ESA/STAT/AC.267/10
1 May 2013

UNITED NATIONS
DEPARTMENT OF ECONOMIC AND SOCIAL AFFAIRS
STATISTICS DIVISION

Expert Group Meeting on International Statistical Classifications
New York, 13-15 May 2013

Use of SDMX in stocking, disseminating and managing the Swiss classification of economic activities and throughout the production process of a statistic

Angelina Dungga Winterleitner, Federal Statistical Office Switzerland

METIS COMMON METADATA FRAMEWORK (CMF)
METADATA CASE STUDY

< SWITZERLAND / FEDERAL STATISTICAL OFFICE FSO>

This template was created to guide the content and format of case studies on statistical metadata systems. It is intended to assist agencies in sharing their experiences, and to encourage the update of information as projects develop. From a reader's perspective, a consistent approach to presenting case studies can facilitate better understanding and comparability between organizations.

Instructions for completion:
The template is divided into seven sections. Each section comprises a table with headings in the left hand column (in bold). On the right is a description of the suggested content, which should be replaced with the appropriate text.

The template is flexible. If a particular heading is not relevant, it can be modified to something more appropriate, or deleted. If there appears to be no place for certain information, new headings/sections should be added. Examples of existing case studies are published at www.unece.org/stats/cmf.

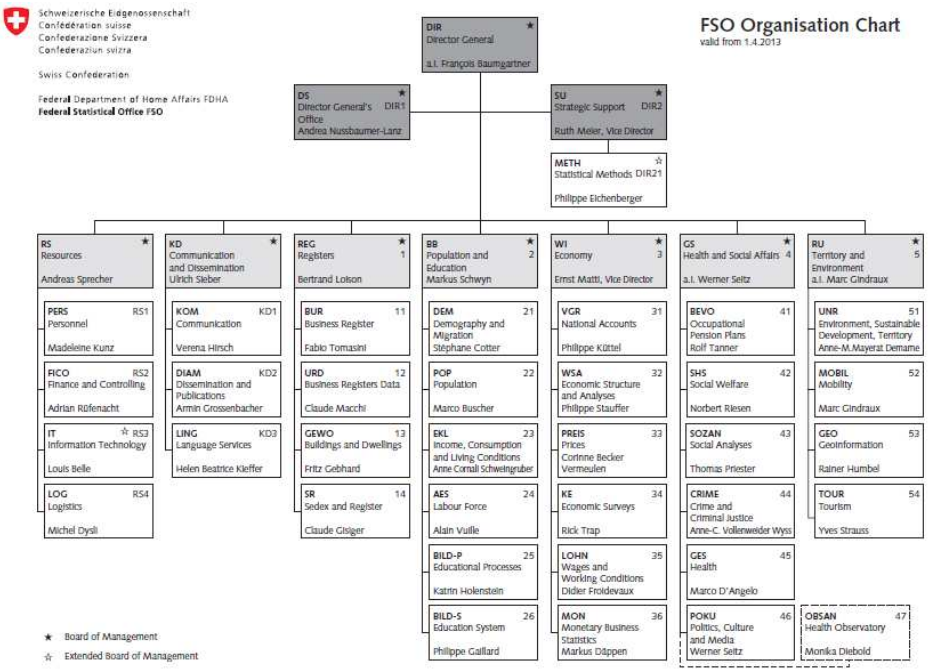
Questions and feedback should be directed to the UNECE secretariat: email steven.vale@unece.org or support.stat@unece.org, or phone +41 22 917 3285.

1. INTRODUCTION

Organization Name Swiss Federal Statistical Office FSO

Number of staff The Federal Statistical Office is part of the Federal Department of Home Affairs. 660 permanent employees hold the equivalent of 560 jobs. 48% of the jobs are held by women and 47% by French speakers. Around 60% of the permanently employed are university graduates. Personnel and materials expenditure amount to a little over 158 million CHF.

