

**UNITED NATIONS STATISTICAL COMMISSION and
ECONOMIC COMMISSION FOR EUROPE
CONFERENCE OF EUROPEAN STATISTICIANS**

**EUROPEAN COMMISSION
STATISTICAL OFFICE OF THE
EUROPEAN COMMUNITIES
(EUROSTAT)**

**ORGANISATION FOR ECONOMIC
COOPERATION AND DEVELOPMENT
(OECD)
STATISTICS DIRECTORATE**

Joint UNECE/Eurostat/OECD work session on statistical metadata (METIS)
(Luxembourg, 9-11 April 2008)

Topic 2(i): Advocating for metadata in a corporate context

VALUE OF META-INFORMATION SYSTEM FOR THE CZECH STATISTICAL OFFICE
Invited Paper

Submitted by the Czech Statistical Office, Czech Republic¹

Executive summary

The Czech Statistical Office (CZSO) has launched in 2005 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 (PDSI). Necessity to identify PDSI 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 (classifications, variables and tasks) is in the phase of a pilot testing.

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 presented in the following parts:

- I. Redesign of SIS
Purposes for SIS redesign, its major goals and global architecture.
- II. Unification of statistical processes
Standardization of statistical processes in CZSO, focus on the life cycle of statistical production.
- III. SMS goals and architecture

¹ Prepared by Ebbo Petrikovits, ebbo.petrikovits@czso.cz

SMS functions and their interpretation in the SMS model and architecture. SMS plans.

IV. SMS management and organization.

Management strategy and organization in the stages of the SMS development, implementation and operational running.

V. Lessons learned

Experiences from the SMS development and implementation. Important findings and lessons learned.

I. REDESIGN OF SIS

1. National statistical institutions (NSIs) are facing increased number of users both on national and international level. Furthermore, a great challenge for NSIs are 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.

2. 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.

3. In accordance with the CZSO strategic goals, a redesign of the statistical information system (SIS) was launched in the second half of 2004 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.).

4. The basic conceptual strategy for statistical tasks² 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.

5. 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 key process as a survey is misleading. Definitions of both terms are located in note 2.

6. 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.

7. Major goals for Redesign have been:

- a) reducing response burden and boosting respondent motivation;
- b) optimising production of statistical information in the CZSO;
- c) designing a conceptual model of Redesigned SIS and of Statistical Metainformation System (SMS);
- d) defining a unified architecture of statistical tasks;
- e) improving quality of statistical information;
- f) increasing users' comfort.

² 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.

8. The **processes of production and dissemination of statistical information** (PDSI) is the main and most important process in the CZSO. It is a unique process in the sense that there is no blue print for it in both governmental information systems and business information systems.

9. In the new SIS architecture, all phases of the process of PDSI (from assessment of users' requirements up to the dissemination) 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 a life cycle of statistical task we applied in this work.

10. 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.

