Linkage of main components of GSBP model through integrated statistical information system.

Helena Glaser- Opitzova [1], Ľudmila Ivančikova [2],

[1] *the Statistical Office of the SR, Bratislava, Slovakia; helena.glaser-opitzova@statistics.sk*
[2] *the Statistical Office of the SR, Bratislava, Slovakia; ludmila.ivancikova@statistics.sk*

Abstract
Development and implementation of new models, modern procedures and tools led to complex assessment of existing statistical system within production of the Statistical Office of the Slovak Republic. Stimulus was given Vision, existence of models such as GSBPM and need of reaction to requirements of internal and external environment. In article we are concerned with main components of integrated statistical system, which was developing in the Statistical Office of the Slovak Republic in 2013 and 2014. The system is described not in terms of information technologies, but from the view of its specific using for statistical production and for creation of individual statistics. Article is focused on system architecture in relation to preparation, data collection and data processing and dissemination. The goal of system components is to contribute to reducing burden on respondents, increase of work efficiency and consequently make data of higher quality at input and output level.

Keywords: integration, statistical information system, standardisation

Introduction[1]

The task of statistics in Slovakia is to carry out activities in systematic and planned way with a view to obtaining, processing, dissemination and evaluation data relating to mass nature phenomena. Standardised working procedures are regarded as major preconditions to meet this

[1] The authors of the article would like to take this opportunity to express their sincere gratitude to Ms Silvia Jenovčíková for drawing up graphical outputs used in this article.

demanding task. The aim is to eliminate specific individual solutions, although it is necessary to emphasise that in practice we cannot exclude them completely. Another precondition is to ensure adequate tools and instruments to perform statistical activities. These tools and instruments should be available to producers of statistics as well as to reporting unites (respondents) and users. With the aim to meet those preconditions and in full compliance with Vision for next decade, the SOSR designed and developed the project Integrated Statistical Information System (IŠIS) in the years 2008 – 2014.

1. Specification of the starting points for the new system building up

Taking into account the starting points in 2008, we have to acknowledge that morally and technically outdated information and communication technologies were applied. At that time Automatized Statistical Information System (AŠIS) was the main production system which was being developed from the mid-90's. In this way, rather satisfying standardisation in business surveys was delivered, however, this fact was not achieved in all domains of business statistics. Besides above mentioned system, the SOSR operated with a number of other systems such as INTRASTAT/EXTRASTAT, Mts-SÚD, Statgen, Regstat or Slovstat working separately. Individual solutions for collection, processing and dissemination of data existed in parallel. The data collection was ensured by various instruments or by various versions of standard instrument - application software DC 2000 and his "clones" for data collection in households and in price statistics. Instruments for ensuring the same activities for different statistical domains were developed in parallel. For example, business statistics were processed in ASIS, however, foreign trade and national accounts developed their own instruments, and social statistics were operating under their own way. Project IŠIS meant cessation of individual solutions and it launched process of integration on the common platform for all statistical domains. The starting points for new integrated statistical information system building up were as follows:

- Model of Statistical Business Process (based on GSBPM)
- Unified information model
- Standard methodological procedures and instruments

- Technology modernisation

1.1 Statistical Business Process Model

GSBP model provided us with standardized framework and harmonized terminology. It helped us to standardize already existing statistical production processes and share methods and components. At the same time this model was used as the basis for statistical computing infrastructure harmonization. Designed model of statistical business process for IŠIS defines statistical sub processes and activities of sub-processes which are needed for official statistics production. The typical statistical process covers sub-processes of preparation, collection and processing of data and the production of statistical outputs and their dissemination. The model covers several overarching processes which are applied within all sub-processes and activities of sub-processes.

- *Quality management* covers quality assessment (as a whole) and control mechanism (incorporated mechanism of validation)
- *Metadata management* – metadata are created or generated and they manage activities in each sub process, therefore the emphasis is given on metadata system (METIS) – to ensure that metadata would remain connected to data during the whole statistical process.
- *Data management* includes general management and data protection, data quality, managed access to data
- *Management of data describing processes* covers management of generated data and metadata with the aim to be informed or to provide information on all parts of statistical process of production (generating of technical projects)

Model is implemented into IŠIS flexibly. IŠIS supports that individual sub processes and activities of sub processes are underway in sequence (e.g. making of statistical survey copy for next year) or parallelly (e.g. testing of collection instruments and correction of algorithms). Some sub process activities may be carried out iteratively (survey preparation and testing and

subsequently return to preparation) or they may be omitted (there are used metadata from the past period).

