

REPUBLIC OF MACEDONIA
STATE STATISTICAL OFFICE

Metadata Strategy 2013-2015

May, 2013

Statistical Metadata is any information that is needed by people or systems to make proper and correct use of the real statistical data, in terms of capturing, reading, processing, interpreting, analysing and presenting the information (or any other use). In other words, statistical metadata is anything that might influence or control the way in which the core information is used by people or software.¹

1. Introduction

As statistical metadata becomes an increasingly important part of statistical information for all parties involved (data providers, data producers and data users), it is vital to ensure that State Statistical Office of the Republic of Macedonia (SSORM) business gets the most benefit out of Metadata System introduction.

In order to achieve this, SSORM needs to make sure that the type of Metadata System chosen supports organization business and adds most possible value to it.

It is important to plan ahead for the development of the Metadata System, which will need to be able to grow and adapt to any changes in statistical production business.

The future of Statistical Metadata lies in a professional approach to system design and software development, producing metadata-aware applications² that can use, capture, present, manipulate and re-use metadata and the associated resources which it describes.

The Metadata Strategy provides a roadmap for how metadata will be structured what metadata standards will be adopted, where and how it will be captured and stored, for what reasons it will be used and in what kind of overall statistical data production system architecture it will be implemented.

2. Mission

SSORM produces and disseminates official statistical data on the Macedonian economy and society as a basis for the process of decision-making and democracy, based on information. Statistical data will be of high quality, impartial, available, simple to use and published in a timely manner, supported by statistical metadata of high quality and according to adopted international standards.

3. Vision

To be recognized as an institution that provides quality, timely and internationally comparable statistical data. This would be achieved by implementing advanced technologies based on statistical metadata driven approach for statistical information production and dissemination.

¹ MetaNet: *Network of Excellence for Statistical Metadata*. See www.epros.ed.ac.uk/metanet

² Software applications recognizing structured metadata items thus allowing adopting applications to metadata content

4. Description of current situation in SSO

4.1. SSORM Statistical Business Process Model

UN/ECE Generic Statistical Business Process Model (GSBPM) has been adopted as SBPM of SSORM with some minor modification in the second and third level of activities. As a common framework, SBPM leads to common understanding across various areas of production and better understanding of processes which could be advantageous in their optimization and integration over statistical surveys. Implementation of the well described and documented model in the IT systems supporting common activities is underway. Fully implementation of the SBPM is a long time-consuming process that requires substantial organizational changes in the statistical production, particularly changes in the daily work of the participants in the process.

4.2. Structural metadata

Referring to structural metadata, SSORM is in the process of collection and analyses of statistical variables from different subject-matter areas. SSORM developed methods and excel tool for standardization, harmonization, documentation and variable inventory in the first phase of the establishment of the variables sub-system... This will be a necessary precondition for the establishment of variables sub-system. In regards to the development of classification sub-system, SSORM intends to adapt the Neuchatel terminology model for classifications³.

4.3. Reference metadata

SSORM has established the system for collection and dissemination of reference metadata based on the Euro-SDMX Metadata Structure (ESMS). The main aim of the system is to document all existing statistics in accordance with the defined time frame in order to fulfil the reporting needs in regards to Eurostat ESMS and ESQRS structures, but also for dissemination purposes i.e. to release reference metadata on the internet in a standardized manner.

State Statistical Office applies the Eurostat recommendations for transmitting reference metadata using the ESS Metadata Handler application (ESS MH). These metadata refer to data sets sent to Eurostat Single Entry Point (SEP) using the ESMS and ESQRS (ESS Standard for Quality Reports Structure) structures. The intention of the State Statistical Office is to accept recent developments of ESS MH and to apply new metadata structures: Eurostat Process Metadata Structure (EPMS) and The Euro SDMX Metadata Structure for Inventories (ESMSINV).

³ <https://unstats.un.org/unsd/class/intercop/expertgroup/2013/AC267-9.PDF>

5. Basic strategic principles

SSO will develop a metadata driven production system. It is a change in the traditional work approach and culture and requires training and adopting best practice methods according to the given internal conditions. For successful implementation of the metadata system, the most important is the support by the management of the Office.

