

Metadata and classification system development Bosnia and Herzegovina

Project BA 08-IB-FI-03



MISSION REPORT


on

Activity 3.2.1: Assessment of the current situation regarding metadata and classification systems

Mission carried out by
Mogens Grosen Nielsen
Statistics Denmark

Denmark, 30 April 2012

Version: FINAL

		 STATISTICS DENMARK
		Statistics Denmark

Author's name, address, e-mail

*Mogens Grosen Nielsen
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
DDI	Data documentation initiative

Executive summary

This mission report was prepared within the state and entity statistical institutions in Bosnia and Herzegovina. It was the first mission to be devoted to topic 2: “Metadata and classification system” 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.

1. Introduction

This mission report was prepared within the state and entity statistical institutions in Bosnia and Herzegovina. It was the *first* 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 consultant would like to express his thanks to all officials and individuals met for the kind support and valuable information which he received during the stay in Bosnia and Herzegovina, and which highly facilitated the work of the consultant.

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

2. Objectives in general

Implementation of the activities foreseen in this project will significantly contribute to the fulfilment of priorities stated in European partnership for BiH:

Short-term priorities:

“- Implement the agreement between the Entities in the statistical system on improvement of the work of the Central Statistical Agency in Bosnia and Herzegovina and improve the range and quality of statistics, in particular at State level.”

Medium-term priorities, European standards, Sectoral policies:

*“- Develop reliable economic statistics and build up institutional capacity to produce and publish basic statistical data harmonised with European standards, in particular in the areas of national accounts, agricultural, macro-economic **and business statistics**, and social statistics, including education, labour and health statistics*

- Carry out the population census” (Contract)

Component 3 in the project - Strengthening the Institutional Capacity of BIH Statistical System - comprises long term strategy and metadata. The objectives for component 3 are: a) sustainable development and b) better communication between the three statistical institutes and between providers and users of the statistics.

This component is planned to strengthen the institutional capacity for the BiH statistical sector as a whole. This horizontal component should

- *improve practical co-operation between the three institutions.*
- *Such co-operation requires the further work towards*
 - *a common IT architecture,*
 - *the organisation of joint meetings, and*
 - *common access for the statistical managers in BiH to experience and know-how in the organisations,*
 - *co-ordination and implementation of statistical work.*

- *staff in the different Statistical Institutions*
- *the availability of joint training and other learning experiences to acquire the same knowledge on harmonised methods and the same practical experience to apply these harmonised methods. (Contract)*

The benchmark for metadata and classifications are

- *Metadata strategy developed by 8th project quarter*
- *Database model with data flow diagram prepared by 8th project quarter*
- *Plan on development of software for classification database by 8th project quarter (Contract)*

This includes technical documentation, including selection of IT tools for Metadata system, prepared Development plan and preparatory activities carried out for Classification server.

Objectives for the mission

The purpose / expected result of mission was

1. *Improve the knowledge of all participants regarding metadata and classification systems;*
2. *Define the basic concept of metadata and classification systems in Bosnia and Herzegovina Statistics*
3. *Report detailing the current state of classifications and metadata. Detailed road map for implementation of other activities. (From ToR)*

The purposes of the activity involved:

1. *Presentation of the basic metadata functionality and classification system from experts' experience*
2. *Assessment of current situation regarding metadata and classification systems (Presentation of the completed activities in metadata and classification system)*
3. *Overview of classifications use, quality systems used, documentation systems, present software. What is the role and status of Metadata Manager (Present the current situation regarding metadata and classification in Statistical Institutions in Bosnia and Herzegovina)*
4. *Determination of possibilities, frame and needs (EU requirements);*
5. *Discussion on location and responsibilities for updating metadata/classifications;*
6. *Evaluation of previous project assessments and results related to metadata and classification systems; (from ToR)*

3. The situation today, metadata-strategy, issues and recommendations

This chapter presents the results of the activities at the mission: Section 3.1 presents results from the discussion of the current situation. Section 3.2 is about metadata-strategy and presents possible ways forward including timetable, standards, business processes etc. Section 3.3 addresses specific issues that were collected during the mission. Each issue is closed with a recommendation.

