WP. 14
ENGLISH ONLY

UNITED NATIONS
ECONOMIC COMMISSION FOR EUROPE
CONFERENCE OF EUROPEAN STATISTICIANS

EUROPEAN COMMISSION
STATISTICAL OFFICE OF THE EUROPEAN UNION (EUROSTAT)

ORGANISATION FOR ECONOMIC COOPERATION AND DEVELOPMENT
(OECD)
STATISTICS DIRECTORATE

Workshop on Statistical Metadata (METIS)
Session IV: The GSBPM and Process Quality Management
Geneva, Switzerland, 05-07 October 2011

STATISTICAL BUSINESS PROCESS IN THE CZECH STATISTICAL OFFICE

Czech Statistical Office[1]
v 19.9.2011

I. Introduction

1. In 2005 the CZSO experts drafted the vision of the content and basic imagination of statistical meta-information system (SMS). The top management of the CZSO approved the conception and launched the SMS project. In 2005-2006 we designed the content and functional specifications of the SMS modules Statistical Classifications (CLASS) and Statistical Variables (VAR). These specifications created the base for implementation of the first two SMS modules.

2. In 2007 we started production run of both SMS modules. The module CLASS we uploaded with existing classifications and code-lists from the previous module METIS-Classifications and saved the newly solved official statistical classifications. Parallel we started to describe statistical variables in the module VAR according to our model of statistical data descriptions by metadata.

3. The SMS module Statistical Tasks was the next module for design. Before starting this activity we realized that this task couldn't be solved without a detailed knowledge of key statistical business process - preparation, processing and dissemination of statistical data and information.

4. In early 2006 the SMS team in cooperation with statistical department representatives drafted description of the statistical business process (SBP) used in the CZSO. Our experiences and practice created the frame for the formal description of the SBP model (SBPM).

[1] Prepared by Ebbo Petrikovits

5. Level of detail of the individual phases and sub-processes was different, for some sub-processes we specified activities as well. Each level of SBPM has a detailed specification of its content. For this paper we would use the abbreviation CZ-SBPM for the SBPM implemented in the CZSO.

6. We have to state here that the practices and experiences of preparation and processing of business statistics influenced our work on SBPM significantly. The experiences with preparation and processing of social statistics the SBPM influenced the model less.

II. Statistical Business Process implemented in the CZSO

7. The originally designed CZ-SBPM consists of six phases:
   A - Requirements of users,
   B - Preparation of a statistical task,
   C - Elaboration of technical and organizational specifications of a statistical task,
   D - Preparation of statistical task processing,
   E – Data collection and data processing,
   F – Dissemination of statistical information.

8. We did not take in consideration the need for a special phase for archiving data. Needs for data archiving were usually a standard part of the technical and organizational specifications of statistical tasks. The phase Evaluation of a statistical task we postponed for the future work on the model.

9. In the **Phase A** - Requirements we focused on the registration, evaluation of the requirement, searching for available sources (data, finance, human resources, time space), the way of possible implementation and draft decision (positive or negative) which would be submitted to the top management.

10. A - Requirements (NEED). This phase consists of following sub-processes:

i. registration of the requirement,
ii. analysis of the requirement and specification of impact on the statistical surveys and processing,
iii. proposal for potential implementation (if possible in variants),
iv. cost estimation,
v. specification of data sources (survey, administrative data, and/or other statistical tasks),
vi. information and decision proposal for the CZSO top management.

In principle the list of sub-processes is very similar to the GSBPM.

11. **B - Statistical task[2] preparation.** The structure of this phase differs form the phase 2 Design of the GSBPM. We have divided the phase B into two main sub-processes and each sub-process into specific activities (or sub-sub-processes). They are:

i. statistical task content specification,

[2] 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.

ii. specification of data sources.

The reason for this division was the need to describe in detail activities of statistical task content specifications and analyse various data sources needed for statistical purposes.

