

# **Implementation of Metainformation System in the Czech Statistical Office**

**Ebbo Petrikovits**

## **Summary**

The Czech Statistical Office (CZSO) launched in the beginning of 2005 year a substantial redesign of its Statistical Information System (SIS). Major goals of the project are: to increase efficiency of statistical production, to improve quality of statistical information and to reduce respondents' burden.

The redesign of SIS is based on the CZSO strategic goals. It is ranked as a first priority task for the whole office. The project development and implementation is directly supervised and managed by the top management of the CZSO.

The model of SIS is focused on statistical production, incorporating all phases of the production and dissemination process of statistical information (further statistical business process - SBP). Necessity to identify statistical business process and its phases called for the unification of statistical processes. The major aim of this exercise was to strengthen the organization and management of statistical work.

Statistical Metainformation System (SMS) is an important basic component of a redesigned SIS. SMS tools shall ensure integration of SIS inside and outside the statistical office. Furthermore, SMS shall contribute to evaluate efficiency of statistical processes and quality of statistical data. Currently, the first part of SMS project (subsystems Classifications, Variables and Tasks have been implemented) is in the phase of semi-production use.

Management and organization of the work, applied for the SMS development, resulted in a significantly increased cooperation and interest of subject-matter statisticians and methodologists. Lessons, learned from the project so far, clearly justified a direct involvement of top management as a necessary precondition to make SMS a success story.

The paper is divided into following chapters:

1. Redesign of SIS – purpose of SIS redesign, main goals and structure
2. SMS goals and architecture – functions and components of the SMS, contents and implementation status.
3. SMS management – management during implementation phase, management in production phase, SMS administration.
4. Conclusions – lesson learned and experiences from design and implementation.

## **1. Redesign of SIS**

National statistical institutions (NSIs) are facing increased number of users both on national and international level. Furthermore, a great challenge for NSIs is changing users' needs and calls for better quality of statistical information. At the same time, however, NSIs are requested to increase efficiency of statistical production and to reduce burden on statistical respondents.

Such development has a significant impact on the scope, contents and quality of observed statistical information and, consequently, on the methods, tools and techniques used for collection, processing and dissemination of statistical information.

In accordance with the CZSO strategic goals, a redesign of the statistical information system (SIS) was launched in the beginning of 2005 year. The first important step in this endeavour was to design a new architecture of SIS. Driving force for a new architecture of SIS is satisfaction of an increasing user's demand for statistical information (private sector, governmental institutions, international organizations, multinational enterprises etc.).

The basic conceptual strategy for statistical tasks<sup>1</sup> called "Model 2008" was approved at the beginning of 2005. Based on this strategy new document "Global Architecture of SIS" (GA SIS) was elaborated. The GA SIS is a basic strategic document for redesigns SIS. The goals and targets of GA SIS will become an integral part of the CZSO strategic documents.

Here it is necessary to explain the term "statistical task" what we introduced and use in our practice and in this document as well. We use this term instead of often-used term "survey". Coming out from the CZSO's practice surveys are usually a part of the spectrum of activities connected with statistical production of statistical data and information. With wider use of administration data for statistical purposes we will replace the set of surveys with administrative data suitable to gain input data for statistical production. Therefore we prefer term "statistical task" instead of "statistical survey" because the first one describes the whole process of statistical production adequate and more precisely. On the other hand it means that we will not work with surveys but with data from other sources. Therefore to name the statistical business process as a survey is misleading. Definitions of both terms are located in note 1.

After the GA SIS approval by the CZSO top management in June 2007, the corporate implementation plan, the budget proposal and the proposal for its financing have been prepared. The first implementation stage of GA SIS is on the way, its finishing is scheduled for 2011.

The **statistical business process** (process on production and dissemination of statistical information) is the main and most important process in the CZSO. The content and structure of this process (phases and sub-processes we designed in spring 2006. Recently, in spring 2009, the Statistical Division of the UNECE presented the generic statistical Business Process Model (GSBPM) developed under the umbrella of the UNECE METIS Work Session. The proposed model has been deeply discussed during the workshop and accepted for further recommended standard applicable in the NSIs.

In the new SIS architecture, all phases of the statistical business process (from assessment of users' requirements up to the dissemination of required information) have been carefully explored in order to design general methods, tools and techniques to support these phases. Such designed global model should have a link to other information systems inside CZSO (like bookkeeping, human resources management, management IS, etc.). The model of the SBP as a life cycle of statistical task we have applied in this work.