3.1 The situation today

Regarding classifications a lot of work has been carried out. The main classifications are available on the Internet (including search-facilities). Differences in classifications and code-lists in BHAS, FIS and RSIS were discussed. Regarding classifications there are only few differences. Regarding code-lists there are more differences.

A short analysis of existing business processes was carried out on the SBS-survey. The analysis followed the GSBPM-processes for development (Need, Design, Build) and operation (Collect, Process, Analysis, Disseminate and Evaluate). For each process the following items were collected: input, output, involved staff, databases and classifications. In addition a box below the table shows: a) Who are the users? b) Metadata used / produced and c) remarks. See result in annex 2.

The analysis and the discussions showed that there are many differences in business processes depending on which people / entities are involved. This makes it difficult to build common metadata-system and other common systems. Metadata today are stored locally with no common structure.

In order to get a full picture of the situation it is recommended to collect summary information: For each survey the following items should be collected: a) database and applications and b) metadata (Methodology and quality metadata, Concepts, variables, Variables and Classifications/codelists). The collection for each survey must show the situation in BHAS, FIS and RSIS. The information should be integrated into an existing database on surveys, that already exists at BHAS.

The on-going work on quality and related elements were presented and discussed. The inclusion of methodological and quality indicators in the coming metadata system was discussed. Today the quality report is produced and used after end of production.

Section 3.3 later in the report gives more detail and recommendations on issues recorded..

3.2 Metadata strategy

The development and implementation of a metadata-strategy was discussed. It was agreed to start with the following draft for outline and timetable.

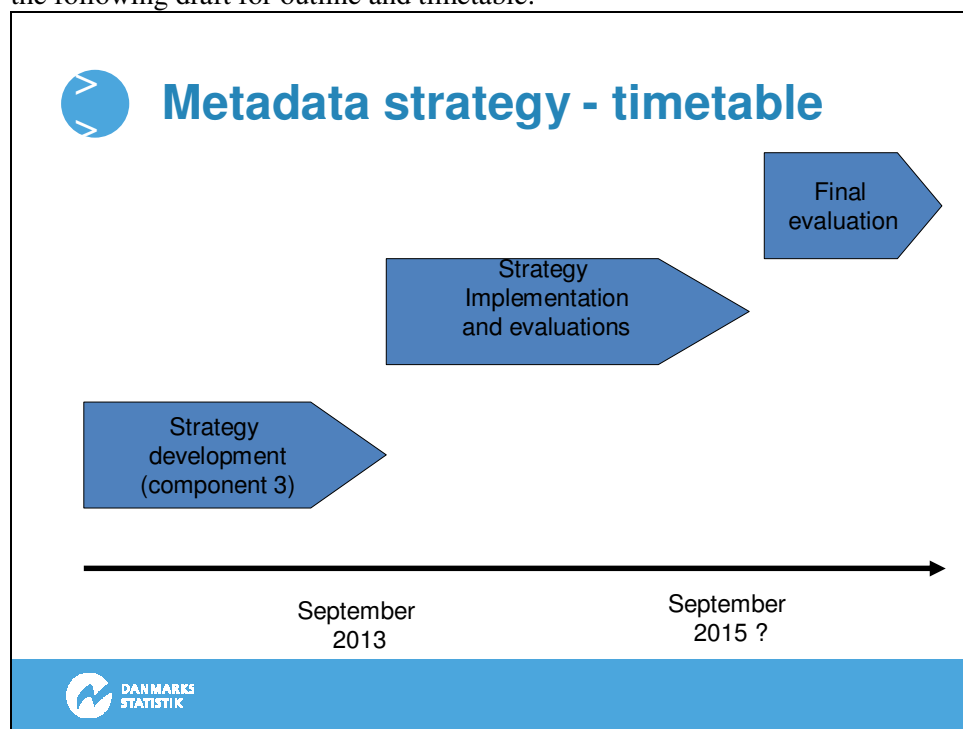


Figure 1. Timetable for development and implementation of the Metadata Strategy.

The development of the strategy should be finalised at the 8th project quarter in September 2013.. It is recommended to implement the strategy using a staged approach. This means that the timetable for the implementation should be divided into 3-6 pieces. Each piece should start with a detailed plan and be closed with an evaluation. Using this approach ensures that the project is adjusted using collected experiences. The approach furthermore ensures (via evaluation and planning) that participants are learning about the process, tools and contents of metadata.