12. The sub-process "**statistical task content specification**" comprise these activities:

i. description of the statistical task content,
ii. specifications of needed statistical variables,
iii. specifications links (relationships) among variables,
iv. specification of necessary data sources,
v. design of requested statistical outputs,
vi. specification of processing algorithms,
vii. list of basic legislation and standards,
viii. list of needed code-lists.

13. The sub-process "**specification of data sources**" we divided into four independent branches:

i. preparation of statistical surveys,
ii. preparation of input data from existing statistical tasks,
iii. preparation of input data from administrative sources[3],
iv. preparation of data from other external sources (public or private data sources).

14. The branch "**preparation of statistical survey**" comprise following activities:

a) for collection via paper or electronic forms:
   i. design of a statistical questionnaire including explanations/instructions for respondents,
   ii. instruction for questionnaire collection and data keying,
   iii. printing of questionnaires;
b) for conducted interviews:
   i. design of conducted dialogue,
   ii. instructions for interviewers.

15. The branch "input data preparation from other statistical tasks" usually comprises only necessary structure transformation of statistical data into suitable format for designed statistical task.

16. The other two branches, preparation of input data from administrative sources and from other sources, consist usually of

i. specifications of links among variables,
ii. description of content and technical transformation rules from input structure of administrative data into structure necessary for statistical processing (including correct identifications of statistical units).

17. **C – Elaboration technical and organizational specifications of a statistical task.** The goal of this phase is to elaborate technical specification of a statistical task processing as a basic document for preparation of adequate program applications and carrying out individual, integration and acceptance tests of he developed applications. For this purpose we have defined several document outline patterns for preparation of the document.

[3] Under the expression „administrative data" we understand data produced by the public administration bodies.

18. **D – Preparation of statistical task processing.** This phase we have divided into two sub-processes:

i. preparation conditions for survey/s,
ii. preparation procedures for administrative data input.

19. The first sub-process comprises specification of sample parameters for sample, own sample generation, questionnaires distribution to the respondents, training of interviewers, instructions for preparation individual processing steps.

20. The second sub-process contains preparation of detail conditions for taking over administrative data and timetable for providing data from the administrative body/ies.

21. **E - Data collection and data processing.** The phase E we divided into following sub-processes:

i. carrying out field surveys,
ii. collection of input questionnaires,
iii. input data keying,
iv. input data validations,
v. imputation of missing data,
vi. production of aggregations,
vii. production of outputs (for result validations).

22. **D – Dissemination of statistical information.** This phase has been divided into six sub-processes:

i. outputs preparation for publication,
ii. preparation of electronic outputs,
iii. printing of outputs for publication (in limited extent),
iv. web presentation update,
v. non-standard outputs preparation (if necessary),
vi. dissemination of all kinds of outputs.

23. This approach to the description of the SBPM for the CZSO was applied in year 2006. After publishing the GSBPM (result of the UN/ECE METIS Working Group in 2009) we made brief comparison of both models. In principle we did not discover basic differences between both models. Our model has been strongly influenced by our practice in business statistics procedures. On the other hand we have to state that the GSBPM has been more sophisticated and generalized. Therefore our intention has been to revise our CZ-SBPM and elaborate it according to the GSBPM. For present activities under the IOP Redesign of SIS Project for the CZSO we have used for SIS architecture design.

Comparison of the CZ-SBPM with the GSBPM

24. When we compare the CZ-SBPM and the GSBPM we have to take in consideration several characteristics:

i. The CZ-SBPM was developed in 2006, the GSBPM in year 2009;
ii. CZ-SPM was prepared by the CZSO experts, GSBP by the statistical community under the METIS Working Session of the UNECE;
iii. The aim of the CZ-SBPM was to describe the statistical business process with the stress on the phase Design an Build; the reason for it was very

simple; we needed a base for preparation of content and functional specifications of the future SMS module TASKS in that time;

iv. The GSBPM has been developed as a comprehensive handbook on statistical business process for statistical community.

