3. Statistical Metadata in each phase of the Statistical Business Process (Czech Statistical Office)

back to case study home page

3.1 Metadata Classification

Statistical metadata include **content oriented** and **technological metadata.** Both groups are needed for design, implementation and running of STs.

Content oriented metadata are presented in the following groups:

1. **Metadata on statistical concepts and models**

This group describes models of statistical classifications and statistical variables.

2. **Metadata on statistical methods**

This group describes imputation methods of missing values/data, grossing up to the whole population of observed units, methods for time series conversions, seasonal adjustment, methods for expert estimates, analytical, mathematical and statistical methods of evaluation, etc.

3. **Metadata on processing procedures**

This group describes processing procedures for individual stages of STs life cycle. For example the data collection, respondent burden measurement, preparation of statistical questionnaires, data validation, quality assessment, and aggregation, preparation of statistical tables, etc.

4. **Metadata on use of statistical information**

This group describes user satisfaction, use of statistical information by respondents, analysis of users' requirements for information, FAQ, users' opinions, use of web pages, etc.

5. **Metadata on SBP assessment and evaluation**

This group provides source materials for assessment and evaluation of effectiveness in individual phases of SBP and source materials for financial controlling in the CZSO.

The table below shows a placement of the above groups of metadata in the SMS architecture.

SMS **Groups of metadata**	**Classifications**	**Variables**	**Tasks**	**Quality**	**Dissemination**	**Users**	**Respondents**	**Time series**	**Data fund**
Statistical models	x	x							

Statistical methods			x						
Processing procedures			x	x	x	x		x	x
Use of statistical information					x	x	x	x	
Assessment and further development	x	x	x	x	x	x	x	x	x

3.2 Metadata used/created at each phase

The aim of the SMS is to support all phases of SBP as set out in point 2.3. The SMS is an end-to-end system.
The subsystems store, update and use metadata items in individual phases of SBP, as shown in the following table.

SBP phases **Subsystems**	**Definition of statistical task**	**Processing preparation**	**Data collection**	**Primary processing**	**Data analysis**	**Dissemination**
Classifications	x			x		x
Variables	x					x
Tasks	x	x	x	x	x	
Quality	x		x	x	x	
Dissemination	x					x
Users	x					x
Respondents	x	x	x			x
Time series	x				x	x
Data fund	x	x		x		

Due to progress of work on the SMS (see point 4.1) we take an example of detailed information on metadata kept in subsystems CLASS, VAR and TASKS that are being developed.

Subsystem **CLASS** is based on the Neuchâtel model of statistical classifications. It allows creation, storage, update and use of statistical classifications, which are necessary for data processing. There is basic metainformation kept on each classification incl. its history, e.g. the title and coordinator of classification, validity and contents of classification/code-list in language versions (CZ, EN) etc.These characteristics on code-lists are stored in the SMS database.

- identification code of the code-list,
- code-list item code,
- full and short name of the code-list item,
- names for presentation,
- validity of the code-list item (from-to),
- definition of the code-list item,
- attributes of the code-list item (voluntary),
- links to the other code-list and its items.

Subsystem **VAR** is based on a unique model for description of statistical variables at micro data and macro data level. The model was developed in the CZSO using experience of Work Session on Metadata UNECE. Metadata are aimed at description of the contents of statistical data. The most important metainformation is the statistical concept, statistical function, title, definition, and unit of measure and subject-matter breakdown. Also metainformation on the coordinator, subject-matter area, validity, etc., is kept on statistical variable.

. Following characteristics of a variable are kept in SMS database:

- identifier,
- structural description,
- full and short name,
- definition,
- validity (from-to),
- set of compulsory and voluntary attributes,
- names for presentation,
- remarks

Subsystem **TASKS** contains metainformation on functional and technological specifications of STs. Mainly the following metainformation is kept:

- basic characteristics of a task,
- statistical questionnaire content a structure definition,
- input data sets,
- annex to the decree on annual programme of statistical surveys (list of surveys with response duty),
- data item validation rules, auto-correction rules, transformation rules, derivation rules
- definition of statistical samples,
- specifications of imputation methods,
- quality requirements,
- aggregations rules,
- estimates procedures,
- specification of users,
- time-tables for preparation of a task, for user tests and for statistical production,
- applied code-lists,
- legislation base for a task,
- data flow and organization of the collection and processing,
- documentation (user and technological).

3.3 Metadata relevant to other business processes

The division of life cycle of the statistical task (part 2.3) will also be used for cost controlling purposes. Metainformation on the history of a statistical task, especially time-table of processing, will be used when considering labour intensity of individual phases/activities of the process.

Information on statistical data quality obtained in the history of processing will be used for quality of work assessment of the CZSO's departments responsible for design and implementation of statistical tasks. This information can also be used to assess the quality of work of the entire Office. The method applied is European Foundation for Quality Management (EFQM).

Subsystem TASKS is designed to allow specification of non-statistical tasks such as controlling, other administrative subsystem or development tasks.