A Statistical Metadata System (SMS) will be developed through the following steps:

- 5.1. The elaboration and organization of a Metadata based work flow;
- 5.2. Analysis and understanding of the metadata creating processes, usage of the appropriate standards as well as data flow analysis will allow for the design of a central element of SMS – Sub-system for structural metadata ;
- 5.3. Metadata will be captured all along the production process;
- 5.4. It will be stored in the metadata repository for running the production system covering the full statistical data life cycle;
- 5.5. SMS implementation will include the most essential parts as follows:
 - Establishing the metadata base as the central information source for the management of the statistical production process, necessary structural metadata (variables and classifications) and reference metadata (methodology, regulations, descriptions, quality issues);
 - Clear division of roles and responsibilities for all processes of metadata management covering the full statistical data life cycle from collection to dissemination;
 - Stepwise development of the contents and functionality of the entire metadata system;
 - Importing metadata from available internal/external sources in order to reuse it in the new environment;
 - Providing metadata browsing, retrieval and analysis tools;
 - Providing wizard applications for important production phases in order to support statisticians entering proper metadata;
 - Transferring metadata to different production tools (data warehouse, data edit processes, questionnaires etc.);
 - Providing tools for employees in order to run the production system properly.
- 5.6. Most important benefits from implementation and usage of SMS:
 - Besides increasing the efficiency of statistical production, the quality of statistical information will increase because metadata provides more background information, and statistical production will speed up and will become more up-to-date;
 - The most important principle will be to demonstrate results achieved within the process of development. This will ensure that the statistical staff understands and see the benefits of the solutions that will be implemented gradually, i.e. solutions that will significantly change the traditional manner of statistics production (stove pipe approach) and increase their efficiency.

6. SMS - core element of the production environment

The new statistics production environment will be based on use of statistical metadata as driving force of all statistics production processes and sub-processes. SMS will support:

- Statistical data collection processes for all collection methods;
- Integrated Metadata Driven Process Oriented Statistical Production System (IMD POSPS);
- Dissemination and interchange system on the Euro-SDMX standard basis.

7. Metadata in each phase of statistical business process

7.1. Metadata classification

Metadata in SSO information system will be classified dependable of its usage (see Figure 1, page 6):

- Structural metadata:
- Reference metadata:

7.2. Use of Internationally adopted Metadata Standards and Models

SSORM adopts ISO/IEC 11179⁴ for all metadata connected with production processes and quality assurance, Neuchâtel Terminology Model for classification database object types and their attributes⁵ and Euro-SDMX⁶ structure for reference metadata creation and usage.

7.3. Metadata created/used at each phase of SSO SBPM

All the metadata will be collected/created and used in compliance with the processes of SBPM adopted in SSORM. It will be done in line with the produced document “Guidelines for Survey Managers” (internal SSO document that defines the input /output metadata in each phase and activity of SBPM).

Mentioned documents are added as Annex 1 of the Strategy.

8. System design

8.1. Architecture of the Metadata system

Structural and reference metadata on the conceptual level will be stored in separated but interrelated metadata repositories like it is reflected in the Figure 1.

Reference metadata collected and stored provides SSO personnel with the competence building opportunity. It is important to organize easy and secure access to it for both SSO personnel internally and external users of statistical information via Web.

⁴ <http://metadata-standards.org/11179/>

⁵ <http://www1.unece.org/stat/platform/pages/viewpage.action?pageId=14319930>

⁶ http://epp.eurostat.ec.europa.eu/portal/page/portal/statistics/metadata/metadata_structure