Core principles for Redesign SIS are as follows:

- a) systematic assessment and evaluation of statistical data requirements,
- b) increasing share of administrative data,
- c) increasing use of data modelling,
- d) implementation of SMS,
- e) implementation of statistical data warehouse,
- f) freeze of statistical surveys for 2-3 years,
- g) avoiding redundancy in statistical surveying.

Global architecture of SIS is composed of three basic parts:

- a) **Content component of SIS** describes statistical tasks system, its basic principles, links to statistical surveys, their composition and administrative data used.
- b) **Metainformation component** ensures systematic use of metainformation inside and outside SIS and provides tools for internal and external integration of SIS. Statistical metainformation system (SMS) is focused on SBP. The model for metainformation definition of a statistical indicator/variable ensures standard description of statistical data in all stages of this process (from statistical task formulation up to dissemination of statistical information).

---

<sup>1</sup> Note: Statistical task - is a set of statistical activities needed to fulfil a users' request for statistical information. The statistical task can be composed of one or more statistical surveys.  
Statistical survey - is a set of activities connected with the proposal of statistical questionnaire, preparing a sample, printing and distributing questionnaires, collecting completed questionnaires, data entry (including electronic collection of data) and data validation. Statistical surveys are always a part of statistical task.

- c) **Information and technological component** supports functionality of individual processes that are necessary for completing a content component. It includes methodological and technological tools enabling to assess and evaluate the existing and the new requirements of users and, after a positive evaluation, to prepare a statistical task. Furthermore, it includes tools for data collection and processing, mathematical models and mathematical and statistical methods, methods for data confirmation, release and dissemination. The target is maximum unification of procedures and tools used in all stages of statistical task design and processing. Statistical information will be centrally stored in data warehouse and released through dependent data marts (including public database).

GA SIS integrates all current and newly developed information systems in the CZSO. It defines new tools and their links, systems of processing, integration tools like SMS and statistical registers, ICT and the structure of data warehouse.

## 2. SMS Goals and Architecture

Effectiveness of a new SIS depends directly on the effectiveness of its SMS. The SMS is an integral part of SIS. Statistical metainformation has two basic functions: to inform about statistical content and statistical processes and to inform about technological processes.

A strong need to develop and implement a coherent SMS came in the CZSO along with the growing globalisation and spreading use of the Internet. Possibilities of electronic data collection, interactive communication with users and the need to provide electronic information for many national and international-information systems brought about the necessity to offer relevant statistical metadata to all participants in this process.

At the same time, a significant shift in priorities of statistical metainformation's functions could be observed. While in the past the first priority was the need of metainformation related to the technology, the current clear priority is functions related to the statistical contents and methodology.

In 2004 the CZSO declared in its Strategic Programme (Mission, Vision, Strategic Goals and Sub-goals) the SMS as a priority task. The CZSO management in February 2005 approved the SMS strategy. This document took carefully into account all important lessons learned from the work with metadata in the past. It was clearly recognized that the lack of a central coordination in the design and implementation of statistical metadata in the past resulted in many duplicate classifications, inconsistent definitions of statistical indicators/variables, etc. Users of metadata faced the lack of coordination in a unified interpretation of statistical information, diverse user's guidelines and tools for the work with metadata and/or duplications in methods and forms of metainformation description of statistical data.

The SMS shall create tools supporting unification and standardization of work processes inside the CZSO and create a knowledge base about SIS. Such knowledge base will enable to share information about design, implementation and running of statistical tasks by the whole statistical staff. It could diminish a risk of the CZSO staff migration.

The main goal of the SMS at present is to support the key task of the CZSO, i.e. **the statistical business process** (production and dissemination of statistical information). In this context the SMS shall support especially the following statistical activities:

- a) management of methodology-related activities,
- b) assessment of statistical data quality,
- c) provision of statistical data to users,
- d) monitoring of respondent burden,
- e) integration of SIS with IS of public administration,
- f) integration of SIS with IS of international organizations,
- g) design and implementation of statistical task (collection, processing and dissemination of statistical data and statistical information),
- h) management of statistical task processing,

- i) assessment and evaluation of statistical processing,
- j) monitoring of statistical system performance,
- k) monitoring of users' satisfaction,
- l) monitoring of costs and benefits of SIS.

