Integration of Multiple Administrative Data Sources and Statistical Surveys – Case of Data Warehouse Implementation at State Statistical Office of Macedonia

Valentina Jankovik Cacorovska
State Statistical Office of Republic of Macedonia, Research and Development Sector
Dame Gruev 4
Skopje, Republic of Macedonia
valec@stat.gov.mk

Lidija Petkovska
State Statistical Office of Republic of Macedonia, Research and Development Sector
Dame Gruev 4
Skopje, Republic of Macedonia
lidijap@stat.gov.mk

1. Introduction

Main tasks of the State Statistical Office of the Republic Macedonia (SSOM) are statistical data gathering, processing, analysis, publishing, data estimations and prognosis and preparation of studies on socio-economic changes in the society. SSOM conducts annually a lot of statistical surveys with different periodicity (monthly, quarterly, annual), Censuses. Besides these tasks, SSOM has also legal obligations for establishing and maintaining Statistical Registers.

2. System of National Accounts (SNA), reforms, automation project

National Accounts are important macro-economic indicator showing the level of development of a country (gross domestic product) and the system which integrates a lot of data in a consistent system of supply-use tables, input-output tables, integrated set of accounts for national economy.

National Accounts Sector is the greatest user of statistical data produced by the Business Statistics Sector and Social Statistics Sector of SSOM, as well as of data from numerous administrative sources.

At present the data are gathered in different formats, on different media and very often are based on incompatible classifications. This situation makes the process of national accounts calculation very complex and imposes the need for additional technical engagement in connection with data re-classification, re-formatting and standardization. All these problems listed above have negative impact on national accounts data quality and timeliness and could be a serious restrictive factor in their future development.

Therefore the automation of national accounts data gathering, processing and analysis is urgent need of SSOM. In connection with the overall process of the reform of the payment system in the Republic Macedonia, this leaded to the definition and implementation of "Project of the SNA automation" for establishing and implementing DW on financial statistics (in order SSOM to continue providing financial statistics data which were provided by Payment Operational Office). This project implementation i.e. automation SNA should be realized in two phases: 1. Automation of financial statistics for which data are mainly gathered from administrative data sources; 2. Integration of basic statistics in the SNA.

3. Project goals

In relation with its obligations in this project related to financial statistics management SSOM should provide: Establishment of Data Warehouse on financial statistics; Standardized data transfer from State Government Organizations – providers of administrative data to Data Warehouse; Automation of financial data processing and analysis; Decision-support system for business analysts at phases of data adjustments and analysis; Automation of the reporting phase.

4. Data warehouse ?

Within the context of a statistical office, in general, a statistical data warehouse can be defined as a single, complete and corporate repository of data and metadata which have been acquired from different sources, assembled, combined to form one structure, documented in a standard format, and stored in a structure that allows users to view, query, combine and download data for analysis at different levels.

To achieve above mentioned goals data warehouse as data storage place should be established, but also should be established and documented the overall data warehousing process in which the enterprise gathers, transforms and loads operational (transactional) data into a separate physical repository optimised for decision support applications.

5. Project solutions on data warehouse organisation and implementation

5.1. History

Up to now employees from the National Accounts Sector have made these complex calculations by using MS Excel and MS Access. Taking in consideration the amount of data and complexity of the processing it is clear that the software used is not suitable for the needed tasks. In the frames of the preparations for the above mentioned project, but especially for the needs of the Annual accounts ' 99 data processing, SAS software was used for data processing, and a

lot of modules of this software are used for data warehouse implementation.

5.2. Data warehouse implementation

Rapid Warehouse Methodology is used, series of steps was defined and followed during the work on the project.

Statistical information system of SSOM enables data warehouse implementation in client/server environment (AIX IBM RISC Servers, Windows NT workstations).

SAS/Data Warehouse Administrator module is used for building of financial statistics data warehouse.

