منظمة الأغذية والزراعة للأمم المتحدة | 联合国粮食及农业组织 | Food and Agriculture Organization of the United Nations | Organisation des Nations Unies pour l'alimentation et l'agriculture | Organización de las Naciones Unidas para la Agricultura y la Alimentación

COORDINATING WORKING PARTY ON FISHERY STATISTICS

Twenty-third Session

Hobart, Tasmania. 22-26 February 2010

SDMX for Fisheries Statistics Eurostat Project

Author: Eurostat

EUROPEAN COMMISSION
EUROSTAT

Directorate E: Sectoral and regional statistics
Unit E-2: Agriculture and fisheries

SDMX FOR FISHERIES STATISTICS
EUROSTAT PROJECT

(EUROSTAT'S CONTRIBUTION to the CWP-23 * FEBRUARY 2010)

1. Introduction

SDMX is an initiative to foster standards for the exchange of statistical information that aims to facilitate more efficient processes for the exchange and sharing of data and metadata. This is being developed by a consortium of major international organisations including, Eurostat, the World Bank, the Bank for International Settlements (BIS), European Central Bank (ECB), International Monetary Fund (IMF), the Organisation for Economic Co-operation and Development (OECD) and the UN. More detailed information on the rationale and benefits of SDMX is provided at Annex 1.

Work to apply SDMX principles to fisheries statistics was started in November 2009 by Eurostat, FAO and ICES under the working title SEIF (SDMX for Eurostat, ICES and FAO). This work is not exclusive to these organisations and the need to involve all bodies involved in fisheries data capture an exchange is recognised.

2. Background

In October 2008 Eurostat began a project to fundamentally re-engineering their processes to collect, store and disseminate fisheries data. This sought to address an acute problem whereby Member States sent statutory statistical reports to Eurostat in a variety of formats which made processing very cumbersome. The project also sought to make the best possible use of available information technology tools, particularly those that were being adopted as standard elsewhere in the organisation.

For data storage and dissemination, new relational databases are being developed on an Oracle platform. It is intended that these will be operational with basic functionality by the middle of 2010. SDMX is being adopted as the data collection format, in-line with Eurostat policy and in recognition of its position as an international standard for data exchange. The efficiencies achieved through improved data handling will be allow more time to be spent on ensuring data quality and integrity than has been possible. It is also hoped that it will provide greater transparency of what data is collected for what purpose and this may help in better targeting of data requests and reduced work for data providers where they are able to avoid sending the same data to different organisations. Together, these benefits should allow Eurostat to be better positioned to meet future challenges for meeting evidence needs such as may be needed for the reformed Common Fisheries Policy. Improved quality and standardisation should also help reduce burdens on Eurostat's partners in this area, including FAO, ICES and regional fisheries organisations.

Commission européenne, L-2920 Luxembourg. Telephone: (352)43011.
Fax: (352)4301 33859.

http://europa.eu.int/comm/eurostat/

Co-operation between FAO, ICES and Eurostat was initiated at a meeting in Copenhagen on 16 November 2009. It was foreseen that other organisations would likely have their own subsets of formal code lists which they would maintain themselves.

3. Discussion

The collaboration of FAO with SEIF indicates its intention to align as much as possible with the standards identified through SEIF co-operation. At the same time, FAO maintains the right to choose its own standards especially when harmonization is found to be difficult.

Eurostat are shortly to publish SDMX data structure definitions (DSDs) and Message Implement Guidance for Member States will be available during February. Adoption of the SDMX reporting format by Member States will initially be by 'gentleman's agreement'. Eurostat also proposes making data available to other international organisations in SDMX format.

Eurostat DSDs have several code lists that originate with the European Commission. These include well established codes such as for fish products presentations and intended uses. New codes have also been generated to meet the needs of aquaculture reports supplied under Regulation (EC) No. 762/2008[1]. Acceptance of these codes will need to be sought from the wider statistical community to facilitate sharing of this data. Eurostat propose continuing to use the ISO 3 alpha code for country, in line with other organisations in this domain.

[1] Regulation (EC) No 762/2008 of the European Parliament and of the Council of 9 July 2008 on the submission by Member States of statistics on aquaculture and repealing Council Regulation (EC) No 788/96 (Text with EEA relevance). OJ L 218, 13.8.2008, p. 1–13.

