



UNSD

**INTERNATIONAL CONFERENCE ON INTERNATIONAL OUTREACH AND
COORDINATION IN NATIONAL ACCOUNTS FOR
SUSTAINABLE GROWTH AND DEVELOPMENT**

6th to 8th May 2008

Luxembourg – Jean Monnet Building - Room M6

Eurostat and UNSD

DOCUMENT RELATING TO

SESSION 9: Common tools



SDMX

Statistical Data and Metadata eXchange

1. WHAT IS SDMX?

SDMX consists of technical and statistical standards and guidelines, together with an IT architecture and IT tools, to be used for the efficient exchange and sharing of statistical data and metadata. Seven European and international organisations (namely the Bank of International Settlements, the European Central Bank, Eurostat, the International Monetary Fund, the Organisation for Economic Co-operation and Development, the United Nations Statistical Division and the World Bank) act as sponsors of SDMX.

Standardised file formats for data and metadata and standardised contents of these files are the pre-condition for the automated production, processing and exchange of SDMX data and metadata files between national and international statistical organisations. For SDMX in general the preferred syntax is XML, but the SDMX standard also includes the GESMES/TS standard using EDIFACT syntax (now renamed SDMX-EDI).

The first version of the SDMX technical standard (V1.0), which was finalised in 2004, has been approved as an ISO standard (ISO/TS 17369:2005). SDMX V2.0 was approved by the sponsors in 2005.

SDMX is not just a format for data exchange. Taken together, the standards and content guidelines, together with the IT architecture and tools, can support improved business processes for any statistical organisation, enabling data to be transmitted, disseminated and shared in the most efficient way.

Forthcoming international meetings with coverage of SDMX

Three international statistical meetings in April 2008 include coverage of SDMX and related issues. The documents and presentations will be available through the following links:

- > [UNECE-Eurostat-OECD Meeting on Management of Statistical Information Systems, MSIS 2008](#)
- > [UNECE-Eurostat-OECD Work Session on statistical metadata, METIS 2008](#)
- > [OECD Expert Group on Statistical Data and Metadata Exchange](#)

Posted in: [events](#) |

UN Statistical Commission: SDMX is preferred standard

At the 39th Session of the UN Statistical Commission, attended by delegations from about 130 countries and 40 international agencies, SDMX was recognized as the preferred standard for the exchange and sharing of data and metadata. The Commission also encouraged implementation by national and international statistical organisations.

[Read the rest of this entry »](#)

Posted in: [events](#) |

Draft SDMX Content-Oriented Guidelines - Comments Welcome

Draft SDMX Content-Oriented Guidelines (February 2008) have been released for a three-month public comment period ending on 31 May 2008.

 [Content-Oriented Guidelines](#)

The Guidelines are supplemented by five annexes:

SDMX is an initiative to foster standards for the exchange of statistical information.

Sponsored by
BIS - ECB - EUROSTAT - IMF -
OECD - UN - World Bank

- > [Home](#)
- > [About us](#)
- > [FAQs](#)
- > [SDMX User Guide](#)
- > [Contact us](#)

Enter your e-mail address to sign up for further information about SDMX

2. WHAT ARE THE MAIN DELIVERABLES OF SDMX AND HOW DO THESE DELIVERABLES GET IMPLEMENTED IN THE EUROPEAN STATISTICAL SYSTEM?

What are structural metadata and reference metadata?

Structural metadata are those metadata acting as identifiers and descriptors of the data, such as names of variables or dimensions of statistical cubes. Structural metadata must be associated with the data, otherwise it becomes impossible to identify, retrieve and browse the data.

Reference metadata are metadata that describe the contents and the quality of the statistical data (conceptual metadata, describing the concepts used and their practical implementation, methodological metadata, describing methods used for the generation of the data, and quality metadata, describing the different quality dimensions of the resulting statistics, e.g. timeliness, accuracy). While these reference metadata exist and may be exchanged independent of the data and its structural metadata, they are often linked (“referenced”) to the data.

2.1. The SDMX standards used for data and structural metadata

Based on the SDMX information model (which is built on the GESMES information model), *data structure definitions* are created for particular statistical domains. This is done not only at the level of the SDMX sponsors (also involving their respective member countries), but also by all statistical organisations implementing SDMX.

When data collections common to several national or international organisations are concerned, the data structure definitions are developed jointly by those international and national organisations. For example, Eurostat and the European Central Bank (ECB) have worked together with their respective working groups to develop data structure definitions for national accounts and many other statistical domains. About half of the incoming data at Eurostat already use SDMX-compliant data structure definitions (as also the GESMES format is covered).