The following outline to a metadata-strategy was discussed:

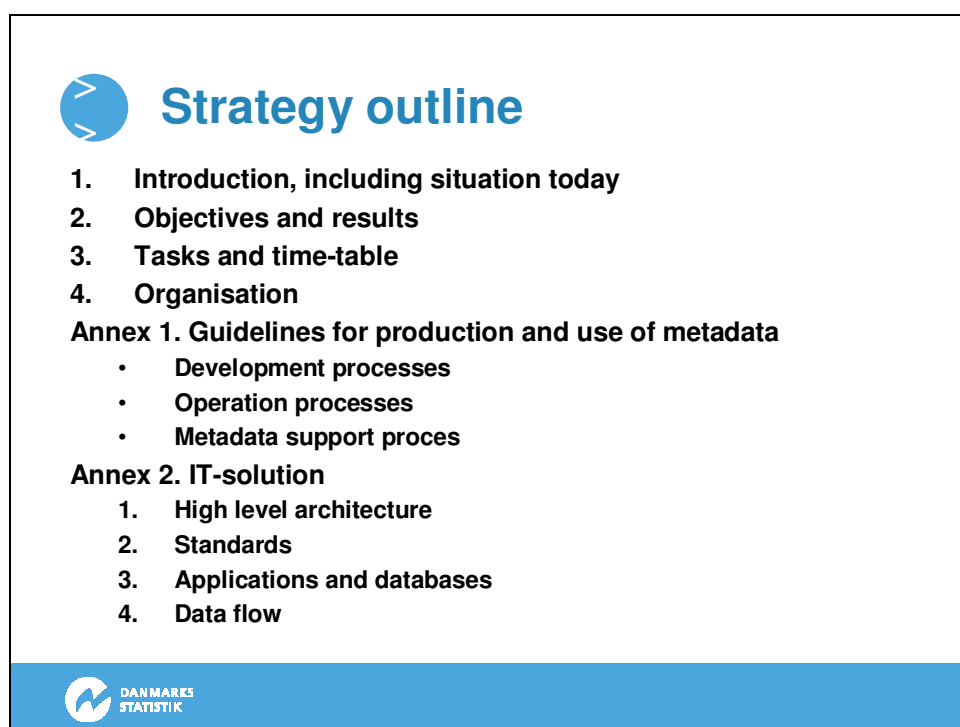


Figure 2. Outline for a metadata-strategy (draft)

Regarding the guidelines for production and use of metadata (annex 1 in the outline) it was agreed to use the GSBPM-model. In order to fulfil objectives regarding user-needs and communication the following model should be used as a help on producing the guidelines. An important element is processes to handle feedback and knowledge in relation to users.