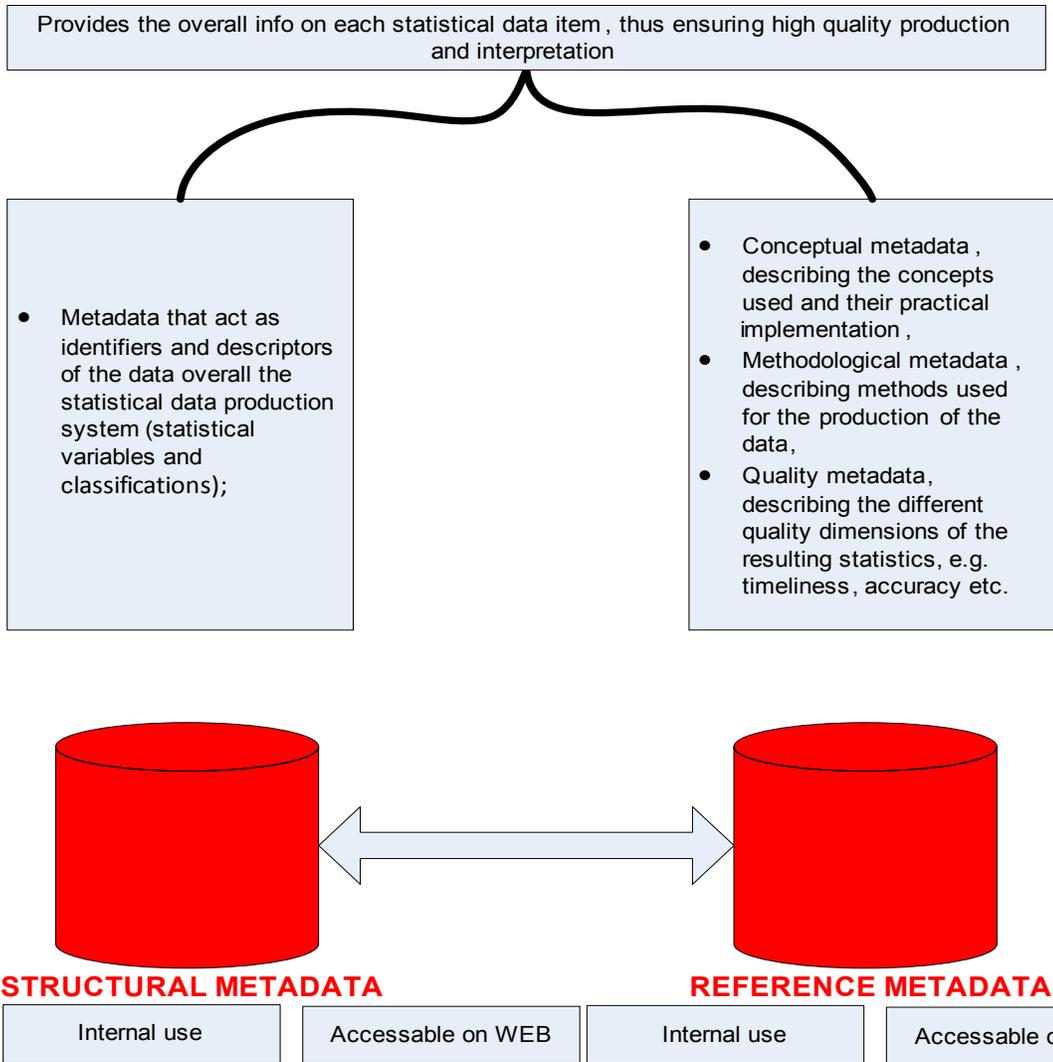


Figure 1 SSO Metadata system

8.2. Integrated Metadata Driven Process Oriented Statistical Production System Architecture

Within the IMD POSPS, SMS is a core element that drives all the system processes and control the data flows to and from system (import and export procedures). High level conceptual architecture of the system is presented on Figure 2.

The entire system has to be gradually built ensuring that Corporative Data Warehouse (CDW) will be the storage of all statistical information (data & metadata) throughout all statistical production and dissemination processes. All the storages of CDW will be interrelated thus ensuring usage of metadata- aware software, automation of main applications development (data entry, raw data validation, aggregation and outputs preparation) and the highest level of the information security.

