Redesign of the Statistical Information System: Czech experience

Marek Rojíček[1]

[1] *Czech Statistical Office, Prague, Czech Republic; marek.rojicek@czso.cz*

Abstract
In the last 10 years, preparation and implementation of the integration of software tools for collection, processing and dissemination of statistical data has been realized. The main tasks of this project were decrease of the administrative burden laid down on respondents together with the increased quality and accessibility of statistical data for every user of the statistical information system (SIS). The main part of the project has been realized in the years 2013 – 2014 and in 2015 the project was transformed into routine statistical production process. After implementation of the new SIS the need for higher cooperation of all statistical as well as IT departments significantly increased. Although most of the roles were anchored into the line organization structure, it was necessary to keep partly project type of the management of the whole process due to its complexity.

Keywords: metadata, digitalization, integration, data collection.

1. Introduction

Since 2005 the Czech Statistical Office (CZSO) has been developing new system of collection, processing and dissemination of data, based on the common data warehouse and using Statistical Metainformation System (SMS). In accordance with the CZSO strategic goals, a redesign of the statistical information system (SIS) was launched after the accession process to the EU had been successfully finished. The first important step in this endeavor 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.). It calls for efficient ways to cope with user's needs in terms of availability, timeliness, comparability and correlation of statistical information. Frequently

required is statistical information in specific layer of a structural and/or regional decomposition. Consequently, determination of sampling criteria and stratification of statistical population in sample surveys is an important target in a new architecture. Improvement of statistical quality is an important strategic intention of the CZSO, thus the system for monitoring and evaluation of statistical data quality is an integral part of a new architecture. Last but not least, a part of a new architecture is statistical metainformation system, needed for interpretation of statistical information in all phases of the statistical production process, namely for the dissemination of statistical information to users.

In the new SIS architecture, all phases of the production and dissemination statistical information (PDSI) processes have been carefully explored (including their links) in order to design general methods, tools and techniques to support these phases. There was a broad cooperation and discussion in the CZSO on the new SIS architecture. The basic conceptual strategy for statistical tasks (called "Model 2008") was approved and further developed in the framework of the document "Global Architecture of SIS" (GA SIS). The GA SIS is a basic strategic document for redesign SIS. The goals and targets of GA SIS became an integral part of the CZSO strategic documents. The global architecture and its implementation plan were approved by the CZSO top management in June 2007. It included also financial plan, which counted with support of EU budget.