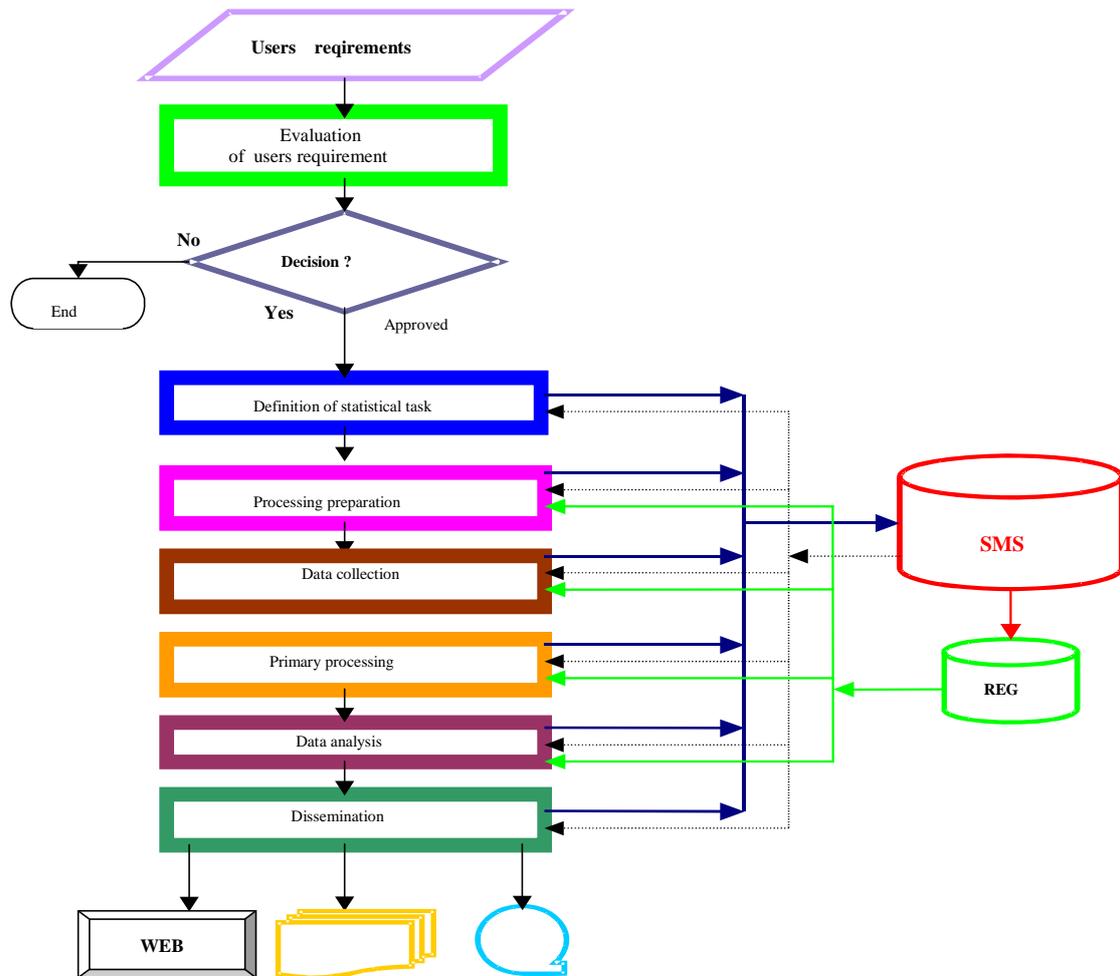
11. Global architecture of SIS is composed of three basic part:

- 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 PDSI. The model for metainformation definition of a statistical indicator/variable ensures standard description of statistical data in all stages of this process (from ST formulation up to dissemination of statistical information).
- 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 ST. 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 ST processing. Statistical information will be centrally stored in data warehouse and released through dependent data marts (including public database).

12. 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, SMS, ICT and the structure of data warehouse.

II. UNIFICATION OF STATISTICAL PROCESSES

13. As already mentioned above, the basis for GA SIS is a unified model of a process of production and dissemination of statistical information. The diagram below demonstrates the structure of the PDSI, its composition of phases and links to major integration tools that is statistical metainformation system (SMS) and system of statistical registers (REG).



The goals of SPP phases

14. **Evaluation of users’ requirements.** Assessment and evaluation of users’ requirements is made with regard to already existing statistical tasks. It should encompass a consideration on capacity requirements (human, technological, financial). This phase is in the responsibility of subject-matter departments and methodology department. They prepare the draft decision for final approval of users’ requirements by the top management.

15. **Definition of statistical task (ST).** In case of a new statistical task, the specification of its content, organisation and technology should be prepared. It includes namely the following: (i) ST concept, (ii) definition of statistical variables, (iii) statistical outputs, (iv) statistical surveys, (v) samples, (vi) questionnaires, (vii) other input data sets (administrative data and others), (viii) calculations (imputation methods, aggregations, derived statistical variables and others), (ix) timetable for ST preparation and implementation, etc. In case that ST already exists, its update should be prepared. This phase is in the responsibility of subject-matter department and methodology department.

16. **Preparation of data processing.** It contains (i) selection of samples, (ii) preparation and distribution of questionnaires, (iii) training of interviewers and staff responsible for individual phases of production, (iv) getting data sets from administrative and other external data sources. This phase is in the responsibility of data processing department and subject-matter department.

17. **Data collection.** It includes (i) collection of questionnaires (ii) input data capture, (iii) data validation, (iv) use of data from other data sources (incl. data validation), (v) production of input data sets (for further processing). This phase is in the responsibility of data processing department and subject- matter departments (they approve the results of data validation).