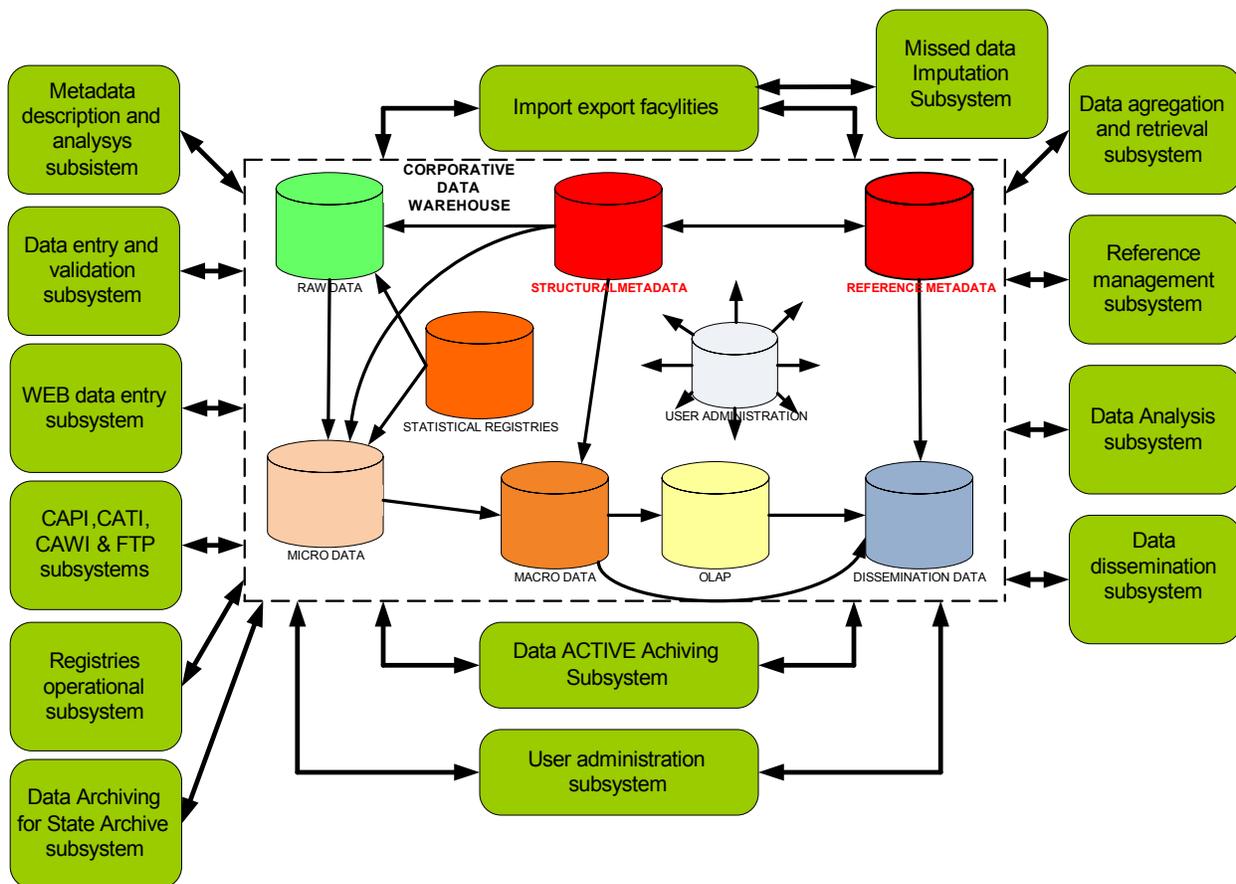


Figure 2 SSO IMD POSPS high level conceptual architecture

8.3. Tools for entire system development

To ensure the possibly lowest costs of entire system development, SSO will continue using Microsoft software platform. Applications that will be elaborated on the principles of metadata aware software will ensure metadata as driving force for entire system.

8.4. Ways of development. Outsourcing versus in house development

The best way of development the IMD POSPS keeping actual efficiency is based on a cost-benefit approach. Depending of the results, there are several possibilities for system development:

- Total outsourcing;
 - Total in sourcing.
- or the composition of both
- Selective sourcing.

Selective sourcing, which capitalizes on the inherent advantages of both internal IT departments' and external vendors' inherent cost advantages, is recommended.

Decision will be taken for which tasks which development method to use dependably of:

- Of resources available (human, technical and financial);
- Complexity of the task;

- Availability of internal experts;
- Time constrains.

8.5. System development phases

System development have to be done in sequence what ensures achieving fast and efficient results convincing statistics producers – subject matter statisticians on high efficiency of the metadata driven process oriented statistics production approach thus motivating them to assist the development and implementation in a most effective way.

Prioritization of the development and implementation will be done in strong accordance with SBPM.

SBPM main processes are subdivided in three main groups:

- New surveys preparation (Specify Needs; Develop & Design; Build);
- Surveys data processing (Collect; Process; Analyze);
- Dissemination and Archiving.

Specifics and interrelations of the main processes are described in the column “IMD POSPS development” in Figure 3. Develop and design process interrelates with Surveys data processing group of processes as for new surveys development as for existing surveys due to necessity of metadata descriptions.

Proposed sequence is presented in the Figure 3.

