Meta Data Driven Integrated Statistical Data Management System

By Karlis Zeila, consultant, karlis.zeila60@hotmail.com
Vice President of CSB of Latvia 1996 - 2009

The aim of this report is to introduce participants with the experience gained within the process of the development and implementation of the new generation statistical data processing system, which integrates several subsystems, and is metadata driven.
The new system in accordance with CSB IT Strategy is developed as centralized system, where all data is stored in corporate data warehouse. The new approach is to unite what logically belongs together by using advanced IT tools, to ensure the rationalizing, standardization and integration of the statistical data production processes. This in practice means movement from classic stovepipe to process oriented data processing approach.
The theoretical basis for system architecture is results of Meta data studies done and published by Professor Mr. Bo Sundgren, Statistics Sweden.
Metadata driven data processing system is based on systemisation and storage in the centralised Meta data base - all metadata on surveys, indicators and classifications used, thus ensuring preconditions for unification and harmonisation of the statistical indicators avoiding duplication in different surveys. Introduction of unified system of statistical indicators leads to expansion of the system functionality and metadata becomes as the key element of the entire system.
System integration is based on direct use of the Business Register data in case of Business statistics surveys data processing. Several statistical registers could be included in the Registers module. The main idea of such kind of integration is to ensure easy access to wide range of the respondent's data; this data is located in the statistical registers.
Most of the system software modules are connected with the Registers module.
All System software modules are connected with the Core Metadata module. We can talk about integrated system because of all data is stored in the corporative data warehouse.
System is divided into following business application software modules, which have to cover and to support all phases of the statistical data processing: Core metadata base module; Registers module; Data entry and validation module; Mass data entry module; WEB based data collection module; Intrastat data processing module; Missing data imputation module; Data aggregation module; Data analysis module; Data dissemination module; User administration module.
WEB based data collection module practically works like extension of paper based questionnaires data collection technology. The layout of the e- form is about the same, as on the paper version, therefore respondents do not meet any changes moving from paper to the screen. Even more they get much more advantages submitting data electronically. They can see data submitted in the previous periods, they could make changes within data submitted in previous periods and etc.
System is successfully working in CSB of Latvia from August 2002 .
Key words: statistical data processing system, metadata, business register, respondents

1. Introduction

From 1997 – 1999 Central Statistical Bureau of Latvia (CSB) experts in cooperation with experts contracted from PricewaterhouseCoopers were prepared General Technical Requirements for the project "Modernisation of CSB – Data Management System". Technical Specification embodied key technical and functional requirements for the new system where statistical Meta data should be used as the key element in statistical data processing. A lot of additional requirements appeared within the process of development
The main business and information technology (IT) improvement objectives that the CSB intends to achieve as the result of project have been identified and are as follows:

Using modern IT solutions:

- Increase efficiency of the main process at CSB, production of statistical information;
- Increase the quality of the statistical information produced;
- Improve processes of statistical data analysis;
- Modernise and increase the quality of data dissemination;

- Avoid hard code programming via standardisation of procedures and use of Meta data within the statistical data processing.

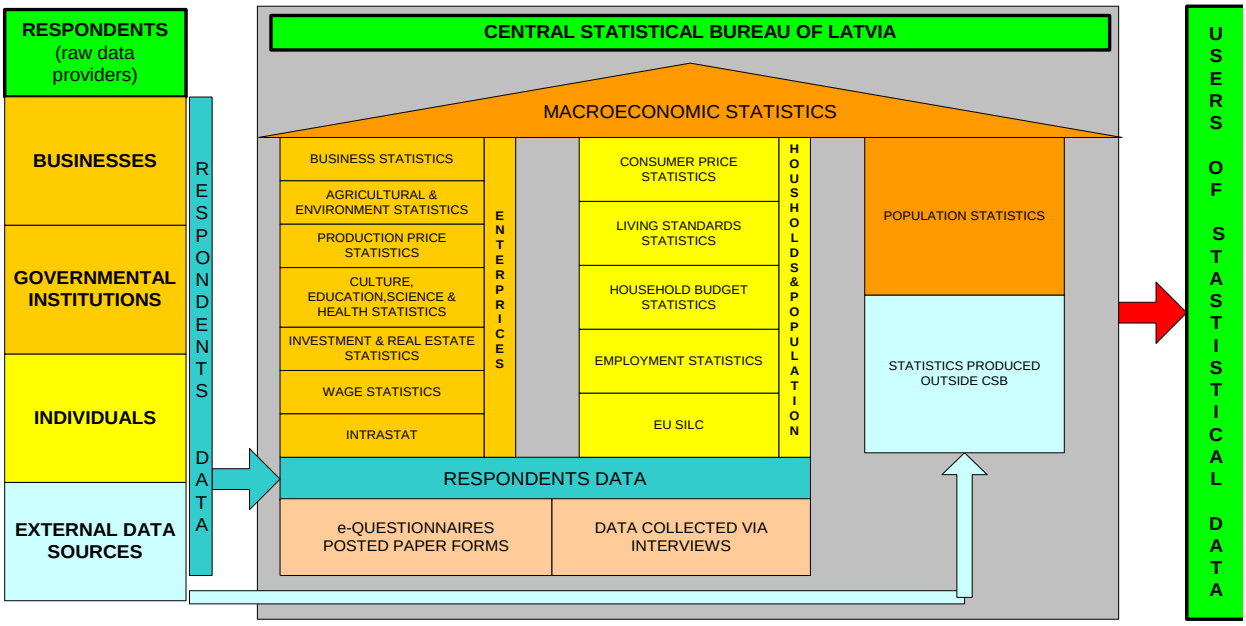
2. Technical platform and standard software used

To be in line with the CSB IT strategy, existing computer and network infrastructure for the system development the Microsoft SQL Server 2000 was taken to handle system databases. All applications comply with the client/server technology model, where data processing performed mostly on server side. Client software applications are developed using Microsoft Access 2000. Other components of Microsoft Office 2000 are used as well. For multidimensional statistical data analysis is used Microsoft OLAP technology. As tool for data dissemination was chosen software product PC-AXIS developed by Statistics Sweden, which is widely used in different statistical organizations in different countries.
System has been upgraded to MS Server 2003, MS SQL Server 2005 and applications reprogrammed in .net in 2008.

