INNOVATION IN DATA COLLECTION AND MANAGEMENT IN STATISTICS LITHUANIA

INTRODUCTION

The strategy of Statistics Lithuania for 2008-2012 spares sufficient attention to the management and development of information and communication technology. The main objective is to interconnect information systems of Statistics Lithuania into a flexible and safe information infrastructure, providing conditions for better management and more rational use of processes and resources.

Aiming at implementation of this objective, the basic goals of the strategic period are the introduction of an Integrated Statistical Information System (ISIS), use of ICT potential for improvement of public services and strengthening the relations with users of statistical information and respondents.

With further introduction of ISIS, it is foreseen to completely interconnect partial processes of statistical information development: collection of statistical data, processing and dissemination of statistical information. Data flows between these processes will be fully automated. With gradual transfer of all statistical survey data into the consolidated ISIS databases, newly introduced application systems will be integrated into these databases.

A new Integrated Statistical Information System (ISIS) was developed and implemented at Statistics Lithuania in 2007. The key drivers for the system development were:

- a mix of centralized and custom-developed IT solutions for statistical systems;
- obstacles in merging data from different surveys;
- efficiencies to be gained from standardized software (especially for common statistical processes);
- need to use metadata and common classifications ;
- efficiencies to be gained from modern technologies.

The following objects were proposed in various areas:

Statistical surveys area: to standardize the survey preparation, data collection, processing and dissemination processes; to minimize the number of developed custom survey applications; to improve the quality and to reduce the costs and time of statistical data processing; to improve the quality of collected statistical data; to provide opportunities for respondents to submit data in various forms and ways; to use administrative data more efficiently.

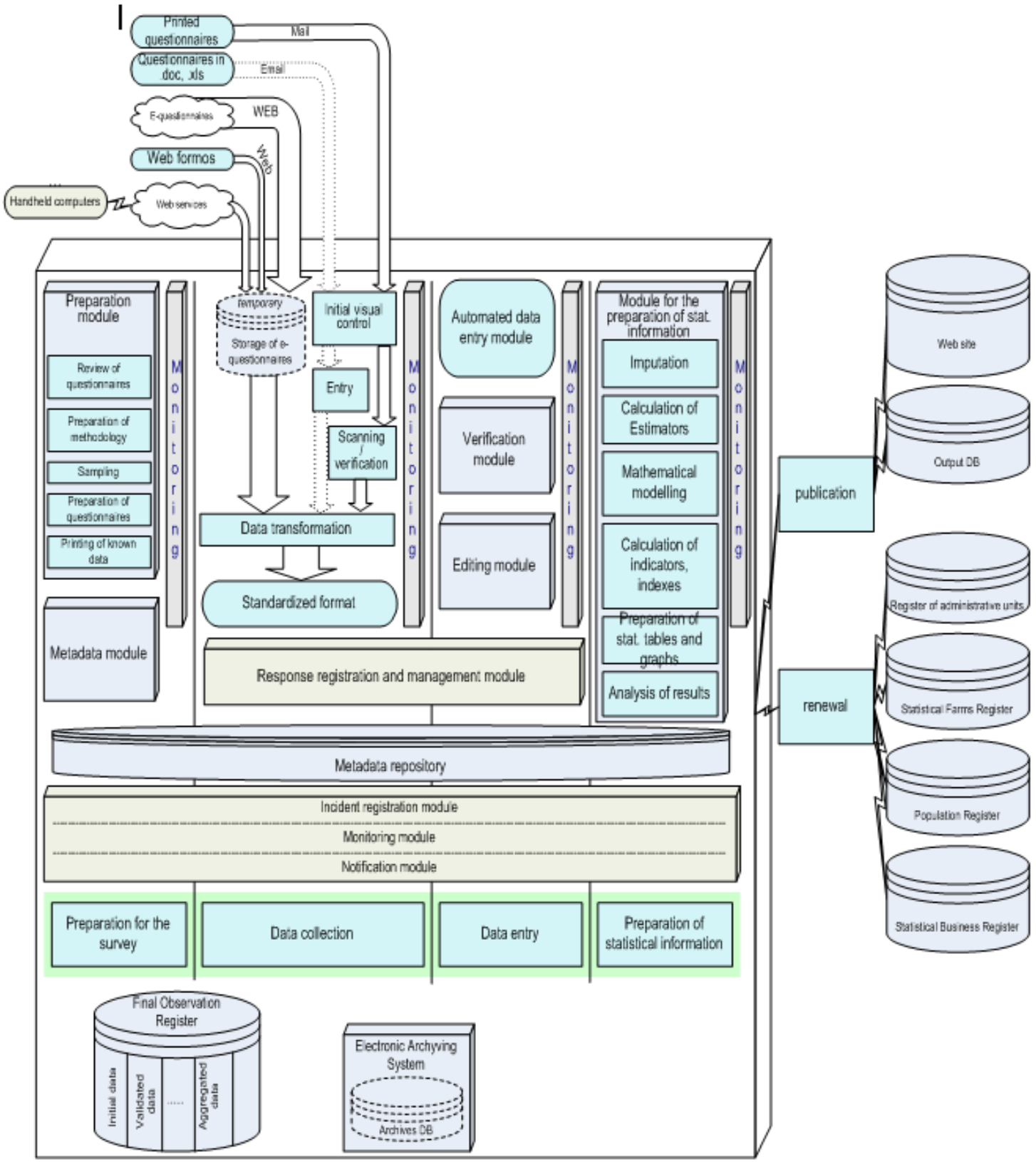
Communication, improvement of staff activity and management area: to ensure the means of monitoring; to create conditions for incident management and analysis; to facilitate communication between SL divisions.

