

TWINNING CONTRACT

**Support to the State and Entity Statistical
Institutions, phase V**

Bosnia-Herzegovina

EU Twinning Project BA 08-IB-FI-03



MISSION REPORT

on

Metadata and classification system development

Component no 3.2.5

Mission carried out by
Søren Netterstrøm, Statistics Denmark
Mogens Grosen Nielsen, Statistics Denmark

26 – 30 August 2013

Version: FINAL



Metadata and classification system development

Author's name, address, e-mail

*Søren Netterstrøm
Statistics Denmark
Farverstræde 6
DK-4850 Stubbekøbing
Denmark
Tel: +45 20451731
sne@viv.dk*

*Mogens Grosen Nielsen
Statistics Denmark
Sejrøgade 11
DK-2100 Copenhagen Ø
Denmark
Tel. +45 3917 3894
mgn@dst.dk*

List of Abbreviations

BiH	Bosnia and Herzegovina
BHAS	Agency for Statistics of Bosnia and Herzegovina
FIS	Federal Institute of Statistics of the Federation of Bosnia and Herzegovina
RSIS	Republic Institute for Statistics of the Republika Srpska
GSBPM	Generic Statistical Business Process Model
GSIM	Generics Statistical Information Model
DDI	Data documentation initiative

Content

- 1. Introduction..... 4
- 2. Activities and expected output of the mission..... 4
- 3. Assessment and recommendations..... 4
 - 3.1 Strategy..... 4
 - 3.2 Classifications 6
 - 3.3 Missions 76

1. Introduction

This mission report was prepared within the state and entity statistical institutions in Bosnia and Herzegovina. It was the *fifth* mission to be devoted to *topic 2* within Component 3 (*Strengthening the Institutional Capacity of BiH Statistical System*) of the project concerning Metadata and Classification System Development for the BiH statistical system.

The consultants would like to express their thanks to all officials and individuals met for the kind support and valuable information which they received during the stay in Bosnia and Herzegovina, and which highly facilitated the work.

The views and observations stated in this report are those of the consultants and do not necessarily correspond to the views of Statistics Denmark.

2. Activities and expected output of the mission

The expected activities are:

- Short presentation of strategy for metadata system and part for classifications
- Introduction of classification – current situation in statistical offices
- Presentation of Neuchatel data model
- Database model for classifications
- Software for classification to be discussed. Must take considerations on standards and integration in the strategy into account
- Presentation of classifications at the web

Expected output

- Strategy revised with review comments
- Database model for classification. The database model should be based on conceptual model of Neuchatel and sufficiently open for implementations. It is assumed that the DDI implementation will be based on Neuchatel including integrations with other metadata-elements.
- Plan for next missions

3. Assessment and recommendations

3.1 Strategy

Bosnia prepared the strategy before mission based on input from previous missions. The strategy gives a short and clear overall framework for the future work metadata in Bosnia: The goal for strategy improved. The relevant international standards are included: GSIM, GSBPM, DDI, SDMX-standards for referential metadata (SIMS, ESMS, ESQRS etc). An appropriate working group structure is included.

The following review-comments were discussed and put into the strategy:

- Introduction improved with regard to audience and expected use of the strategy and references to international objectives of having a well-functioning metadata-system
- Improvement of relation between the 4 main components: methodology, concepts, variables and classifications and standard. Conceptual level: GSIM – selected areas in Concepts group
- Update of the section on quality declaration. SIMS should work as an inventory of quality concepts. ESMS and ESQRS should be used as reporting standards.

3.1.1 Architecture

RSIS stated, as they have done previously, that in case the architecture would only include a single server (to be placed in the Agency), they would not be able to adopt the strategy as a common strategy for all three institutes. It was, after some discussion clarified, that it was the actual location of the server that constituted a problem, not the sharing of metadata or other issues.

Two new proposals were put on the table by the consultants.

One option would be to limit the system to only contain common metadata stored and updated by Agency. This option is based on requirements in the contract between the Delegation of the European Union to BIH and the involved institutions in Bosnia: *“Metadata and classifications have to be stored and updated in one place: the Agency. RSIS and FIS will have web access to this information and use it in their compilation and dissemination of statistics. RSIS and FIS will, based on this information, be able to report to the Agency and ensure that the same classifications etc. are used in all three statistical organisations”* The needed changes of metadata-environment at entity-level must be implemented as part of future projects. Common metadata would be the classifications, concepts and variable description provided by the agency as the common basis for statistical surveys. This solution would solve the issue and fulfill the requirements in the contract.

