SDMX-HD

Introduction

SDMX (Statistical Data and Metadata eXchange) is an ISO standard for exchanging and sharing statistical data and metadata among organizations.

SDMX-HD (Health Domain) is a WHO implementation of the SDMX standard to allow medical facilities to share and exchange medical indicators and metadata between medical organizations.

The rest of this page will provide some documentation and important links that will allow people to become more familiar with the SDMX-HD standard.

High-Level Overview

The following is taken from the SDMX user guide:

ISO/TS 17369:2005, Statistical Data and Metadata Exchange (SDMX) describes statistical metadata through a data structure definition, or DSD, which defines concepts, dimensions, attributes, code-lists and other artefacts necessary to describe the structure of data. In a parallel manner, a metadata structure definition, or MSD, describes metadata associated with data at observation, series, group, and dataset levels.

When these SDMX structures are combined with metadata attributes recommended in international standards and harmonized across programmes, a message that will satisfy a majority of requirements can be developed.

Useful SDMX-HD links and documentation

- SDMX user guide
  - This large document and Section B.9 is good for a brief understanding.
- SDMX-HD Google Group
- SDMX-HD documentation and examples: SDMX-HD1-0_draft03.zip
  - Sample 1 and Sample 2 are good examples to have a look at. Sample 1 shows a SDMX-HD message without actual observation values whereas Sample 2 shows the same example with observation values. (see sample1.html and sample2.html for more information)

SDMX-HD structure overview
Useful SDMX terminology

This excerpt was taken from the official SDMX user guide:

A.3.4 The SDMX information model for data in a nutshell

The SDMX standards are based on the SDMX information model (SDMX-IM) which represents statistical data and metadata. The list below describes the minimal knowledge needed about the SDMX information model so that we can start developing data structure definitions based on the SDMX standard:

Descriptor concepts: In order to make sense of some statistical data, we need to know the concepts associated to it (for example, the figure 1.2953 alone is pretty meaningless, but if we know that this is an exchange rate for the US dollar against the euro on the 23 November 2006, it starts to make more sense).

Packaging structure: Statistical data can be grouped together. The following levels are defined: the observation level (the measurement of some phenomenon), the series level (the measurement over time of some phenomenon, usually following a regular interval), the group level (group of series. A well-known example is the sibling group which contains a set of series which are identical except that they are measured with different frequencies) and the data set or data flow level (made up of several groups, for instance to cover a specific statistical domain). The descriptor concepts mentioned in point 1 can be attached at various levels in this hierarchy.

Dimensions and attributes: There are two types of descriptor concepts: the ones which both identify and describe the data are called dimensions, and those which are purely descriptive are called attributes.

Keys: Dimensions are grouped into keys, which allow the identification of a particular set of data (for example, a series). The key values are attached at the series level, and are given in a fixed sequence. By convention, frequency is the first descriptor concept, and the other concepts are assigned an order for that particular data set. Partial keys can be attached to groups.
**Code lists:** Each possible value for a dimension is defined in a code list. Each value on that list is given a language-independent abbreviation (a code) and a language-specific description. Attributes are sometimes represented with codes, but sometimes represented by free-text values. This is fine as the purpose of an attribute is solely to describe and not to identify the data.

**Data Structure Definition:** A Data Structure Definition (key family) specifies a set of concepts which describe and identify a set of data. It tells which concepts are dimensions (identification and description), and which are attributes (just description), and it gives the attachment level for each of these concepts, based on the packaging structure (Data Set, Group, Series, Observation) as well as their status (mandatory versus conditional). It also specifies which code lists provide possible values for the dimensions, as well as the possible values for the attributes, either as code lists or free text fields.