18. **Data processing.** It contains (i) building of input database, (ii) imputation of missing records, (iii) processing of aggregates, (iv) seasonal adjustment, (v) database update, (vi) data warehouse update. This phase is in the responsibility of data processing department in cooperation with subject-matter departments.

19. **Data analysis and output production.** This phase includes (i) application of mathematical and statistical methods, (ii) processing of required outputs, (iii) data quality assessment, (iv) setting up the data confidentiality, (v) approval of data for publication, (vi) data marts update. This phase is in the responsibility of subject-matter departments together with data processing department.

20. **Dissemination.** It focuses on (i) web pages, (ii) public output database, (iii) other subject-matter oriented data marts, (iv) printed publications, (v) electronic outputs, (vi) ad hoc outputs. Other important activities in the Dissemination phase are as well users' satisfaction surveys, analyses of the use of disseminated statistical information and statistical data confidentiality. This phase is in the responsibility of dissemination department, selected subject-matter departments in cooperation with data processing department.

Links of SMS to other processes

21. Parallel with the standardization of the key process PDSI we have taken in consideration the links to other supporting processes in the CZSO. Special attention has been paid to the following processes: (i) cost controlling, (ii) work efficiency evaluation using the attributes on data and processes quality recorded during the production process. Some significant progress has been made in links between SMS a cost controlling process. Common classification of statistical themes and statistical task has been elaborated as a base for monitoring the cost of individual activities inside the key process PDSI.

III. SMS GOALS AND ARCHITECTURE

22. 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.

23. 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.

24. 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 are functions related to the statistical contents and methodology.

25. 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.

26. 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 STs by the whole statistical staff. It could diminish a risk of the CZSO staff migration.

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

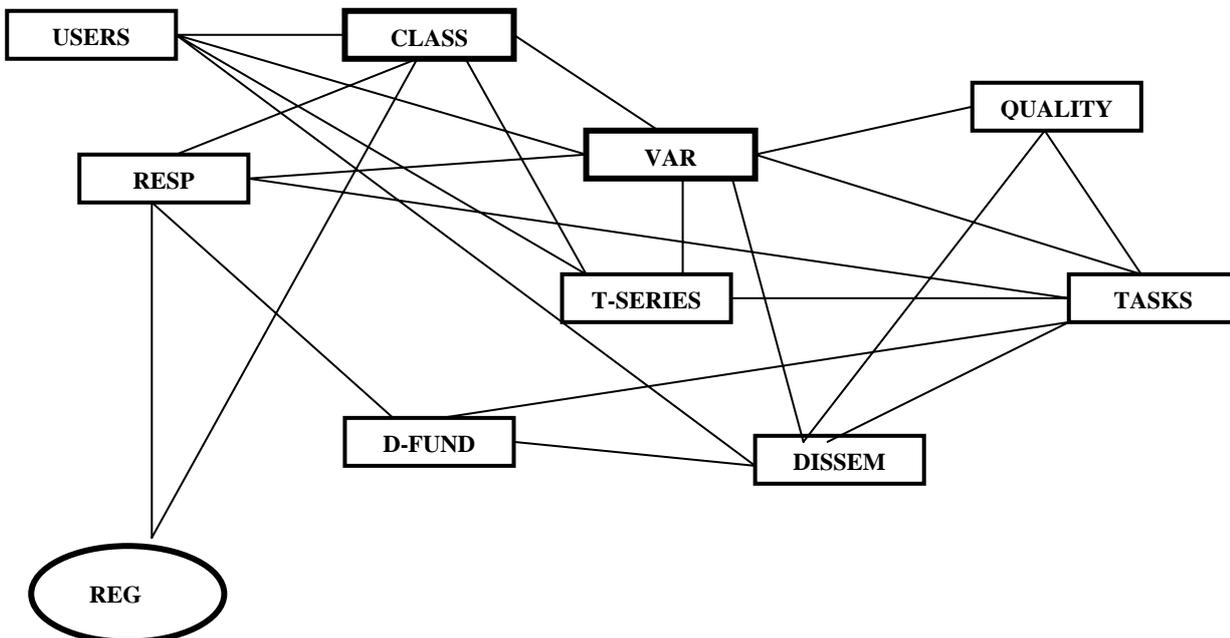
- a) management of methodology-related activities,
- b) assessment of statistical data quality,
- c) provision of statistical data to users,
- d) monitoring of respondent burden,
- e) integration of SIS with IS of public administration,
- f) integration of SIS with IS of international organizations,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

SMS architecture

28. 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.