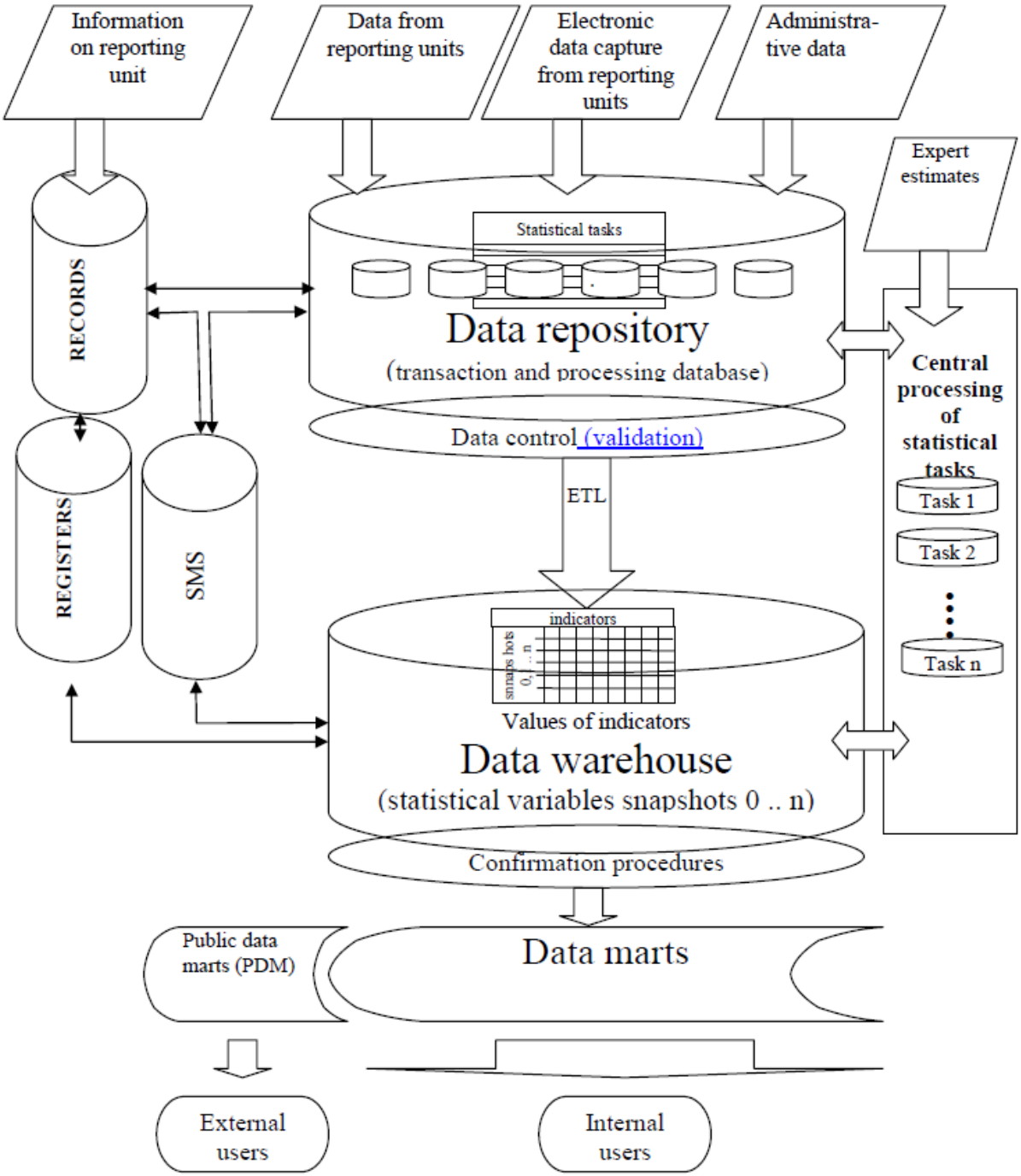
2. The main pillars of the Redesign of SIS

Identification of PDSI and its phases issued a call for a redesign of statistical activities in the CZSO with the major aim – to strengthen organization and management of statistical work. The whole global architecture had several aspects (parts) – content (what data are collected and from whom), processes (how the data are collected, processed and disseminated) and modernization of IT infrastructure, which have been realized in different stages. The scheme of the new SIS is described in the Figure 1.

The first part of the project was realized in the years 2005 and 2006, when the basic part of SMS was created (statistical classifications and indicators). During the second part in 2007 -

2009 (co-financed by European Transition Facility project) the system of statistical tasks (surveys) was implemented into SMS and data warehouse was built in the CZSO.

Figure 1: Simplified description of data flows



The third phase of the RSIS was focused mainly at the modernization of the software tools for collection, processing and dissemination of statistical indicators and IT infrastructure and started in 2011 and was realized in 2013 and 2014. The budget of this project (in total 200 mil. CZK = 7,4 mil. EUR) was covered from 85% by the European structural funds. Three quarters of the total sum were investment expenditures. The project was realized mainly by external suppliers in close cooperation with experts from CZSO. During this stage of the Redesign of SIS the most of the new statistical tools were developed and the new IT infrastructure (computing centre) was built.

2.1. SIS subsystems and their main functions

The aim of subsystem REQUIREMENTS is to support and secure activities related to acceptance, recording and administration of requirement for provision of statistical information. It consists in evaluation of individual requirements and their comparison with statistical tasks in the process of implementation.

In the subsystem PREPARATION statistical tasks are described including the list of used and newly designed statistical variables.

PROGRAMME subsystem deals with the preparation of ST processing, pre-print preparation, the print itself and distribution of statistical printed forms, programs of electronic data capture from a respondent and information on reporting duty. The subsystem deals with the program preparation for input and central processing.

INPUT subsystem covers primary processing of statistical task (including administrative data processing). The subsystem tools ensures source material for securing and solving of the issue of response to statistical reports and questionnaires, their records and acceptance, validation of input data and processing of control and qualitative outputs. Confirmation procedure based on qualitative and quantitative parameters makes part of the subsystem.

CENTRAL subsystem includes all activities related to central processing. This applies especially to the creation of statistical task input data, detection of extreme values of statistical variables, imputations, seasonal adjustment, modeling and statistical data imputations. An

important activity is the data confirmation and release including the preparation and assignment of expert estimates. The subsystem also includes statistical data analysis and protection of data confidentiality.