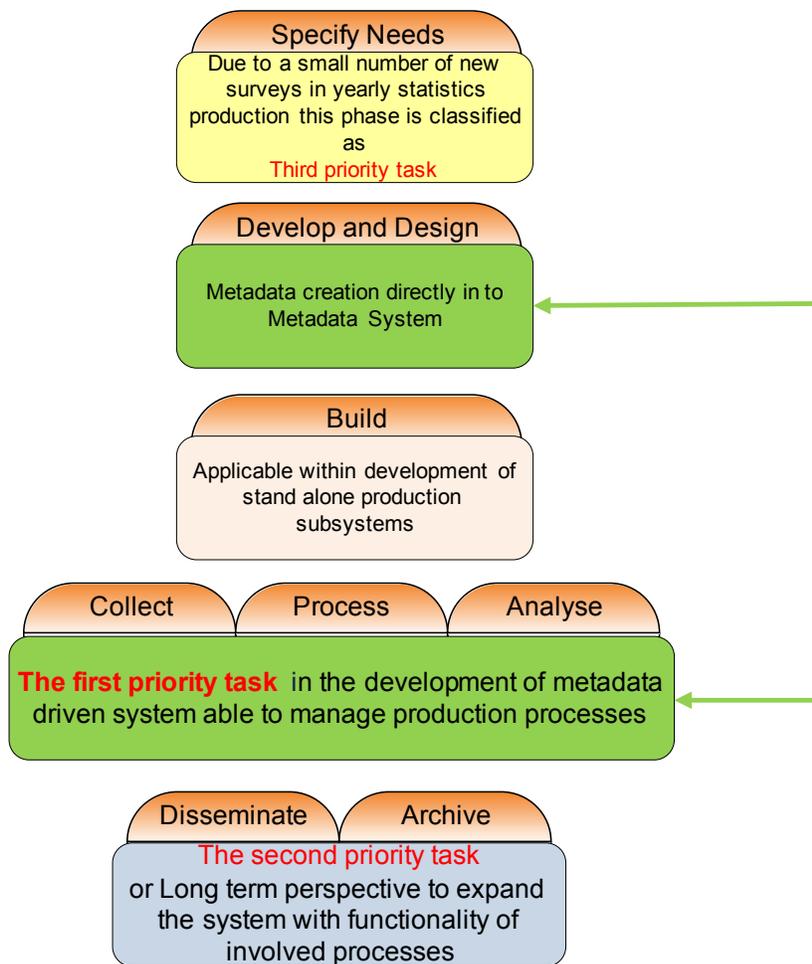


Figure 3 System design according grouped phases

9. Organizational and workplace culture issues

Transition of whole statistics production from Surveys production system to Process orient one will start when the core elements (the first priority task). Production metadata base and Micro data base will be developed, tested and launched in production environment.

To achieve the success of this reorganization and increase culture of the entire work the listed below activities will be implemented:

- Staff of statistical subject matter units will be regularly informed about the main activities within the system development process;
- Special seminars for employees will be organized to explain the advantages of new approach of statistics production and new division of tasks and responsibilities which will lead to significant changes in the organizational structure of the SSO;
- Statisticians will be involved in the defining of the development tasks, testing of interim solutions and completed applications as much as possible;

- Training programs will be timely elaborated and implemented;
- ICT infrastructure (particularly workstations and networking) will be tested on compliance of new production system requirements;
- The first system modules and/or sub modules successfully implemented will be promoted for wider usage.

9.1. Partial process oriented working environment

Development of metadata driven statistical data production system is connected with the transition from “stove-pipe” to process-oriented data production approach. It is proven in practice that the best results are achieved implementing partial process oriented solution in which most time and resources consuming processes (collection, entering & editing and outputs dissemination) are integrated as for all surveys conducted. Data analysis (micro and macro), outputs preparation as well as new surveys design would still be a task of competent statisticians in subject matter divisions thus ensuring higher quality results.

The schema presenting partial process oriented organizational structure is presented on Figure 4.