Operational data (from all administrative sources and basic statistics) are in different formats (.txt, SAS, DB2/6000) and stored in groups according the administrative evidence they take the origin from. SAS/Access software to DB2 is used to extract data from DB2 databases. All data sets – source .txt files, as well as source and produced SAS data sets are stored on the AIX servers, and can be accessed by using the SAS/Connect software from the Windows clients.

DW is organised in subjects i.e. logical groups of data sets related to certain topic, area. There is one subject per input data source in our DW organization, because process flows are different for all business subjects (especially at data corrections and analysis). Within subjects, data sets are organised in groups (Data Groups) depending on data source, purpose, format (detail, OLAP), permanency (permanent/intermediate). DW data are kept in relational databases (DB2/6000) and multidimensional databases (SAS/MDDB), which lead to HOLAP (Hybrid On-Line Analytical Processing) solution.

Central role at using of this software have metadata – input sources, targets and processes which describe transformation from inputs to outputs are defined in metadata.

Graphical user interface, which enables schematic view of processes, makes their definition and re-definition easier. User-written code is used for defining the processes, although possibilities of SAS Data Warehouse Administrator software for code generation make the work easier.

Client/server environment organization has many good sides i.e. data can be held on AIX servers and designed and tuned for efficient access; users of the application work within a familiar Windows environment of Windows styles screens; any processing required by the application can be performed on by the client or the server, depending on the complexity and size of the data involved (in our solution all data processing is performed on the servers).

Having constructed such a vast data repository imposes the need for access to data and reports in the most flexible way possible. SAS/EIS module offers just this possibility and was chosen for the starting point of our data exploitation phase. Creation of multidimensional reports by using this software is an easy process which doesn't require any technical expertise (business analysts can prepare reports themselves).

SAS/Insight module graphical user interface is used for interactive data analysis (outlier's), while object-oriented custom SAS/AF applications are designed for decision-support at data analysis and correction phases. Output reports have look and feeling the users are used to, bilingual version (macedonian/english) of the reports is provided.

Implementation of a data warehouse as large and complex as this one, and which involves numerous sources of data necessities the provision of very clear documentation of the process involved.Metadata can be exported in other formats and we hope that this will enable establishment of integrated metadata system.

6. Conclusions/Expected benefits

Our expectations are: Increase of the work efficiency; Improvement of data organization; Establishment of central repository of sources, classification tables and statistical output; Foundation of time series for economic indicators - 1997-2000 at the moment; IT infrastructure providing easy adoption to business changes; Documentation and reuse of the implemented process; Availability of the data for various users.

This leads to the realisation of the basic idea: To shorten time for data preparations and spend more time on data analyses.

REFERENCE

1. Document : "Project : System of National Accounts Automation" - State Statistical Office of R.M;
2. "Implementation of the Data Warehouse and automation of the national accounts" by Biljana Apostolovska and Dimitar Bogov – written for the Meeting "Statistical days 2000", Radenci, R.Slovenia, 13-15/11/2000.
3. Just in time processing as one of the requirements for input data warehouses, by M.H.J.Vucsan, Statistics Netherlands, Joint ECE/Eurostat Meeting on the Management of Statistical Information Technology, 02/2001.
4. Strategic data warehousing principles using SAS software, by Peter R. Welbrock, 1998, SAS Institute.

RESUME

Dans ce document sont expliquées les fonctions fondamentalles de système des comtes nationoux, les états et les problemes de projet d'automatisation de système de comtes nationoux, c'est a dire d'établir un depôt des données de statistiques financiéres, dans le contexte de la reforme de la circulation monétaire dans la Republique de Macedoine. On montre aussi les pincipes generaux d'établissement du depôt des données (de l'aspect d'organisation des données et de l'aspect du procés), aussi que les aspects pratigues de l'organisation et l'execution de depôt des données des statistiques financiéres dans le Bureau statistique de l'Etat de la Republique de Macedoine.