Good metadata management is crucial during performance of statistical activities. The Metadata are present in each sub process and they are either build up or transferred from previous sub process. The key challenge was the availability of metadata to data at each moment. One consolidated database (one data source) with connected metadata represents a support for simple reporting by using analytical instrument for work with data (COGNOS) and subsequently a direct export to Excel, csv, xml, pdf format. Users work with data and metadata in system on the basis of assigned access rights.

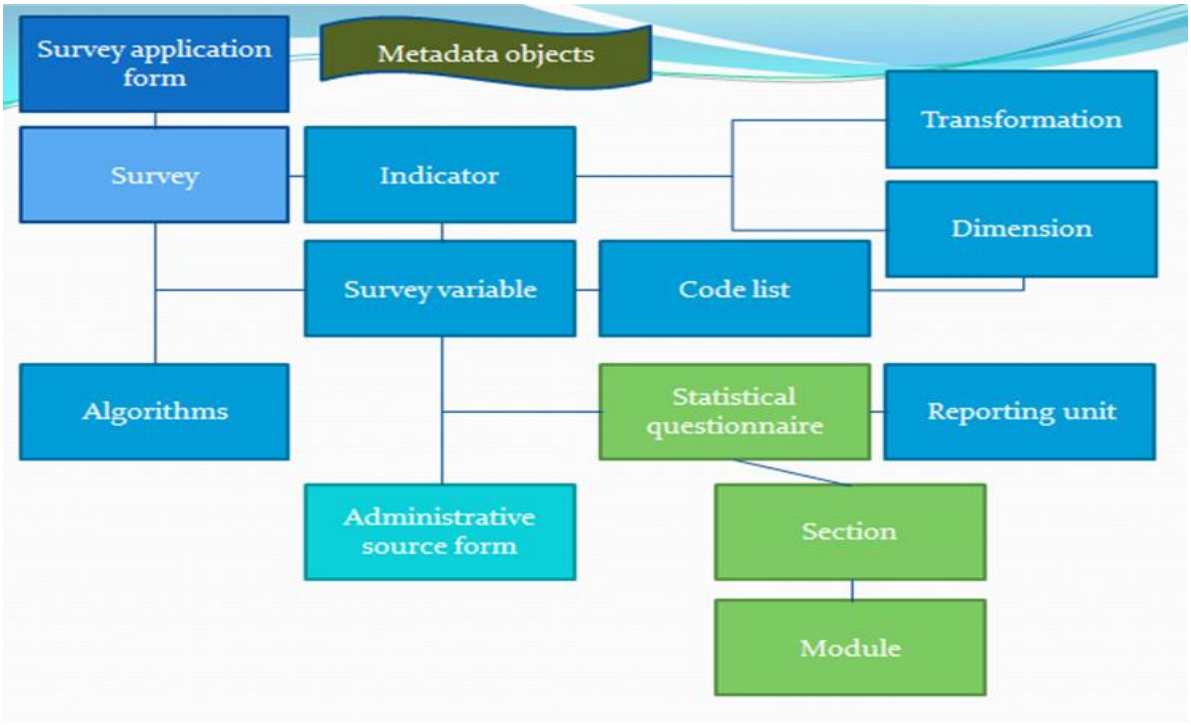
Fig. 1 Part of the statistical business process model directly supported by IŠIS

Building and testing, data collection, processing, analysis, dissemination are sub processes of the statistical business process model which are directly supported by IŠIS.

1.2 Information model – data, metadata and system of registers

Information model describes such objects which are inputs and outputs of particular statistical activities. It represents statistical data and metadata which are created and they further proceed throughout the statistical business process. Within IŠIS, model is composed of objects.

