The Basics of SDMX and Possible Applications

Alex Blackburn, UNSD

Oslo Group 9th meeting

Abu Dhabi, 7th May 2014

United Nations
Statistics Division

Structure

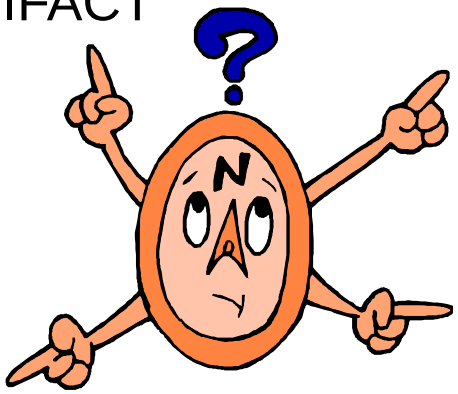
- What SDMX is
- A practical UNSD example
- Why apply to energy statistics? How??

Lack of standardisation in data exchanges or across organisations

Different **formats** of data and metadata

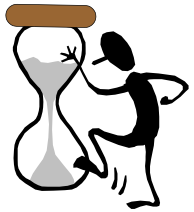
EDIFACT

XML



paper form

SQL



Different **places** to store data and metadata

Different **media**

Email

Web-form

dial-up

removable media

file upload

Paper

SDMX

ISO IS 17369

World Bank

WHAT SDMX IS

- ✓ **Standard Data and Metadata eXchange**: A model to describe statistical data and metadata
- ✓ A standard format for automated communication using XML, helping the exchange of data and metadata
- ✓ SDMX aims to ensure that data and metadata remain linked to each other

To do this:

- ✓ Statisticians agree to use a common description for data and metadata
- ✓ Data descriptions are made available for everybody who wants to understand and reuse the data

SDMX as an international standard

- In February 2008, the 39th session of the **United Nations Statistical Commission** recognised and endorsed the use of SDMX:

> *"the preferred standard for exchanging and sharing data and metadata in the global statistical community."*

Advantages from the points of view of producers and users of statistics

- Easier to automate collection, processing and dissemination of data files
- Ability to generate and manipulate data and metadata using generic IT tools
- More agile transmission of statistical data, data transfer agreements
  - Less waiting time (from days to seconds)
- Harmonisation of statistical variables - Better international and cross-domain comparability

Better data quality (hopefully)

DSDs

- SDMX Data Structure Definition (DSD) describes the characteristics of data to be exchanged, similar to a database structure
- Major components:
  - Dimensional structure
  - Concepts
  - Code lists
- In SDMX, a DSD is a prerequisite for any data exchange.
  - Just like a database structure needs to be created before a database can be used
- Each dataset to be shared needs a DSD associated with it.

Are there any DSDs available?

- A number of DSDs have been developed and approved for global use
  - National Accounts, Balance of Payments, and Foreign Direct Investment
  - Millennium Development Goals
- Some have been derived from global DSDs
  - CountryData
  - UIS Dataset
- Work is underway on other global DSDs
  - IMTS (Expected 2015)

DSD: IMTS Example

Role	Id	Name	Example	Description	Presentation
Dimension	FREQ	Frequency	Annual	The time interval at which observations occur	CODE LIST: CL_FREQ
Dimension	REF_AREA	Reference area	France incl. Monaco	The country or geographic area to which the measured statistical phenomenon relates.	CODE LIST: CL_AREA
Dimension	TIME_PERIOD	Reference period	1999	The period of time to which the measured observation is intended to refer.	Gregorian time period: Represented as YYYY, YYYY-MM or YYYY-MM-DD,
Dimension	TRADE_FLOW	Trade flow	Exports	Trade flow or sub-flow (exports, re-exports, imports, re-imports, etc.)	CODE LIST: CL_TRADE_FLOW
Dimension	COMMODITY	Commodity	Gadget	Commodity code or commodity group code (its composition includes a prefix that identifies the commodity classification)	CODE LIST: CL_COMMODITY
Dimension	COMMODITY_CUSTOM_BREAKDOWN	Custom commodity breakdown		"Dummy" identifier for country-specific commodity code beyond 6-digit HS	CODE LIST: CL_COMMODITY_CUSTOM_BREAKDOWN
Dimension	COUNTERPART_AREA_1	Primary partner area	USA	The primary partner country or geographic area for the respective trade flow	CODE LIST: CL_AREA
Attribute	COUNTERPART_AREA_1_TYPE	Type of primary partner area	Destination	Type of primary partner country or area	CODE LIST: CL_PARTNER_TYPE
Dimension	COUNTERPART_AREA_2	Additional partner area	Netherlands	A secondary partner country or geographic area for the respective trade flow	CODE LIST: CL_AREA
Attribute	COUNTERPARTS_AREA_2_TYPE	Type of additional partner area	Consignment	Type of secondary partner country or area	CODE LIST: CL_PARTNER_TYPE
Dimension	TRANSPORT_MODE	Mode of transport		The mode of transport used when goods enter or leave the economic territory of a country	CODE LIST: CL_TRANSPORT_MODE

What if a DSD is not available?

- Someone has to develop a DSD when there isn't one
  - … and maintain it!
- The DSD should comply with agreed standards as much as possible
  - Cross-domain concepts and code lists

What is required of data providers?

