Data Warehouse Implementation in Financial Statistics using SAS at State Statistical Office of Republic of Macedonia

Authors:
Lidija Petkovska (e-mail: lidijap@stat.gov.mk)
Valentina Jankovik (e-mail: valec@stat.gov.mk)

State Statistical Office of Republic of Macedonia
Agenda

- Introduction
- About the project of SNA automation
- Design and implementation of Statistical Data Warehouse
- Exploitation interfaces
- Overview of DW organisation and used SAS modules
- Concluding remarks
- Future plans
Introduction

State Statistical Office of Republic Macedonia (SSOM)

Main tasks:
- data collection and processing
- data analysis and publishing
- data estimations, prognosis and projections
- studies on socio-economic changes in the society

Statistical surveys (~300 conducted annually)

Censuses

Statistical Registers
System of National Accounts (SNA)

- National Accounts - important macro-economic indicator

- National Accounts Sector at SNA implementation uses:
  - data from numerous administrative sources
  - statistical data produced by SSOM

- Complex calculation process
  - data gathered in different formats, on different media, often based on incompatible classifications

- Additional engagement needed
  - for data re-classification, re-formatting, standardisation
Project of SNA automation

- Problems
  - negative impact on NA data quality and timeliness
  - restrictive factor for future development

- Urgent need identified
  - automation of NA data gathering, processing and analysis

=> New project defined and implemented

“Project of Data Warehouse implementation in Financial statistics”
About the Project of SNA automation

Project phases and Data Warehouse content

1. Automation of Financial Statistics for which data are mainly gathered from administrative data sources
   - (end - June 2001)

2. Integration of basic statistics in SNA
   - (June 2001 - April 2002)

Administrative Data Sources

- Payment Operations Office
- Ministry of Finance
- Public Revenue Office
- National Bank
- Health Insurance Fund
- Employment Office
- Pension and Invalidity Insurance Fund
- etc.

Basic Statistics

- Foreign Trade Statistics
- Internal Trade Statistics
- Labour Force Survey
- Household Budget Survey
- Investments Survey
- Industrial Production
- etc.

Data Warehouse Content
Basic idea of the project

To shorten the time of compilation of integrated economical accounts to increase efficiency of business users

Spend less time for data preparation,
Use more time for data analysis
Project goals

- Establishment of **Data Warehouse** on financial statistics
- **Standardized data transfer** from data providers to DW
- **Automation of processes** of financial data transformation, processing, analysis and dissemination
- **Decision Support System** for business analysts
  - at phases of data adjustments and analysis
- Automation of the **reporting phase**
- Establishment of **consistent system of**:  
  - Input-output tables  
  - Supply-use tables  
  - Integrated Economic Accounts
Project team

STATE STATISTICAL OFFICE OF REPUBLIC OF MACEDONIA

SAS

EUREKA Informatika - Skopje
Project methodology

Rapid Warehouse Methodology

- series of steps defined and followed
Designing and implementation of Statistical DW

Data warehouse?

- Statistical Data Warehouse
  - single, complete and corporate *repository* of data and metadata
  - acquired from different sources
  - assembled, combined to form one structure
  - documented in a standard format
  - structured in a way that allows users to
    - view
    - query
    - combine
    - download data for analysis at different levels

- Data Warehouse (storage)  <-->  Data Warehousing Process
- Transactional Systems  <-->  Decision Support Systems
3. **Statistical Information System in SSOM**

### Hardware

1. **Mainframe Unisys environment**
   - Central SSOM building:
     - IBM RISC/6000 Servers - AIX OS
     - Windows NT Servers
     - Windows NT/2000 work-stations

2. **Client/Server environment**

### Software

- **SAS**
- **MS ACCESS**
- **MS Visual Basic 6.0**
- **COBOL**
- **DB2/6000 as RDBMS**

3. **8 Regional Offices:**
   - Windows NT server
   - Windows NT work-stations
SAS as a solution

Yesterday

- MS Excel (MS Access)
- Complex calculations & large amount of data
  => not suitable

Now

- Basic SAS training of NA Sector
  => Base SAS
  SAS/STAT
  SAS/INSIGHT
  SAS/Enterprise Guide
- Project of SNA automation
  - implemented by using SAS
Physical Model of Data Warehouse