The subsystem DISSEMINATION is dealing with the tools for providing of statistical information and its presentation to the final users, supplying information to internal users and arranging for discussion forums. The subsystem tools enable to prepare analyses on users of statistical information and on use of statistical information.

Integration tools of SIS are designed to support functions of individual sub-systems.

The subsystem REGISTERS, as an integration tool, includes mainly administration of registers and other supporting databases, proposed system of the CZSO registers, administrative, descriptive and statistical functions of individual parts of registers and proposal for building up sample frames for statistical tasks.

An important integration tool is SMS, and especially its sub-systems CLASS, VAR, TASKS and QUALITY.

From the aspect of data administration and data store the whole SIS strategy is based on an integration tool DATAWAREHOUSE which deals with the definition and administration of data warehouse (data stored in the warehouse will be used prevailingly for central processing, data analyses and dissemination of statistical information) and data sources.

2.2. Content part of Redesign of SIS

Simultaneously with change in technology and processes the substantial change in the system of business surveys was prepared and implemented. It was based on "Model 2008". Major goals of redesigned SIS were improvement of statistical quality, and reducing of respondent burden (focused on better use of administrative data and modelling/stratification change of the respondents' samples). There was a significant shift in the content component of SIS from the statistical survey approach towards statistical object oriented approach (statistical tasks and variables). It was closely connected with the permanent effort to reduce respondent burden as follows:

- maximum use of modelling;

- maximum use of administrative data;

- use of data from one statistical task in other;

- coordination of surveys samples;

- "rotation" of an extended sample for individual CZ-NACE activities (only a certain group of industries will have wider selection, the next year the group of industries will change);

- "rotation" of variables for which a detailed structure is required;

- maximum stabilization of statistical tasks system;

- direct e-use of data from respondents' information systems for statistical purposes.

Very important was also the aspect of statistical coherence. So called principal statistical tasks were defined with the aim to determine an absolute value of surveyed (estimated, modeled) variables (by calibration or confrontation) in all relevant tasks or to be binding for determination of more detailed structure of these variables;

- different statistical units shall respect consistency of published data (single figure principle).

Very important was also the principle of completeness, which means that published outputs of core and standard variables cover the whole population (not only a fraction, e.g. only businesses with 10+ employees). If statistical survey covers only a fraction of the population it is supposed that the below-threshold part estimate will be determined by modeling (e.g. on the basis of administrative data or other surveys).

3. The main strengths and weaknesses of the new SIS

The main strength of the new statistical information system is its integrity role. All data are stored in the warehouse only once, they can be easily interlinked via metadata description notwithstanding what is their source (statistical survey, administrative data, etc.). The SMS system fulfils also the checking role of the uniqueness of statistical variables, because they all have to be described and duplicities can be easily revealed and then either eliminated or

confirmed as meaningful. The project helped also significantly in integration and modernization of tools used in the process of collection, processing and dissemination of data from business surveys. The new system brought a lot of user level tools and enabled to move a lot of operations previously conducted by skilled programmers to qualified users (statisticians).

The high level of integrity has on the other hand significant risks, because the newly developed tools are closely interrelated and if something fails in one part of the system, the functioning of the whole production process is threaten. Significantly higher demands are put on IT services (infrastructure as well as application support) and there is high dependence on the cooperation with external suppliers. All in all it creates pressures on human capacities and financial sources, which is in contrast with the ability of the statistical office to fully satisfy them (in the situation of decreasing budget and competition at the labor market).

4. Conclusions

The new statistical information system developed in different phases during 10 years came is now more or less finished. The statistical office acquired a lot of experience during this process and the results for users (quality, comfort and range of data provided) are significant as well as for respondents (lower burden, modern ways of communication). There is still much to do in efficiency of using existing data sets stored in public as well as private institutions (administrative data, business information systems). The world is changing and it brings new opportunities for statistics for using so called “big data”. The new tools are running in the routine production for more than one year and there are still a lot of “operational” problems to solve. It is not realistic to expect that so unique and sophisticated system would work perfectly. We are also continually collecting ideas for partial improvements of the system especially in the sense of user-friendliness.

5. References

Fischer J. (2007), A statistical system for future generations. Conference “Modern Statistics for Modern Society”, EUROSTAT, Luxembourg.