

Distr.
GENERAL

Working Paper No.21
3 April 2008

ENGLISH ONLY

**UNITED NATIONS STATISTICAL COMMISSION and
ECONOMIC COMMISSION FOR EUROPE
CONFERENCE OF EUROPEAN STATISTICIANS**

**EUROPEAN COMMISSION
STATISTICAL OFFICE OF THE
EUROPEAN COMMUNITIES (EUROSTAT)**

**ORGANISATION FOR ECONOMIC COOPERATION
AND DEVELOPMENT (OECD)
STATISTICS DIRECTORATE**

Meeting on the Management of Statistical Information Systems (MSIS 2008)
(Luxembourg, 7-9 April 2008)

Topic (ii): Statistical information systems architecture

METADATA DRIVEN INTEGRATED STATISTICAL INFORMATION SYSTEM OF CSB OF LATVIA

Supporting Paper

Prepared by Karlis Zeila, CSB, Latvia

I. INTRODUCTION

1. The aim of this report is to introduce participants to the Metadata Driven Integrated Information system (further in text - system), which replaces the existing Metadata Driven Integrated Statistical Data Management System (MD ISDMS) of the Central Statistical Bureau of Latvia (CSB), implemented in August 2002.
2. Successful implementation of the MD ISDMS formed the basis for the CSB regional restructuring implemented in the two-year period from 2003-2004 – 5 Data Centres with 115 employees took place of 26 Regional offices with 180 employees. From the time of implementation of MD ISDMS in August 2002 in Metadata base are described 102 surveys; in active use in year 2008 are 77. For electronic submission are available 57 surveys. E-surveys are integral part of all surveys. Designs for them are identical with usual survey. There is a tendency that number of them is growing up. Maximal rate of electronically submitted data for supplied surveys goes up to 46% and average in year 2007 – 23, 1%. Classifications for usage in metadata base are 129.

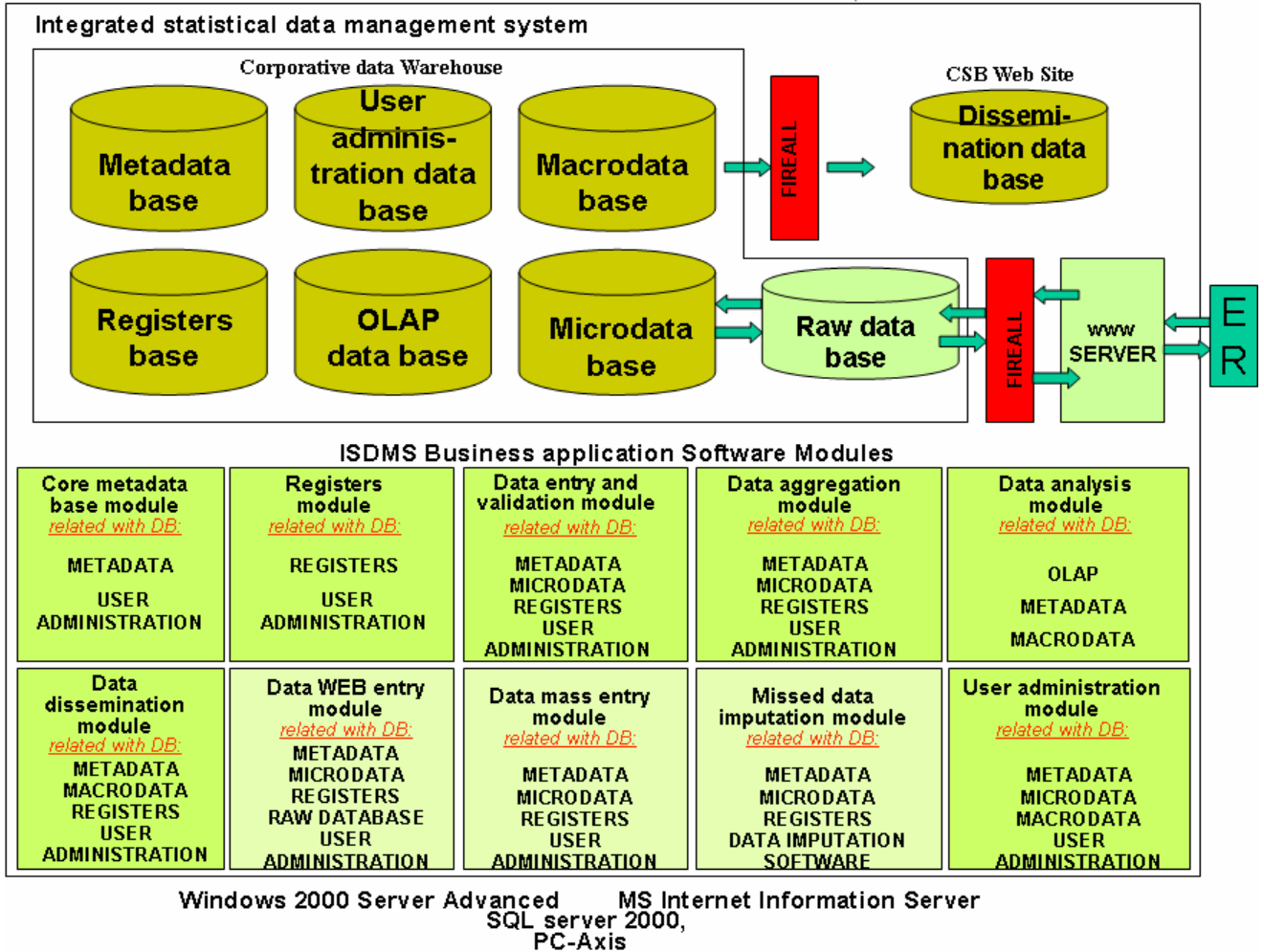
II. OVERVIEW OF THE METADATA DRIVEN INTEGRATED INFORMATION SYSTEM OF CSB OF LATVIA