- Commitment
- Patience
- Willingness to compromise on non-vital issues
  - E.g. by convention SDMX uses ISO codes for reference areas.

SDMX Implementation in UNdata

- Web Service
  - http://data.un.org/WS
- Web Client
  - http://data.un.org/SdmxBrowser
- 3 Datasets online
  - International MDG Database
  - CountryData
  - UIS database on education, literacy, science, culture, and communication

UNSD-DfID (UK) Project

- ✓ To improve the coherence and clarity of dev't indicators
  - ✓ Improve coordination in the NSS
  - ✓ Collate development data in one place
  - ✓ Explain differences between intl. & nat. data

- ✓ Improve accessibility and visibility
  - ✓ Make access to national data easier
  - ✓ Draw attention to wider set of indicators
  - ✓ Reduce data request burden

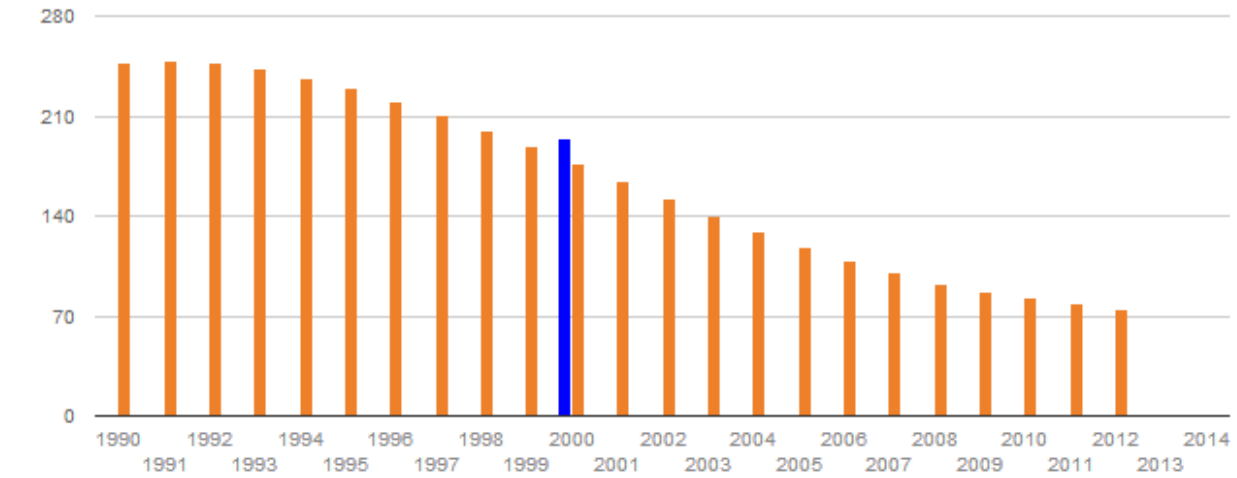
- ✓ Enhance knowledge
  - ✓ Strengthen IT support
  - ✓ Training & skills development

countryData

- Web portal developed for the project, part of the UNdata platform
- Automatically receives and publishes data from countries' registries
  - Provides re-dissemination and visualization of national data
  - Facilitates, and displays the result of, analysis of discrepancies between data from national sources and international agencies.
- Focus on reference metadata as the key to understanding the nature of national data and sources of discrepancies.
- Countries control what is published at CountryData
  - Data is published by the countries without any intervention by UNSD

Under-five mortality rate

Age group: All age ranges **Location:** Total (national level) **Sex:** Both sexes **Units of measurement:** Per 1,000 live births



Differences

Different data sources

Different methodologies

Last update: 14-Feb-2013 @ 05:54

Why is there a difference?

Country data for 2000 comes from Demographic and Health Surveys (DHS). International figures are calculated jointly by UNICEF, World Bank,WHO and UNPD based on models fitting all available data/sources for a given country. Data sources used for Liberia were 1986 and 2007 DHS.

Data | Definition | Method of computation | Comments / limitations | Discrepancies | Collection Method | Release calendar

SDMX Implementation

- ✓ Perceived to be overly complex and difficult to learn
  - ✓ New tools, e.g. Eurostat's SDMX-RI, greatly simplify implementation and lower cost.
  - ✓ Proficiency in SDMX is still required.
- ✓ SDMX exchange established, for the most part, between international organisations or advanced countries and international organisations.
- ✓ So this project gave participants their first SDMX experience with well-known and easy-to-use platform and tools, such as DevInfo software or their own existing systems

How Can We Apply SDMX to Energy? Why would we want to?

- ✓ Would provide a standard way of exchanging information between international organisations and countries
- ✓ Would reduce the reporting burden of countries and make international cooperation easier
- ✓ Data sharing develops a standard language, and becomes quick and simple

SDMX implementation in Energy

- ✓ A common DSD needed! Takes time and compromise
- ✓ Taking IRES as a base would be a good start
- ✓ An area for Oslo group (or InterEnerStat) cooperation?
- ✓ Any countries with experience?
- ✓ For other global DSDs, UNSD is the maintenance agency/secretariat. Could also be for energy (or another OG/InterEnerStat member)
- ✓ Needs to be driven by a genuine need from users

Further reading

countryData	http://data.un.org/countrydata
UNdata A world of information	http://data.un.org
UNdata SDMX API	http://data.un.org/WS