29. SMS architecture is modular. It is composed of relatively self-sustainable, mutually interlinked subsystems as presented in the diagram below.



CLASS – Classifications
VAR – Variables
T-SERIES – Time Series
REG - Registers

RESP – Respondents
D-FUND – Data Fund
DISSEM - Dissemination

30. SMS is composed of the following 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 STs. 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 PDSI. Metainformation from this subsystem is used for data collection, data processing, analysis, 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 TASKS (basic characteristics, statistical questionnaire definition, other input data sets, decree on annual programme of statistical surveys, data validation, definition of statistical samples, imputation methods, quality requirements, aggregations, specification of users, time-tables for data collection, applied code-lists, legislation, provider of ICT services, specification of ICT services, etc). The subsystem is a basic source of information on TASKS from which the metadata for the design, implementation and running of STs are drawn. It keeps history of TASKS.
- 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.

31. 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, and Population Register.

SMS implementation strategy

32. **Implementation approach.** The main condition for introduction of SMS into the SIS operational running is its functionality in all stages of the PDSI. Effective and viable interlink of SMS subsystems interpreted in a single metadata-base is a necessary precondition for that. This requirement predefines priorities in design and implementation SMS subsystems implementation strategy.

33. **The first stage of the SMS introduction into the practice (2008-2009).** The first stage requires that SMS subsystems CLASS, VAR, TASKS and D-FUND be able to function simultaneously. Those subsystems will be tested in the statistical task “Annual labour costs survey“ as a pilot project.

34. This pilot project pre-requires the following:

- a) To complete fully a database on statistical classifications (SMS-CLASS),
- b) To unify methodologically contents of statistical survey(s) for the pilot project,
- c) To complete a description of statistical variables relevant to the pilot and to ensure their storage in the database (SMS-VAR),
- d) To create a database for designing of STs (SMS-TASKS),

- e) To accomplish an application program package which will be based on the work with SMS metadata
- f) To create and launch the work with the statistical data warehouse,
- g) To provide the **SMS operational administration** (for SMS- CLASS, VAR, TASKS),
- h) To accomplish training of personnel for all professions needed for the pilot project (methodology, subject-matter departments, SMS administration, project preparation, IT applications).

35. Building up/filling of the database for SMS subsystems is for the CZSO an entirely new task. The exception is the SMS-CLASS. In the newly established SMS-CLASS database, the links to the existing e-system of statistical classifications should be maintained so that STs processed under the existing conditions could smoothly continue their activities until a complete transition of STs into the new SIS is accomplished.

36. **The second stage of the SMS introduction into the practice (from 2010 on).** It shall be focused on the development, implementation and gradual implementation of SMS sub-systems for monitoring of quality, time series, dissemination, respondents and users of statistical information into practice. The second stage will comprise also the completion of SMS-CLASS, VAR, TASKS and D-FUND namely in terms of their links to the newly prepared SMS subsystems.

IT strategy of SMS

37. IT architecture of SMS is an integral part of IT architecture of SIS. The SMS is a necessary precondition for all statistical data warehouse operations. Data warehouse will finally become the only place to store all statistical data with their completely structured metadata description.

38. **Technological infrastructure.** Computing centre is aimed at use of servers with UNIX operation system and Oracle database technology. Technological equipment is grouped into Unix clusters on which Oracle database and application servers operate. Within the implementation of the SIS new architecture the applied technological tools will be extended by data warehouse technology.

39. SMS application programme package

- a) will not depend on users work stations platforms while operating on MS Windows system (version 2000 or higher) or Linux.
- b) viewing of metadata by the Internet browser without installation of supplementary products at the internal user station.
- c) for metadata administration "big client" solution can be used.
- d) access to individual sub-systems will be unified via SMS access portal while this portal will make part of the CZSO internal portal.
- e) for metadata presentation at the Internet Java Server Pages (JSP) technology will be used.