Organization structure



Contact person (for Metadata)
Jean-François Fracheboud
Head of Standards, Architecture and Metadata / IT Division
Jean-Francois.Fracheboud@bfs.admin.ch
+41 32 71 36734

Contact person (for classification on General Economic Activities NOGA)
Angelina Dungga Winterleitner
Head of Nomenclatures / Register Division
angelina.dungga@bfs.admin.ch
+41 32 71 36523

Introduction

This paper gives only a rough description of the metadata system of the FSO to focus on the experiences made and the issues raised while starting the maintenance work of the General Classification on Economic Activities (NOGA) in the Statistical Metadata System (SMS) of the FSO.

Metadata strategy

The metadata strategy of the FSO is composed of four elementary points:

- A centralised system for a decentralised management of metadata
- A model of governance for managing and harmonising metadata
- Standardised processes and generic access to metadata via Web Services
- Data driven by metadata

In order to achieve these aims, the FSO has implemented a Statistical Metadata System based entirely on SDMX in 2010. Forthcoming releases of the SMS are listed below.

Table 1: Forthcoming releases of the SMS

Releases	Timing
Developing and implementing SDMX Registry based on SDMX 2.1	2013
Web application for accessing metadata on internet (via Web Services)	2013
Domain specific plug-ins	2013 - 2014
Deployment of Web services in the chain of statistical production	starting 2013
Integration of metadata from external producers	2014 - 2015

The main challenge related to classifications in the SMS lies in promoting the use of Web Services for the integration of classifications in the IT-applications. The idea is to create the metadata once, and as early as possible in the statistical production process, and to use and reuse it during the statistical cycle. A strong tendency to get the classification from the SMS (via Web Services) and to store them in domain specific databases can be observed. As a consequence, the classifications are duplicated and the strength of Web Services are not fully exploited.

Current situation

Domain specific metadata are being integrated in the SMS progressively since 2011. The implementation of the system uses a stepwise approach depending on individual timing and needs of the producers rather than on a strict schedule. At present around 25% of FSO metadata are integrated in the SMS.

2. THE GENERAL CLASSIFICATION ON ECONOMIC ACTIVITIES (NOGA) AND THE STATISTICAL BUSINESS PROCESS

2.1 Statistical business process model

The statistical business process model in use is the Generic Statistical Business Process Model (GSBPM). It is used to pinpoint the moment when a metadata is *created* and when it is *used* during the statistical cycle. It illustrates the notion of reusing a metadata once generated and allows an understanding of the overarching task of metadata management. For statistics producers the generic model structures their activities in a suitable way, facilitating them to conceive changes in the current system, notably to plan change processes separately for different phases of the generic model.

2.2 Metadata

Before the introduction of the SMS, a large variety of tools for managing metadata were

management before the SMS

applied and storage took place in a decentralised manner. The system fostered a decentralised metadata management. Each unit has its proper solutions for maintaining, stocking and disseminating its own metadata.

The system used for the maintenance of the General Classification of Economic Activities (NOGA), is described in table 2. As table 2 shows the current system counts a number of applications used for storage, maintenance and dissemination of each version.

Table 2: Metadata system used for NOGA before SMS

Metadata	Storage	Maintenance	Dissemination
Nomenclature	NOGA repository (in SDMX-registry since 2011)	Access database	• On print • In pdf-format on the Internet • In SDMX-format on the internet since 2011 (via Classweb until 2011)
Correspondence Tables	NOGA repository	Excel-sheet	• Excel in shared repositories • Excel on the internet
Aggregation Levels for Statistical Dissemination	NOGA repository	Excel-sheet	• Excel in shared repositories • Excel on the internet
Classification Index	NOGA repository	Access database	• Online tool KUBB • Excel in shared repositories • Excel via E-mail
Caselaws / Operational rules	NOGA repository	Access database	• Online tool KUBB • via E-mail • Trainings • Meetings

