CountryData
Technologies for Data Exchange

MDG Data Structure Definition and CountryData

Nature of MDGs

- Covers topics from across the statistical spectrum (i.e. social, environmental and economic) as well as across time
- Attracts some users with limited experience in interpreting statistics
- This makes it difficult to standardize and present concisely to users

History of SDMX versions

Version 1.0 GESMES/TS
ISO/TS 17369

Version 2.0
SDMX-EDI
SDMX-ML
SDMX Registry

2008
SDMX accepted
at UN level

Version 1.0	Version 2.0	SDMX recognised and supported as the preferred standard	SDMX 2.1
September 2004	November 2005	February 2008	April 2011

MDG Data Structure Definition

- Supports exchange of MDG Indicator data between international agencies (UN, UNICEF, UNESCO, …)
- Developed by SDMX Task Team of Interagency and Expert Group on Development Indicators (IAEG)
- Implemented in SDMX 2.0

MDG DSD (cont'd)

Concept
Age group
Footnotes
Frequency
Location
Nature of data points
Observation Value
Reference Area
Series
Sex
Source details
Source Type
Time Period
Time period details
Unit multiplier
Units of measurement

MDG DSD (cont'd)

Type	Concept	Format
Dim	Frequency	i.e. Annual, Quarterly, etc.
Dim	Series	Indicator title
Dim	Units of measurement	i.e. Percent, number
Dim	Location	i.e. Total, Urban, Rural
Dim	Age group	i.e. 15–49 yr olds, under 5 years old
Dim	Sex	i.e. Total, male, female
Dim	Reference Area	Country name
Dim	Source Type	i.e. Survey, census, admin.
Dim	Time Period	i.e. 1990, 1991, etc.
Mea	Observation Value	-
Att	Unit multiplier	i.e. per 10,000, per 1,000 etc.
Att	Time period details	i.e. 2001 – 2003, Q1 2010 – Q3 2011
Att	Nature of data points	i.e. Estimated, Modelled, Adjusted etc.
Att	Source details	Source name & date
Att	Footnotes	Details of methodology & other notes etc.

MDG DSD (cont'd)

- Single DSD used for all MDG indicators
  - Support for diverse indicators means not all dimensions are applicable in all cases
    - E.g. Age Group is not applicable to indicator "Telephone lines."
    - Value NA is used when an dimension or attribute is not applicable.

CountryData DSD

- Based on the MDG DSD, uses the same dimensional structure (dimensions/attributes)

- Some codelists have been extended to support non-MDG development indicators in the project

Dimension: Frequency (FREQ)

- "Indicates rate of recurrence at which observations occur (e.g. monthly, yearly, biannually, etc.)."
- Both CountryData and MDG currently support annual frequency.

Dimension: Frequency (FREQ)

- However for development indicators, it is customary to set as annual (A).
- From “Cross-Domain Code Lists”;
- “This can also serve cases of multi-annual data (data that appear once every 2, 3 or, possibly, 5 years).”

Dimension: Series (SERIES)

- "The phenomenon or phenomena to be measured in the data set ... SERIES are all the official MDG series and background series currently in the MDG Database."
- Equivalent of Indicator
  - Not to be confused with SDMX time series
- Codelist extended to country-specific indicators in CountryData

Dimension: Unit of Measure (UNIT)

- "Dimension by which the series are described (e.g.: percentage, USD, etc.)"
- It may not be obvious which is the correct unit in some cases. In CountryData, units to be used in time series, are agreed in advance.
- Changes to the code list have recently been approved by the IAEG

Dimension: Age group (AGE_GROUP)

- “Age - or age range - of the individuals the observation refers to.”

- Use NA where not applicable

Dimension: Sex (SEX)

- "Gender condition: male or female. This dimension applies only if data can be disaggregated by sex."

- Use NA where not applicable

Dimension: Reference Area (REF_AREA)

- "Specific areas (e.g. Country, Regional Grouping, etc) the observed values refer to. Reference areas can be determined according to different criteria (e.g.: geographical, economic, etc.)."

- In MDG and CountryData, only country areas are used. Sub-national or regional data is not currently supported.

Dimension: Urban/Rural location (LOCATION)

- "Refers to a disaggregation within the Reference Area the data alludes; normally National (Total), Urban or Rural…"

- Has 3 codes: Total (National Level), Urban, Rural

Dimension: Source Type (SOURCE_TYPE)

- "Type of data source: survey, administrative records, census or other. Details about the source, if available, could be provided in the free-text attribute SOURCE_DETAIL."

- Little used in CountryData, but can be provided where available (default = NA)

Time Dimension: TIME_PERIOD

- "Reference date - or date range - the observed value refers … For MDG data exchange it is usually expressed as a four-digit year (e.g.: 1995) … details about such period should be specified in the attribute TIME_DETAIL."

- CountryData requires a four-digit year. Further info must be placed in TIME_DETAIL.

Primary Measure: Observation value (OBS_VALUE)

- Used to convey the value of a variable at a period of time
- Should be a floating-point number. Textual values are not accepted in CountryData

Attribute: Unit Multiplier (UNIT_MULT)

- "Exponent in base 10 that multiplied by the observation numeric value gives the result expressed in the unit of measure."
- If the observation value is in millions, unit multiplier is 6; if in billions, 9, and so on. Where the number is simple units, use 0.
- Mandatory attribute (default = 0)

Attribute: Time Period Details (TIME_DETAIL)

- "When TIME_PERIOD refers to a date range, this attribute is used to provide METADATA on the actual range the observation refers to (e.g. for period '2001-2003' TIME_PERIOD would be 2002 but the actual dates --2001-2003-- would be expressed here)."

- Optional free-text attribute

Attribute: Nature of data points (NATURE)

- "Information on the production and dissemination of the data (e.g.: if the figure has been produced and disseminated by the country, estimated by international agencies, etc.)"
- For exchange in CountryData, should be set to C (Country Data).
- Optional attribute

Attribute: Source details (SOURCE_DETAIL)

- "Type of data source: survey, administrative records, census or other. Details about the source, if available, could be provided in the free-text attribute SOURCE_DETAIL."

- Optional free-text attribute

Attribute: Footnotes (FOOTNOTES)

- "Additional information on specific aspects of each observation, such as how the observation was computed/estimated or details that could affect the comparability of this data point with others in a time series."

- Optional free-text attribute

Attribute Attachment Level

- MDG DSD specifies that all attributes are attached to observations

MDG/CountryData DSD: Mappings

- Due to its support for heterogeneous indicators, it's not always obvious which values should be used in some dimensions

- What should be SEX in indicator "Births attended by skilled personnel":
  - Not Applicable? Total? Female?

MDG/CountryData DSD: Mappings (2)

- Inconsistent mappings lead to duplications and other anomalies
- In CountryData, mappings for indicators/time series are agreed before data exchange

MDG/CountryData DSD: Developments

- A new version of the MDG DSD, with modified codelists, was approved by IAEG in February 2013.
- Upgrade to SDMX 2.1 is being considered

CountryData DSD maintenance

- Currently, CountryData codelists are maintained by UNSD
  - Eventually, countries will be offered to maintain their own codelists for CountryData project
- Periodically, the DSD needs to be updated

Reference metadata

ID	Description
DATA_PROVIDER	Country Name
SERIES	Indicator Title
STAT_CONC_DEF	Definition of the MDG official indicator or background series provided
METHOD_COMP	Method of computation
COMMENTS_LIMITATIONS	Comments and limitations
DISCREPANCIES	Sources of discrepancies between global and national figures
COLL_METHOD	Process of obtaining data
REL_CAL_POLICY	Expected time of release