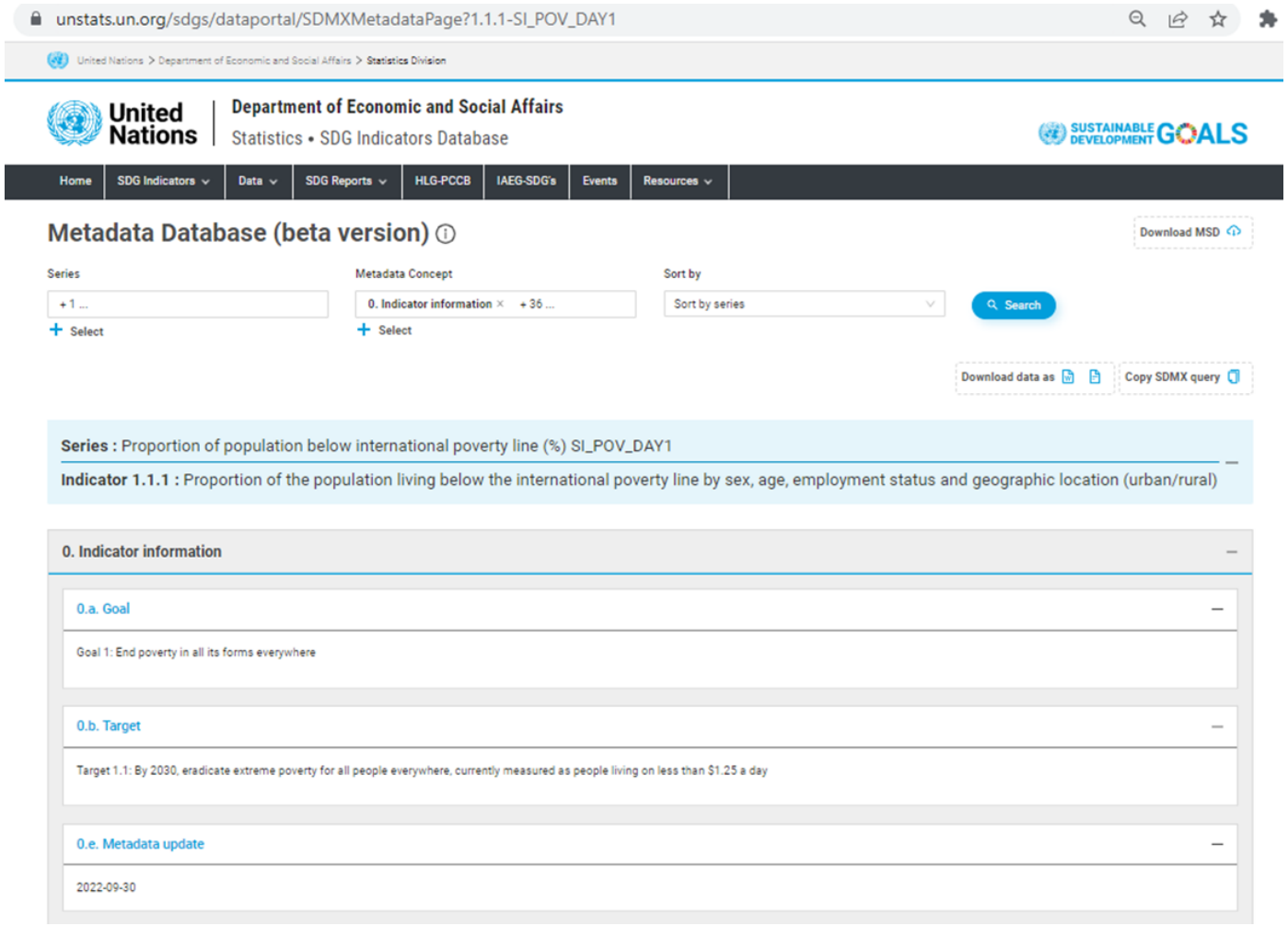
Standardization was needed to ensure that the process worked!!

Accessing metadata through the interface

Metadata for each indicator can be now accessed from the Global SDG indicators database: https://unstats.un.org/sdgs/dataportal/database

Thanks to the APIs, data and metadata can now be linked and shown together.