Dissemination of duplicates

At the time being, duplicates of NOGA are used throughout several statistical business processes, by the business register and by producers of statistics as well. A copy is also sent to the Communication and Dissemination department of the FSO for dissemination via internet. Changes effected in the explanatory notes of a classification version (for instance, the correction of typing errors) are communicated to the BR, the metadata unit and the Communication and Dissemination department, so they can undertake the necessary adaptations in their respective software. Effected changes cannot be passed over to users after they have accomplished the download via *Classweb* (the classification disseminating channel based on CODAM-BRIDGE used until 2010 for maintaining classifications) or the Internet. This leads to the existence of multiple duplicates of the same metadata used not only in different business processes but also within one statistical business process.

Planned changes

All three versions of the NOGA are available for use in the SMS since 2011. It can be accessed and referenced by all subject matter statisticians throughout the statistical life cycle.

Following sub-projects are in the pipeline:

- Integration of convened NOGA-aggregation standards for the publication of statistical data (standard aggregations);
- Integration of correspondence tables between versions of NOGA;
- Eventual integration of the classification index;
- Eventual integration of caselaws.

2.3 Costs and Benefits

The costs relate to the ones caused by:

- Conceptual understanding of opportunities and limits of the SMS;
- Learning basic skills to work with the editor;
- Conceptual and implementation work on planned sub-projects.

The benefits of working with the SMS are:

- A centralized tool for maintaining and stocking all versions of the classification;
- Convenient method for dissemination
    - A single source for dissemination;
    - Instantaneous passing over of changes to users;
- Possibility to integrate a higher level regrouping of NOGA sections for data dissemination purposes (standard aggregations);
- Possibility to specify correspondences between classification versions.
- Eventual possibility to also maintain index and caselaws in the same tool as the classification versions;
- Better understanding of user needs through the feedback of SAM: Individual supervision of transition activities in the FSO, enables SAM to see how the NOGA is used. Needs expressed by users are discussed on *ad hoc* bases with the classification unit.

Currently, following in-house evolutions multiplies the gains for maintaining the NOGA in the SMS:

1. The start of the business register to maintain its metadata in the SMS.
2. The start of several statistic producers in using the SMS.
3. By the end of 2013, the SMS should provide direct access to external users through Web Services. Once completed, the NOGA should be accessible in SDMX-ML format via internet.

2.4 Implementation strategy

A stepwise implementation is intended for the transition of the NOGA-maintenance in SMS. Priority is given to the versions and variations that are used in-house, to assure re-use of the centrally maintained objects and to avoid duplication. The integration of the correspondence tables are not regarded as an important priority, since most of statistics have accomplished the transition to the newest version of NOGA. The integration of the index and the caselaws are treated as long-term projects, since conceptual questions need to be clarified first before the technical solution can be provided.

3. NOGA IN EACH PHASE OF THE STATISTICAL BUSINESS PROCESS

3.1 Metadata Classification

The SMS builds on international terminology of metadata categories and distinguishes three categories of metadata: structural, reference and process metadata. The NOGA is seen as a structural metadata as it contributes to the description of statistical data.

3.2 Metadata used/created at each phase

The General Classification on Economic Activites (NOGA) is a metadata used in a number of statistics and throughout the whole statistical cycle, as it is used in collecting, processing, analyzing and disseminating statistical data. The creation and maintenance of this specific metadata is situated outside of the statistical business process model, as its maintenance is not an integral part of any specific statistical business process.

The SMS, in general, allows for an end-to-end classification management. The implementation strategy chosen for the SMS, authorizes a cutting down of the transition process into the nine phases of the GSBPM and a phase-by-phase implementation. A statistic producer can, for example, start implementing its classifications in the SMS only for the designing- and analyzing-phase, and continue to work with the current system for the collecting- and processing-phase.