Fig. 2 Metadata objects



In Fig. 2, we can see basic metadata objects. Besides them there are many others, e.g. legends and explanatory notes to modules of forms. Objects are designed in such way that they could be assembled into logical units (modules, sections, questionnaires, surveys) and at the same time they could be shared among different questionnaires and surveys. The same variable or the same code list can be used in all questionnaires by which the same phenomenon is

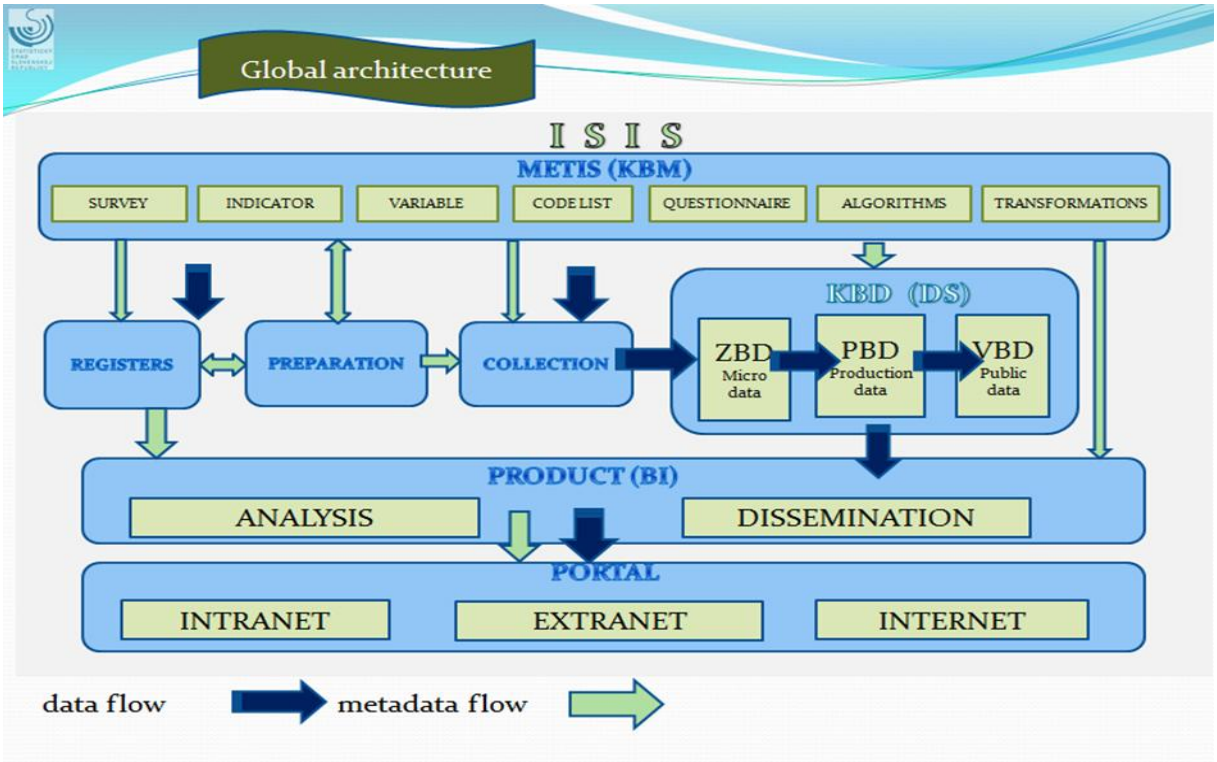
surveyed. The procedural aspect is also taken into account. Survey building in IŠIS starts by creating „Survey application form“ which is followed by building up „Survey“ and building up „Questionnaires“ and all their components in such way that a specimen of questionnaire could be generated subsequently and then created electronic questionnaire (eQuestionnaire) for purpose of electronic data collection (eCollection).

The system of registers is a part of information model. We can see the register as a set of records with regular registration of data and information on concrete set of objects. The SOSR manages the system of registers which is constituted by 11 registers: an agriculture register, a register of local units, a register of non-residents, a business register, a register for metasystem of statistical - accounting data, a register of spatial units, a enterprises groups register, a register of foreign trade subjects, a register of statistical units, a register of accommodation facilities, a register of plants.

1.3 Model of the Global architecture of IŠIS.

Integration of statistical business process model and information model produces global architecture of IŠIS.

Fig. 3 Description of total global architecture of the IŠIS



As we can see in Fig. 3 a process view as well as an object view are built in the model. The model shows inputs and outputs of data and metadata into/from individual sub-processes and their flows. The process view is associated with data collection, data procession, data analysis, production of products, dissemination. Management of the registers means the process itself which operates parallelly and at the same time, it provides information for survey preparation. Production Data Base - virtual (PBD) is a base of calculated indicators values. Public Data Base (VBD) is represented by data cubes. The web site is composed of three parts for communication and dissemination – INTRANET, EXTRANET within NSS, and INTERNET for public.

