

Distr.
GENERAL

CES/AC.71/2004/11 (Summary)
27 February 2004

Original: ENGLISH

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

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

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

Joint ECE/Eurostat/OECD Meeting on the Management of Statistical Information Systems (MSIS)
(Geneva, 17-19 May 2004)

Topic (ii): Development of IT strategies in statistical offices

METADATA DRIVEN INTEGRATED STATISTICAL DATA MANAGEMENT SYSTEM

Invited Paper

Submitted by the Central Statistical Bureau of Latvia¹

Summary

1. The aim of this paper is to present the experience gained in the development and implementation of the new generation statistical data processing system, which integrates several subsystems, and is metadata driven.
2. In accordance with the Central Statistical Bureau's (CSB) IT Strategy, the new system is a centralized system, where all data are stored in a corporate data warehouse. The new approach is to unite what logically belongs together by using advanced IT tools, to ensure the rationalization, standardization and integration of the statistical data production processes. This in practice implies a movement from a classic stovepipe approach to a process oriented data processing one.
3. As the theoretical basis for the system architecture, the results were taken from metadata studies done and published by Prof. Bo Sundgren, Statistics Sweden. A metadata-driven data processing system is based on the systemization and storage in the centralized meta database of all metadata on surveys, indicators and classifications used, thus ensuring the preconditions for unification and harmonization of the statistical indicators, while avoiding duplication in different surveys. The introduction of a unified system of statistical indicators leads to the expansion of the system functionality and metadata becomes the key element of the entire system.
4. System integration is based on the direct use of the Business Register data for business statistics surveys data processing. The Registers module of the system foresees the use of several different statistical registers. Most of the system software modules are connected with the Registers module.

¹ Prepared by Karlis Zeila (karlis.zeila@csb.gov.lv).

5. All system software modules are connected with the Core Metadata module. We can call it an integrated system because all data are stored in the corporative data warehouse.
6. The system is divided into the following business application software modules, which have to cover and to support all phases of statistical data processing:
 - Core metadata base module;
 - Registers module;
 - Data entry and validation module;
 - Mass data entry module;
 - WEB-based data collection module;
 - Missing data imputation module;
 - Data aggregation module;
 - Data analysis module;
 - Data dissemination module;
 - User administration module.
7. The WEB-based data collection module functions as an extension of the data collection technology for paper based questionnaires. The layout of the e-form is almost the same as the paper version, therefore respondents do not encounter any changes when moving from paper to the screen. In fact, the respondent obtains more advantages submitting data electronically: he can see data submitted in the previous periods, he can make changes within data submitted in previous periods, etc.
8. The main conclusions of the system development and implementation process are as follows:
 - Design of the new information system should be based on the results of a profound analysis of the statistical processes and data flows;
 - Clear objectives of achievements have to be set up, discussed and approved by all parties involved: statisticians, IT personal, administration;
 - As the result of the feasibility 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 for this exact survey data processing;
 - To solve problems with the non-standard procedures interfaces for data export/import to/from, a system has been developed to ensure the use of standard statistical data processing software packages and other generalized software available in market;
 - Within the design and implementation process of a metadata driven integrated statistical information system, both parties - statisticians and IT specialists - should be involved from the very beginning,
 - A clear division of the tasks and responsibilities between statisticians and IT personal is the key to achieving successful implementation;
 - Both parties have to have a clear understanding of all statistical processes, which will be covered by the system, as well as metadata meaning and role within the system from the production and user sides;
 - The initiative to move from a classical stove-pipe production approach to a process-oriented one have to come from the statisticians side not from IT personal or the administration;
 - The motivation of statisticians to move from the existing to new data processing environment is essential;
 - Improvement of knowledge about metadata is one of the most important tasks throughout the entire design and implementation process phases of the project;
 - It is necessary to establish and train a special group of statisticians, which will maintain the meta database and which will be responsible for the accuracy of metadata;
 - For the best performance of the entire system it is important to organize the execution of the statistical processes in the right sequence;
 - For the administration and maintenance of the system, it is necessary to have well-trained IT staff, who are familiar with the MS SQL Server 2000 administration, MS Analysis Service, other MS tools, PC AXIS family products and system Data Model, system applications.

- - - - -