3. System architecture

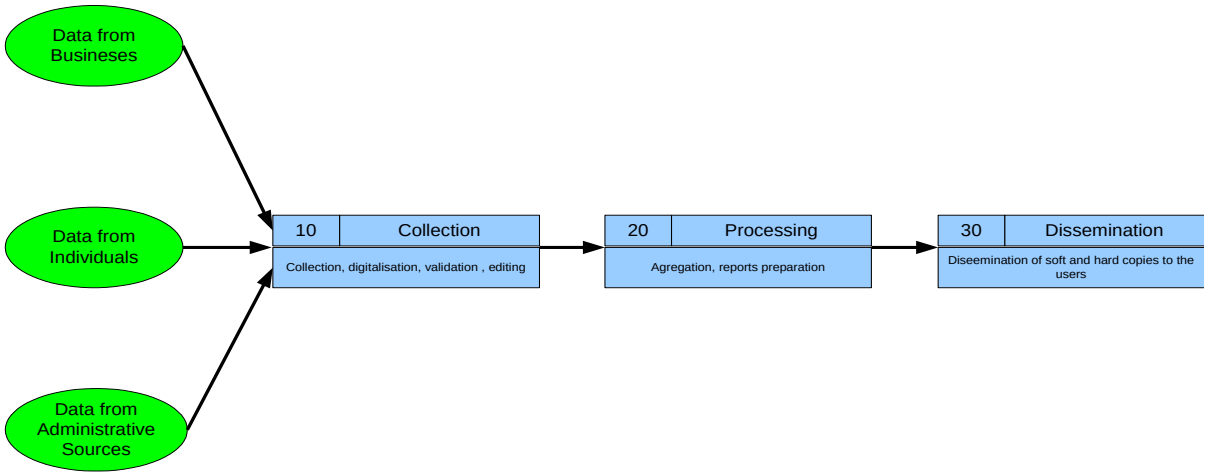
Before development and implementation of the system classic Stove Pipe data processing approach with all appropriate technical incompatibilities existed as a consequence of the wide range of technology solutions was in use.
As the result of the analysis of processes, data flows, user's requirements and situation mentioned above it turned out that most of statistical surveys have the same main steps of data processing starting with survey design and ending with statistical data dissemination. The separation was necessary in between surveys filled in by respondent and surveys filled in with assistance of interviewer. The main difference was found in both: data obtaining methods and data aggregation algorithms obtaining data from businesses and from persons & households. Business respondents are filling in questionnaires in most cases mailing them to CSB. Data from persons & households we are obtaining via interviewers service. Statistics structuring in the Central Statistical Bureau of Latvia taking into account all written above is presented on a high level diagram as it is shown on the Figure 1.

Figure 1. Statistics Structuring in CSB based on the Process Oriented data processing



And typical statistics production high level workflow we would see as very simple diagram on Figure 2.

Figure 2. Typical statistics production high level workflow



Looking deeper in the statistical processes taking place in Statistics Latvia we can define them as in Figures 3. and 4.

Figure 3. Statistical data flow covered by the data processing system

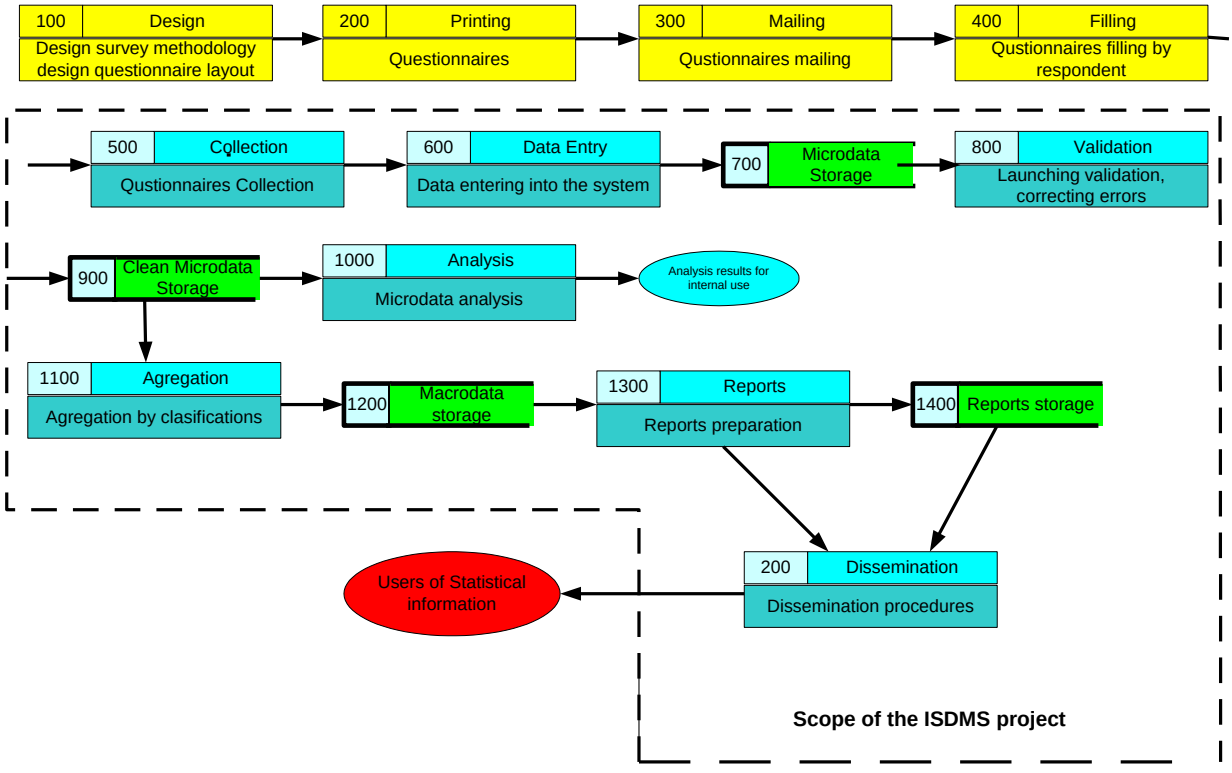
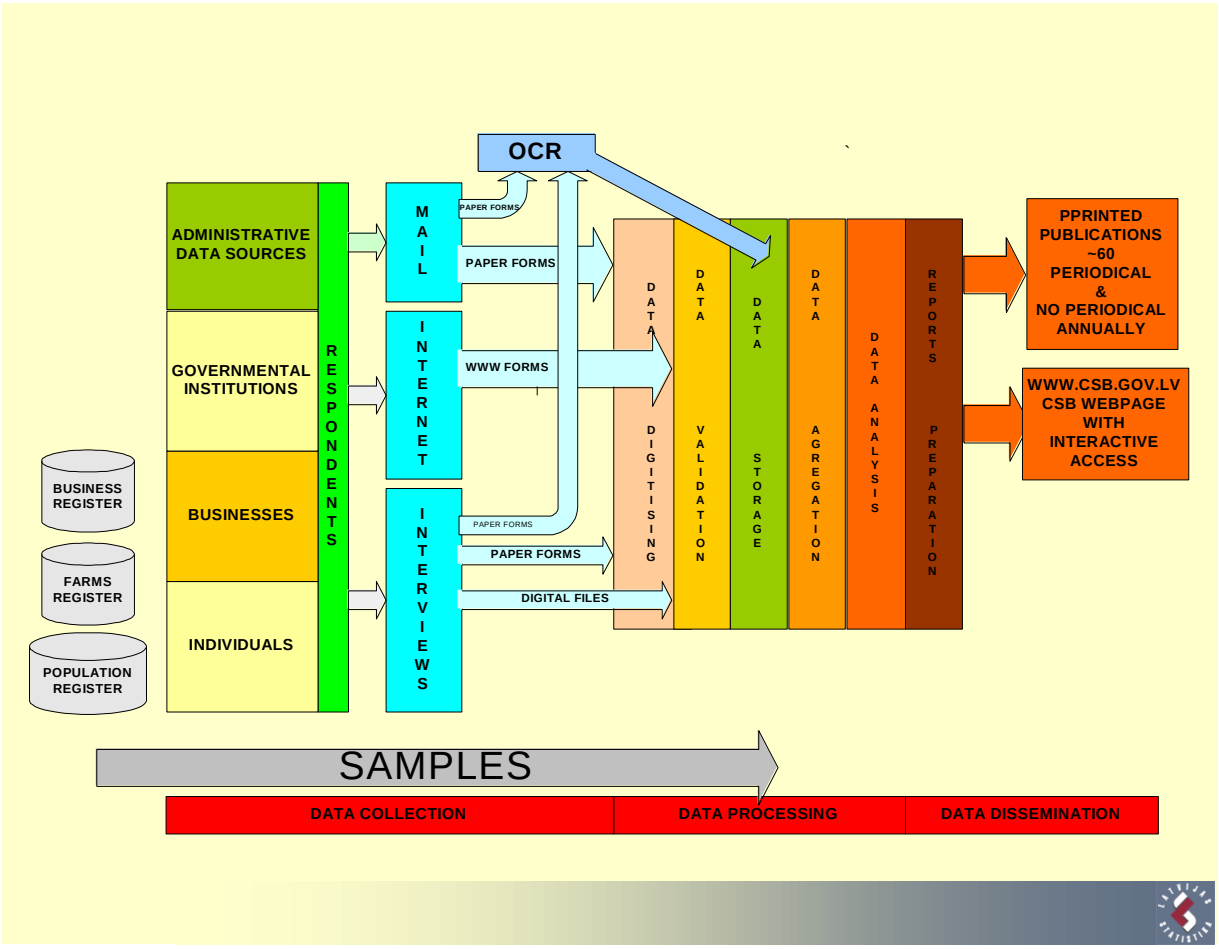