Technology modernisation was an inseparable part of the IŠIS design. This modernisation is seen by statisticians in present days only from the perspective of IŠIS instruments users. The IŠIS was build up on principles of service oriented architecture with using the newest frameworks. Conception of the system was built up with the aim to address the vital situation of official statistics, broader integration of administrative sources to ensure maximization of

the principle: the provision of data one time, last time by legal entities towards public administration. Built up interface of administrative data enables to provide data by 4 ways via:

1. Web interface;
2. Client application which is installed at data provider and which automatically, in set intervals, downloads data from predefined addresser and provides them to web service;
3. Passive interface of web services by which it is possible to write down data;
4. Active interface where the SOSR through defined web service inquires on the interface of provider`s web services.

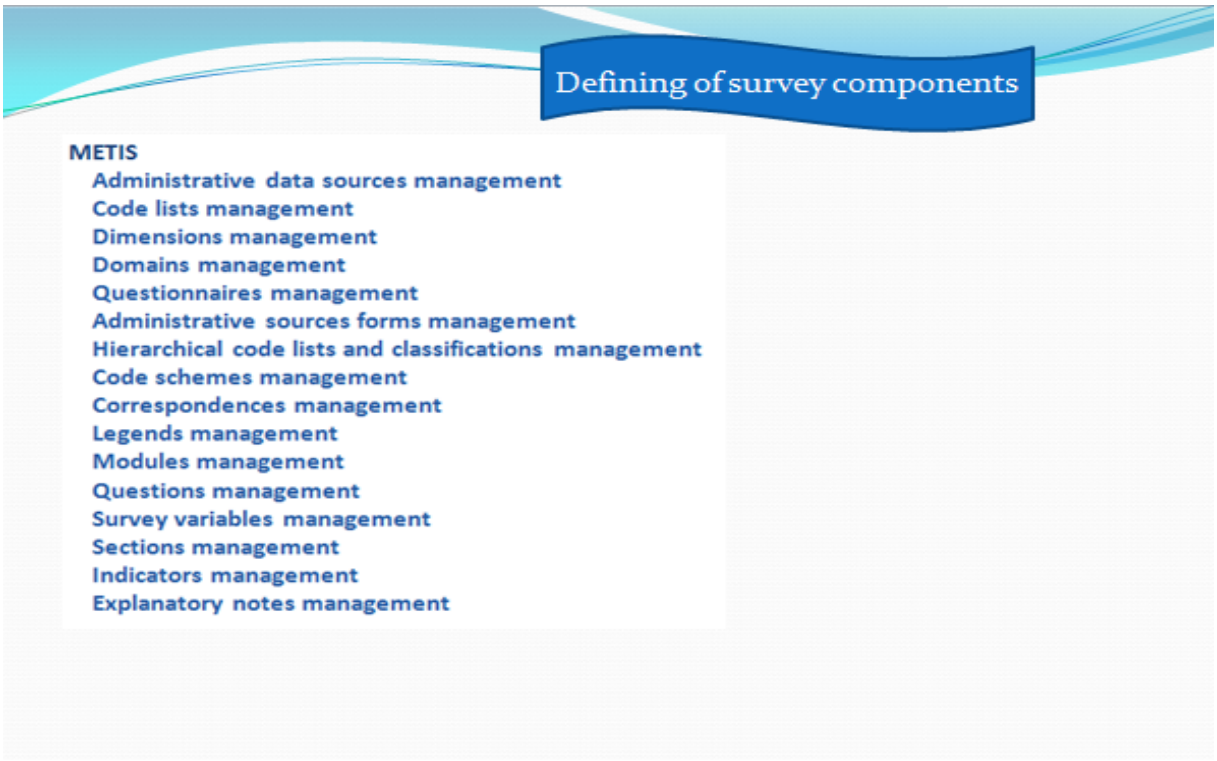
As a part of the system there were built up following services:

- electronic services for the provision of specific statistical products,
- notification on setting up code lists, updating the values and their validity termination,
- notifications on reporting obligation,
- notification on questionnaire availability,
- notifications on coming collection deadlines and reminders.