25. Pros and contras for the CZ-SBPM:

Pros:

i. it was developed for specific aim, for design of SMS module TASKS;
ii. we have had in the CZSO the first comprehensive document on the statistical business process; the document was officially discussed and approved by the top management of the CZSO;
iii. we have used it for partial improvement of selected processes existing at that time,
iv. we discussed in very detail the preparation of data sources, special attention was paid to use of administrative data.

Contras (in comparison with the GSBPM):

i. the content of CZ-SBPM phases has been not balanced very well; the GSBPM has showed us the way for improvement of out model;
ii. we focused our activity mainly on the phases Design and Build,;
iii. we underestimated the importance of division data processing into phases Collect and Process;
iv. we did not take in consideration the phase Archive; it our previous practice the activity "archiving" was the part of the statistical task specifications;
v. the phase Evaluate we left out for practical reason, the approach to this topic was a bit controversial.

26. After publishing the GSBPM we have revised our CZ-SBPM. We practically followed the content and structure of the GSBPM. Still we have some problems how to incorporate activities concerning preparation of administrative data for statistical processing. We have felt that this kind information has been missing in the GSBPM.

27. The revised CZ-SBPM we have applied in the design of the SIS architecture for the Redesign of SIS Project, which was launched under the umbrella of the EU Integrated Operational Programme by the end of 2009. All activities connected with future SIS architecture have been strongly influenced by the GSBPM. It means that we have introduced the CZ-SBPM/GSBPM into our everyday practice. We consider this document as useful and very practical handbook for statisticians and ICT staff in design and development of concrete statistical process for preparation, processing and dissemination of statistical information and data.

III. Metadata for Process Quality Management

28. In compliance with the ESS Handbook of Quality Reports we distinguish two components of the quality model:

i. output quality components (with other words statistical data quality),
ii. process quality components.

29. Both components are closely connected and mutually influence each other. The quality of processes is a precondition for achieving high quality outputs.

30. Process quality has two broad aspects:

i. Effectiveness – which leads to outputs of good quality;
ii. Efficiency – which leads to outputs production at minimum cost to the NSI and to the respondents that provided input data.

31. The European Code of Practice (CoP) provides and specifies in detail process quality components. The components are broken into two groups: components of institutional environment and components of individual statistical sub-processes or activities.

32. **Components of institutional environment.** Institutional and organisational factors have a significant influence on the effectiveness and credibility of a statistical authority producing and disseminating European Statistics. The relevant issues are: professional independence, mandate for data collection, and adequacy of resources, quality commitment, statistical confidentiality, impartiality and objectivity.

33. **Professional Independence;** the professional independence of the staff (responsible for the process and output dissemination) from other policy, regulatory or administrative departments and bodies, as well as from private sector operators, is required to support the credibility of outputs.

34. **Mandate for Data Collection;** the organisation has a clear legal mandate to collect the particular information required. Where a survey is conducted under the statistics act providers can be compelled by law to allow access to or to deliver data.

35. **Adequacy of Resources;** the resources available are sufficient to meet systems and processing requirements.

36. **Quality Commitment;** staff commit themselves to work and cooperate according to the Principles stated in the Quality declaration of the European statistical system.

37. **Statistical Confidentiality;** the privacy of data providers (households, enterprises, administrations and other respondents), the confidentiality of the information they provide and its use only for statistical purpose is absolutely guaranteed.

38. **Impartiality and Objectivity;** production and dissemination of statistics respect scientific independence and are conducted in an objective, professional and transparent manner in which all users are treated equitably.

39. **Components of individual statistical sub-processes**. European and other international standards, guidelines and good practices must be fully observed in the processes used by the statistical authorities to organise, collect, process and disseminate official statistics. The credibility of the statistics is enhanced by a reputation for good management and efficiency. The relevant aspects are sound methodology, appropriate statistical procedures, non-excessive burden on respondents and cost effectiveness.

40. **Sound Methodology;** sound methodology, including adequate tools, procedures and expertise underpins quality statistics.