Figure 4. Typical data flow through the statistical processes



As the theoretical basis for system architecture was taken "Information systems architecture for national and international statistical organizations" elaborated by professor Mr. Bo Sundgren (Statistics Sweden) and issued by UNSC and ECE and approved by Conference of European Statisticians as Statistical Standard.

New system contributes harmonization and standardization and is developed as centralized system, where all data are stored in corporate data warehouse. The approach is by using advanced IT tools to ensure the rationalizing, standardization and integration of the statistical data production processes.

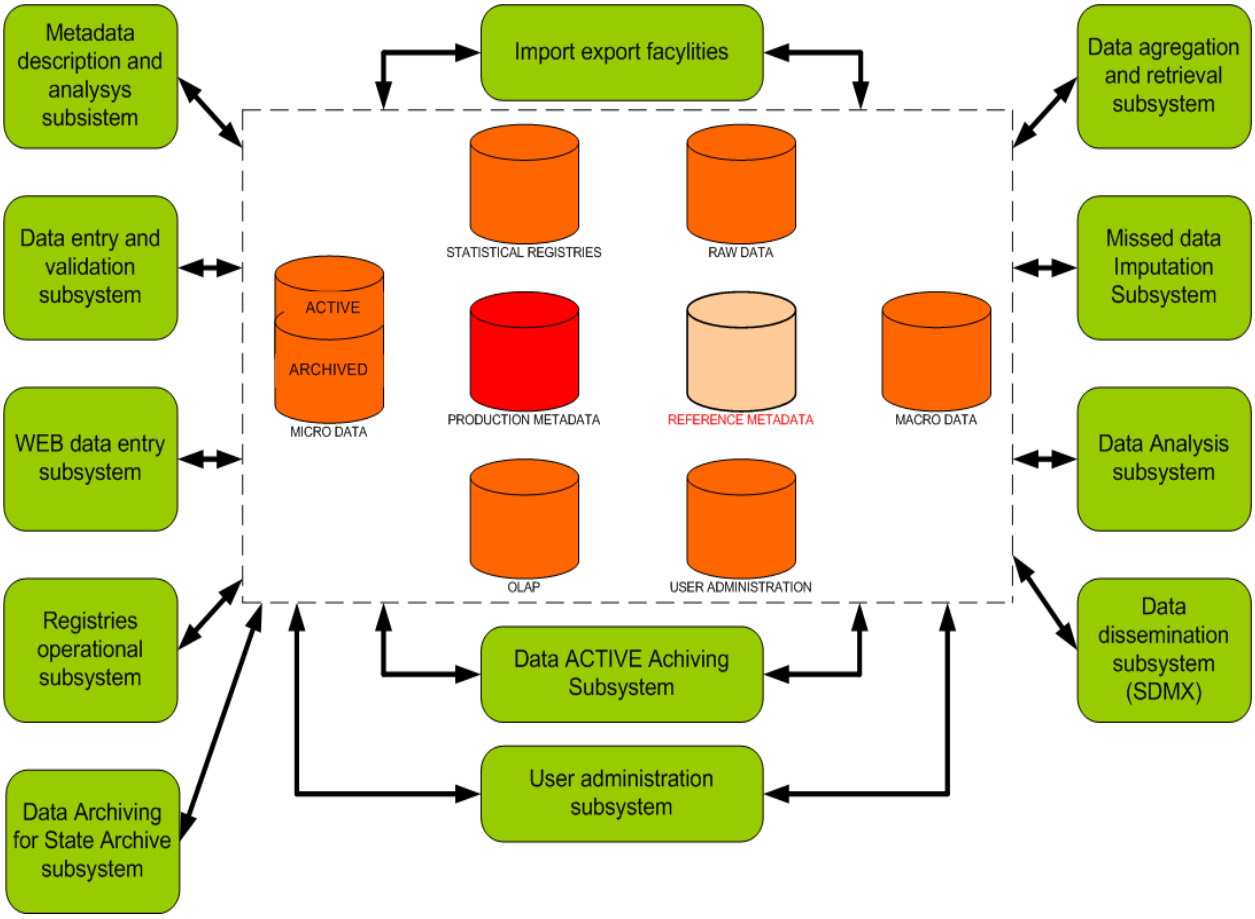
Important task during design of the system was to foresee ways and to include necessary interfaces for data export/import to/from already developed standard statistical data processing software packages and other generalized software available on market, which functionality was irrational to recode and include as the system component.

System is divided into following business application software modules, which have to cover and to support all phases of the statistical data processing:

- Core Meta data base module;
- Registers module;
- Data checking, editing and derivation module;
- Missing data imputation module;
- WEB based data collection module;
- Data aggregation module;
- Data analysis module;
- Data dissemination module;
- User administration module.

System architecture is represented in Figure 5.

Figure 5. ISDMS as corporative data warehouse of CSB



4. Core Metadata base module

The Core Meta data base module is one of the main parts of the new system and can be considered as the core of the system. Meta data base data handled by this module are used by all other modules of the system.