The other option would be to create a system of three servers and a virtual server combining the data in the three servers. All servers would share the same architecture. This would allow not only for the storage of common metadata (at the Agency server), but also for metadata local to each entity.

Both options were in principle acceptable for all parties, but after some discussions it was agreed to use the second model. An improved model (replacing the virtual server with data exchange between the servers) was proposed by the agency and accepted by all parties. It should be emphasized that all three servers will be coordinated.

The Agency will shortly change chapter 5 of the strategy proposal to reflect the decisions taken at this meeting. The strategy should then be finalized and ready for approval by the directors.

As a consequence the decision on software including financial aspects will be out of scope for the actual project.

3.1.2 Implementation

Recommendations regarding timetable for implementation

- It is recommended to revise the strategy at the end of 2014.
- Due to the work on census it is recommended to split up the time-table in work carried out in 2013 and 2014 and focus on less resource-demanding activities in 2013. Possible missions in 2013 would be training on metadata, continuation of work on classifications. See section 3.3 below.
- The plan for 2014 could cover the following topics:
 - Discussion and implementation of software solution and organizational setup (see chapter in strategy)
 - Prepare and review of documentation on common dataset, common variables and common code-lists on 6 surveys. (each institution should be responsible for 2 surveys).
 - Prepare and review of documentation on quality declaration and concepts (each institution should be responsible for 2 surveys).
 - Prepare and review of documentation on work-processes on 6 surveys using the GSBPM-model (each institution should be responsible for 2 surveys)
 - Review and decide on how to continue the work.

3.2 Classifications

3.2.1 Current situation

All three institutes have on their WEB site a collection of classifications as pdf-documents. The Agency has also two classifications that are shown on the web-side based on a small database-application. However, it was noted, the each classification was residing in a specific set of tables in the database. The consultant noted, that this would require the building of a specific interface for each new classification to be added, where with a proper classification database (Neuchatel), all classifications are served by the same database-structure and presentation software.

3.2.2 Neuchatel model

The Neuchatel model was presented in detail and the reasoning behind the objects and some of the attributes was given.

As the Neuchatel model is a conception model (and a terminology) it is open to different implementations. One of these implementations is a Danish-Norwegian collaboration based on a relational database. The database structure was presented (in some detail) together with an XML-schema used for import/export. Finally the application used to edit classifications and a WEB interface was briefly demonstrated.

The connection between classifications and other elements of the metadata (concepts, variables) was briefly discussed. It should be part of a future modelling of a complete metadata system.

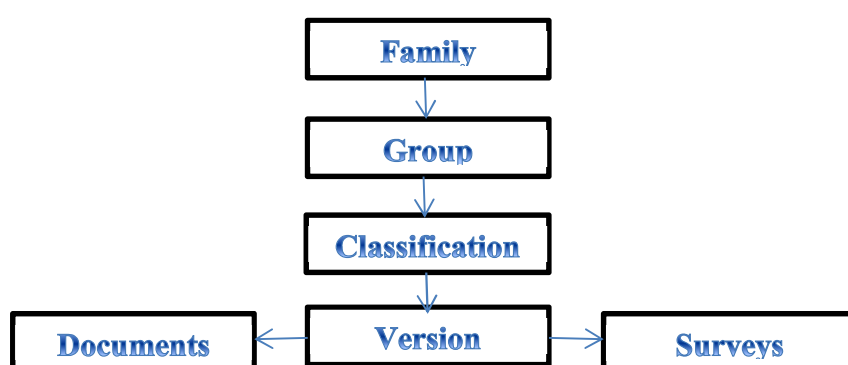
3.2.3 Implementation of classifications

The way forward was discussed.

It was agreed that a full implementation of Neuchatel will depend on the choice of software for metadata in general. If classifications are covered in such a software, it would not make sense to develop (or look for) a separate solution for classifications.

One of the primary goals identified for the classification server was to provide an overview of classifications actually in use in all three statistical institutes. Having such an overview would be an important first step towards harmonisation where needed.

A small data model was developed by the consultant and the participants.



The purpose is to hold information about all active classifications and the surveys in which they are used.

Family is used for a rough division of classifications according to major use.