41. **Appropriate Statistical Procedures;** appropriate statistical procedures are implemented from design through data collection to data validation and evaluation.

42. **Non-Excessive Burden on Respondents;** the reporting burden is in proportion to the needs of the users and not be excessive. It is monitored over time and targets are set for its reduction.

43. **Cost Effectiveness;** resources are effectively used.

44. Te CoP specifies a set of possible indicators for each criterion of process quality measurement. These indicators can be understood as general recommendation. For practical measurement of process quality we have to choose and specify a set of relevant attributes for measurement process quality in practice. We try to explain our approach to this topic in several paragraphs below.

45. At present we are in the state of elaboration of content and functional specification of the SMS module Quality. One part of the specifications will be devoted to the selection and specification of relevant attributes on process quality. Nevertheless for some process quality criteria we have defined several attributes, which have been already practically used for measuring process quality.

46. Concerning the principles for institutional environment the answer we can find in the statistical act. When we compare the content of the Czech act on state statistical service we can to give positive answers to all six criteria (or on recommended indicators). In other words the institutional conditions for carrying out official statistics are provided by law.

47. The situation in statistical processes is a bit different. The given criteria coming out from practical experiences of statisticians and of application of "best practices". We may specify concrete attributes, which can be measured and recorded during the processes (manually or automatically) and evaluated later.

48. Criterion **Sound Methodology**. For recommended indicators we may replay yes or no. It seems to us that we can propose following attributes to this criterion:

i. Quality of the content and functional specifications of a statistical task,
ii. Quality of the documentation for staff carrying data input and data validations,
iii. Quality of specifications of imputation procedures.

49. We can measure these attributes via organising a local survey among statistical staff responsible for data collection and processing. The answers will very subjective, but they will show us the weak points in this field. They will show us which parts of documentation should be improved.

50. Criterion **Appropriate Statistical Procedures**. This criterion expresses the rate of transformation of sound methodology into right described procedures for preparation and processing of a statistical task. We can specify for example these attributes on existence of following manuals, handbooks, guidance and tools

i. Tools for metadata descriptions of statistical tasks,
ii. Manual for design of statistical questionnaires,
iii. Computer aided tool for drawing statistical questionnaire in paper and electronic form,
iv. Manual for specifications of validation rules and data editing,

v. Computer aided tool for preparation of statistical task specifications,
vi. Tools for generation pieces of applications based on metadata description,
vii. Handbook of transformation procedures for using administrative data for statistical purposes,
viii. User manuals of applications for preparation, processing and dissemination of statistical data.

51. Criterion **Non-Excessive Burden on Respondents.** There is very important criterion from the point of view of respondents. In our practice we have measured and recorded following attributes

i. for questionnaires:
   - Number of approached respondents,
   - Percentage of entrepreneurs,
   - Amount of filled-in items in the questionnaire,
   - Estimated time for filling in the questionnaire
   - Periodicity of submitting the questionnaires,

ii. for statistical tasks: rate (in %) of surveyed items according to the response duty origin
   - A – content and form have been taken from the EU regulation,
   - B – content has been taken from the EU regulation, form has been set up by the NSI,
   - C – content and form have been set up by the NSI.

52. These attributes have been recorded for each statistical task and questionnaire/s used in the statistical task. We calculate global response burden for each task in hours per year and cost in CZK.

53. Criterion **Cost effectiveness.** For this criterion we can recommend several possible activities or attributes only. For example

i. rate of automated operations a manually carried out operations,
ii. recording of scheduled start/end time (in timetables) and actual time of start/end for all activities
iii. recording of time consumption by the staff for specified activities,
iv. recording machine time consumption for individual applications/sub-processes.

Specification of attributes for this criterion should be accompanied with evaluation procedure/s.

54. Our present situation in the field of metadata on quality we can characterize in following way. Our expert team responsible for quality implementation has elaborated functional specification of the SMS module QUALITY including the frame of quality evaluation. Implementation of the module will be under the IOP Redesign of SIS Project. In time being we have finished the specifications for the invitation to the public tender. According to the preliminary schedule the evaluation of bids should take place at the beginning of next year.

