**Redesign of the Statistical Information System: Czech experience**

Marek Rojíček, Czech Statistical Office, 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. In the next few years it would be necessary to do strategic decision about the further progress and renewal of the HW and SW in the final stage of their lifetime.*

**Keywords:** digitalization, integration, data collection, customization

**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 endeavour was to design a new architecture of SIS.

Driving force for a new architecture of SIS was satisfaction of an increasing user’s demand for statistical information from private sector as well as governmental institutions, international organizations, multinational enterprises etc. It called for efficient ways to cope with user’s needs in terms of availability, timeliness, comparability and correlation of statistical information.

**2. Global architecture of SIS**

In the new SIS architecture, all phases of the production and dissemination statistical information (PDSI) processes were carefully explored (including their links) in order to design general methods, tools and techniques to support these phases. 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.

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 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.

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 majority (85%) of the total budget (200 mil. CZK = 7,4 mil. EUR) was covered 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.

**3. The main goals achieved**

The main strength of the new statistical information system is its integrity role (see figure 1). 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).

**Figure 1. The basic architecture of data flows in SIS**



Source: Fischer (2007)

Alongside with the reform of processes in the SIS there was also “content” part of the reform. New principles were set up, such as maximum use of modelling, maximum use of administrative data, use of data from one statistical task in other, coordination of surveys samples or estimation of the variables for the whole population. Altogether this lead to significant reduction of the respondent’s burden as well as increase of statistical quality.

**4. The lessons learned**

At the beginning it was expected that the tools developed under this project would be implemented in the whole range of statistical task within the office and this was the reason why they were designed and programmed as universal tools based on the GSBPM scheme. The practical experience showed that universal solutions have except their benefits also significant risks. The system was designed using experience mostly with the “production line” of the business statistics. Redesign of SIS solved the problem of out-of-date (and non-supported) applications and non-unified approach to the processes of collection, processing and dissemination of data. On the other hand the household surveys are significantly different and the decision to leave it separately appeared as good solution. Separate system of data processing was kept also in some other statistical domains, e.g. foreign trade statistics, where applications had been developed and supported by own capacities of the CZSO.

After some time delay since implementation the positive and negative aspects of the whole project can be objectively assessed. One of the main problem is extreme customization of the tools which makes most of the user requirements expensive to do because of the outsourcing of the support. In practice it is also very complicated to get support from other suppliers than those who developed the applications (even if the CZSO owns the source codes).

The high level of integrity has (alongside with its benefits) also 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. 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. 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 sharp competition at the labour market).

**5. The way forward**

In the next few years it would be necessary to do strategic decision about the further progress and renewal of the HW and SW in the final stage of their lifetime. It is difficult task because the strategic decision have to take into account the development in technologies as well as in the methodology of statistics and user needs.

The world is changing and it brings new challenges for official statistics. We face the competition of many data sources which are attractive for wide public although do not guarantee the quality. In the future it is necessary to further reflect the trends in digitalization and users’ needs of working with data. There is also much to do in efficiency of using existing data sets stored in public as well as private institutions for creating official statistics.

The main trends which would be probably reflected in the future redesign of SIS are:

* Direct link to ERP[[1]](#footnote-1) systems of business companies;
* Using different tools for collection of different types of data;
* Using standardized software tools enabling user’s customization;
* Extensive use of modelling using administrative data sources and reduction of the role of statistical surveys;
* Using big data and appropriate tools of data mining;
* Modernization of dissemination tools.

**5. References**

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

Rojíček M. (2016), Redesign of the Statistical Information System: A Czech experience. European Conference on Quality in Official Statistics, Madrid.

1. ERP = Enterprise Resource Planning [↑](#footnote-ref-1)