Data Transmission
Item 8 of the agenda

Eurostat

ESTAT-B5: Luca Gramaglia
ESTAT-F5: Giuliano Amerini

Statistical **D**ata and **M**etadata E**X**change

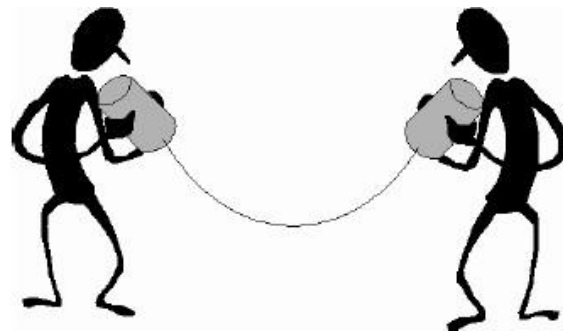## SDMX

ISO TS 17369

SDMX is a standard to...

Describe and model data:

- *Describes the concepts underlying a dataset and the relationships between them*
- *Common vocabulary and modelling principles across statistical domains*

Transmit data:

- *SDMX allows the construction of an IT infrastructure for data exchange*

Common language...   **...within and across statistical domains**

Describing data using SDMX: the DSD

Identify the concepts involved in the datasets and their role:

- **Dimensions**: Identify and describe the data

- **Attributes**: Provide additional information about data. May be attached at different levels (observation, dataset)

- **Measure**: Represents the phenomenon or phenomena to be measured

COUNTRY

FREQUENCY

TOURISM TOPIC

TIME

Number of touristic establishments in Italy, annual data

Indicator Time	A100 Hotels and similar	B010 Tourist Campsites	B020 Holiday dwellings
2002A00	33411	2374	61479
2003A00	33480	2530	58526
2004A00	33518	2529 E	56586
2005A00	33527	2411 P	68385

OBS_VALUE

OBS_STATUS

DIMENSIONS   ATTRIBUTES   MEASURES

Transmitting data using SDMX

- Transmission files must be structured according to the DSD

- The preferred transmission format is SDMX-ML

- Eurostat makes a variety of tools available to Member States to implement SDMX for data transmission

SDMX for ESSPROS

- SDMX to be implemented for ESSPROS data only (metadata postponed)

- SDMX conceptual framework (DSDs) proposed for ESSPROS, along with a reorganisation of the dataset identifier

- Definition of the new format for the datasets

SDMX for ESSPROS: DSDs

- ESSPROS data collection subdivided into 19 thematic units, 14 thematic units covered by SDMX implementation

- Conceptual analysis of the 14 thematic units: all in all, 30 concepts identified

- Code lists associated to the concepts

- 4 DSDs created to cover the 14 thematic units: a single DSD can describe multiple datasets if they share the same conceptual structure

SDMX for ESSPROS: DSDs

- Once finalised the DSDs will be uploaded on the SDMX registry, where they will be available to all users

- Updates once published are possible, but it is important the first version is of high quality

- The finalisation of the DSDs and associated code lists is a precondition for all subsequent implementation steps

Datasets and datasets identifiers

- ESSPROS composed of 7 datasets, 3 of which for data. Each has a specific DataSet Identifier (DSI) in eDAMIS:
  - **ESSPROS_QUANTI_A**
  - **ESSPROS_BENEF_A**
  - **ESSPROS_NET_A**

- Reorganisation and rationalisation of the datasets and DSI's proposed based on DSD structure, transmission deadlines and legal status

- 8 new DSIs in total

	Thematic units (see Table 1)	Legal deadline for transmission	Data set identifier	DSD Id
1	Classification of schemes	30th June N+2	1	SCHEME
2	Receipts data	30th June N+2	2	REC
3	Expenditure data by residency status (optional)	30th June N+2	3	EXP
4	Compulsory expenditure data	30th June N+2	4	
5	Compulsory expenditure data by detailed benefits	30th June N+2	4	
6	Optional expenditure data by detailed benefits	30th June N+2	5	
7	Compulsory net expenditure data	31st Dec.N+2	6	
8	Compulsory net expenditure data by detailed benefits	31st Dec.N+2	6	
9	AISCR data	31st Dec.N+2	6	
10	AITR data	31st Dec.N+2	6	
11	Residual fiscal benefits data	31st Dec.N+2	6	
12	Optional net expenditure by detailed benefits data	31st Dec.N+2	7	
13	Pension beneficiaries compulsory data	31st May N+2	8	BENEF
14	Pension beneficiaries optional data	31st May N+2	8	