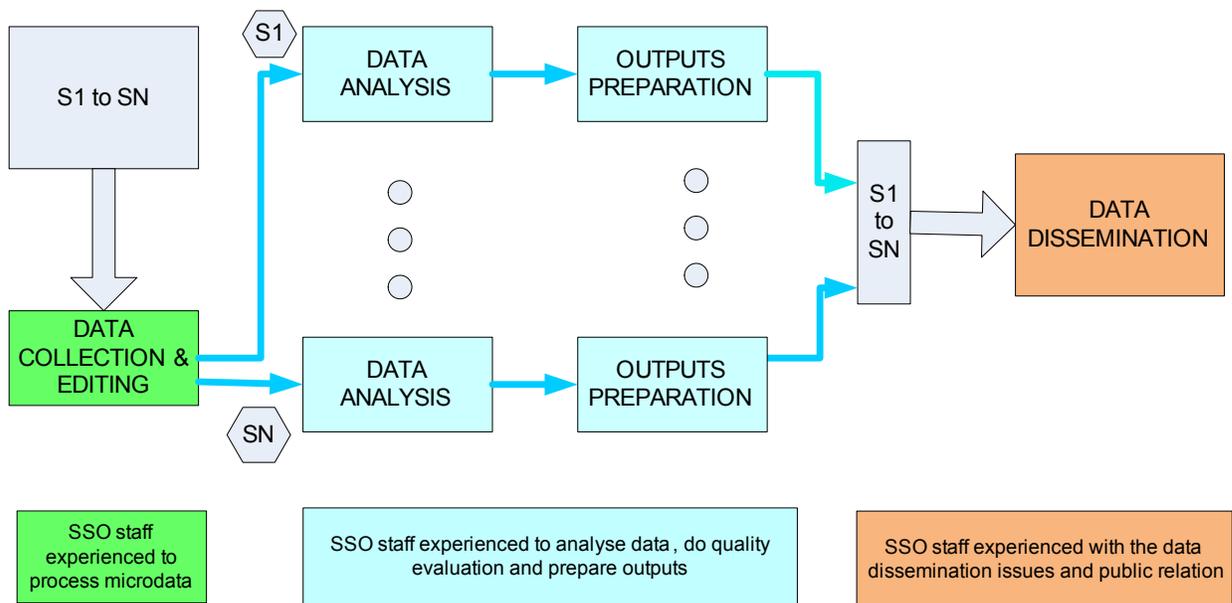


Figure 4 Proposed schema of partial process oriented production organizational structure

9.2. User types and roles

To ensure high level data security within the entire system and effective access to the necessary data within the data production processes special subsystem of the user’s administration will be created.

Roles of the users with different access rights will be defined in this subsystem. Number and types of roles will be elaborated and defined during the process of system design.

9.3. System governance and competence building (knowledge management)

System governance is a comprehensive framework for measuring and improving the system qualities. System quality is a very broad concept. It covers purposes, viability, supportability, risks, compliance to regulations, and compliance to technical standards.

The benefits of implementation of the System governance approach will be as follows:

- Improved business and IT relationship because IT infrastructure requirements are translated to measurable goals;
- Improved return on investment on IT infrastructure because management can track and enforce the use of new infrastructure;
- Reduced project failures because technical and compliance issues are identified earlier;
- Reduced cost for compliance with regulations and internal standards because these are managed within an efficient framework;
- Reduced long-term risks and costs due to fragmentation of standards because the management has visibility of the degree of compliance;
- Improved measurement of IT systems' performance, which is particularly valuable for governing outsourced contracts;
- In outsourcing contracts, system governance can define and measure the service provider's stewardship of the systems. It can ensure that the systems are of high quality and up to date at the end of the contract period. It can even be used to ensure that systems improve under the care of the outsourcer.

Technological development and diversity are placing large demands on the competence of both ICT staff and subject matter statisticians. SSO metadata competence must build on insight and experience of SSO key production processes, according to SBPM adopted

Close connections with other national statistical institutes in regards to the development of common solutions and exchange of competence are desirable.

Technological developments, based on metadata driven approach, open standards and platform independent systems, directing to new electronic services and forms of data and metadata collection, processing, dissemination and interchange nationally and internationally is a real working environment requiring a high competence level. SSO will increase its competence in order to make the best use of new developments.

The aim of the ICT operations are to contribute to the simplification, improvement and increased use of common working processes. Therefore, it will be necessary to increase the ICT staff's competence concerning methods and tools to be used for solution modelling and project management. Thus, the risk for projects failure can be avoided.

In order to meet the objectives of this Metadata strategy, ICT development projects must to a greater degree than before be organized across the organisational structure of SSO. Therefore, there will be a need for competence associated with help and support systems for coordinating and collaborating across the boundaries of different structural units.

Based on the above, a competence building plan will be developed. It will describe SSO current individual and collective metadata competence and set up the actions for its improvement.

The plan will provide specific actions for developing relevant competence required to reach the strategic goals.