Metadata can also be accessed in the Metadata Database interface https://unstats.un.org/sdgs/dataportal/SDMXMetadataPage



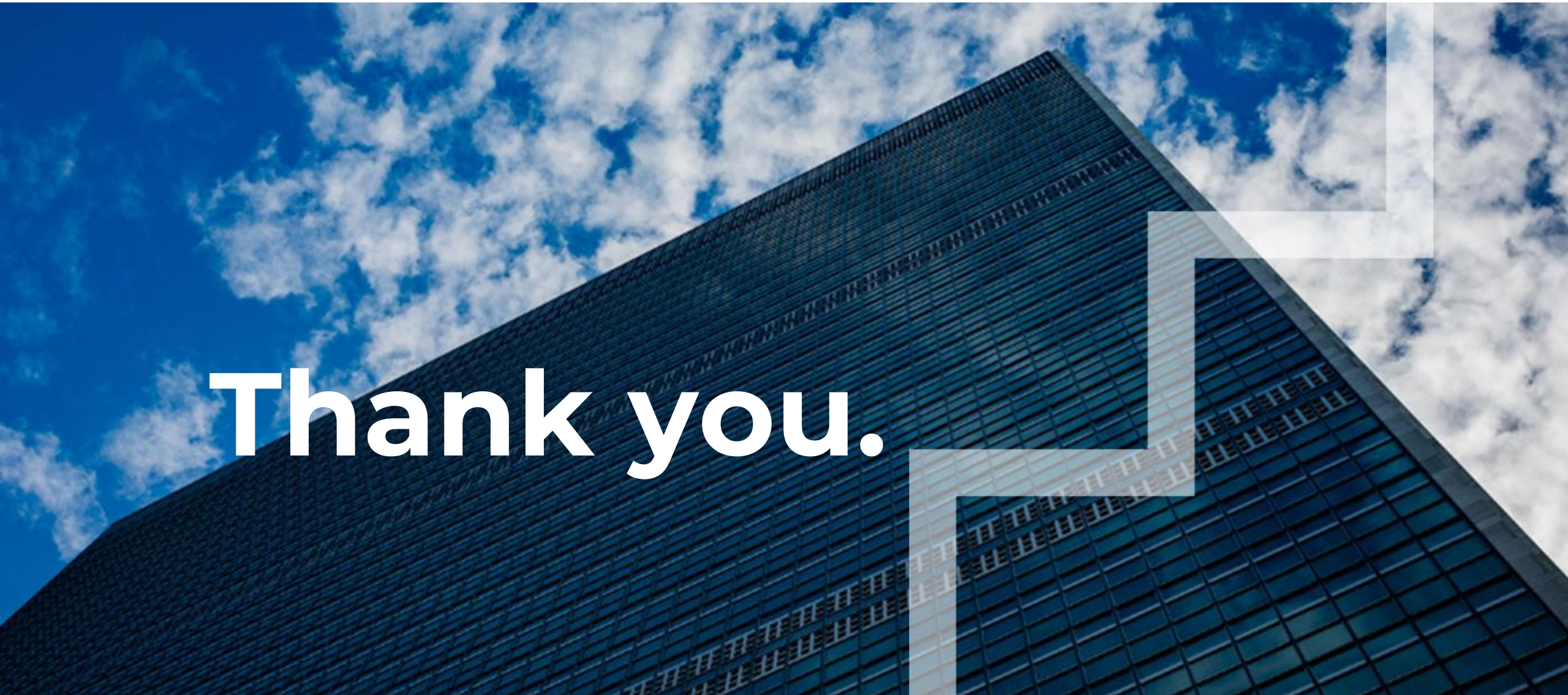
Lessons learnt

- Validation of datasets by reporter is a key benefit to both the reporter and collector
- Need of coordination and constant communication
- Availability of tools – metadata for example needed to be developed
- Structures, supporting documentation and software tools must be maintained as an integrated set
- Training for reporters is also needed
  - Some agencies do not have the capacity to adopt SDMX

What needs to be addressed

- Improve support for highly multi-dimensional datasets
  - Reduce disruption to data exchange caused by updates to dimensionality of a data structure
  - A paper based on the SDG data exchange experience proposes a solution to the long-standing problem

- Implement VTL in commonly used SDMX tools
  - Make content validation available to the reporter, sparing both reporter and receiver from follow-up on issues with the data


Thank you.