Indonesia's Statistical Data Standards:

A Practical Combination of Statistical Ontology Implementation with SDMX Structure

SDMX Global Conference

Tri Listianingrum, Timotius Vincent Gunawan, David Setya Prehandoko
BPS--Statistics Indonesia

Backgrounds

Indonesia’s One Data Policy in 2019: aimed to create integrated, good-quality government data that can be easily **accessed** and **exchanged**.

4 key indicators:

- compliance with data standards → **statistical data standards (SDS)**, constructed by BPS
- metadata reporting
- data interoperability
- use of reference codes and/or master data.

Statistical Data Standards

A set of standardised concepts, definitions, classifications, measurement, and unit of measurement to produce standardised statistical data.

Concepts (Statistical Terms)
Generic idea of a data.

Definitions
Detailed description of a data, defining the scope and the boundaries of a data.

Classifications (Breakdown)
A standardised level of disaggregation to present data in a detailed manner.

Measurements
A unit which is used to measure a data.

Units of Measurement
Standard quantity which are used as a standard of measuring.

SDS and SDMX

SDS	SDMX
• Statistical data standards is used as a guideline for data producers about how the data should be produced. • Focuses more on the consistency of concepts and definitions • Used across government institutions in Indonesia	• SDMX is used as a standard format for reporting and dissemination to make the exchange of statistical data and metadata much easier. • SDMX helped the data structure more consistent because of the standardised format • Used internationally

The integration of SDS with SDMX will enhance the impact in standardising statistical data production process.

SDS-SDMX Artefacts Correlations

- Concepts and definitions will be treated as reference metadata
- Classifications will refers to SDMX Global Registry, if exist. If it does not exist yet, BPS will create code lists corresponding to the existing classifications. For disaggregation, it can use the code lists combined with content constraints.
- The measurement and its unit will refer to SDMX’s code lists as well, and they will be combined with the unit multiplier.
- The overall SDS will be constructed as general DSD, meanwhile the content constraint for more detailed view will be created as dataflow in metadata reporting.

How SDS are used in One Data Policy

- Ideally, SDS should be proposed around 2 years before the data collected.
- At the end of the year, BPS will set the accepted SDS proposal so it can be used by all government institutions.
- The data does not refer to the SDS directly, but the SDS will be **referred in the metadata**. If the metadata refers to the SDS, then we can say that the data are standardised.
- If the data are standardised, then it’s **easier to interoperate** the data across all government institutions.

Some Example of SDMX Implementation in SDS

Classification of sex (**332**20007) →
SDMX:CL_SEX(2.1)
M. Male
F. Female

Classification of institutional sector (**230**20021)
→ ESTAT:CL_SECTOR(1.14)
S11. Non financial corporations
S12. Financial corporations
S13. General government
S14. Households
S15. Non profit institutions serving households
S1N. Not sectorised

Classification of marital status (**101**20063) →
SDMX:CL_CIVIL_STATUS(1.0)
S. Single
M. Married
D. Divorced
W. Widowed

Example of data from IAEG-SDGs:CL_SERIES(1.21) that have been standardised:

CL_SERIES	CL_SERIES Description	SDS Code	SDS Title
SE_TRA_GRDL	Proportion of teachers with the minimum required qualifications, by education level and sex [4.c.1]	**103**10028	*Persentase Guru yang Memenuhi Kualifikasi Sesuai Standar Nasional*
SE_ACS_ELECT	Proportion of schools with access to electricity, by education level [4.a.1]	**103**10071	*Persentase Satuan Pendidikan dengan Akses Listrik*
SH_DYN_MORT	Under-five mortality rate [3.2.1]	**104**10007	*Angka Kematian Balita (AKBa)*/Under-Five Mortality Rate (U5MR)
SH_DYN_IMRT	Infant mortality rate [3.2.1]	**104**10008	*Angka Kematian Bayi (AKB)*/Infant Mortality Rate (IMR)
SG_XPD_HLTH	Proportion of total government spending on essential services, health [1.a.2]	**104**10133	*Persentase Pengeluaran Pemerintah untuk Layanan Pokok Kesehatan*
SP_DYN_MRBF15	Proportion of women aged 20-24 years who were married or in a union before age 15 [5.3.1]	**104**10144	*Persentase Perempuan Berumur 20—24 Tahun yang Usia Kawin Pertama atau Usia Hidup Bersama Pertama Sebelum Umur 15 Tahun*
AG_PRD_FIESMS	Prevalence of moderate or severe food insecurity in the population [2.1.2]	**105**10016	*Prevalensi Penduduk dengan Kerawanan Pangan Sedang atau Berat*
AG_LND_FRST	Forest area as a proportion of total land area [15.1.1]	**241**10226	*Persentase Kawasan Hutan*
SI_POV_NAHC	Proportion of population living below the national poverty line [1.2.1]	**331**10020	*Persentase Penduduk Miskin*

Challenges

- Determining the scope of data standardisation in SDS
- Increasing awareness and participation of other government agencies/institutions to propose and implement national standards for the data produced
- Developing standard codes for classification outside the SDMX
- Updates on classification domain values. e.g. ISIC rev. 5 to replace ISIC rev. 4 or SDMX code lists update.
- Maintaining standards comparability between SDS regulations
- Implementing the whole SDMX standards until DSD implementation and data constraint through collaboration among units in BPS

Thank you!