IV. CZSO metadata-system – the state of implementation

55. As it has been mentioned in the chapter I, we began the development and implementation of the SMS modules in the middle of year 2005. In time being we have implemented three

basic modules – Statistical Classifications (CLASS), Statistical Variables (VAR) and Statistical Tasks (TASKS). These three modules practically create the base of the SMS. And everyday practice has convinced us that the TASKS module represents the kernel of the SMS.

Statistical Classification module

56. The **CLASS module** (in Czech we use abbreviation KLAS) was implemented as the first one. Conceptually, the module has been based on the principles specified by the Neuchatel Group. The module contains two main objects – statistical classifications and statistical code-list. Code-lists are stored physically and classifications create virtual objects. It means that we have for each classification the description of the classification hierarchy in the classification catalogue and physically stored individual code-lists for all hierarchical levels. The list of individual code-list is kept in the code-list catalogue.

57. We distinguish three types of code-lists:

i. Basic code-list which fully follows the Neuchâtel Terminology Model for statistical classifications; the content of the basic code-list is exhausting and items are mutually exclusive (not overlapping);
ii. Aggregation code-list – its items are composition of basic code-list/s items; there are direct links between aggregated item and source items in basic code-lists;
iii. Combined code-list – its items can be an combination of items from basic and aggregation code-lists.

58. Basic characteristics of a code-list are as follow:

i. Item code,
ii. Short name,
iii. Full name,
iv. Set of presentation texts,
v. Definition of the items (if possible and necessary),
vi. Attributes (can be specified on the request of the response person of the code-list),
vii. Links to the other code-lists.

59. At present this module works in production regime. It comprises cca 1200 code-lists and all international classifications for statistical purposes. The module creates only one source of code-lists and classifications for the whole statistical information system and for the SMS as well. Selected part of the stored classifications and code-lists has been presented on the Internet.

Statistical Variables Module

60. The **VAR module** (in Czech abbreviation UKAZ) was the next implemented module. It contains only one object of description – statistical variable (STATVAR). The STATVAR describes the content part of the statistical data description model. The model is based on the identification metadata description of the statistical data.

61. The metadata identification set is divided into four complex variables:

i. statistical variable,
ii. time variable,

iii. statistical object (space variable)
iv. complementary variable.

The complex variable consists of elementary variables.

62. The statistical variable has structural description and reference description. Structural description is created by compulsory and optional elementary variables:

i. compulsory elementary variables:
   a) statistical concept,
   b) statistical function,
   c) measurement unit category
ii. optional elementary variables:
   a) 1-n subject-matter breaks down.

63. In time being we have analysed a described over 4 000 statistical variables of business statistics and for Population Census 2011. Stored metadata on statistical variables we use for description of the statistical tasks content.

Statistical task module

64. The **TASKS module** (in Czech abbreviation ULOHY) has been the third implemented module. The goal of the module is to prepare formalized and textual descriptions of all objects required for the description defined in the module. These objects we utilize for full description statistical task, it means statistical content (variables and items of variables), required functions for processing (validation rules, imputation), questionnaires, data structures, outputs, users, technological requirements, organization of data collection and processing, time tables for phases BUILD, COLLECT, PROCESS and DISSEMINATE. The metadata stored in the SMS database create a source for activities under the phase BUILD. From metadata description we can generate patterns of questionnaires, applications for data validations, document describing content, functional, technical and organizational specifications of a statistical task and applications for data entry.

65. The CZ-SBPM was the starting framework for the design of metadata attributes for selected phases of the SBPM, mainly for phases DESIGN, COLLECT and PROCESS.