Databases improvement area: to improve the accessibility of statistical data and information; to integrate all databases into a common system; to create a common metadata base integrated into the system; to reduce data duplication in the system; to enable saving intermediate states of data; to enable data archiving.

General system functionality improvement area: to provide the means for the staff of regional statistical offices to work with the system; to analyze and evaluate possibilities of the redistribution of human resources; to ensure integrity with the existing components of the system and those

planned to be developed; to develop software satisfying the system's objectives and users' requirements; to implement hardware and software necessary for proper functioning of the ISIS; to implement means to improve data reliability, integrity, confidentiality.

ISIS


Printed questionnaires
Mail
Questionnaires in doc, xls
Email
E-questionnaires
WEB
Web formos
Web
Handheld computers
Web services
Preparation module
Review of questionnaires
Preparation of methodology
Sampling
Preparation of questionnaires
Printing of known data
Monitoring
temporary
Storage of e-questionnaires
Initial visual control
Entry
Scanning / verification
Data transformation
Standardized format
Metadata module
Automated data entry module
Verification module
Editing module
Module for the preparation of stat. information
Imputation
Calculation of Estimators
Mathematical modelling
Calculation of indicators, indexes
Preparation of stat. tables and graphs
Analysis of results
Response registration and management module
Metadata repository
Incident registration module
Monitoring module
Notification module
Preparation for the survey
Data collection
Data entry
Preparation of statistical information
Final Observation Register
Initial data
Validated data
.....
Aggregated data
Electronic Archyving System
Archives DB
publication
renewal
Web site
Output DB
Register of administrative units
Statistical Farms Register
Population Register
Statistical Business Register

MAIN COMPONENTS

- Core Metadata Repository for Storage of unified metainformation of all statistical surveys (variables, code lists, editing and tabulation rules, etc.);
- Software for metadata management;
- Uniform software for data manipulation using metadata (survey management and operation);
- Interfaces to basic registers and Dissemination DB;
- Interface to Electronic Archive (under development);
- Data quality management tools.

ACHIEVED RESULTS

Survey preparation

Selection/description of variables and code lists: classifications and indicators are managed and used in a centralised way. Needs for change of variables and records classifications are registered in the system. Described variables are used in the data processing process. There is an opportunity to monitor the usage of metadata.

Loading of population/sample records: an opportunity to describe population attributes and their link with code lists. Users can load records themselves; automated control of correspondence to classifications and uniqueness. There is an opportunity to use the same populations/ samples in different surveys.

Preparation of a questionnaire template: users create data entry templates themselves. There is an opportunity to generate templates of e-questionnaires and a possibility to have versions of created templates.

Description of data validation and editing rules: users create data validation and editing rules themselves using the tools embedded in the system. There is an opportunity to form/describe respondents' and questionnaires' packages, allocate tasks to operators.

Description of tabulation rules: users describe tabulation and validation rules themselves. There is an opportunity to use micro- or macrodata of other surveys.

Configuration of Reminders templates: users create texts of reminders themselves and describe the way of sending and schedule thereof.

Survey management/operation

Data collection: received questionnaires are registered (mode and date of submission). There is an opportunity to record the history of communication with the respondent. Respondents receive automated reminders about data submission.

Data entry/import and validation: data import can be performed by users from various formats. All data entry forms of different surveys have a common user interface. There is an opportunity to view the scanned questionnaire and data of other periods or surveys. Users can view data errors while editing history. Data validation can be performed for all or part of microdata.

Imputation: imputation is carried out by users based on the established rules. There is an opportunity to have versions of data.

Data export for further processing: a responsible user can export microdata from the system to various formats.

Monitoring of the response rate: there is an opportunity to view the response rate by sampled cells and other parameters. Users can monitor one respondent's response rate in all surveys where he participates.

Data quality and overall operation state: metadata on data quality and processing stored in a common database. Users have an opportunity to define and form the required quality report. They can analyse obtained information at different levels and in different ways.

Aggregation and preparation for dissemination: for microdata aggregation, the prepared tabulation rules are used. Users can form tables and analyse data. There is an opportunity to import an out-of-system formed table. Users can harmonize code lists of formed tables with code lists of the Data Dissemination System.

DEPLOYMENT

For acceptance testing of the system, 10 surveys were selected: household budget survey, trade and catering enterprises activity and network survey, mortality statistics, survey on retail prices of consumption goods and services, carriage of goods by road transport survey, survey on earnings, survey on insurance enterprises, quarterly survey of main financial indicators of non-financial enterprises, crop area and harvest on farmers' and family farms, labour force survey. Next year it is planned to implement 80 % of all statistical surveys in the new system.