<table>
<thead>
<tr>
<th>ZPP Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>ODD</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Dimension Classification Tables</th>
</tr>
</thead>
<tbody>
<tr>
<td>SDW</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Business Definitions</th>
</tr>
</thead>
<tbody>
<tr>
<td>ODD</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Detail Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>SDW</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Summarized Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>SDW</td>
</tr>
</tbody>
</table>

1. Detail tables (business units, sectors)
2. Corrections made and effect on detail level

- Business Units
- Sectors
- Reporting Tables
- Analysis Tables
- MDDB Tables
- Archived Data
Prosess Model of Data Warehouse

Warehouse Environment

- ETL Processes on yearly basis
- Updating dimension tables
- Inserting Adjustments
- Archiving of historic detail data

Input data
- ZPP data
- Dimension Tables
- External data

DW Administrator
- DW Archive

DW Users
- Analysis
- Reporting
- Personal data marts
POO Business subjects:

1. Non-financial enterprises – big
2. Non-financial enterprises - small and medium
3. Users of Budget sources and social funds
4. Budgets of Local Units
5. Others
6. Banks
7. Insurance companies
8. Self-employed workers

Sectors:

1. S11 - Non-financial
2. S12 - Financial
4. S14 - Households
5. S15 – Non-profit inst.
### Naming Conventions and Business Rules

#### Enterprises

<table>
<thead>
<tr>
<th>Short name</th>
<th>Macedonian Name</th>
<th>English Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>BAN</td>
<td>Banki</td>
<td>Banks</td>
</tr>
<tr>
<td>BUX</td>
<td>Buxet</td>
<td>Budget</td>
</tr>
<tr>
<td>BUK</td>
<td>Buxetski korisnici</td>
<td>Users of the Budget</td>
</tr>
<tr>
<td>DRU</td>
<td>Drugi Subjekti</td>
<td>Others</td>
</tr>
<tr>
<td>OSI</td>
<td>Osiguritelni Društva</td>
<td>Insurance</td>
</tr>
<tr>
<td>SAM</td>
<td>Samostojni</td>
<td>Unicorporated</td>
</tr>
<tr>
<td>SGO</td>
<td>Stopanski Golemi</td>
<td>Non-financial Big</td>
</tr>
<tr>
<td>SMA</td>
<td>Stopanski mali</td>
<td>Non-financial Small</td>
</tr>
</tbody>
</table>

#### Sectors

<table>
<thead>
<tr>
<th>Short name</th>
<th>Macedonian Name</th>
<th>English Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>S11</td>
<td>Sector S11</td>
<td>Non-financial</td>
</tr>
<tr>
<td>S12</td>
<td>Sector S12</td>
<td>Financial</td>
</tr>
<tr>
<td>S13</td>
<td>Sector S13</td>
<td>General Government</td>
</tr>
<tr>
<td>S14</td>
<td>Sector S14</td>
<td>Household Sector</td>
</tr>
<tr>
<td>S15</td>
<td>Sector S15</td>
<td>Non profit institutions servicing households (NPISH)</td>
</tr>
<tr>
<td>Type</td>
<td>Data Table</td>
<td></td>
</tr>
<tr>
<td>----------------</td>
<td>------------</td>
<td></td>
</tr>
<tr>
<td>Operational Data</td>
<td>ODD</td>
<td></td>
</tr>
<tr>
<td>Temporary</td>
<td>TT</td>
<td></td>
</tr>
<tr>
<td>Detail</td>
<td>DT</td>
<td></td>
</tr>
<tr>
<td>Export</td>
<td>OUT</td>
<td></td>
</tr>
<tr>
<td>Import</td>
<td>Cor</td>
<td></td>
</tr>
<tr>
<td>Dimension</td>
<td>DIM</td>
<td></td>
</tr>
<tr>
<td>External</td>
<td>EXT</td>
<td></td>
</tr>
<tr>
<td>Definitions</td>
<td>DEF</td>
<td></td>
</tr>
</tbody>
</table>
3 types of **business definitions**: 
- Definitions for calculating categories
- Definitions for correction of the categories
- Definitions for building the reports

**Definitions** - in MS Excel file, which in different columns contains formulas for:
- Calculation of *basic categories* of gross domestic product
- Calculation of *corrected categories* on the basis of data analysis and comparisons with other administrative and statistical survey’s data (depends from the business subject)
  - on each inclusion of some data source in corrections, process interrupt is made and business analysts are included by using application for data corrections and analysis;
- Calculation of *ESA variables* needed for reporting.
**Exploitation interfaces**

Based on the experiences in the past, Business users express some requirements, which are solved by integrating an appropriate GUI in the DW.