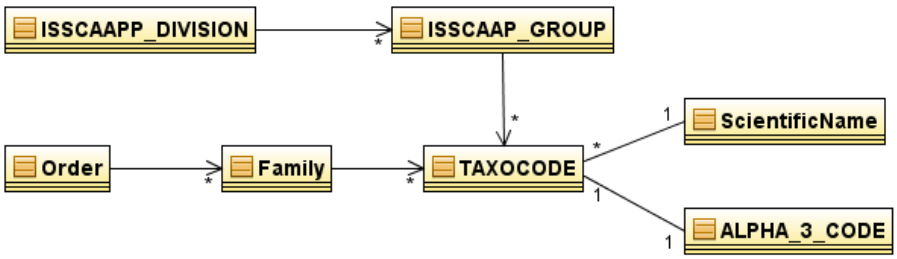
ASFIS SDMX Code list

The basic structure of the ASFIS species code list is:

ISSCAAP	25
TAXOCODE	1020100101
3A_CODE	LAU
Scientific_name	Petromyzon marinus
English_name	Sea lamprey
French_name	Lamproie marine
Spanish_name	Lamprea marina
Author	Linnaeus 1758
Family	Petromyzontidae
Order	PETROMYZONTIFORMES
Stats_data	1

FAO analysis of the ASFIS list proposes generation of an SDMX structure as shown in figures 1 and 2.

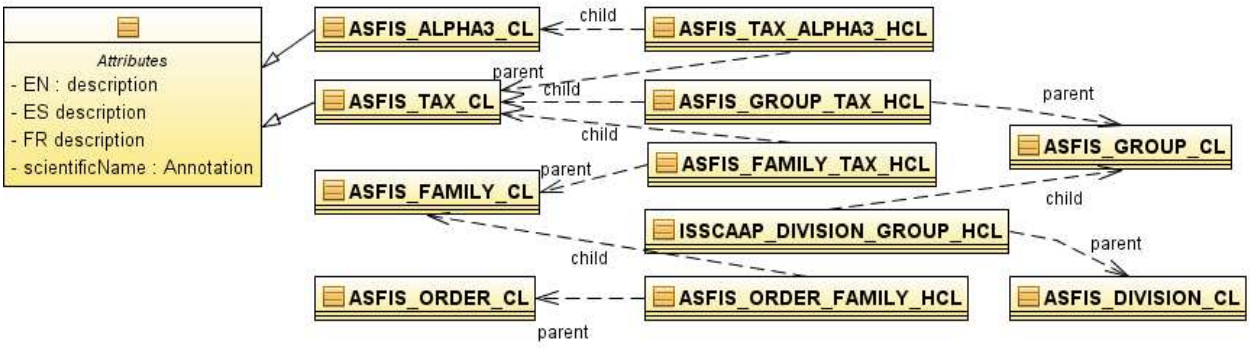
Figure 1 – analysis of ASFIS code structure



Notes:

1. This diagram suggests that TAXO_CODE is more important than ScientificName and ALPHA_3_CODE, which is not the case. Though one had to be chosen in order to avoid generating many extra relationships. This ordering is also applied in the SDMX design.
2. Author is not explicit, it can be found as in an annotation in the ASFIS_TAX_CL
3. 3A_CODE S is changed in ALPHA_3_CODE because it is not SDMX conform to start a code name with a numeric value.
4. The scientific name of the species cannot be published as a SDMDX Code list. The scientific name has not a one to one relation with the taxonomical code. The scientific name is therefore added as an annotation.

Figure 2: design of SDMX code list



Major Fishing Area SDMX Code list

The basic structure of this code list is:

OBJECTID	11
OCEAN	Atlantic
SUBOCEAN	2
F_AREA	27
F_SUBAREA	27.3
F_SUBUNIT	27.3d
F_DIVISION	27.3.d.30
F_SUBDIVIS	
Shape_Leng	23.4252297
Shape_Area	25.4533505

FAO analysis of the list proposes generation of an SDMX structure as shown in figures 3 and 4.