The data in the Meta data base, in essence, is information about Micro and Macro data i.e. description of the numerical data within the statistical production process and the real world meaning of this numerical data. Also the system Meta data base contain description of statistical surveys itself, their content and layout, description of validation, aggregation and reports preparation rules.

The system ensure that Meta data base is used not only as knowledge base for statisticians and users, but also as the key element for the creating universal, common, programming-free approach for different statistical surveys data processing instead of development of software specially for certain survey, where every change of the survey will require a corresponding adaptation of the programs source code, and where it will be also necessary to develop new software for every future survey.

System users can easy query necessary data form Micro data /Macro data bases navigating via Meta data base. Meta data are widely used for data analysis and dissemination.

4.1 Structure of Micro data (observation data) [Bo Sundgren model]

Objects characteristics:

$$C_0 = O(t).V(t) \quad (1)$$

where: **O** - is an object type; **V** - is a variable; **t** - is a time parameter. Every results of observations is a value of variable (data element) – C_0

All variable values have object (respondent) requisites added, which can be called vectors or dimensions. By analysing all the respondents' population, these dimensions are used for creating different groupings and for data aggregation.

In business statistics the following respondents requisites (vectors) for example can be added to each value of variable:

- Main kind of Activities (NACE classification);
- Kind of Ownership and Entrepreneurship (IUFIK classification)
- Regional location (Regional classification - ATVK)
- Employees group classification
- Turnover group classification.

4.2 Structure of Macro data (statistics)

Macro data are the result of estimations (aggregations). The estimations are made on the basis of a set of Micro data.

Statistical characteristics:

$$C_S = O(t).V(t).f \quad (2)$$

where: **O** and **V** - is an object characteristics; **t** - is a time parameter, **f** – is an aggregation function (sum, count, average, etc) summarizing the true values of **V(t)** for the objects in **O(t).**

The structure for Macro data is referred in Meta data base to as box structure or "**alfa-beta-gamma-tau** " structure.

For data interchange **alfa** refers to the selection property of objects **(O)**, **beta** – summarized values of variables **(V)**, **gamma** – cross classifying variables, **tau** – time parameters **(t)**.

4.3 Structure of Surveys (questionnaires)

New survey should be registered in the System. For each survey a questionnaire version should be created, which is valid for at least one year. If questionnaire content and/or layout do not change, then current version and it description in Meta data base is usable for next year.

Each survey contains one or more data entry tables or chapters (data matrix), which could be constant table - with fixed number of rows and columns or table with variable number of rows or columns.

For each chapter it's necessary to describe rows and columns with their codes and names in the Meta data base. This information is necessary for automatic data entry application generation, data validation e.t.c.

Last step in the questionnaire content and layout description is cells formation. Cells are the smallest data unit in survey data processing. Cells are created as combination of row and column from survey version side and variable from indicators and attributes side.

As an example of the fixed structure table on Figure 6. we could look at Retail Trade Statistics Questionnaire structure from Meta data point of view

Figure 6. Example of the fixed structure table from Meta data point of view

Structure of trade statistics questionnaire (data matrix - fixed table)

Name of Questionnaire, index, code, corroboration date, Nr.
Respondents (object) code, name and address;
Period (year, quarter, month)
Name of chapter

INDICATOR 1 + ATTRIBUTE

Metadata repository: common table of statistical indicators, table of attributes (classifications) and table of created variables

Goods and commodity groups	Row code	Total turnover (∑ 2,3,4)	Retail trade turnover	Public catering turnover	Wholesale trade
			Indicators →		
A	B	1	2	3	4
Goods, in total (∑ 2010, 2020, 2030-2190)	2000	*15000*	*9000*	*5000*	*1000*
Food products (except alcoholic beverages and tobacco goods)	2010	CELL VARIABLE 1 [2010,1]	*5600*	*6000*	*400*
Alcoholic beverages, in total	2020	*3000*	*2000*	*400*	*600*
of which: spirits and liqueurs, whisky, long drinks	2021	*500*	*300*	*100*	*100*
wines	2022	*1000*	*500*	*200*	*300*

Attributes ↑

4.4 Creating of variables

INDICATOR + ATRIBUTE (Classification) = VARIABLE

ATRIBUTES = dimensions or vectors of INDICATORS
Vectors always are classifications and they could be as follows:

- Kind of activity – NACE
- Ownership and entrepreneurship – IUFIK
- Territory and etc.

Example:

Number of employees + no attribute	= Number of employees total
+ kind of activity (NACE)	= Number of employees in breakdown by kind of activities
+ location (Territory classification)	= Number of employees in breakdown by territories

The creation of the core Meta data base structure model and Micro data / Macro data base structure model was the primary task of the system development. Meta data base is linked at database structure model level with Micro data base and Macro data base (see figure 7). Correct and carefully deliberative databases structure model design is the basis for successful further system development and implementation.
All survey values from questionnaires are stored in Micro data base and each value have relation to cell (from Meta data base), which describes value meaning. Also each value in Micro data base has additional information about respondent, which gives current value and time period. The same situation is in Macro data base, where are stored aggregated values. Each aggregated value has reference to cell (from Meta data base), reference to each value aggregation conditions (from Meta data base) and correspondent time period.

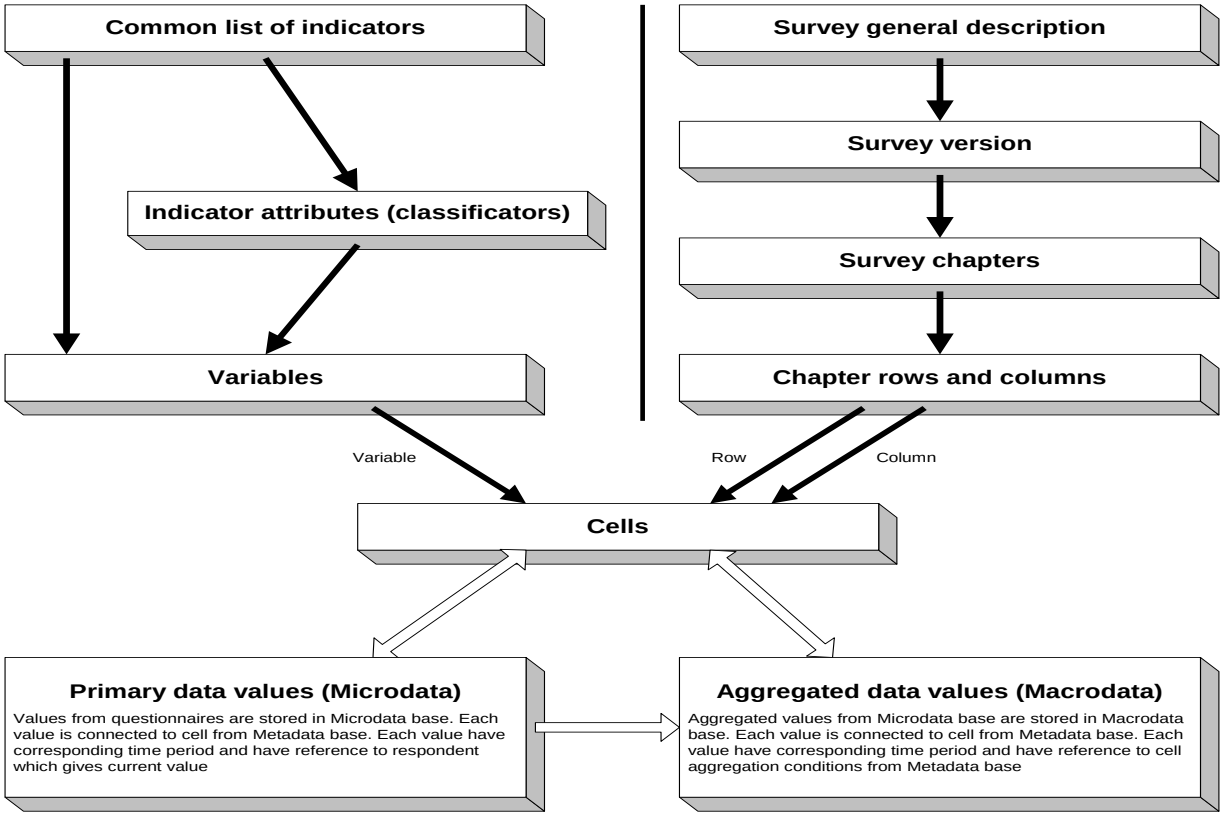
Meta data base module contains following main applications:

- General description of statistical survey;
- Description of survey version;
- Description of indicators and attributes of statistical survey;
- Description of content of statistical survey chapters;
- Maintenance of validation rules of statistical survey;
- Description of aggregation conditions of statistical survey;
- Grouping of classifications records;
- Description of reports;
- Common Meta data base data browsing;

A specially trained personnel (4 persons) of the Section of Meta data Maintenance under umbrella of Information Technologies Department operates the Meta data base module. They have rights to perform Meta data entry, actualisation, changing and are responsible for accurateness of Meta data. It is very important, that Meta data entered into Meta data base are carefully checked and correct, because these Meta data are used for automatic generation of data entry applications, validation, aggregation, reports preparation procedures as well as during data conversion for OLAP and PC-AXIS needs.

Meta data base link with Micro data and Macro data bases are showed on the Figure 7.

Figure 7. Meta data base link withMicro data /Macro data bases



5. Registers module

Within the system at the time being are integrated two statistical registers: Business Register and Intrastat Register. Both registers is stored in separate databases. Maintenance of the registers is stand alone function. Both registers are linked with the entire system trough the respondents lists (samples) which would be considered as the main output from statistical registers. Respondent's

data from the registers are available for statisticians working with Data Entry and Validation Module.

6. Data validation, editing and derivation module

Module provides standardised approach to different statistical surveys data processing. For surveys data processing is implemented automated data entry forms generation, automated generation of data validation, aggregation and reports creation procedures by using information entered into Meta data base.
If it is necessary to make any changes for survey content and/or layout, then it is necessary only to change survey description in Meta data base.
Such approach makes much more easily system maintenance and reduces costs. It is not necessary in case of changes re-write program codes, which can be done only by IT professionals.
For each selected survey for selected period following main functionality is available:

- respondents list maintenance;
- Data entry and validation;
- Editing;
- Derivation;
- Data aggregation;
- Reports creation;
- Data export/import.

The system ensures linking of the statistical survey to a particular list of respondents obtained from Business register. Each survey version for each period has its own list of respondents which is necessary to maintain.
Data entry and validation application is displayed on Figure 8.
Data entry screen shows selected respondent survey chapters and allows select, enter, update and delete data. There are three types of survey chapters:

- With fixed rows and columns number;
- With fixed rows and variable columns number;
- With fixed columns and variable rows number.

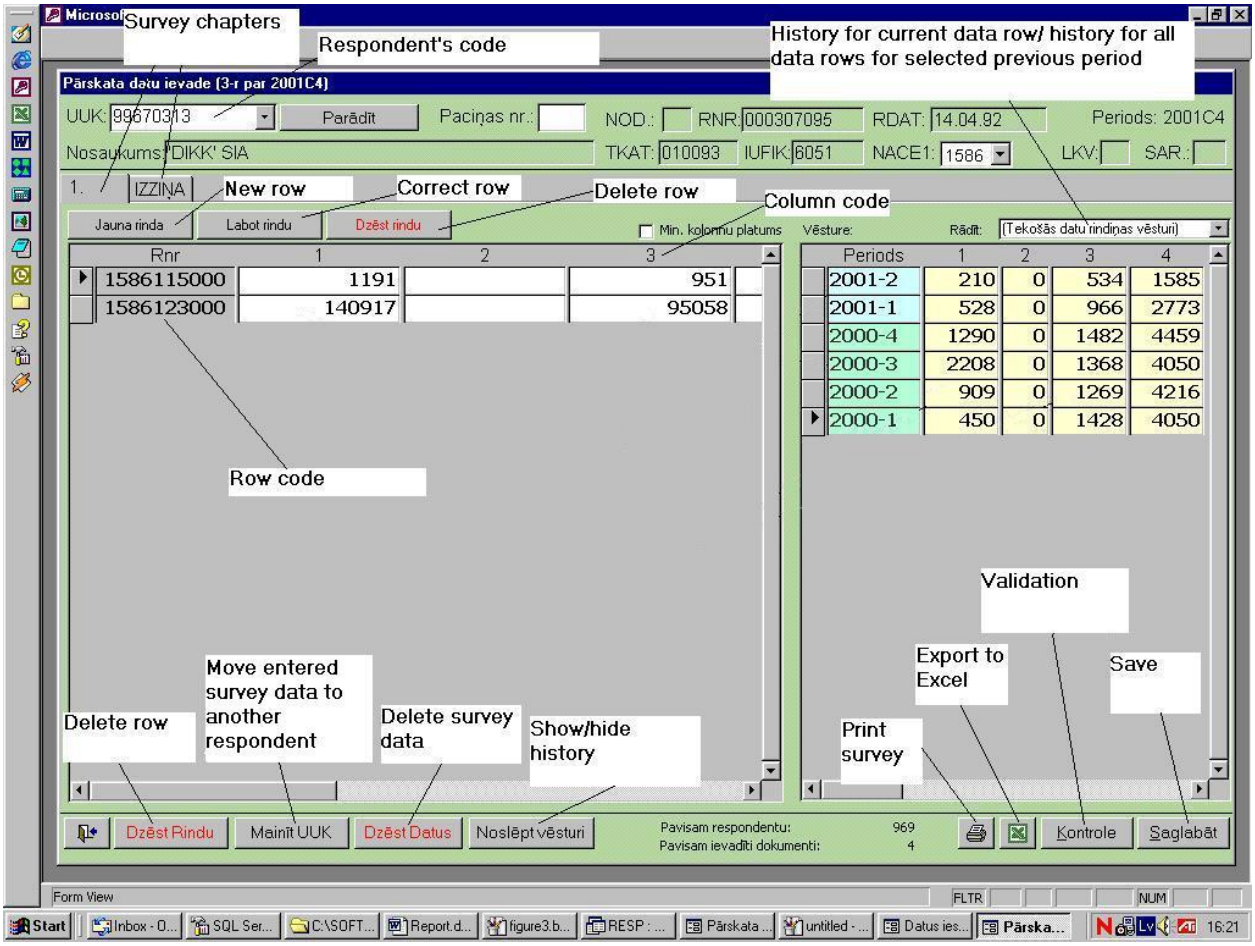
For example, if we have survey with chapter where is several indicators in chapter rows and as indicators attributes are used values from NACE classification in chapter columns, then we have chapter with fixed rows and variable columns number, because each enterprise can have different number of economic activities, which are classified in NACE classification.
Application manually allows adding or deleting rows or columns, when it is necessary.
Historical data (previous period's data) can be displayed in two ways:

- Historical data of all previous survey periods for currently selected row in data grid;
- Historical data for all data rows at the same time for one of selected previous periods.