## 2.1. SMS architecture

To assure unified approach to the design of the SMS subsystems specific document "Global architecture of SMS" was elaborated. It specifies main principles for global SMS architecture are as follows:

- a) SMS must be equipped with an access portal ensuring a unified user's communication ;
- b) use of metainformation in the SIS processes should be ensured,
- c) unified user interface for internal users (search, update, administration),
- d) unified interface for external users (navigation, selection, interpretation),
- e) unified data interfaces between individual subsystems of SMS,
- f) keeping history on the SMS objects,
- g) unified storage and update of metadata,
- h) recording time of update and its author,
- i) unified user's documentation,
- j) unified technical documentation,
- k) unified model of data protection,
- l) consistency of metadata inside the SMS subsystem and between them,
- m) unified technological tools for implementation.

SMS architecture is modular. It is composed of following, relatively self-sustainable, mutually interlinked subsystems:

- a) **Statistical Classification (CLASS)** – maintenance and update of statistical classifications/nomenclatures. It is the only source of classifications/nomenclatures for design, implementation and running of statistical tasks. From this source the classifications/nomenclatures are provided to internal and external statistical users.
- b) **Statistical Variables (VAR)** – maintenance and update of the catalogue of statistical variables. Description of VAR is based on the **metadata model** valid and used for VAR in all stages of the SBP. Metainformation from this subsystem is used for data collection, data processing, analysis and preparation of final outputs and dissemination of statistical information. Furthermore, the metainformation is used for description of statistical data stored and maintained in statistical data warehouse
- c) **Statistical Tasks (TASKS)** – maintenance of metadata necessary for the design and processing of statistical tasks (basic characteristics, statistical questionnaire definition, other input data sets, decree on annual programme of statistical surveys, data validation rules, definition of statistical samples, specifications of imputation methods, quality requirements, aggregations, specification of users, time-tables for data collection and processing, applied nomenclatures/code-lists, legislation, provider of ICT services, specification of ICT services, etc). The subsystem is a basic source of information on statistical tasks from which the metadata for the design, implementation and running of statistical tasks are drawn. It keeps history of individual statistical tasks specifications.
- d) **Statistical Quality (QUALITY)** – maintenance and update of qualitative characteristics and methods for statistical data assessment.
- e) **Statistical Time Series (T-SERIES)** – maintenance and update of metadata on current statistical time series. Description of principles of time series maintenance. Specification of variables, which are to be kept in time series, levels of aggregations, etc.

- f) **Dissemination (DISSEM)** – maintenance and update of metadata linked with activities related to dissemination of statistical information (statistical publications, electronic outputs, web site, etc.).
- g) **Respondents (RESP)** – maintenance and update of metadata on respondents, respondent burden, respondent opinions, links to statistical survey etc.
- h) **Users (USERS)** – maintenance and update of metadata on the SIS external users (users' opinions, FAQ, etc.). Outcomes from the user satisfaction surveys. Users' proposals for further SIS development.
- i) **Data Fund (D-FUND)** – maintenance and update of metadata on contents and structure of files included in SIS and provided to external users.

SMS is interlinked with the Statistical Registers system. The main registers in this system are the following: Business Register, Register of Census Districts and Buildings, Register of Farms and Population Register.

## 2.2. Implementation status

In time being of writing this paper the implementation state of individual subsystems was as follows:

**Subsystem CLASS.** The subsystem has been put into production operation in July 2009 and replaced the old subsystem METIS-Code-lists. All statistical tasks take the needed nomenclatures/code-lists directly from the this subsystem.

**Subsystem VAR.** For pilot test the statistical task Annual labour statistics variables were successfully described. The subsystem has been put in semi-production regime. In this year we have described about 2500 statistical variables from the domains of structural business statistics and short-term statistics. At present we have begun description of the variables of population census 2011 and demographic statistics.

**Subsystem TASKS.** Based on the description of the annual labour statistics variables description of the statistical task Annual labour statistics was fully described. The results of the 1<sup>st</sup> phase of subsystem development were proved. It means also that the phase two, design and development of applications for interpretation of metadata description, can start.

**The other subsystems** have been in preparatory phase, the content and functional specification have been elaborated or will be elaborated in next future.