<table>
<thead>
<tr>
<th>Requirement</th>
<th>Solution</th>
</tr>
</thead>
<tbody>
<tr>
<td>1  Corrections and adjustments-</td>
<td>GUI (Graphical User Interface) – Application which will enable users to</td>
</tr>
<tr>
<td>To be able to make corrections on the</td>
<td>change values in the detail DW tables</td>
</tr>
<tr>
<td>detail level (companies) after first</td>
<td></td>
</tr>
<tr>
<td>calculation of analytical variables</td>
<td></td>
</tr>
<tr>
<td>(categories) and to make notes about</td>
<td></td>
</tr>
<tr>
<td>changes</td>
<td></td>
</tr>
<tr>
<td>2  Analysis (due to corrections) -</td>
<td>Objects in the application –</td>
</tr>
<tr>
<td>To be able to display data on</td>
<td>multidimensional reports</td>
</tr>
<tr>
<td>summarization level about sectors</td>
<td></td>
</tr>
<tr>
<td>3  Reports –</td>
<td>GUI (Graphical User Interface) – Application for reports definitions</td>
</tr>
<tr>
<td>Design of the reports is defined by the</td>
<td>and creation is developed</td>
</tr>
<tr>
<td>National Accounts Team</td>
<td></td>
</tr>
</tbody>
</table>
Exploitation interfaces - 1. Application for corrections/adjustments

### Table: Classification Variables

<table>
<thead>
<tr>
<th>mbpodzav</th>
<th>golemina</th>
<th>sektor</th>
<th>podsektor</th>
<th>skoi</th>
<th>skop</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Table: Analytical Variables

<table>
<thead>
<tr>
<th>melufazna potrošučka</th>
<th>Neto dodadena vrednost</th>
<th>podsektor</th>
<th>data12</th>
<th>data14</th>
<th>data16</th>
<th>data18</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>S.1313</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>S.1313</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>500</td>
<td>0</td>
<td>S.1313</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1985</td>
<td>0</td>
<td>S.1313</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>572</td>
<td>1743</td>
<td>S.1313</td>
<td>70</td>
<td>49</td>
<td>17</td>
<td>92</td>
</tr>
<tr>
<td>384</td>
<td>0</td>
<td>S.1313</td>
<td>0</td>
<td>319</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2271</td>
<td>2746</td>
<td>S.1313</td>
<td>213</td>
<td>556</td>
<td>0</td>
<td>578</td>
</tr>
<tr>
<td>1081</td>
<td>0</td>
<td>S.1313</td>
<td>92</td>
<td>319</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3343</td>
<td>0</td>
<td>S.1313</td>
<td>188</td>
<td>392</td>
<td>170</td>
<td>681</td>
</tr>
</tbody>
</table>
Exploitation interfaces - 2. Application for data analysis