66. For each activity we have proposed a group of metadata, which would describe features of the activity. In the next step we have organized these groups of metadata into so called "objects of description". Each object of description has had own set of metadata characteristics: attributes, items of statistical variables, structure specifications, subordinated objects, and responsible persons. The content of the metadata items may be free text, structured text, coded value from code-lists, mathematical expression. The proposed description objects and their metadata we use for the structured description of statistical tasks.

67. At present we can describe following object types needed for full description of a statistical task:

i. statistical task - identifies the task, specifies a responsible person and department, briefly describes the task, sets up the initial and target values of quality attributes;
ii. item of variable (so called VIP)– detailed description of basic variable used for concrete survey or data file;

iii. data structure components (DSC) – this common term covers a set of data structure object types (explanation see below);
iv. validations rules - rules for check procedures on input data;
v. reporting duty specification - defines the rules for the selection of statistical units from sample frame to the survey sample for concrete survey;
vi. time tables for programming, testing and production;
vii. document in various data format (doc, docx, rtf, xls, xlsx, pdf);
viii. program module and program module run - for keeping information on production process.

68. For data structure components (**DSC)** we have defined following object types for data structure specifications:

i. Super-section (= VIP container),
ii. Super-Questionnaire,
iii. Simple questionnaire and combined questionnaire,
iv. Annex to the combined questionnaire,
v. Enclosure to the combined questionnaire,
vi. Section,
vii. Section variant.

69. The DSC **Super-questionnaire** consists of one or more combined questionnaires, one or more annexes and/or one of more enclosures. The super-questionnaire is a virtual object, which we need for specification of its real components. This kind of input form construction solves for preparation of individual questionnaires content based on the group of NACE codes. Our goal is to send to the respondents only that part of the combined questionnaire, which is relevant for him accordingly to his NACE codes as the respondent.

70. The **simple questionnaire or combined questionnaire** consists of

i. questionnaire heading specifications,
ii. set of questionnaire sections and reduced sections,
iii. explanatory notes to the individual sections and section variants.

71. The **annex** comprises

i. annex heading description,
ii. sections and section variants,
iii. explanatory notes.

The annex must be always connected with a super-questionnaire. It cannot be used independently.

72. The **enclosure** is a specific variant of an annex. It contains one section only. It is always connected with a combined questionnaire.

73. The **section** creates a basic building block of a questionnaire/kernel-questionnaire, annex or an enclosure. It has two main parts:

i. statistical content description,
ii. structure description.

74. The **section variant** is derived form the section; it contains selected rows or columns of the source section.

75. The content description of the section consists of a set of attributes, list of used variable items (VIPs) and list of additional objects. The structure component consists of attributes, dimensions, structure objects, legends for rows and columns, footnotes and remarks to the section.

76. Now we should explain the concept of special object – **item of variable**. In our practice we call this object as "internal field identifier" and we use for it the abbreviation **VIP**.

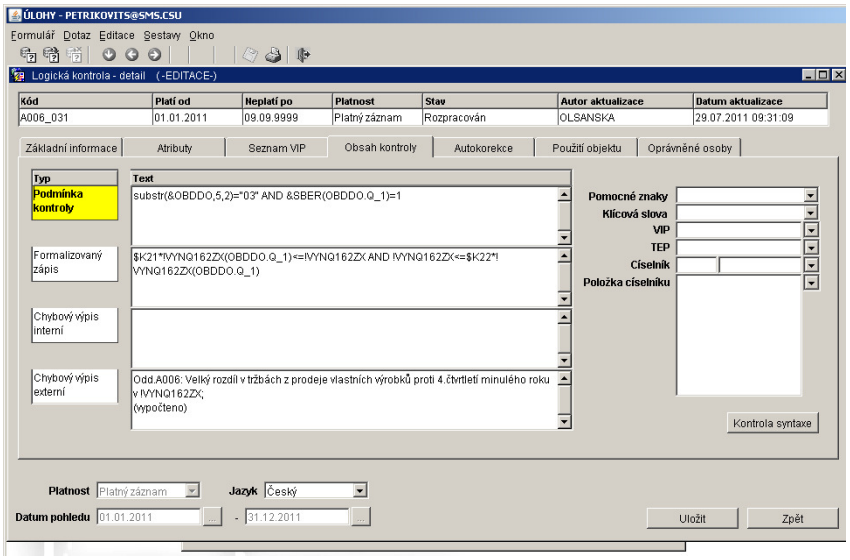
77. The VIP has a structure and an identifier. The VIP structure represents the structure of statistical variable extended with additional elementary variables describing its actual structure used in concrete survey or data file. The structure consists of statistical variable identifier (IDENT; it is taken from VAR module), compulsory elementary variables "time characteristic" and "concrete measurement unit" and optional elementary variables depending on the concrete VIP structure. The optional elementary variables may be additional object/s, seasonal adjustment or constant prise level.