3.3 Metadata relevant to other business processes

The NOGA is also used for the coding of statistical and administrative entities in the business register (BR), whose relevant output is the creation of the sampling frame, which is used as an input to statistical processes. The BR is viewed as an input provider for statistical production and stands at the edge of the statistical business process model.

4. SYSTEMS AND DESIGN ISSUES

4.1 IT Architecture

The IT-architecture of the SMS is composed of five components:

- A registry, including Web Server, for metadata storage;
- Web Services for reading and writing metadata;
- REST Services for technical access (through IT-applications);
- SDMX Editor for managing metadata (create, maintain, visualise, navigation);
- Plug-Ins (for personalised interfaces or particular functionalities).

4.3 Standards and formats

The SMS is based exclusively on SDMX 2.0.

NOGA artefacts in the SMS

The NOGA is conceived as a Hierarchical CodeList (HCL_) consisting of several CodeLists (CL_) representing each level of the classification. Each CodeList and each Hierarchical CodeList is viewed as an artefact.

A unique artefact is identified by three parameters:

1. The AgencyID: identification of responsible unit for metadata maintenance
2. The ID: identification of the artefact itself
3. Version of the artefact: All changes or updates to an artefact lead to the creation of a new version.

In case of NOGA, the organizational unit ‘Nomenclatures’ is defined as the maintenance unit CH1_BUR. The IDs for the NOGA levels read as follows:

- CL_NOGA_SECTION
- CL_NOGA_SUB-SECTION
- CL_NOGA_DIVISION
- CL_NOGA_GROUP
- CL_NOGA_CLASS
- CL_NOGA_TYPE

The ID for the NOGA as a classification reads as follows:

- HCL_NOGA

When adding information about the version of an artefact, you will know exactly what classification you're looking at in the SMS.

	NOGA 2008	**NOGA 2002**	**NOGA 1995**
AgencyID	CH1_BUR	CH1_BUR	CH1_BUR
ID	HCL_NOGA	HCL_NOGA	HCL_NOGA
Version	3.0	2.0	1.0

The same applies for a CodeList e.g. a classification level:

	NOGA 2008	**NOGA 2002**	**NOGA 1995**
AgencyID	CH1_BUR	CH1_BUR	CH1_BUR
ID	CL_NOGA_DIVISION	CL_NOGA_DIVISION	CL_NOGA_DIVISION
Version	3.0	2.0	1.0

A specific version of a Hierarchical CodeList consists of a number of different CodeLists of the same version.

Elements of a NOGA artefact

Following elements describe an artefact:

1. Name: denomination of the artefact
2. Description: text describing the artefact, can contain hyperlinks to external sources.

These elements can be maintained in several languages. As a multi-linguistic country this aspect is very important. For example, the name is provided in German, French, Italian and English.

Name_de	Allgemeine Systematik der Wirtschaftszweige 2008
Name_fr	Nomenclature générales des activités éconoiques 2008
Name_it	Nomenclature generale delle attività economiche 2008
Name_en	General Classification of Economic Activities 2008

Optional elements:

1. Validity period: defines the time period, when a specific version of an artefact is valid.
2. Annotations: additional information
3. Tag "is final": defines if this specific artefact can be used for production and can be made publicly accessible

The 'valid from' date informs about the date the classification started being valid, the 'valid to' date informs about the date the classification ends to be valid, because a new

version of the classification starts being valid. Currently, the ‘valid from’ date for the NOGA 2008 is set at ’01.01.2008’, the ‘valid to’ date is not informed yet.

Annotation types for NOGA

For the NOGA-CodeLists following annotation types have been built:

- Abbreviation;
- Includes;
- Includes_also;
- Excludes.