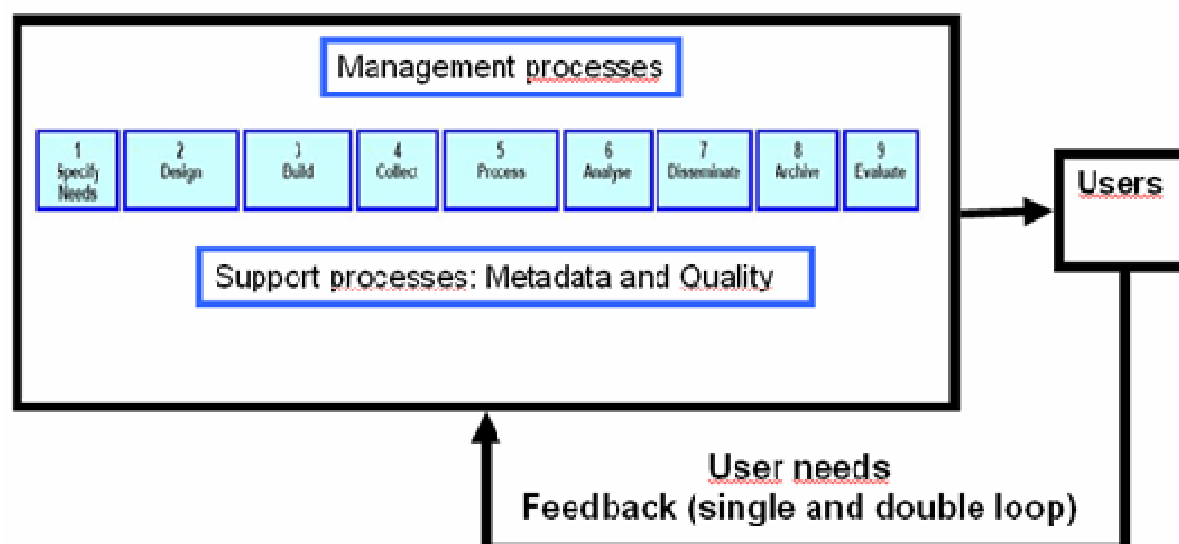


Figure 3. GSBPM with users and feedback

The use and definition of metadata were discussed. It was agreed to use the following high level model:

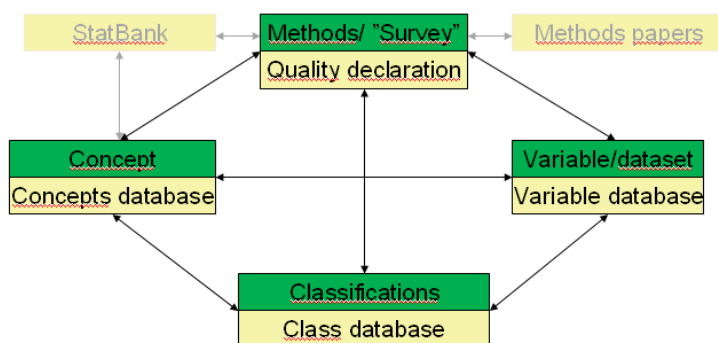


Figure 3. High level metadata model

This implies that metadata consist of two main elements:

- Quality report (methodology, relevance, accuracy, comparability etc)
- Other metadata (concepts, variables and classifications)

The use of standards and tools (DDI, Neuchatel, Metapplus etc) was discussed. Solutions in Sweden, Portugal, Denmark and Australia were presented and discussed. It was apparent that all standards in more or less complicated ways can handle information on statistical units, questions, variables, concepts and classifications. It is important to use international standards (and not develop proprietary standards and tools) for the following reasons:

- Sustainable development
- No 'lock-in' to software
- Support integration / exchange
- 3 layer software model / many tools
- "Division of labor" in the international statistical community
 - National Statistical Organisations should share the work on developing standards and tools
 - Standard tools should be flexible and easy to implement

The work internationally follows two main tracks: a dynamic process model GSBPM (Generic Statistical Business Process Model) and GSIM (Generic Statistical Information Model); the information model supplement each other. The report from global GSIM working group states:

- *"Operationalizing GSIM highest priority ... for example, a commonly agreed to representation in XML"*
- *"The workgroup identified [SDMX](#) and [DDI-L](#) (DDI-Lifecycle) as the key starting points in this regard"*

Initiatives should follow the work on standards and benefit from the work internationally. This is now supported by software. Both free and not free.

3.3 Issues and recommendations

This chapter summarises the discussion on various topics. Each topic is closed with a recommendation.

1. The *use of standards and tools* (DDI, Neuchatel, Metapplus etc) was discussed. Solutions in Sweden, Portugal, Denmark and Australia were presented and discussed. It was apparent that all standards in more or less complicated ways can handle information on statistical units, questions, variables,

concepts and classifications. It is recommended to investigate the DDI-standard. A pilot study on using DDI on SBS should be conducted.

2. There are many *differences in business processes* depending on which people are involved. This makes it difficult to build common a metadata-system and other common systems. It is recommended to adopt GSBPM as a common generic business process model. As a first step the GSBPM model should be translated. On the next mission a strategy for implementing GSBPM should be prepared.

3. The *current situation on metadata* was discussed. Metadata today are stored locally with no common structure. A template for describing business processes was tested for the purpose of getting an overview of metadata. SBS was used as pilot. It is recommended to use a simple template with the purpose of getting an overview of metadata, data and applications. It is recommended to collect the following information for every survey: a) short description of which data/databases are produced and short description of which applications are used and b) short description of metadata used and produced (Methodology and quality metadata, Concepts, variables, Variables and Classifications/codelists). The collection for each survey must show the situation in BHAS, FIS and RSIS.