Figure 3 – analysis of code structure

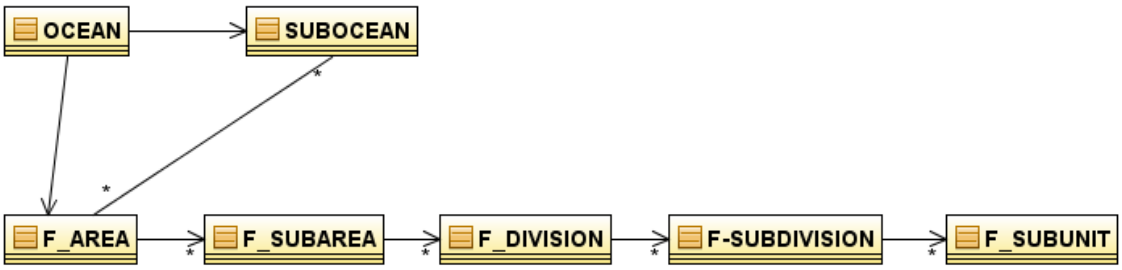
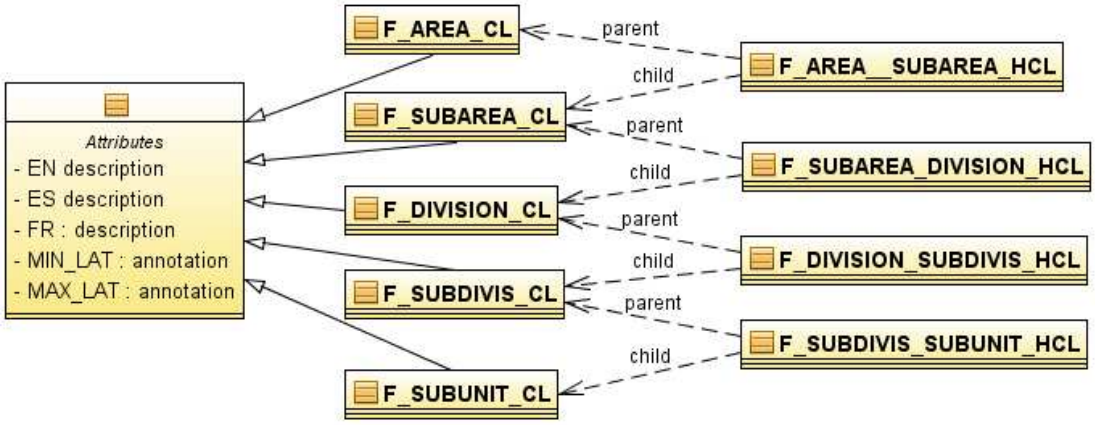


Figure 4: design of SDMX code list



Notes:

OCEAN and SUBOCEAN will not be expressed in SDMX.

4. Summary/conclusions

The value of FAO, ICES and Eurostat collaboration to develop SDMX code list standards for improving the efficiency and transparency of data collection and sharing is recognised. CWP members are invited to comment on this paper and the content of the Data Structure Definitions and Message Implementation Guidance for the Eurostat Reports available on request, particularly the code lists made by FAO with European Commission and ICES comments (please see pages 3 and 4).

5. Next steps

The adoption of SDMX for Eurostat fisheries reports is to be discussed at a workshop on 5 March 2010 in Luxembourg to which EU Member States and Eurostat's partner organisations have been invited.

Annex 1
Brief Introduction to SDMX in Eurostat

1. Why SDMX?

SDMX is a standard designed to describe statistical data and normalise their exchange in public administration, education, health, tax and customs administrations, regional authorities, statistical authorities, etc.
The driving force behind SDMX is reduction of development and maintenance costs, elimination of human errors and faster, more reliable and simpler data processing through the following:

- Reduction of diversity across statistical data production processes by appropriate use of the standard.
- Minimization of manual interventions through machine to machine communication;
- Unification of data stored inside one organisation and across national and international organisations by harmonization of the statistical metadata (e.g., concepts, code lists);
- Standardisation of statistical information by development and use of standard objects (e.g., schemes, data structure definitions);
- Standardisation and sharing of IT tools by development of generic IT tools and architecture.

2. Structure of data and the SDMX objects

Data represent concrete observations of particular statistical phenomenon at a certain moment of time. **Data set** is a collection of related observations, organized according to a predefined structure which the SDMX objects presented in this section describe.
The structure of a data set is determined by statistical concepts describing the meaning of the observations contained in the data set and their roles in the structure.