New data transmission format

Data must be structured according to DSD

SOCIAL PROTECTION EXPENDITURE AND RECEIPTS

Czech Republic, 2010 Currency: Millions national currency

							Scheme 01
D	1120000	Social protection benefits					42,612.24
I	1121000		Non Means-tested				42,612.24
S	1121100			Cash benefits			42,612.24
A	1121110				Periodic		42,612.24
B	1121111					Disability pension	42,308.31
I	1121112					Early retirement benefit due to reduced capacity	303.93

```
2010,A,CZ,_T,EXP,_T,SPR,DIS,_T,_T,_T,C001,EXP_GRS,A,F,,CZK,6,3,,,P1Y,42809.240
2010,A,CZ,_T,EXP,_T,SPR,DIS,NMT,_T,_T,C001,EXP_GRS,A,F,,CZK,6,3,,,P1Y,42809.240
2010,A,CZ,_T,EXP,_T,SPR,DIS,NMT,CASH,_T,C001,EXP_GRS,A,F,,CZK,6,3,,,P1Y,42809.240
2010,A,CZ,_T,EXP,_T,SPR,DIS,NMT,CASH_PER,_T,C001,EXP_GRS,A,F,,CZK,6,3,,,P1Y,42809.240
2010,A,CZ,_T,EXP,_T,SPR,DIS,NMT,CASH_PER,DIS,C001,EXP_GRS,A,F,,CZK,6,3,,,P1Y,42493.488
2010,A,CZ,_T,EXP,_T,SPR,DIS,NMT,CASH_PER,ERRC,C001,EXP_GRS,A,F,,CZK,6,3,,,P1Y,315.752
```

New data transmission format

- Recommended format is SDMX-ML (XML-based)

- Implementation of SDMX-ML could be more problematic for countries where the final compilation of the questionnaires prior to transmission to Eurostat is not performed in National Statistical Offices

- Support for SDMX implementation always available at
  ESTAT-SUPPORT-SDMX@ec.europa.eu

Pilot exercise

- **Implementation of SDMX for ESSPROS at Eurostat:**
  - Started 2014, Planned completion 2015
  - Development of necessary infrastructure to start with the net module.
  - End-2014 Eurostat should be able to treat net module SMDX files.
- **Pilot transmission exercise for net module data**
  - Possible timing: end of 2014, in parallel with the official exercise
  - Precondition: finalisation and installation of DSD, code lists, DSIs etc.
  - Participants will be supported through specific bilateral contacts
  - Possible implementation of ad-hoc on-line training sessions
- **If pilot is successful:**
  - Other pilots could be organised (Core system, Pension beneficiaries)
  - SDMX framework could be gradually introduced as the official data transmission method for a number of countries.

Gentlemen's agreement

Implementation of SDMX represents a big investment for Eurostat and countries in terms of time and resources.

As many aspects are still to be analysed and assessed, Eurostat prefers that this is undertaken according to

- **A step-by-step approach in co-operation with the countries**
- **A Gentlemen's Agreement**

Transitional periods when both the traditional transmission method (Excel files) and the new one (SDMX framework) co-exist should be minimised.

SDMX will simplify statistical operations for all parties but advantages can only be achieved if all parties participate.

Items for discussion

- *Comments on the DSDs and associated code lists, with the aim of adopting them*

- *Comments on the new DSI structure, with the aim of adopting it*

- *Comments on the use of SDMX-ML*

- *Volunteers for pilot data transmissions starting in 2014*

- *Discussion on Gentlemen's agreement for SDMX implementation*

- ***EEV****: 14 countries replied (NSI, **non-NSI**) (**BE,DE,FR**,LV,LT,NL,AT,RO,SI,SK,SE,UK,NO,CH)*

For discussion: DSD and code list

Countries are invited to comment on the first draft of the DSD and associated code lists for ESSPROS, under the SDMX framework, with the aim of adopting them. *(EEV)*
(Box 2 of DOC SP-2014-08)

DE, FR, LV, LT, RO, SI, SK, UK, NO agreed with the draft proposal.
CH trusts Eurostat experts.

> ➢ **FR** proposes **:**
> 1) more details (S.11 S.12) in the CL_SECTOR to allow links with National accounts
> ➔ *in principle not required by ESSPROS manual (optional?: to evaluate)*
> 2) Code "unclassified" for the Classification of schemes
> ➔ *in principle NO (all schemes have to be classified)*

Other comments ?