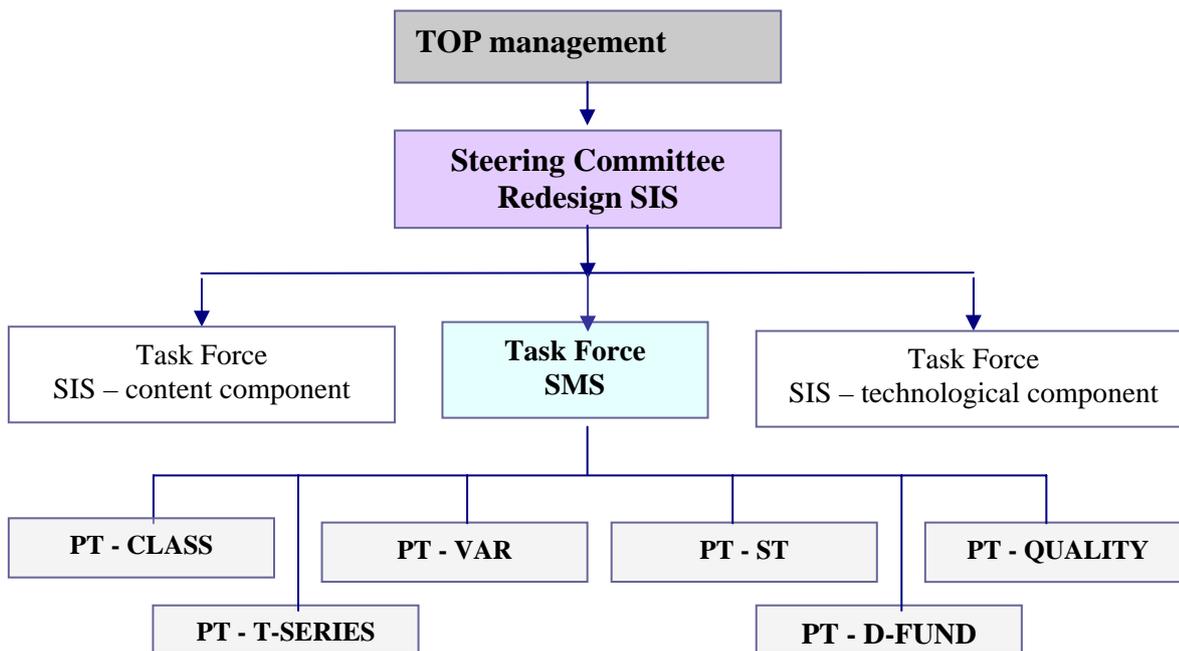
40. Interlinks between data and metadata will take place in data warehouse. Structured metadata added to statistical data will be drawn from SMS database.

IV. SMS MANAGEMENT AND ORGANIZATION

41. 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 (TF): TF for SIS-content component, TF for SIS-metadata component (SMS) and TF for SIS- technological component. Each TF is composed of project teams (PTs). The top management appoints chairpersons and members of all above-mentioned bodies.

42. 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.

43. Organisation of SMS management is shown in the scheme below:



A. Management of SMS development and implementation

44. Steering Committee

The First Vice-President heads SC. 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.

45. 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.

46. 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.

47. Project Teams SMS

At present, the following PTs are operational: PT-CLASS, PT-VAR, PT-ST, PT-Quality, PT-T-SERIES, PT-D-FUND.

48. The PTs 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 BSO documents and cooperate on the preparation of TSO documents with external suppliers.

49. The nature of SMS project requires participation of diverse professions in the PTs. Members of the PTs 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 of departments and subject-matter experts of the CZSO.

B. Management of SMS operational running

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

51. 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.

52. During the development of individual SMS subsystems we have proposed main functions, basic rules and roles of users within the SMS administration. In time being 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.

53. 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.

54. 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.

V. CONCLUSIONS

55. The experiences with the design, development and the first phase of testing the applications CLASS and VAR 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);

- g) Financial costs of the operational running of the SMS should be covered from the Office budget.

Abbreviation used in text

CZSO	Czech Statistical Office
GA SIS	Global architecture of statistical information system
ICT	Information and communication technology
IS	Information system
NSI	National statistical institute
PDSI	Production and dissemination of statistical information
PT	Project team
REG	Registers
SC	Steering committee
SIS	Statistical information system
SMS	Statistical metainformation system
ST, STs	Statistical task, Statistical tasks
TF	Task force