Existing data structure definitions reflect established practice in the different statistical domains. In future, the SDMX compliant data structure definitions should use more and more harmonised structural metadata (e.g. a harmonised list of countries, a harmonised list of economic activities, etc.). These lists will be successively released by the SDMX sponsors. It will however remain difficult to harmonise the codes linked to these lists between the SDMX sponsors.

A data file in SDMX-ML for Short Term Statistics:

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<CompactData xmlns="http://www.SDMX.org/resources/SDMXML/schemas/v2_0/message"
  xmlns:common="http://www.SDMX.org/resources/SDMXML/schemas/v2_0/common"
  xmlns:compact="http://www.SDMX.org/resources/SDMXML/schemas/v2_0/compact"
  xmlns:cross="http://www.SDMX.org/resources/SDMXML/schemas/v2_0/cross"
  xmlns:generic="http://www.SDMX.org/resources/SDMXML/schemas/v2_0/generic"
  xmlns:query="http://www.SDMX.org/resources/SDMXML/schemas/v2_0/query"
  xmlns:structure="http://www.SDMX.org/resources/SDMXML/schemas/v2_0/structure"
  xmlns:utility="http://www.SDMX.org/resources/SDMXML/schemas/v2_0/utility"
  xmlns:estat_sts="urn:sdmx.org.sdmx.infomodel.keyfamily.KeyFamily=EUROSTAT:STS"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://www.SDMX.org/resources/SDMXML/schemas/v2_0/message SDMXMessage.xsd">
  <Header>
    <ID>STSRTD_TURN_M</ID>
    <Test>true</Test>
    <Name xml:lang="en">TRANS0001</Name>
    <Prepared>2006-01-01T00:00:00+02:00</Prepared>
    <Sender id="NSI_GR">
  </Sender>
    <Receiver id="ESTAT">
  </Receiver>
    <DataSetAgency>NSI_GR</DataSetAgency>
    <DataSetID>STSRTD_TURN_M_GR_0001</DataSetID>
    <ReportingBegin>2002-01-01T00:00:00</ReportingBegin>
    <ReportingEnd>2003-03-01T00:00:00</ReportingEnd>
  </Header>
  <estat_sts:DataSet>
    <estat_sts:Series FREQ="M" REF_AREA="GR" ADJUSTMENT="N" STS_INDICATOR="TOVV"
  STS_ACTIVITY="NS5201" STS_INSTITUTION="1" STS_BASE_YEAR="2000" TIME_FORMAT="P1M">
    <estat_sts:Obs TIME_PERIOD="200201" OBS_VALUE="84.5" OBS_STATUS="A" OBS_CONF="F"/>
    <estat_sts:Obs TIME_PERIOD="200202" OBS_VALUE="85.6" OBS_STATUS="A" OBS_CONF="F"/>
    <estat_sts:Obs TIME_PERIOD="200203" OBS_VALUE="95.4" OBS_STATUS="A" OBS_CONF="F"/>
    <estat_sts:Obs TIME_PERIOD="200204" OBS_VALUE="106.2" OBS_STATUS="A" OBS_CONF="F"/>
    <estat_sts:Obs TIME_PERIOD="200205" OBS_VALUE="98" OBS_STATUS="A" OBS_CONF="F"/>
    <estat_sts:Obs TIME_PERIOD="200206" OBS_VALUE="95.3" OBS_STATUS="A" OBS_CONF="F"/>
    <estat_sts:Obs TIME_PERIOD="200207" OBS_VALUE="105.4" OBS_STATUS="A" OBS_CONF="F"/>
    <estat_sts:Obs TIME_PERIOD="200208" OBS_VALUE="107.1" OBS_STATUS="A" OBS_CONF="F"/>
    <estat_sts:Obs TIME_PERIOD="200209" OBS_VALUE="105.2" OBS_STATUS="A" OBS_CONF="F"/>
    <estat_sts:Obs TIME_PERIOD="200210" OBS_VALUE="109.4" OBS_STATUS="A" OBS_CONF="F"/>
    <estat_sts:Obs TIME_PERIOD="200211" OBS_VALUE="104.5" OBS_STATUS="A" OBS_CONF="F"/>
  </estat_sts:Series>
  </estat_sts:DataSet>
</CompactData>
```