The latter three form the explanatory notes of the classification. They are maintained on the Level of the Codes. Annotations can be maintained in several languages. Each annotation is accompanied by subsequent elements:

- ID
- Type
- Title
- Text
- Urn

Only the elements type and text are actually informed for the NOGA annotations in the SMS. Example:

Annotation Type	**Text**
Abbrev	Grow.of sugar cane
Includes	This type includes: - growing of sugar cane
Excludes	- growing of sugar beet (see 011300)

These annotation types are associated to (or belong to) following code, …

CodeValue	011400
ParentCode	
ParentName	
Name	Growing of sugar cane
Description	
URN	
URI	
Annotation	

… of following artefact

Elements	**Artefact**
AgencyID	CH1_BUR
ID	CL_NOGA_TYPE
Version	3.0
valid from	2008-01-01
valid to	
Name	Type
Tag 'Is Final'	checked

4.4 Version control and revisions

The three versions of NOGA available in the system are all finalised and are hence publicly accessible. This is marked by checking the ‘is final’ tag. Once the ‘is final’ tag is checked, no more changes can be effected. The idea is to enable traceability of the

metadata in the statistical system in creating a new version each time modifications are affected.

At state, a 3.1. version of the NOGA has been created to allow minor changes, like for instance corrections of typing errors. Each time, a modification has been made, the classification unit informs the metadata unit, so the modifications can be integrated in the officially finalised 3.0 version of the NOGA.

To enable documentation of preparatory work with regard to the next revision, a 4.0. version of the NOGA has been created. The tag 'is final' is not checked, and therefore the artefact not visible for users yet. Once the work is finalised, the tag should be checked.

5. ORGANIZATIONAL AND WORKPLACE CULTURE ISSUES

5.1 Overview of roles and responsibilities

Standards, Architecture and Metadata (SAM), IT Division
Role: System administration, highest level in SMS
Responsibilities:

- IT-Architecture
- Security management
- Management of the SMS (metadata strategy, harmonisation, administration, access rights, support)
- Development of the SMS (IT development, maintenance, plug-ins, user interfaces)
- Supervision of metadata maintenance processes

Statistical production units, Classification units, Register units
Role: Owner of a metadata
Responsibilities: metadata creation and maintenance

Reporting is conducted whenever a work topic might be of interest for the other unit, and collaboration takes place whenever the expertise of the other unit is needed.

5.2 Training and knowledge management

The training is provided by SAM and is designed to fit individual needs.

6. LESSONS LEARNED

6.1 Flexible implementation strategy

Beginning to maintain the NOGA in the SMS puts us before conceptual and technical issues to be solved. The implementation strategy of the SMS chosen by SAM allows to address these issues in a collaborative manner and to find individual solutions.

Issues raised

The transition of maintaining NOGA to the SMS has raised several issues:

"Is Final" notion

To make a classification publicly accessible for users, finality has to be granted by checking the "is final" box. Once a classification is final, no more changes can be effected by the metadata owner. This

blocking system makes sense in theory. In practice, for example typing errors need to be remediated, even after a classification has started being valid. While it is important to keep a security blocking, once a classification is final, the metadata owner should be able to make minor modifications within the text, even after a classification is officially made public - without having to create a new version of the classification.

Conceptual issues

The maintenance of a classification like the NOGA is not restricted to the maintenance of the actual valid version and the respective correspondence tables. In addition to the official classification an index, standard aggregations, a list with title-abbreviations and a list of rulings is provided. While the standard aggregations and the title-abbreviations will not be subject to updates as long as the actual valid version of the classification is not superseded by a new version, the index and the rulings are frequently updated. Questions arise when it comes to mapping these elements into the metadata system:

- Should the additional elements be treated as a distinct entity from the actual classification or should, for example, an index entry be directly associated to the code value concerned?
- If treated as a distinct entity, can a link to the concerned code value be drawn?
- Is versioning maybe a good solution for traceability of changes in general?
- Should they be designed as CodeLists or another type of artefact?