A. Architecture of MD ISDMS

3. Integrated statistical data management system has 3 versions. In the first version it had seven modules: Data dissemination module, Core Metadata module, Registers module, Data entry and validation module, Data aggregation module, Data analysis module and User administration module.

4. In the second version there were added 3 new modules - Data Web entry module (e-Survey), Data mass entry module and Missed data imputation module. In the latest third version, year 2006, systems structure was not changed. In this version is added new functionality – Meta database in data processing use classificatory as a value or using Professor Bo Sundgren terminology the “ γ ” variable or cross classification variable were introduced MD ISDMS architecture is shown below on the figure.1

Figure 1 Integrated statistical data management system

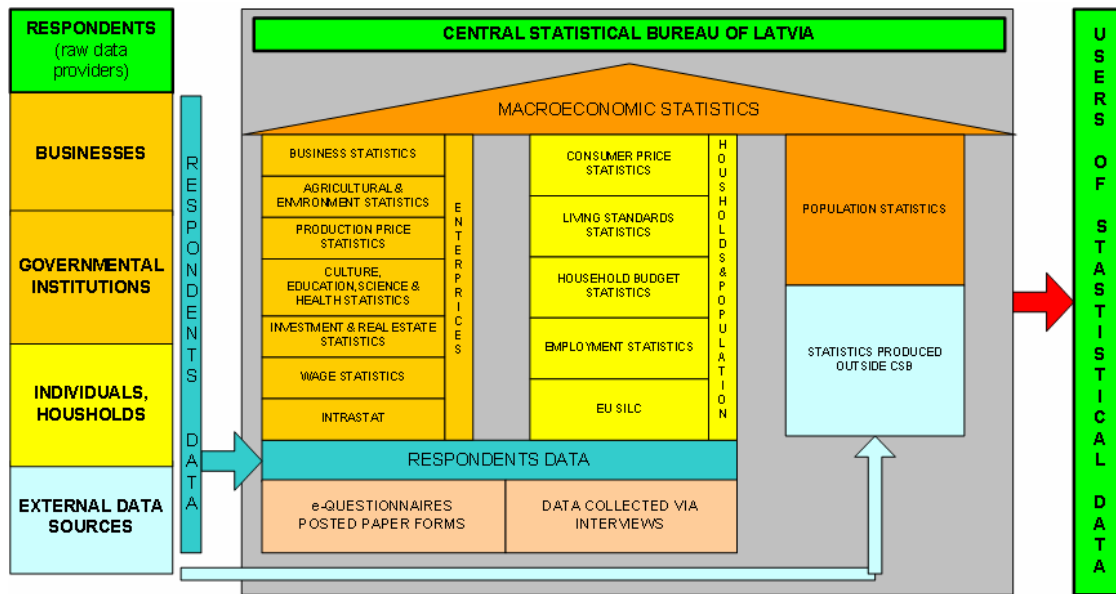


B. Statistic structuring in CSB

5. Statistic structuring in CSB is based on the Quasi Process Oriented data processing approach. This structure in the CSB is a result from implementation of the Metadata Driven Integrated statistical Data Management System and regional restructuring. MD ISDMS at the time being is able to work with the Business statistics surveys. The most difference with the second largest statistical domain - Social statistics is in the way how the primary data are obtained.