```

<estat_sts:Obs TIME_PERIOD="200212" OBS_VALUE="111.9" OBS_STATUS="A" OBS_CONF="F"/>
<estat_sts:Obs TIME_PERIOD="200301" OBS_VALUE="89.1" OBS_STATUS="P" OBS_CONF="F"/>
<estat_sts:Obs TIME_PERIOD="200302" OBS_VALUE="88.3" OBS_STATUS="P" OBS_CONF="F"/>
<estat_sts:Obs TIME_PERIOD="200303" OBS_VALUE="96.1" OBS_STATUS="P" OBS_CONF="F"/>
</estat_sts:Series>
</estat_sts:DataSet>
</CompactData>

```

2.2. The SDMX Content-Oriented Guidelines used for reference metadata

The SDMX Content-Oriented Guidelines comprise Statistical Cross-Domain Concepts, Cross-Domain code lists, Statistical Subject-Matter domains and the Metadata Common Vocabulary. The SDMX Content-Oriented Guidelines are used for defining a new structure for reference metadata for the European Statistical System: the Euro SDMX Metadata Structure (ESMS) which should be successively implemented within the European Statistical System.

This ESMS based reference metadata structure uses the following 21 statistical concepts:

1. Contact	8. Release policy	15. Timeliness and punctuality
2. Metadata update	9. Frequency of dissemination	16. Comparability
3. Statistical presentation	10. Dissemination format	17. Coherence
4. Unit of measure	11. Accessibility of documentation	18. Cost and burden
5. Reference period	12. Quality management	19. Data revision
6. Institutional mandate	13. Relevance	20. Statistical processing
7. Confidentiality	14. Accuracy and reliability	21 Comment

A number of those concepts are further subdivided into sub-concepts. The ESMS concepts also include quality criteria which will allow the better measurement of data quality in statistical domains.

2.3. The SDMX IT architecture and tools

SDMX supports two complementary modes for data exchange and data sharing: the “push” mode (where data are transmitted from one organisation to another) and the “pull” mode (where one organisation retrieves data from another organisation’s server). SDMX also supports the “hub” concept, where users obtain data from a

central hub which itself automatically assembles the required dataset by querying other data sources.

To support the use of SDMX, many IT tools have been developed by the SDMX sponsoring organisations or by other bodies. These tools can generally be freely downloaded via the SDMX website. The source code is available so that they can be used as components for building IT systems in statistical organisations. Eurostat has provided several of these tools, including an SDMX converter (for converting data to and from SDMX formats) the Data Structure Wizard (for creating and viewing data structure definitions) and an SDMX registry (for storing and interacting with SDMX structural metadata).

Eurostat has put in place its own SDMX infrastructure, based on the Eurostat SDMX Registry, which will cover harmonised structural metadata, including the data structure definitions designed for statistical domains, and the ESMS. This application – acting as a general metadata registry - will be accessible within the European Statistical System and beyond.



2.4. Capacity building

The seven sponsoring organisations carry out different capacity building activities such as international conferences on SDMX (2007 in Washington, 2009 in Paris), the creation of self-learning packages for SDMX and various training activities.

For the European Statistical System, numerous training activities are organised directed towards statisticians, managers and IT staff. These courses and tutorials cover many different aspects of SDMX, including the basic principles of SDMX and the construction of data and metadata structure definitions, as well as the techniques for building IT applications to work with SDMX. In addition Eurostat also provides bilateral support to countries for implementing SDMX.

3. SDMX COSTS AND BENEFITS

The benefits of SDMX include the provision of standardised data and metadata file formats allowing the automated production and processing of the underlying files. In addition, the SDMX Content-oriented Guidelines facilitate the mutual understanding of the content of the data and metadata files through the use of common statistical concepts in structural and reference metadata. Furthermore, the XML technology means that there is a wide selection of IT applications and tools readily available for users of SDMX.

On the other hand, the implementation costs of SDMX are modest for many statistical organisations, especially those which are already experienced with GESMES/TS. For many statistical organisations, standardisation based on SDMX can be expected to offer efficiency gains over time, which will reduce costs in the medium term. The reusable SDMX software components will help organisations to reduce IT development costs.

4. WHAT ARE THE NEXT WORK STEPS?

To achieve the improvements to business processes which SDMX is designed to support, SDMX needs to get more broadly known and implemented. This is true on a world-wide basis, but also for the European Statistical System. EU Member States first adopted a strategy for the implementation of SDMX in 2005, which has already led to a number of projects using SDMX in various statistical domains. Following the adoption of a detailed implementation plan in 2007, Eurostat aims to accelerate the systematic use of SDMX to achieve optimal data transmission, sharing and dissemination for a number of additional important statistical domains.

For more information

<http://www.sdmx.org>