4. The *location and responsibility of the common metadatabase* was discussed. According to the contract: "Metadata and classifications have to be stored and updated in one place: the Agency." RSIS stressed that the database should not be centralized, since not all metadata would be shared among the three parties. The Agency and FIS stressed the importance of having metadata stored only one place. There was no agreement on this. It is recommended to build a centralised common database in order to move towards a situation where metadata can be established as a common resource. It should be possible to update metadata by and from all three parties.

In order to make sure that all three parties work in the same direction is recommended that each part is chairing the process of data and processes for 1 year and after one year the responsibility moves to the next part.

If one of the parties has their own surveys they can use the common metadata system but it is not mandatory. It is recommended that the same standards are used as much as possible.

5. *Differences in classifications and code-lists in BHAS, FIS and RSIS* were discussed. Regarding classifications there are only few differences. Regarding code-lists there are more differences. It is recommended to set up a working-group that first describes the current situation and afterwards does the harmonization.

6. Due to *differences in legislation* in BHAS, FIS and RSIS, there are differences in variables in administrative sources. It is recommended that statistical organizations are involved in definition of administrative files including discussion of metadata in order to fulfill user needs from statistics. It should also be noted that using administrative sources requires work on defining statistical concepts and doing necessary harmonization.

7. The problem of having *old and new coding systems* was presented. What to do with historical surveys. It must be possible to include historical metadata in the coming metadata-system. It is recommended to assess the needs for historical metadata and hereafter decide on the extent of historical metadata to include.

8. *How does implementing metadata standards affect our work.* There is both an organizational and a technical challenge. The organizational challenge should be handled in the coming organization and the implementation of GSBPM. Regarding the technical challenge: In the short run: metadata should be extracted from existing systems and put into a common metadata-system. In the long run it is recommended to establish a library with concepts, variables, statistical unit-types, questions etc. and use these as building blocks when you create a survey.

9. The *human resource situation* was discussed. There is a lack of resources in all three organizations. It is recommended to ensure knowledge is shared among IT-staff. A list of activities should be drafted and discussed and put in the strategy. The top-management must approve the strategy and ensure that necessary resources are available. As part of the establishing a metadata-system, subject-matter and methodological staff are important players as they define requirements. In addition, requirements from external users should be investigated and described.

10. The *inclusion of methodological and quality indicators* in the coming metadata system was discussed. Today the quality report is produced and used after end of production. It is recommended to include a top-level methodological document and quality-declarations in the coming metadata-system. The work on methodology and quality indicators should be an integrated part of needs, design and build processes. The reason for this is that quality and methodology information is valuable both in relation to discussion with external and internal users. For the technical solution it is recommended to publish / present metadata in an integrated way using ideas of presentation from Portugal (www.ine.pt) and others.

11. The *reports from SIDA* were evaluated. The recommendations have been taken into account during the mission, including setting up an outline of the coming metadata-strategy. SIDA suggested an organisation with content- and metadata-staff forming groups. It is recommended to further discuss the recommendations from SIDA. Regarding organisation it is recommended to establish a metadata-process as a supporting process to the statistical processes in GSBPM. Discussions on how to do this will continue on the next mission.

4. Plan for work until next mission

1. DDI-pilot on SBS (recommendation 1)
2. Plan for improving code-list – tasks, timetable and participants - draft version. (recommendation 5)
3. Translation of GSBPM document (recommendation 2)
4. Overview of metadata (recommendation 3)

5. Topics for the next mission

1. Follow up on work decided (see chapter 2)
2. Elaboration of strategy for implementing GSBPM
3. Elaboration of metadata-strategy
4. Discussion, training etc
5. Detailed road-map for implementation of activities in component 3.2 (including planning of future missions)

Annex 1. Terms of Reference

Metadata and Classification System Development

23th to 27th of April 2012

Terms of Reference - Activity 3.2.1

EU Twinning Project BA 08-IB-FI-03

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 8th project quarter

Database model with data flow diagram prepared by 8th project quarter

Plan on development of software for classification database by 8th project quarter

3.2.3. Activity

Assessment of the current situation regarding metadata and classification systems

Purpose of activity

The expected activities are:

- Presentation of the basic metadata functionality and classification system from experts' experience
- Assessment of current situation regarding metadata and classification systems (Presentation of the completed activities in metadata and classification system)
- Overview of classifications use, quality systems used, documentation systems, present software. What is the role and status of Metadata Manager (Present the current situation regarding metadata and classification in Statistical Institutions in Bosnia and Herzegovina)
- Determination of possibilities, frame and needs (EU requirements);
- Discussion on location and responsibilities for updating metadata/classifications;
- Evaluation of previous project assessments and results related to metadata and classification systems;

Expected output

- Improved the knowledge of all participants regarding metadata and classification systems;

- Defined the basic concept of metadata and classification systems in Bosnia and Herzegovina Statistics
- Report detailing the current state of classifications and metadata. Detailed road map for implementation of other activities.