The concepts called "**dimensions**" determine the data set's "physical" structure. The **code lists** are linked to the dimensions listing the possible values the concepts can take. Other concepts do not affect the data set structure itself, but give additional information about the concepts used and they are called "**attributes**". Attributes can be coded or not coded. The actual reported value ("**measure**" in SDMX

language) is also considered a concept. **Concept Scheme** is a maintained list of statistical concepts that are used in **DSD**.

In the cases of actual transmission of data between organisational entities, the data are retrieved from their storage environment at the **Data Provider**, put in data files with predefined format (by the DSD) and sent to the receiver. The transmission can be accompanied on both ends of the transmission channel by operations like validation, encryption/decryption.

To parameterise data exchange the **SDMX Information Model** defines several objects (incl. DSD and Concept Scheme) and their features allowing their identification, maintenance and versioning: **(Hierarchical) Code list** is a (hierarchical) inventory of codes used in a DSD listing values to be used in the representation of dimensions or attributes.

Data Flow Definition links DSD to one or more data exchange (flow).

The **Provision Agreement** describes the way in which data are provided by **Data Provider**.

Other objects define the organisations involved in the data exchange, including maintainers of the metadata defining the objects (e.g., **Organization Scheme)**.

To allow for general access and ease of use the objects are stored in a special database accessible on Internet - the **SDMX Registry**.

All the SDMX objects are defined in the **SDMX Technical Specifications** (www.sdmx.org).

3. A simplified example of a data table showing relation with some SDMX objects.

This rudimentary example describes the first two steps of the compliance procedure as described in the previous section.

Let's assume a statistical data table entitled:
Monthly amounts (tons) and value (M€) of traded goods among EU MSs, arrivals and dispatches, in 2008.

Then **Data Structure Definition (DSD)** contains the detail description of concepts and their roles:

The concepts that are **dimensions**:
- **Reporting MS** (27 countries in its **code list** – EU MSs)
- **Partner MS** (27 countries in its **code list** – EU MSs)
- **Flow** (with 2 elements in its **code list** – arrivals & dispatches)
- **Reference Month** (12 months in its **code list** – January till December 2008)

The concept that are **measures**:
- **amounts**
- **value**

These two measures could be also grouped in one dimension containing these two concepts (measures) as elements in its code list)

The concepts that are **attributes** are:

- **Unit of Measure**, attached to **amounts** (tons)
- **Currency**, attached to **value** (M€)

The DSD also contains:
- the **code lists** used in the table
- the **Concept Scheme**, which is the list of all concepts used in the **DSD**

There could be other SDMX objects, e.g.:
- If Eurostat provides this table to OECD then Eurostat is the **Data Provider.** The **Provision Agreement** might state that the table should be provided before the end of January each year.
- **Organisation scheme** making Eurostat the maintainer of the objects (related to the data table): concepts, code lists, Provision Agreements, Concept Schemes, etc.

4. Compliance & implementation

To achieve **SDMX compliance** the responsible authority appreciating the expected benefits from using SDMX decides to implement it. **SDMX Technical Specifications** define how to prepare and use the SDMX objects related to particular data exchange processes in practice. Generally the following four steps need to be done:

1. analyse the data and their exchange processes to have the needed knowledge to
2. Develop all the necessary objects according to the rules in the SDMX Technical Specifications.

Then
3. Test the data and their exchange process using the developed objects. Finally
4. Prepare methodological and implementation documentation (accompanying SDMX objects) for stakeholders, e.g., data providers and those who maintain the objects.

The steps produce specifications allowing and assuring **compliance**. If the developments are applied in practice so the data and metadata exchanges within the domain are carried out according to SDMX compliant specifications this constitutes an **SDMX implementation.**

Note that SDMX and its implementation also apply to reference metadata.

For the exchange of the data sets (often referred to as tables, files or cubes), two file formats are recognized by SDMX:
- SDMX-ML (preferred) based on the XML to describe and structure the data inside a file
- SDMX-EDI (also called GESMES/TS). Due to inherent constraints, this format is not always able to respond to all requirements

For exchange of statistical micro data, geographical data and updates to registers and for data exchange via web services SDMX do not provide specific formats. The work to set up such standards is ongoing.