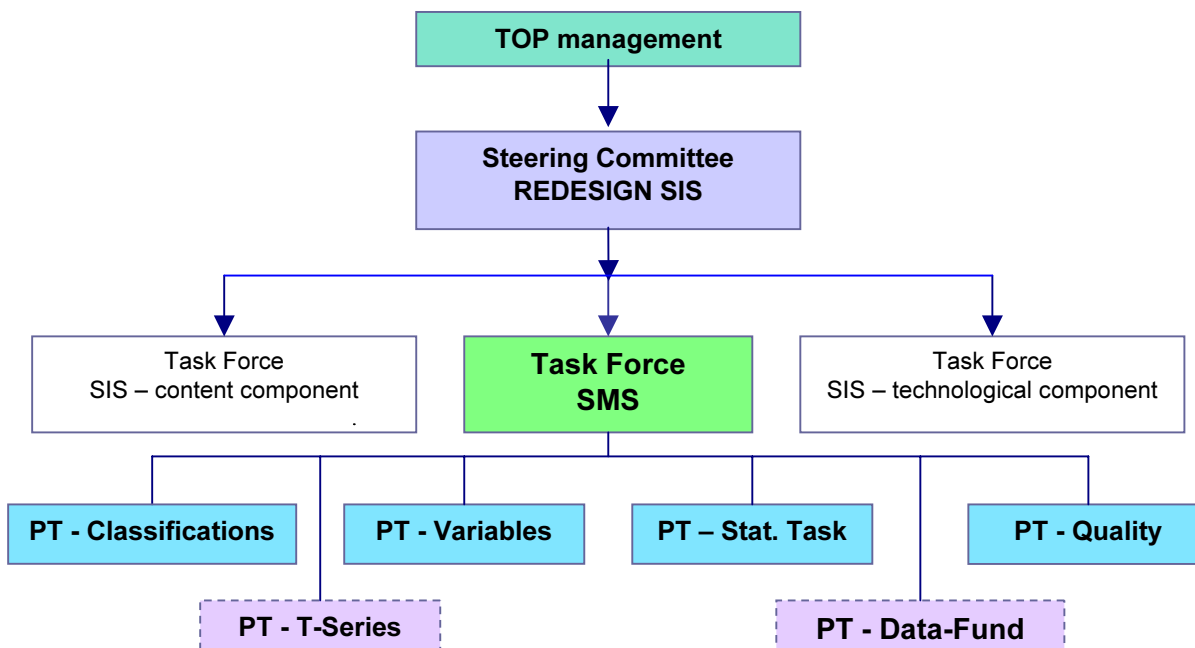
## 3. SMS Management and Organization

### 3.1. Project management in the development phase

The organisation structure for SMS development and implementation is the integral part of the corporate organizational structure of Redesign SIS. The Steering Committee (SC) heads it. The SC chairs the following Task Forces: Task Force (TF) for SIS-content component, TF for SIS-metadata component (SMS) and TF for SIS- technological component. Each TF is composed of project teams. The top management appoints chairpersons and members of all above-mentioned bodies.

The SC is supervised directly by the top management of the CZSO. The top management reviews progress reports of SMS subsystems, submitted by the SC. Achieved results and/or proposals for changes are subject for the top management approval.

Organization of SMS management is shown in the scheme below:



### The roles of the managerial bodies

**Steering Committee (SC)** is headed by the First Vice-President. Members of SC are: vice-president, director of methodological branch, directors of selected subject-matter departments, director of methodology department, director of ICT department, head of TF-content component, head of the TF-SMS, head of the TF-technological component and advisor to the President of the CZSO for the Redesign SIS. The SC supervises conceptually the SMS activities. It regularly controls (in three months intervals) the progress of work and takes decisions for further SMS development.

**Task Force SMS (TF SMS)** coordinates and monitors the work in SMS subsystems. For each subsystem a PT was established and its head appointed. TF cooperates with the heads of PTs closely. It is responsible for preparation and controlling of PT timetables. It prepares and/or coordinates numerous training courses, workshops and seminars related to SMS. It cooperates with the heads of other TFs within the project Redesign SIS and coordinates working activities in the SMS-PTs accordingly.

TF cooperates with external suppliers, monitors their working plans and progress of work. Prepares SMS progress reports for SC consideration. Based on the SC decisions prepares a final progress report for the top management.