Today: Datasets and datasets identifiers (DSI)

Today ESSPROS composed of 7 datasets, each has a specific DataSet Identifier (DSI) in eDAMIS:

- **Core System – Data**   **ESSPROS_QUANTI_A**
- **Core System – Qualitative information**   **ESSPROS_QUALI_A**
- **Core System – Quality report**   **ESSPROS_QUALREP_A**
- **Pension Beneficiaries data**   **ESSPROS_BENEF_A**
- **Pension Beneficiaries Quality Report**   **ESSPROS_QRPENB_A**
- **Net Benefits**   **ESSPROS_NET_A**
- **Net Benefits Quality report**   **ESSPROS_QRNET_A**

Note: ESSPROS = eDAMIS domain

A = Annual data

Data datasets; metadata datasets

Future: Datasets and datasets identifiers (DSI)

DSIs "QUANTI", "BENEF and "NET" will eventually be replaced with eight new DSIs

- **Illustrated in Table 3 of SP-2014-08-Annex 2**
- **New DSIs are currently numbered but have no names. These will be set once approval to adopt SDMX has been granted.**

Some data will be transmitted using the same DSD ("EXP") but different DSIs (3 to 7):

- **Distinguish data transmitted at different times during the year: gross end of June (DSIs 3-5), net is end of December (DSIs 6-7).**
- **Distinguish data with different legal statuses: compulsory data (DSIs 4 and 6), optional data (DSIs 3, 5 and 7).**

For discussion: new 8 DSI (for data)

Countries are invited to comment on the first draft proposal of the new DSIs to be used for future "data datasets" transmissions, under the SDMX framework, with the aim of adopting them. *(EEV)*
(Box 3 of DOC SP-2014-08)

DE, LV, LT, RO, SI, SK, UK, NO CH agreed with the draft proposal.

> ➤ **NL** proposal :
> call the core system data ESSPROS_CORE.
> All the look alikes in 'QUA..' is avoided

Other comments ?

For discussion: SDMX-ML

Countries are invited to inform Eurostat about the current knowledge and use of SDMX-ML in their Statistical Office and (where relevant) the Institution in charge of data transmission. (EEV)
(Box 5 of DOC SP-2014-08 – Part 1)

LV, RO have good knowledge of SDMX-ML

SI, SK, SE have knowledge of SDMX-ML

LT, NL,UK limited knowledge of SDMX-ML

BE, DE , FR, CH don't have knowledge of SDMX-ML

For the countries for which the Institution in charge of data transmission are the Ministries: please contact the NSI where there is knowledge (used in other domains)

Other comments ?

For discussion: SDMX strategy

Countries are invited to inform Eurostat about problems envisaged in the implementation of the SDMX strategy in general and, more specifically, concerning the use of the SDMX-ML file format. (EEV)
(Box 5 of DOC SP-2014-08 – Part 2)

FR not clear vision of the implied changes

CH worries for impact on statistical production

NL necessary time unpredictable; converter excel ➔ SDMX ?

UK significant costs for implementation (any EU funding?, information about the EU approach to meet the costs)

Pilot exercise

*Implementation of SDMX for ESSPROS **at Eurostat**:*

- **Started 2014, Planned completion 2015**
- **Development of necessary infrastructure to start with the net module.**
- **End-2014 Eurostat should be able to treat net module SDMX files.**

Pilot transmission exercise for net module data

- **Possible timing: end of 2014, in parallel with the official exercise**
- **Precondition: finalisation and installation of DSD, code lists, DSIs, ...**
- **Participants will be supported through specific bilateral contacts**
- **Possible implementation of ad-hoc on-line training sessions**

If pilot is successful:

- **Other pilots could be organised (Core system, Pension beneficiaries)**
- **SDMX framework could be gradually introduced as the official data transmission method for a growing number of countries.**

For discussion: volunteers for pilot

Countries are invited to inform Eurostat about their availability to participate in a pilot data transmission exercise for the Net Benefits data, using the SDMX infrastructure. (EEV)
(Box 6 of DOC SP-2014-08)

DE, LV, LT, RO, SI, SE, available for the pilot exercise according to the proposed timetable.

UK, NL interested if resources available.

SK in favour but no resources for end 2014

Other comments ?

Gentlemen's agreement

Implementation of SDMX represents a big investment for Eurostat and countries in terms of time and resources.

As many aspects are still to be analysed and assessed, Eurostat prefers that this is undertaken according to

- **A step-by-step approach in co-operation with the countries**
- **A Gentlemen's Agreement**

Transitional periods when both the traditional transmission method (Excel files) and the new one (SDMX framework) co-exist should be minimised.

SDMX will simplify statistical operations for all parties but advantages can only be achieved if all parties participate.

For discussion: Gentlemen's agreement

Countries are invited to express their engagement to implement the SDMX strategy as soon as possible. *(EEV)*
(Box 7 of DOC SP-2014-08)

DE, LV, LT, RO, SI, SK, SE, CH, agreed

BE, FR, NL, AT, UK, NL agreed in principle, provided that they dispose of the necessary time and resources

NO disagreed

Other comments ?

Way forward: 2014

1) To finalise DSD, code lists and DSIs

2) Upload in SDMX Registry

3) Organise tests with volunteers (training, right contacts)