2. Major milestones of the whole process in the IŠIS

The whole process in the IŠIS starts by making "Survey application form" in accordance with statistical program. The survey itself is generated on the basis of „Survey application form". Via this interface we can manage all objects related to survey (indicators and variables, a questionnaire, a section, a module, algorithms etc.). Defining of survey components (Fig. 4) encompasses functional instruments as well as key objects of which survey and its questionnaire or questionnaires are composed.

Fig. 4 Survey components in metainformation system (METIS)



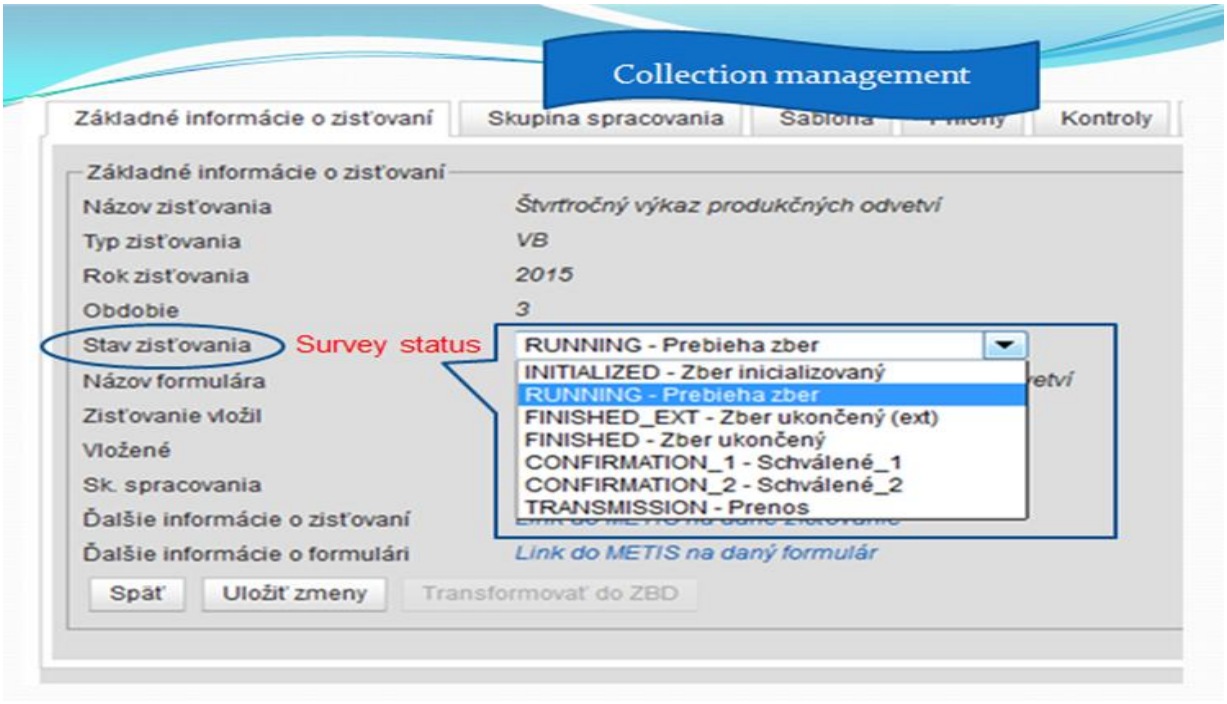
Algorithms are special metadata in the IŠIS. Algorithms are significant metadata by which the process of data collection is managed. Validation rules are defined by them under which the examinations are carried out during data collection or after completing data collection. Algorithms for application of mathematical and statistical methods are also defined. Parameterisable algorithms are stored in a library of algorithms and can be shared by different surveys. This is the way by which standardised validation rules are developed.

The system enables to define directly a sampling frame and select a sample, where appropriate. This activity is connected directly with system of registers.

The system enables to generate a specimen of statistical questionnaire and this one is subsequently released on the web site of the SOSR in Metadata section. It is also used as a template for questionnaire printing on the purpose of potential paper collection.

In the system there is testing algorithms prior to collection initialization and validation of formal aspect of the eQuestionnaire. There is still possibility to return to sub-system for preparation of statistical survey and perform correction needed. The whole testing simulates real data collection process.

Fig. 5: The stages of data collection and the procedure of validation



Within the questionnaire finalisation, after completing testing, a potential corrections the state of survey will be set up on the value "PUBLISHED".