List of Participants

Experts

Agency for Statistics of BiH

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

Institute for Statistics of Federation of BiH (FIS)

Institute for Statistics of Republika Srpska (RSIS)

Mr. Aleksandra Zec, Senior Official for the DTP Design

Ms. Nevenka Bozalo, Officer - Administrator for metadata system

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

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

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

Haris Imamovic, Interpreter

Agenda

23 April (Monday) – 27 April (Friday)

Agency for Statistics of Bosnia and Herzegovina – Zelenih beretki 26, Sarajevo

Time	Day	Place	Event	Purpose / Detail of event
12:00-16:00	Monday	BHAS	Meeting with the BC Component Leader and BC Experts	<p>Presentation of the basic metadata functionality and classification system from experts' experience with accent on Neuchâtel model;</p> <p>Expert, BHAS, FIS, RSIS</p>
9:30-16:00	Tuesday	BHAS	Meeting with the BC Component Leader and BC Experts	<p>Presentation of the completed activities in metadata and classification system;</p> <p>Overview of classifications use, quality systems used, documentation systems;</p> <p>BHAS, FIS, RSIS</p>
9:30-16:00	Wednesday	BHAS	Meeting with the BC Component Leader and BC Experts	<p>Presentation of materials for metadata system in Bosnia and Herzegovina;</p> <p>Determination of possibilities, frame and needs (EU requirements);</p> <p>Expert, BHAS</p>
9:30-16:00	Thursday	BHAS	Meeting with BC Experts	<p>Discussion on location and responsibilities for updating metadata/classifications;</p> <p>Expert, BHAS, FIS, RSIS</p>
9:30-12:00	Friday	BHAS	Meeting with BC Project Leader and Experts	<p>Presentation of results, Agreement on report. Implications for the next activities (missions);</p> <p>Expert</p>

Annex 2. Metadata and classification system development in Bosnia and Herzegovina: survey name: SBS

Development

GSBPM Proces	Input	Output	Involved / responsible
Needs / requirements	a) EU regulations b) Requirements from nat. account / c) Business register / and other internal requirements d) Internal initiatives	a) Workplan b) Decision	a) Work group with 3 entities involved b) Internal users
Design	b) Needs	a) Methodology b) Questionnaire c) Controls d) Definition of variable, code-list, indicators etc	Work group with 3 parties involved
Build	a) Entities requests b) Agency: Request from entities	a) Applications and databases	a) Methodologist and IT b) Entities

<i>Operation</i>				
GSBPM	Input	Output	Involved /responsible	Applications and databases(archive)
Proces				
Collect	a) Sample b) Regulations c) Empty questionnaire d) Request for logical controls e) Changes in classifications Code lists	a) Filled in questionnaires b) Data in database	a) Entities (collect, key in, control and process data)	a) Web-application b) Sql server a) Client server application – centralized – sql server
Proces	Data in database Request for logical controls Outlier defined values	Clean database Lists with errors Variables at several levels: (survey, unit etc Indicator – unit, section etc	Entities	same
Analyse	Entities: Clean database indicators etc Agency Clean databases	Entities: Reviewed results Imputation etc Published paper Agency	Entities (analysis department) Sample department Agency countrylevel – all parties involved	Entity Same R SPSS Agency SPSS
Disseminate	RSIS: aggregated data Agency: Microdata Report from entities	Entities Reports Agency: Reports	Entities	Basic sw Entities: Agency:
Evaluate			Agency	

Additional information and remarks

Who are the users?

Metadata used / produced Methodology and quality metadata: produced and included in methodology report
 Concepts: only internal in system produced
 Variables: only internal in system produced
 Classifications / codelists: only internal in system produced

Remarks