It is possible to hide historical data section to increase screen space for data entry and later again to show it.
After executing of validation procedure of respondent survey data (we can run this procedure from data entry and validation application) opens form with list of validation errors. In this form for each error is following information:

- Error number;
- Error description;
- Error type:

Figure 8. Data entry and validation application (example with varying number of rows – Prodcom classification)



- Acceptable – error can be marked with “A”;
- Unacceptable – error is marked with “X” and its type cannot be changed (defined in Meta data base);
- Not marked.

- Error type reason – for some errors it is necessary to describe the acceptability reason.

Errors marked with “X“are critical and there is no way to pass them during validation process. All other errors in the list of validation errors can be marked with letter “A”. Errors marked with “A” will not appear in error list next time (to see in errors list also acceptable errors checkbox “show all errors” must be marked). These are less important errors in surveys, which can be passed as not significant for further survey data processing.

Also it is possible to run validation procedure for all respondents of one survey. Then we receive list of respondents, which surveys data have validation errors.

To define exceptions in validation rules defined in Meta data base two additional applications were created, which interacts with data validation process. First application allows to “switch off” one or more validation rules defined in Meta data base for survey. Then during validation process such validations rules will be not checked. Second application allows to “switch off” one or more survey validation rules defined in Meta data base for concrete respondent and again during validation process such validations rules for this respondent will be not checked.

Survey data for specified period can be exported to fixed format Microsoft Access database file for external use. Also fixed format Microsoft Access database file can be imported into the system.

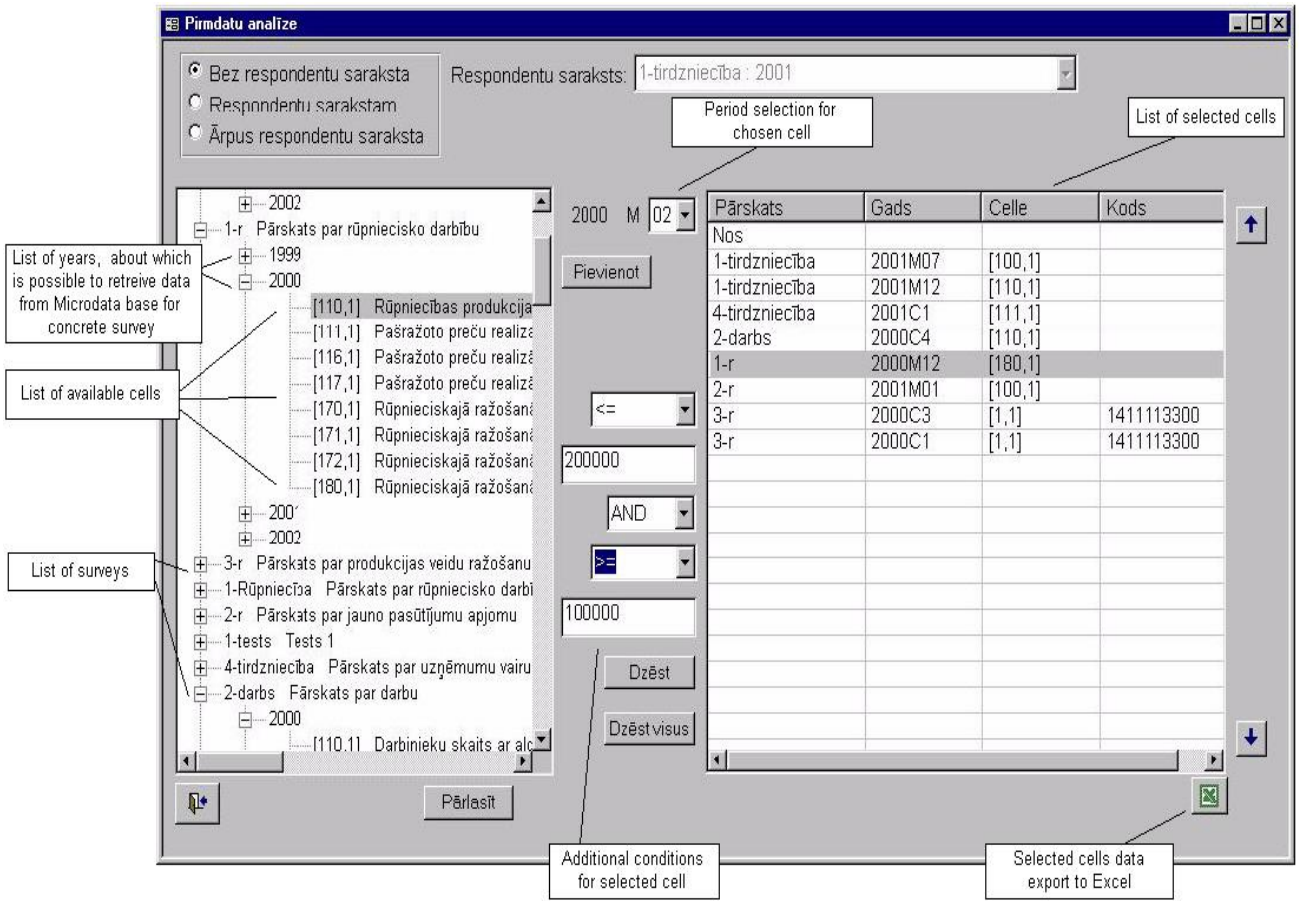
Other features, which are included in functionality:

- To ensure possibility for each department using module applications to work only with their set of respondents, in application is included facility to set filter, which select necessary set of respondents for department;

- For safety needs was included functionality for data locking. When it is necessary, responsible statistician can lock survey data, which forbid further data changes. Before data can be changed it is necessary to unlock survey data;
- Statisticians can create lists, which helps to follow to surveys data input process. There are there types of lists:
    - List of respondents who have sent in survey;
    - List of respondents who have not sent in survey;
    - List of respondents, which survey has one or more errors after validation.