Group (not in Neuchatel) is used to bind together all classifications that cover the same object type and classify the objects according to the same base criteria. All classifications in a group are potential candidates for being harmonised, but may not need this. As an example, NACE rev 2 and the BiH version of NACE are in the same group according to this, but even if they are not identical, they are fully harmonised.

Classification and version (taken for Neuchatel) just contains an identifier, title, short description and information about ownership and contact person (within BiH).

Surveys is a list of the surveys using a given classification version

For all classification versions, any available electronic description of content (pdf, xls, doc etc) should be kept in a single place and the URL should be given in the model. These documents can then be used to determine the actual content and need for harmonisation.

It was agreed to start the work on such a model, where each institute would they collect classifications within their domain. A small survey on this may be used to get information from the surveys. The information gather will be shared among the institutes.

3.3 Missions

The upcoming census has put a large pressure on the resources of all three institutes and is likely to do so for the remaining period of this project. It was not considered likely that this working group would be able to allocate time for preparations and participation in any missions for the rest of the year. Some options for future missions were briefly discussed without any conclusions.

It was decided to postpone the question. In case resources can be found the Agency should be responsible for drawing up a terms of reference to be discussed to the RTA and potential consultants for one or two missions to take place November-December 2013.

Metadata and Classification System Development
26th to 30th of August 2013
Terms of Reference - Activity 3.2.5
EU Twinning Project BA 08-IB-FI-03
Institute of Statistics of Republic of Srpska, Veljka Mlađenovića 12d,
Banja Luka

- Component 3 Institutional Capacity
- 3.1. Long term strategy
 - 3.2. Metadata and Classification System Development
 - 3.2.1 Assessment of the current situation regarding metadata and classification systems
 - 3.2.2 Metadata system I
 - 3.2.3 Metadata system II
 - 3.2.4 Classification systems I
 - 3.2.5 Classification systems II**
 - 3.2.6 Classification systems III
 - 3.2.7 Follow-up on the work done on metadata and classification systems
 - 3.3. Promotion of dissemination

3.2. Benchmark

Metadata strategy developed by December 2013

Database model with data flow diagram prepared by December 2013

Plan on development of software for classification database by December 2013

3.2.5. Activity

Metadata system IV

Purpose of activity

The expected activities are:

- Short presentation of strategy for metadata system and part for classifications
- Introduction of classification – current situation in statistical offices
- Presentation of Neuchatel data model
- Database model for classifications
- Software for classification to be discussed. Must take considerations on standards and integration in the strategy into account
- Presentation of classifications on the web

Expected outputs

- Strategy revised with review comments

- Database model for classification. The database model should be based on conceptual model of Neuchatel and sufficiently open for implementations. It is assumed that the DDI implementation will be based on Neuchatel including integrations with other metadata-elements.

List of Participants

Experts

Mogens Grosen, Statistics Denmark (participated Monday, Tuesday and Wednesday until 11 AM)

Søren Netterstrøm, Statistics Denmark

Agency for Statistics of BiH

Mr. Sulejman Hasanović, Head of Department for applications and database management

Mr. Alen Mrgud, Senior officer for the development and maintenance of classification server

Institute for Statistics of Federation of BiH (FIS)

Amer Korić, Head of Department for database maintenance

Amel Sikirić, Senior officer for designing and programming

Institute for Statistics of Republika Srpska (RSIS)

Mr. Aleksandra Zec, Senior Official for the DTP Design

Twining Project Administration 00 (387) 33/911-992

Bente Dyrberg, RTA, 00 (387) 66/657-967

Djemka Sahinbasic, RTA Assistant, 00 (387) 63/721-104

Svjetlana Pavicic, Interpreter

Agenda

Time	Day	Place	Event	Purpose / Detail of event
13:00-15:30	Monday	BHAS	Meeting with the BC Component Leader and BC Experts	Strategy for metadata system and classifications Introduction in classification – current situation in BiH
8:30-15:30	Tuesday	BHAS	Meeting with the BC Component Leader and BC Experts	Presentation of Neuchatel data model
8:30-15:30	Wednesday	BHAS	Meeting with the BC Component Leader and BC Experts	Results in report regarding strategy and discussion of next missions. Database model for classifications Software for classifications
8:30-15:30	Thursday	BHAS	Meeting with BC Experts	Database model for classifications Presentation of classifications on web site
8:30-12:00	Friday	BHAS	Meeting with BC Project Leader and Experts	Presentation of results, Agreement on report. Next missions