By that, the survey preparation is completed and data collection in sub-system "COLLECTION" is open for respondents of a statistical survey. In the system we move to COLLECTION management. COLLECTION management means that during each separate survey it is possible to monitor the state of the survey at each moment.

The stages of data collection and procedure of their validation are depicted in the Fig. 5. The data collection is open and closed by regional office of the SOSR responsible for concrete survey performance. Data are validated by survey manager and subsequently it is possible to carry out transfer of data into Source Data Base.

Let's imagine, the survey is in the stage "RUNNING". A respondent logs in to electronic data collection via the web site of the SOSR.

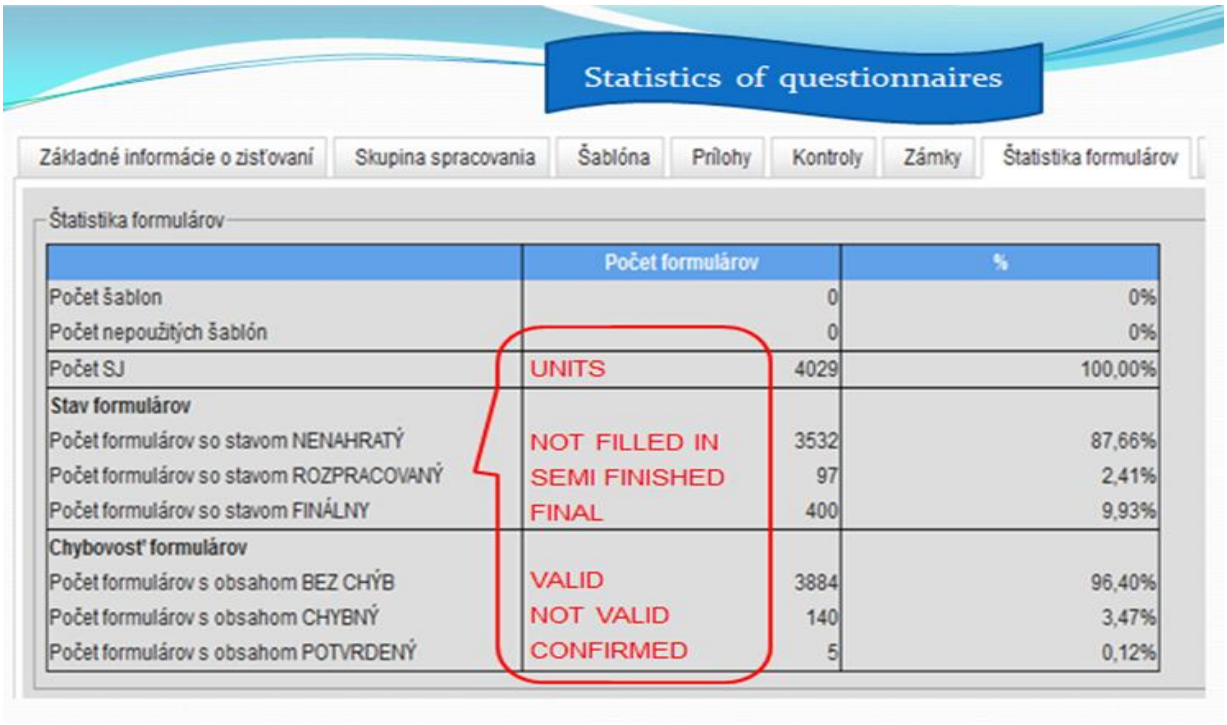
In order to facilitate work of respondents in new eCollection, the SOSR prepared and since 1 January 2016 has put into operation the targeted web site to electronic data collection with new visual and using modern visualisation elements.

Following logging in to eCollection, respondent carries out data records while respondent's visual is composed of metadata and a data record itself is managed by algorithms. The respondent can launch controls of filled in data. If he/she does not do it, the system will carry out that activity at the time of saving of a semi-finished questionnaire or at the time of transmitting finished questionnaire. There is a possibility for the respondent to print the form (or to save generated pdf). By this activity the work of respondent is completed.

It is possible to monitor the state of data collection in the course of collection. Monitoring can by carried out via coloured visual (information on the state of the questionnaire) or via statistics concerning the questionnaires (Fig. 5). This information is used mainly by the Regional Office of the SOSR which is responsible for the collection and initial data processing for survey in question.