6. As it shown on the figure 2, all statistics we can show as the virtual building in which Macroeconomic Statistics lays on the other statistical pillars Business, Social, Demography etc . Collection of primary data is different and in most cases we could distinguish in two - Data collection via posted questionnaires and data obtained via interviews

Figure 2
Statistics Structuring in CSB



B. Reengineering

7. As mentioned in the introduction, the described data range is growing all the time. Major data range is a reason of systems performance decrease. Microsoft SQL Server 2000, which handles system databases, became a weak point of the MD ISDMS from the performance point of view. At the time being we are in the process of MD ISDMS reengineering project in the frames of which we are moving to new stable and approved platform – Microsoft SQL Server 2005. There were found one more solution to increase performance – establishment of active archive storage. Active storage will store data, which is less than 3 years old; the other data will be stored in active archive storage.

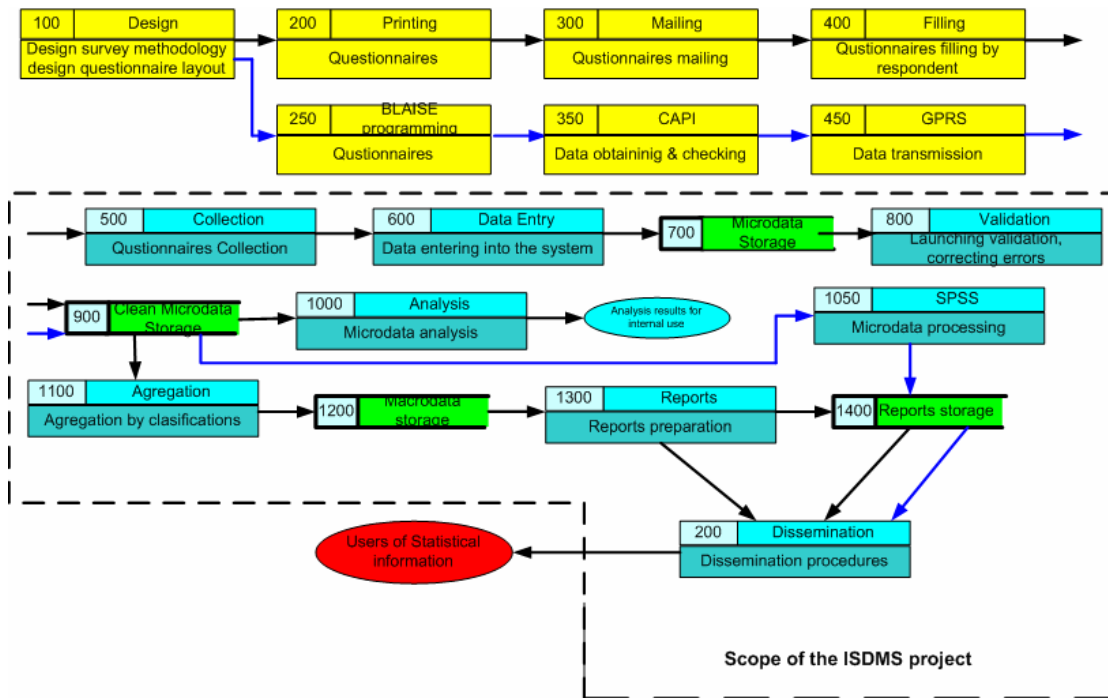
8. There is ongoing realization of MD ISDMS adaptation towards Social statistics data processing, analysis and disseminations, thus moving from MD ISDMS to MDISIS (Metadata Driven Integrated Statistical Information System).

9. Within the frames of the mentioned reengineering project we have to improve usage of different Internet browsing Software packages as number of E-surveys is growing up and many of respondents use Open source browsing software packages.