For now, a first step in addressing these conceptual issues has been made in a first draft mapping of the NOGA structure in the SMS with the Neuchâtel Terminology Model for classifications (see Annex 1). Further work will take place within the planned sub-projects mentioned in section 2.2. of this paper.

Lessons learned

From an overall perspective, the SMS as a whole and its SDMX data model seem to be very pragmatic tools for metadata harmonisation and exchange. The inherent logic seems to be compatible with the NOGA comprehension of a classification and “classification by-products”. The big challenge seems to lie in addressing conceptual questions of translation and mapping of classification concepts into the SMDX-language rather than in the technical implementation.

Other

Editing of large text blocks

The SMS was conceived to maintain simple CodeLists (such as gender, age, activity status, etc.) that don’t usually contain large text blocks. To edit large text blocks, it is recommended to copy the SMS content into a Word-format, edit the text there, and copy it back into the SMS. Hence, the system is not very user friendly when it comes to the maintenance of classifications that contain large text blocks.

7. ATTACHMENTS & LINKS

7.1 NOGA-website

Website with SDMX-versions

Online Tool KUBB

*** END ***

ANNEX 1:
1st draft mapping of NOGA structure in SDMX with Neuchâtel Terminology Model for Classifications

Object Types_Neuchâtel Terminology	Attributes_Neuchâtel Terminology	NOGA artefacts and its elements in SDMX
Classification Family	Identifier Title Classifications	Not in use
Classification	Identifier Title Description Context Units classified Subject areas Owners Keywords Family Versions Current version Footnotes	Not in use
Classification version	Identifier Title Description Release date Termination date Current version Introduction Maintenance unit Contact persons Legal base Publications Languages available Floating Updates possible Changes from previous version Updates Copyright Dissemination allowed Classification Predecessor version Successor version Derived from Levels Items Case laws Correspondence tables Classification indexes Footnotes	HCL_NOGA Elements: Id Version Agency Valid From Valid To Creation Diffusion Name Name_de Name_fr Name_it Description Is Final Last modification Updated by Annotations
Classification variant	Three kinds of variants: extension variants, aggregate variants or regrouping variants Identifier Title Description Release date Termination date Current variant Introduction Maintenace unit Contact person Legal base Publications Title types Languages available Floating	Under construction

	Updates possible Changes from previous variant Updates Copyright Dissemination allowed Base classification version Predecessor variant Successor variant Levels Base version levels Items Case laws Correspondence tables Classification indexes Footnotes	
Classification index	Identifier Release date Maintenance unit Contact person Publications Languages Floating Corrections Classification version/variant Index entries Footnotes	Under construction
Correspondence table	Identifier Title Description Owners Maintenance unit Contact persons Publications Source version Target version Source level Target level Relationship type Source complete Target complete Floating Footnotes	Under construction
Classification level	Level number Level name Description Number of items Code type Code structure Level coverage Dummy code Foreign level Items	CL_NOGA_SECTIONS CL_NOGA_SUB-SECTION CL_NOGA_DIVISIONS CL_NOGA_GROUP CL_NOGA_CLASS CL_NOGA_TYPE Elements: Id Version Agency Valid From Valid To Creation Diffusion Name Name_de Name_fr Name_it Is Final Last modification Updated by
Classification item	Code Official title Alternative titles Explanatory notes	Elements: Code value Code name_de / _fr / _it / en Parent code_de / _fr / _it / en

	General note Includes Includes also Excludes Level number Generated Currently valid Valid from Valid to Past events Future events Changes from previous versions Updates Classification version/variant Parent item Sub items LInked items Case laws Index entries Footnotes	Parent name_de / _fr / _it / en Link to resource with additional information Annotations_de / _fr / _it / en: ID Type Title Text
Item Change	Date of change Type of change Descriptions Predecessors Successors	Not in use
Case law	Description Items	Under construction
Classification index entry	Text Item Valid from Valid to	Under construction
Correspondence item	Source item Target item Valid from Valid to Footnotes	Under construction