From Data entry and validation module users can create reports using report preparation description form Meta data base. Also from Data entry and validation module users can run application, which is used to maintain aggregated data and another application for common Macro data analysis (see description of these applications in the next section of report – "Data aggregation module").
Additionally in Data entry and validation module is included application for common Micro data base analysis navigating via Meta data (see Figure 9). For end-users it is not necessary to know where and how Micro data are physically stored in Micro data base.
Using this application, users can select and combine data from different surveys for different time periods and to define different conditions for data, which should be queried. All selected data will be exported to Microsoft Excel.

Figure 9 Application for common Micro data base analysis



7. Electronic data collection module

Electronic survey data collection module is based on WEB data entry applications, including survey design and preparation, special data validation algorithms, automatic request sending, response check and management.

Traditional methods of data collection currently at CSB are based on paper questionnaires, which can be completed by interviewers or respondents, who receives a paper questionnaire via post, which has to be completed and returned to CSB.
Current CSB bottleneck during processing of paper form questionnaires, which received from respondents, are:

- Management of requests sending to respondents via post;
- Incoming paper forms data initial analysis and validation;
- Data retyping into CSB Data management system;
- Response control and management.

The core element for Electronic Data Collection Module is WEB based data entry and validation forms that is used for different surveys. These features are available to respondents and can replace ordinary paper questionnaires. Responses (completed forms) are transferred to the CSB through the Internet. Certain features of electronic surveys contribute to increasing of data quality data can be checked immediately.
The business and design goals for Electronic survey data collection module were:

- Design and preparation of electronic surveys in automated mode;
- Collecting questionnaire data in electronic form from respondents;
- To improve the quality of collected data by using on-line validation rules that were missing in paper questionnaires;
- Automatic request sending to respondents and automatic response control, implementation of reminders system to respondents;

The system will remove unnecessary stage of retyping of information by CSB personal.
The Electronic data collection software module key features are:

- Statisticians use the same design tool for the design of paper questionnaires as well as for WEB forms;
- Use of the Meta data provides universal approach for generation of the WEB based data entry forms, which can be done without participation of IT professionals;
- Management system for WEB forms is created, including version control;
- Response control application allows defining automatic reminders/requests sending timetable and checking response;
- Software module ensures registration of respondents and defines detailed access rights for them;
- WEB forms offer following new features for the respondents:
    - Pre-loaded data: respondent or survey specific data (e.g. respondent's name and address);
    - Feedback data: historical data is available;
    - Auto-fill in fields: some fields can be automatically filled in, depending on values of previously completed fields;
    - Auto calculation: columns or other fields can, for example, be summarised.
    - Automatic validation: WEB forms could include validation rules;
    - For periodical surveys in WEB based applications respondent have a possibility to see and prove previous periods data previous periods data;
    - During data entry process in-form validation could be provided;
    - Where necessary respondents are able to search and use classifications such as NACE, PRODCOM, etc;
    - To see the list of surveys to be completed both electronically and on the paper as well as information about reminders;
    - The respondent is able to print questionnaires, for internal use;
    - Help facilities.
- System provides means of security by using user access rights control and data transfer via HTTPS protocol;

- CSB collects all survey data in the raw data base, ensure transfer of the data to the Micro data base to continue processing in the system.
- Subsystem work is ensured with clients using Netscape Navigator and FireFox and MS Internet Explorer.
- It is possible to upload the subsystem with prepared in advance exact format files containing the survey data
- There are ensured possibility to export data from e-Questionnaire to MS Access.

8. Software for missed data imputation

Non-response in statistical surveys can never be totally prevented, but it can be considerably reduced. To achieve this, an optimal data collection design is necessary. The optimisation of the survey design is one of the main activities for reducing the number of missing data. Nevertheless, more modern strategies to cope with missing data are based on use of modern and user-friendly software for missing data imputation.
In case of our system approach is as follows:

- Extraction of the data files with missing data;
- Using of standard software packages for missing data imputation like WIDE, SOLAS or SPSS ensure imputation;
- Import of the data files/supporting Meta information back into Micro data base/Meta data base.

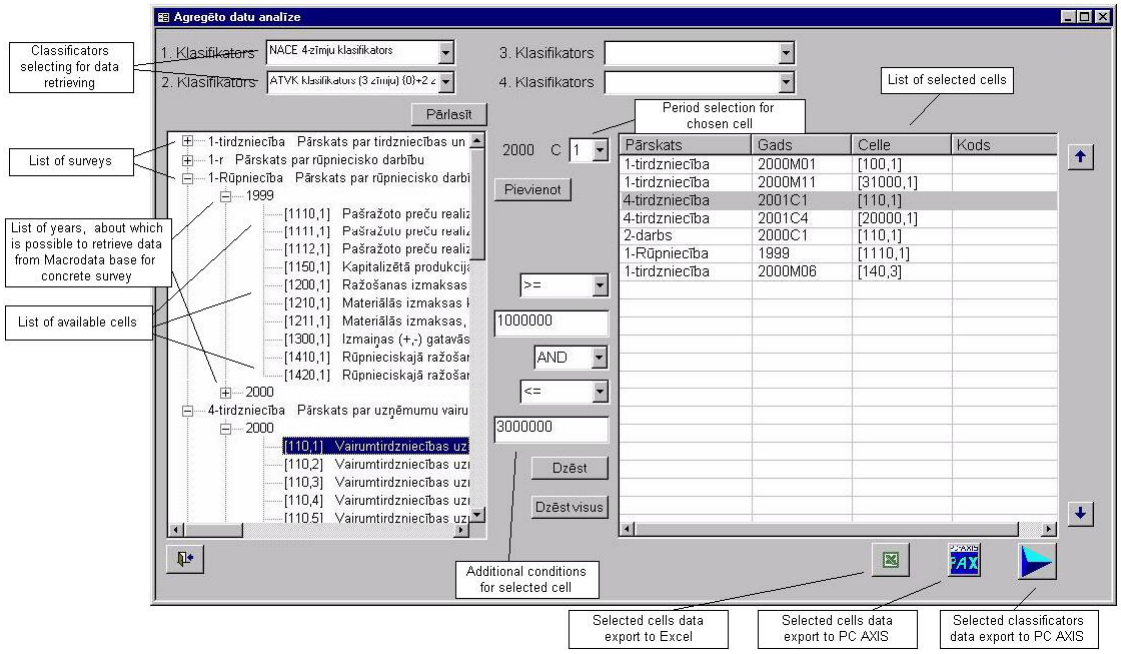
9. Data aggregation module

As it was mentioned before from Data entry and validation module users can run application, which is used to maintain aggregated data. For each survey version for selected period is possible to store several aggregated data versions in the system. Using mentioned application users maintain aggregated data versions, i.e., create new versions or delete unnecessary versions of aggregated data. When users create new version of aggregated data, application analyze Meta data information about current survey version aggregation conditions and create data aggregation procedure, which after running store aggregated data in Macro data base. If an exact survey for selected time period has several data aggregation versions, then one of them must be selected as active. Active aggregated data version is used in all applications, which works with aggregated data (reports creation, aggregated data analysis).
From Data entry and validation module users can also run application for common Macro data base analysis (see Figure 10). Using this application, users can extract from Macro data base any data, which they need. Data selection starts from classifications selection. Then application display whole surveys list, which have aggregated data for selected classifications. For each survey it is possible to see years when aggregation has been done as well as to see cells list for current survey version. Using this application, users can select and combine aggregated data from different surveys for different time periods and to add additional selection conditions. All selected data will be exported to Microsoft Excel. Selected data also it is possible to convert to PC AXIS file format for dissemination purposes.