10. Implementation of SDMX standard requirements within data dissemination processes is a part of ongoing project.

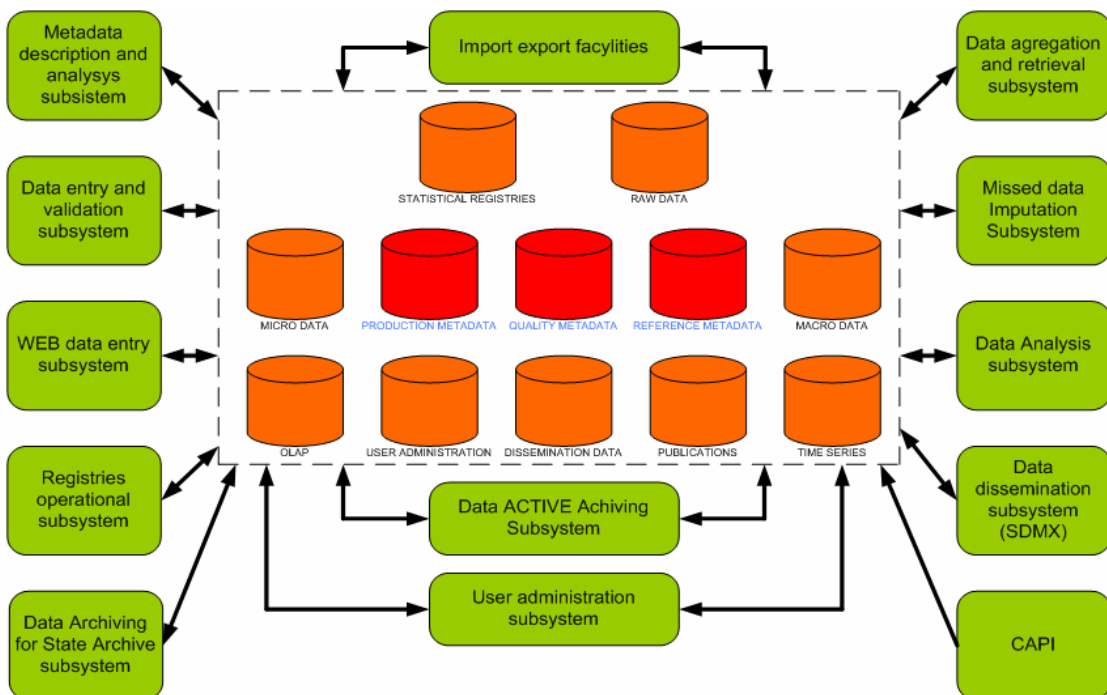
11. Figure 3 shows processes covered by MD ISDMS and Business and Social Statistics Domains. This system development is started and planned to be finished in the second part of this year. There is a not included survey preparation processes like methodology, preparation of survey's form, etc. in the scope of the system. With black lines are shown processes which are realized in business statistics and with blue lines – Social statistics. Business statistics includes processes like Data collection, data entry and validation, but social statistic is included starting with Micro data storage, because the data collection and cleaning is covered by Blasé based CAPI system. System have highly developed data export and import facilities, it support software's which can use Microsoft Excel, Access and Fox Pro data analyzing tools.

Figure 3 Processes covered by the system



12. Figure 4 shows the system structure that will be implemented during 2008-2009. In 2008 existing system functionality is already expanded with data archiving for the state archive subsystem, will be added mentioned above functionality of data active archiving subsystem. The part of the Reference Metadatabase including methodology and quality metadata will be implemented in this year as well. Other part or reference metadata as well as Time Series database will be added to the system in the next step in Year 2009, thus making system compatible with SDMX standard requirements.

Figure 4 MDIIS as CSB Cooperative Data Warehouse



II. NEW MD – INTEGRATED STATISTICAL IS OF CSB OF LATVIA

13. As shown in figure 5, all structure is divided in 3 subsystems – Metadata subsystem, Data processing subsystem ISDMS+ (where is used CAPI, FTP, e-respondents and mail respondents data) and Data and Metadata dissemination subsystem.

14. In the next is planned to introduce and finish working on Reference Metadata systems, improving data collections and processing processes and introducing statistical data and metadata exchange standards.

Figure 5 MD – Integrated IS of CSB of Latvia