78. The picture below shows the metadata structure for description of a statistical data.

Statistical variable (STATVAR)				**Complemntary variable**									**Statistical object (space variable)**					**Time variable**					
Concept	Function	KMU	S-M breaks down	**Concreta MU**	**Time characteristic**	**Price level**	**Seasonal adhjustment**	**deflation**	Way of data capture	Statistical task ID	Version of processing	Process run	**Main SO**	Complementary SO1	Complementary SO2	.....	Complementary SOn	Reference time period				Time period of the processing	
																		current		basic			
																		B	E	B	E	B	E
STATVAR Ident																							
VIP = Item of STATVAR														VIP									

79. In present application we can derive three types of VIP identifier. The application TASKS generates primary the basic identifier; it consists of the concrete VIP container identifier (so-called prefix of the VIP code) and six-digit number. User may change the 6-digit part into an abbreviation with some meaning or into a structured code derived from the VIP structure.

80. The **validation rule** object type describes condition for checking correctness of input data. Special language of for validation rules specification has been developed and implemented in the application TASKS. The picture below illustrates an example of validation rule specification.



81. The **response duty** object type specifies al kind of conditions needed for derivation of samples for individual statistical tasks. At present we can specify response duty for economic subjects from the Business Register (BR) only. The application of BR can read the specifications, transforms the specifications into retrieval procedure and generates the final sample file.

82. The **program module** object type describes group of applications, which are needed for carrying out set of activities of a concrete sub-process. It contains information of application components, short function description, responsible person and technological address for the initiation of the module.

83. The program module object type is closely connected with the **program module run** object type. In this object we want to record starting an ending date and time of the program module, result of the run (OK or failure), person who initiates the run, reason of the failure (if occurs). This information will be used in process quality evaluation procedure.

84. The **timetable** object type has been used for description of individual activities in time axis. The timetable can have hierarchical structure; it means that a step in timetable level 1 can has several detailed steps in timetable level 2. Time can be calculated in days ad hours, for calendar days or for working days only.

85. The **document** object type allows to store in the SMS database any kind of dataset in various formats. Stored document has an identifier, name, and characteristic of the content.

86. Where we are in statistical tasks descriptions. In time being we have fully described the STS[4] statistical tasks and main part of the SBS[5] statistical tasks. Other statistics will be described in next year 2012. We schedule putting the module in full production regime in the second half of year 20013, for surveys and statistical tasks of year 2014, after finishing the IOP project Redesign of SIS.

[4] STS – Short Term Statistics
[5] SBS Structural Business Statistics

V. Lessons learned

87. The GSBPM we consider as the useful and fruitful guidance supporting the implementation of statistical business process or for analysis of of individual phases or sub-processes effectiveness. The final version of the GSBPM reflects everyday practice and experiences of the NSI.

88. The GSPM describes in detail phases of the statistical business process, their division into sub-processes. The model is open, allows adding additional levels, which describe concrete sub-process in deeper detail accordingly the need of the practical implementation. The reasons for deeper details may be usually improvement of the sub-process or analysis for specification of relevant metadata for description of the sub-process or metadata for sub-process quality measure.

