Global SDG Data Compilation with SDMX
A case Study

17th April 2018

Eduardo Belinchon
United Nations Statistics Division

What is SDMX

- "SDMX, which stands for Statistical Data and Metadata eXchange is an international initiative that aims at standardising and modernising ("industrialising") the mechanisms and processes for the exchange of statistical data and metadata among international organisations and their member countries." *Source Wikipedia*

sdmx
Statistical Data and Metadata eXchange

SDMX sponsors

- Bank for International Settlements (BIS)
- European Central Bank (ECB)
- Eurostat (the statistical office of the European Union)
- International Monetary Fund (IMF)
- Organisation for Economic Co-operation and Development (OECD),
- United Nations Statistics Division (UNSD)
- World Bank

What is SDMX

- Data modeling standard :
  - How we structure our data
- Technical Standard
  - How do we exchange our data
- Concepts
  - Dimensions, Attributes..
- Code Lists
  - Series, dimension values, attribute values…
- Data flows
- Validation, transformation, etc..

Data modeling:

1.2.1 - Proportion of population living below the national poverty line, by sex and age

USA : Percent of US population living below the US poverty line

https://sdg.data.gov/

year	tot_proportion_belownatpov_line	male	female	under_18_years	18_64_years	65andover_years	Note
2000	11.3	9.9	12.6	16.2	9.6	9.9	Implementation of 2000 …
2001	11.7	10.4	12.9	16.3	10.1	10.1	
2002	12.1	10.9	13.3	16.7	10.6	10.4	

1.2.1.a - (A)Proportion of population living below the national Upper poverty line, by region

Bangladesh

http://www.sdg.gov.bd/

Time Period	National / Disaggregation	Value/Description	Remarks
2000	Total	48.9	
2010	Total	31.5	
2016 (base Year)	Total	24.3	
	Urban	18.9	
	Rural	26.4	

- Data representation
- Naming
- Dimensions & Series Codes

Data modeling in SDMX:

REF_AREA	SERIES	URBANISATION	TIME_PERIOD	OBS_VALUE	TIME_DETAIL	NATURE	SOURCE	COMMENT_OBS
AF	SI_POV_NAHC	R	2007	38.2	2007	CA	World Bank Source	Poverty figures are not comparable due to seasonality and changes in the questionnaires used The national poverty line is absolute. Data are produced by World Bank. Most Recent Comparable Series:2007,2011. The complete series can be downloaded from the databank, http://databank.worldbank.org/data/reports.aspx?source=poverty-and-equity-database.
AF	SI_POV_NAHC	U	2007	28.9	2007	CA	World Bank Source	Poverty figures are not comparable due to seasonality and changes in the questionnaires used The national poverty line is absolute. Data are produced by World Bank. Most Recent Comparable Series:2007,2011. The complete series can be downloaded from the databank, http://databank.worldbank.org/data/reports.aspx?source=poverty-and-equity-database.
AF	SI_POV_NAHC	_T	2007	36.3	2007	CA	World Bank Source	Poverty figures are not comparable due to seasonality and changes in the questionnaires used The national poverty line is absolute. Data are produced by World Bank. Most Recent Comparable Series:2007,2011. The complete series can be downloaded from the databank, http://databank.worldbank.org/data/reports.aspx?source=poverty-and-equity-database.

Global SDG Compilation

Resolution adopted by the General Assembly on Work of the Statistical Commission pertaining to the 2030 Agenda for Sustainable Development (A/RES/71/313):

"Requests the Secretary-General to continue to maintain the Sustainable Development Goals global indicator database to inform the yearly progress report on the Goals and to ensure transparency on the data, statistics and metadata presented on countries and used for the regional and global aggregates "

Global SDG Compilation

- 50+ Agencies
- 132 Indicators with data
- ~ 460,000 observations

First wave

- Short time solution : Use of the MDG DSD
- Excel based exchange
- Excel templates manually generated
- SDMX Tools for data upload but manual validation
- Quick set up but not user friendly
- Data providers not sure about non-required dimensions and other fields

Second Wave

- Custom build solution
- User friendly interface for Excel upload
- Integrated version control and release system for archival
- Based on SDG DSD with extra dimensions
- Excel templates auto generate from DB with previous data
- Not fully aligned with SDMX Standard but can be converted without data loss.

Lessons learnt

- SDG is far more complex than MDG
- Indicator updates my pose a challenge for SDMX implementation
- Tools are good but custom solution is ideal (70% increase of productivity)
- Focus on the use of agreed naming conventions : Code Lists
- Not always obvious which values should be used in some dimensions. (_T, NA, F) vs No dimension
- SDMX has limitations for data dissemination (lower & upper bounds, XML, Json, …)

SUSTAINABLE DEVELOPMENT GOALS

Thank you !