Following data validation by the statistical survey manager, a transfer from the collection database is carried out into the source database. Data stored in the source database are the source for statistical outputs production, e.g. via analytical and reporting instrument COGNOS by which ad hoc outputs or predefined reports in time series can be generated. These ones require to build up models over this database at first (to build in calculations and other transformations) and prepare a report itself.

Fig. 5 Statistics of questionnaires



Statistical output production means entry into the phase of dissemination. Besides statistical data, there are also published selected metadata which are disseminated via the web sites of the SOSR directly from database through dynamic sites. Static sites production was eliminated in the new system. For example, if the questionnaire specimen is saved into the IŠIS and its publication on the web site is allowed in the IŠIS, the system makes it automatically accessible on the web site. Any updating in the system is reflected on the web site.

It is particularly worth mentioning the work with administrative data sources. The SOSR obtains data via statistical surveys as well as from owners of administrative data sources. For that purpose, it uses within the IŠIS Administrative Data Sources Communication Interface (KRAZ) which serves for input and data storing into the database COLLECTION. Storing into Source Database (ZBD) is carried out by the same instruments as in the case of statistical surveys.

3. The Directive for Statistical Business Process and IŠIS Integration

Statistical business process was fundamentally changed in all domains of statistics by launching the IŠIS into routine operation. Therefore it was necessary to overhaul more internal documents, specifically the Directive for Statistical Business Process. Since the IŠIS supports the majority of sub processes and activities GSBPM (see Fig. 1), directive structure was designed on the basis of this model and its implementation into the IŠIS.

The directive covers activities in the entire process of statistical outputs production which are being ensured by respective departments of the SOSR. In the directive, there is established that production is carried out in accordance with requirements towards Quality Management, GSBPM and the requirements of ES CoP. The directive specifies sub processes, activities of sub processes and defines fundamental responsibilities.

Fig. 6 Activities and quality of statistical product

Section I - Specification of the needs, HTP 01, Article 4 - Basic information about the sub-process

7. In relation to the output quality, the sub-process has a decisive impact on the relevance of the final statistical output what is reflected in the efficiency of the production as a whole. To ensure abovementioned relevance, the focus is put on:

a) Getting suggestions (from external and internal environment, as well as our reflections), to ensure that employees could identify needs of users,

b) Consistent communications with users with the aim not only to understand their needs thoroughly but also to recognize the intended usage of the statistical output, as well as the importance and significance of this usage.

Source: the Directive for Statistical Business Process, 2015, the SO SR

This directive does not define normatively detailed descriptions of activities performed – it cannot be regarded as a work instruction. It accepts and promotes application of implicit knowledge of the SOSR's employees, explicit knowledge set out in respective documents of the SOSR and the SR, best practises as well as international documents and procedures regarding statistical output production. The directive can be regarded as a Quality guidelines.

The directive is broken down into eight sections. Section I (one) represents customers` needs specification; sections II and III include design and building; sections IV – VI cover the output production; section VII dissemination and section VIII covers evaluation in compliance with GSBPM.

For each sub-process, there is defined a target and key performance indicators which are on regular base evaluated and measures needed are adopted. There is also laid down, in what way the activities of a respective sub-process affect the quality of statistical product (see Fig. 6).

The Directive entered into force 15 September 2015. Its implementation into practice was ensured by trainings and discussions at the level of cross-sectional domains.

Conclusion

The IŠIS implementation into routine operation fulfilled a requirement for standardisation of statistical production processes with direct impact on quality of outputs. There were eliminated individual solutions and statistical production was gradually integrated into unified system.

The system enabled to bring into practice electronic data collection for all surveys which are integrated in it and reflects the modern statistics requirements. Electronic data collection was also underpinned by the Amendment to the Act on state statistics. The Amendment brings into practice mandatory electronic data collection for all respondents with effect from 1 January 2016.

At the same time, there were developed technical and technological conditions which enabled the SOSR to respond on requirements of the business environment and to align deadlines laid down for the reporting obligation with deadlines set out for VAT statements delivery – without an impact on dissemination deadlines.

The system is still being developed. In present days, it is focused on implementation in the field of price statistics and national accounts. In the field of the Statistics of demography we are going on ensuring data exchange between information systems of the SOSR and the bodies of state administrative and local administration.