89. Accordingly to our practice and experience with CZ-SBPM (improved in compliance with the GSBPM) we see following areas of GSBPM specified in the paragraphs 90 - 93.

90. The GSBPM represents **basic guidelines** for practical application for a statistical business process implementation:

i. The GSBPM divides the SBP into phases and sub-process. This division gives its full overview and description. The model creates practical manual for users who need to design and implement a statistical survey, processing of surveyed data and disseminate statistical information to the end users.
ii. The model we do not consider as a dogma. User can choose sub-processes which he/or she needs for the own implementation. The model also allows adding specific details reflected the concrete survey and processing.
iii. As a typical example we can mention three variants of the statistical task:
    a) a statistical task with statistical survey as a source of input data,
    b) a statistical task with administrative data a source of input data,
    c) a statistical task with using data form other statistical tasks as a source of input data;

Each of above mentioned three variants of statistical task would have own concrete SBPM. These three SBPM will differ in the phase COLLECT and small differences in the phases DESIGN and BUILD.

91. **Analysis of the sub-processes** or their details. The results of the analysis can help us in technical specification elaboration for program applications. In the same time the analysis my lead to the standardization of sub-process or its detailed activities and proposal for standardization of future program applications.

92. **Derivation of metadata**. When we want to specify set of metadata for individual sub-processes or their activities the GSBPM becomes necessary tool for this activity. We can specified next groups of metadata:

i. Metadata needed for specification of program application functions,
ii. Metadata for gaining information on data quality,
iii. Metadata for gaining information on process quality,
iv. Location and way of recording required metadata (automatically or manually),
v. Design of evaluation procedures for data and processes,
vi. Design of evaluation procedure for the statistical task.

93. **Effectiveness analysis** of sub-processes and activities, mainly following characteristics:

i. Time of activity run,
ii. Requirements on communication with users and operators,
iii. Consumption of ICT resources,
iv. Possibility of replacement of manual activity by automated activity,
v. Communications among applications carrying out given activity/sub-process.

VI. Conclusion

94. According to the work plans of the CZSO, a lot of activities in further development of the SIS and next SMS modules are scheduled for next three years. The kernel of these activities we consider the IOP Redesign of SIS Project, which is concentrated on further development of the SIS and SMS. The top management of the CZSO considers the IOP project as the most important for the future activities of the office and therefore it gives it the highest priority.

95. Coming back to the content of this paper we have tried to present out practical experience with the GSBPM, the way, how we have utilized this model for solving specific parts of the SIS and SMS, preferably in specifications of metadata suitable for recording some features of statistical business process and its quality. Special attention has been paid to the information on the implemented SMS modules used in our practice.

96. At the end, we consider the document describing the GSBPM as very useful and practical handbook for statisticians and ICT experts. It describes the phases and sub-processes of the statistical business process in comprehensive way. User can work in accordance with it and can extend proposed three level structure of the model with additional structure level/s, which will describes the level of sub-processes in details needed by the users. In our opinion the model can by successfully used for derivation of applications for sub-processes implementation and for metadata specifications for formal and textual descriptions of the SIS objects as well. Based on our relatively short experience and work with the model we may recommend its use in statistical practice.

List of literature

1. European Code of Practice, COM(2005) 217
2. ESS Handbook of Quality Reports, Eurostat, 2009
3. ESS Standard of Quality Reports
4. Generic Statistical Business Process, version 4.0, UNECE April 2009

List of abbreviations

CLASS	Application Statistical Classification
CoP	European Code of Practice
CZ-SBPM	Statistical Business Process Model of the CZSO
CZSO	Czech Statistical Office
DSC	Data Structure Component
ESS	European Statistical System
GSBPM	Generic Statistical Business Process Model
ICT	Information and Communication Technology
IOP	Integrated Operational Program

MU	Measurement Unit
SBP	Statistical Business Process
SMS	Statistical Metainformation System
STATVAR	Statistical Variable
TASKS	Application Statistical Tasks
VAR	Application Statistical Variables
VIP	Internal field identifier