10. Data analysis module

Data analysis module is based on On-Line Analytical Processing (OLAP) techniques. OLAP is a set of technologies that takes data in a data warehouse and transforms that data into multidimensional structures, called, cubes, to allow for better response to complex queries.

Figure 10. Common Macro data base analysis application



Data analyses module is realized using Microsoft SQL 2000 component – Analyses Service for multidimensional data cube formation and storing and include facility for multi-dimensional statistical data analysis.

Work with Data analysis module we can divide in two main parts:

- Statistical surveys data preparation for analysis;
- Browsing and analysis of surveys data.

The first part, "statistical surveys data preparation for analysis", is performed by data administrators, who are familiar with existing surveys and well introduced in Meta data base, Macro data base data structures. This part is divided in two steps:

- Data transformation form Meta data base and Macro data base using Microsoft Data transformation service to special data structure format, which is required for Microsoft SQL Analyses service;
- Formations of multidimensional cubes, which administrators create with Microsoft Analyses service manager.

The second part, "browsing and analysis of surveys data", will be done by the system end-users using Microsoft Excel 2000 component – PivotTable, which allow easy to view, rearrange, regroup data in different ways. It is possible to use also other Microsoft Excel component – PivotChart, which allow creating diagrams from PivotTable data.

11. Data dissemination module

Data Dissemination functionality embedded in the Dissemination module are as follows:

- System make available publication of statistical data in the WEB environment ;
- Publication tool –PC-AXIS;
- Publications are performed from PC-AXIS files, not directly from the system data bases;
- PC-AXIS files are prepared by special application extracting and transforming data from Meta data base /Macro data base (changing formats from SQL to PC-AXIS);
- Data for publication has passed confidentiality checks;
- Data for publication are grouped by themes.

As basis for Data dissemination module is used standard software PC-AXIS, developed by Statistics Sweden. Module provides functionality for Meta data and Macro data conversion from Microsoft SQL databases (Meta data base and Macro data base) format to PC-AXIS format.

Data for conversion to PC-AXIS format can be selected from aggregated data using previous described application for aggregated data analysis or can be selected from stored reports.
When data are converted, it is easy for administrator to publicise these data in CSB WEB page using PC-AXIS family tool PX-WEB. Publicised data are not just static data tables – users can interactively, using PC-AXIS tools, select necessary data by setting up different data query, grouping and aggregation conditions.

12. User administration module

User's administration module was designed and developed with the aim to provide easiest way for the system user's administration. It makes available information registration on all activities with system applications. There are lists of roles for users to support correct functioning and data security. For each application is possible to specify a separate application administrator.

System application administrators are responsible for granting/cancelling rights to users or user groups and setting up the functions for each exact user group to be available.

13. System profile at the time being

System implemented in August 2002 on predefined 25 different statistical surveys.
Successful implementation formed bases for the CSB regional restructuring, which has been implemented within the period of two years from 2003 to 2004. Five Data Collection and processing centres replaced previously existing 26 Statistical Regional offices an city Riga office thus taking on responsibility for overall data collection and editing and decreasing amount of necessary statisticians working with data collection and editing from 180 to 115.

System at the time being we can characterize with the indicators listed below:
Amount of Surveys:

- described in Metadata base from 2002 130
- in active use in 2009 71
- Available for electronic submission 59

Classifications in Metadata base:

- For systematic usage available 171

Rate of electronically submitted data:

- Maximal for supplied surveys up to 57%
- Average in 2008 23%

14. Vision and future expectations

After the seven years of the exploitation of the ISDMS we found that system functionality would be reasonably increased. With some restructuring of the existing version of the Meta data base and Meta data related application set we would expand system functionality to process Social statistics data. The most difference between business statistics and social statistics are in the data collection and data aggregation methods used.
There are elaborated technical requirements for the CAPI, CATI and CAWI technologies embedding within the system thus allowing replacing hard code programming in Blaise with the metadata descriptions and automatic data entering forms generation as it is done for Economic statistics.

Another direction for the further improvement of the system is adjustment to SDMX standard for data and meta data exchange.

15. Conclusions

- Design of the new information system should be based on the results of deep analysis of the statistical processes, data flows and user requirements;
- Clear objectives of achievements have to be set up, discussed and approved by all parties involved: Statisticians, IT personal, Administration;
- As the result of case study we clearly understood, that all steps of statistical data processing for different surveys allows standardization, while each survey may require complementary functionality (non standard procedures), which is necessary just for this exact survey data processing;
- For solving problems with the non-standard procedures interfaces for data export/import to/from system has been developed to ensure use of the standard statistical data processing software packages and other generalized software available in market,
- Within the process of the design and implementation of Meta data driven integrated statistical information system both parties statisticians and IT specialists should be involved from the very beginning,
- Clear division of the tasks and responsibilities between statisticians and IT personal is the key point to achieve successful implementation,
- Both parties have to have clear understanding of all statistical processes, which will be covered by the system, as well as Meta data meaning and role within the system from production and user sides,
- Motivation of the statisticians to move from existing to the new data processing environment is essential,
- Improvement of knowledge about Meta data is one of the most important tasks through out of the all process of the design and implementation phases of the project,
- To achieve the best performance of the entire system it is important to organize the execution of the statistical processes in the right sequence,

Summing up improvement goals and IT strategy realised in the system, there are mainly the following targets achieved by the system implementation:

- Increased quality of data, processes and output;
- Integration instead of fragmentation on organisational and IT level;
- Reduced redundant activities, structures and technical solutions wherever integration can cause more effective results;
- More efficient use and availability of statistical data by using common data warehouse;
- Users provided (statistics users, statistics producers, statistics designers, statistics managers) with adequate, flexible applications at their specific work places;
- Tedious and time consuming tasks replaced by value-added activities through an more effective use of the IT infrastructure;
- Using Meta data as the general principle of data processing;
- Use electronic data collection and dissemination;
- Making extensive use of a flexible database management for providing internal and external users with high performance, confidentially and security.

References

- Guidelines for the Modelling of Statistical Data and Meta data. Conference of European Statisticians, Methodological Material

- Sundgren B. (1999).An information systems architecture for national and international statistical organizations. Meeting on the Management of Statistical Information Technology (Geneva, Switzerland, 15-17 February 1999);
- Sundgren B. (2003). Developing and implementing statistical metadata systems. Metanet WG 3 Deliverable
- Sundgren B. (2003). Statistical Metadata and Data Warehouse. Metaware IST-1999-12583
- Terminology on Statistical Meta data. Conference of European Statisticians, Statistical Standards and Studies No 53.
- Willeboordse A. Towards a New Statistics Netherlands. Blueprint for a process oriented organisational structure.