**SMS Project Teams** - at present, the following project teams (PT) are operational: PT\_CLASS, PT\_VAR, PT\_ST, PT\_QUALITY, PT\_T-SERIES, PT\_D-FUND.

The project teams work on design of individual SMS subsystems and cooperate with external suppliers in development, testing and putting the subsystem into the CZSO practice. They prepare functional specifications and cooperate on the preparation of technical specifications with external suppliers.

The nature of SMS project requires participation of diverse professions in the project teams. Members of the project teams are methodologists, subject-matter statisticians, IT specialists, programmers, and specialists on statistical dissemination, users, external providers and external experts. Composition of working teams is flexible, depending on the nature of problems to be solved. Members of the project teams are heads of sections, selected directors and subject-matter experts of the CZSO.

### **3.2. Project management in operational phase**

Operational running of the SMS will be incorporated in the CZSO's organisation structure, which will meet the requirements of SMS administration.

An integral part of the SMS development work has to be preparation of conditions for utilization of SMS database. It means to specify rules, roles and responsibility of SMS users. In practice it means to design and introduce SMS administration in every day work.

During the development of individual SMS subsystems we have proposed main functions, basic rules and roles of users within the SMS administration. At present we have proposed following components of the SMS administration:

- a) central SMS administration,
- b) administration of the CLASS subsystem,
- c) administration of VAR subsystem,
- d) administration of QUALITY subsystem
- e) administration of the TASKS subsystem
- f) technological administration.

In the frame of the SMS administration we have defined these roles of users:

- a) administrator of metadata content – he/she defines the schema for description of an object, elaborates the description of objects kept in the DMS database. Subject matter statisticians carry out this role usually;
- b) methodologist – he/she provides consultancies to the all kinds of users, confirms the description made by the administrator of metadata content. The members of general methodology department carry out this role;
- c) subsystem administrator – he/she provides consultancy to the users, sets up the roles of individual users, takes care on smooth work of the subsystem application, gathers proposal for functional improvement of the application. The staff of the SMS Unit carries out the role of subsystems administrators.

Approval procedures are the standard part of the rules for use and update of SMS database. They describe detailed activities among persons and units responsible for update and approval of the object descriptions. The proposal for SMS administration needs very detailed discussion with statisticians, IT experts and methodologist to reach the right division of work and responsibilities among all participants.

## **4. Conclusions**

### **4.1. Lessons learned**

The experiences with the design, development and the first phase of testing the applications CLASS, VAR and ASKS allow us to derive following finding and recommendations for the implementation of SMS:

- a) SMS strategy in terms of contents and methodology shall be fully in the responsibility of the statistical office;
- b) SMS design and implementation should be organized in the multidisciplinary working teams;
- c) Design and implementation of the SMS project shall be managed and systematically monitored by the top management;
- d) It is necessary to persistently obey the SMS system principles and to maintain a positive motivation of most wide circle of subject-matter statisticians and methodologists; in this respect the CZSO benefited from involvement of external expert as a consultant to the Office President;
- e) Consistent co-ordination of time-scheduled workloads in the SMS project, the Redesign SIS project and current activities of the Office;

- f) Purchasing of financial funds must be systematically monitored by the statistical office in relation to the stage of the project implementation, on the basis of functional specification and qualified estimate of man-hours. It is important to use all potential sources of funding (external and internal sources);

#### 4.2. Future work

Three implemented subsystems (CLASS, VAR, TASKS) create the result of the 1<sup>st</sup> phase of the SMS implementation. The developed applications have been used in everyday statistical practice.

In the 2<sup>nd</sup> phase of the SMS implementation we plan to develop two main areas:

- a) development of applications which will interpret metadata created in the subsystem TASKS, mainly applications for generation of statistical outputs (tables and reports) and drawing the graphical form of questionnaires in automated way;
- b) design, development and implementation of further SMS subsystems, mainly Quality, Respondents, Users and Dissemination.

- \*\*\* -

#### List of abbreviations

CLASS	Subsystem Classifications
CZSO	Czech Statistical Office
GA SMS	Global architecture of statistical metainformation system
GSBPM	Generic statistical business process model
ICT	Information and communication technology
NSIs	National statistical institutes
PT	Project team
REG	Registers
SBP	Statistical business process
SC	Steering Committee
SIS	Statistical information system
SMS	Statistical metainformation system
TF	Task Force
VAR	Subsystem Variables