## Adjustments — Data Analysis

### Data Analysis after Adjustments

<table>
<thead>
<tr>
<th>Sektor</th>
<th>Bruto deloven višek - izvorni</th>
<th>Bruto deloven višek</th>
<th>Bruto vrednosti na proizvodstvo - izvorni</th>
<th>Bruto vrednosti na proizvodstvo</th>
<th>Megufa</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sum</td>
<td>Sum</td>
<td>Sum</td>
<td>Sum</td>
<td></td>
</tr>
<tr>
<td>S.11001</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E.40</td>
<td>-10505.00</td>
<td>10505.00</td>
<td>575.00</td>
<td>575.00</td>
<td></td>
</tr>
<tr>
<td>Vkupno SKOI</td>
<td>-10505.00</td>
<td>10505.00</td>
<td>575.00</td>
<td>575.00</td>
<td></td>
</tr>
<tr>
<td>S.11002</td>
<td>3041717.00</td>
<td>3041717.00</td>
<td>11361450.0</td>
<td>12125751.0</td>
<td>62</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S.11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S.11003</td>
<td>DB</td>
<td>1070.00</td>
<td>34953.00</td>
<td>34953.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>DE 22</td>
<td>11504.00</td>
<td>87812.00</td>
<td>90926.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>DK</td>
<td>9586.00</td>
<td>15938.00</td>
<td>15938.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>DL</td>
<td>7463.00</td>
<td>38285.00</td>
<td>48261.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>DM</td>
<td>-1867.00</td>
<td>105543.00</td>
<td>105543.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>41193.00</td>
<td>174879.00</td>
<td>174879.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>G</td>
<td>371329.00</td>
<td>1081679.00</td>
<td>1132094.00</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>I.60</td>
<td>29158.00</td>
<td>294557.00</td>
<td>294557.00</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>I.63</td>
<td>1463.00</td>
<td>4426.00</td>
<td>4426.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>K</td>
<td>26271.00</td>
<td>253211.00</td>
<td>298455.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Vkupno SKOI</td>
<td>497170.00</td>
<td>2092293.00</td>
<td>2201032.00</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>Vkupno Podsektor</td>
<td>3528382.00</td>
<td>13454318.0</td>
<td>14327358.0</td>
<td>80</td>
</tr>
<tr>
<td></td>
<td>Vkupno Sektor</td>
<td>3528382.00</td>
<td>13454318.0</td>
<td>14327358.0</td>
<td>80</td>
</tr>
</tbody>
</table>
Exploitation interfaces - 3. Dissemination Application
Overview of DW organisation and used SAS modules

- **SAS/Warehouse Administrator** software
  - for building NA Data warehouse

- **Operational data sources** *(ODD’s)*:
  - administrative data sources
  - basic statistics.

- Organised in **ODD groups**
  - depending on origin and type

- **Different data formats** *(.txt, SAS, DB2/6000)*

- **SAS/ACCESS** software to **DB2**
  - to extract data from DB2 databases

- **SAS/CONNECT** software from Windows clients
  - for accessing data sets on the AIX servers
**Subjects**
- logical group of data sets related to certain topic

**Within subjects data are organised in** *Data Groups*
- source
- purpose
- format (detail, OLAP)
- permanency (permanent, temporary)

**Data storage in DW in different formats**
- SAS
  - *Relational databases* (DB2/6000)
  - *Multidimensional databases* (SAS/MDDB on UNIX)

=> **HOLAP** (Hybride On-Line Analytical Processing) solution
**Metadata** ("data about data") - central role
- input sources, targets, processes

Implementation of a large and complex data warehouse
- necessity of very **clear documentation** used
  - when performing operational tasks
  - for future system development and planning

**SAS/Warehouse Administrator** and its **metadata infrastructure** enable:
- automatic documentation
- metadata management
- easy management of all DW changes
- **Graphical user interface**
  - schematic view of processes (definition and re-definition)

- **User written code**
  - used for processes definition

- **Automatic code generation** (by SAS/Warehouse Administrator)
  - makes the work easier
Having the **client/server environment** in SSOM imposes the need to use all the possibilities it offers:

- **data sets** - on AIX servers

- **users of the applications** - work within a familiar Windows environment (Windows styled screens)

- **processing required** - performed by the client or the server
  - depends from the complexity and size of the data involved
  - in our solution - all data processing is performed on the servers
A vast data repository
- need for **access to data and reports** in the most flexible way

**SAS/EIS** module
- starting point of our data exploitation phase

Creation of **multidimensional reports** in SAS/EIS
- simple process - doesn’t require extensive technical expertise

**Custom SAS/AF applications**
- for decision support at phases of data analysis and corrections
Concluding remarks

- Increase of the **work efficiency**
- Improvement of **data organization**
- **Central repository** of sources, classification tables and statistical output
- Foundation of **time series** for economic indicators
- IT infrastructure providing **easy adoption to business changes**
- **Documentation** and **reuse** of the implemented processes
- **Availability of the data** for various users
Future plans

Our planes for the future are as follows:

- To continue with the development in this direction in Module 2 of the Project (July 2001 - July 2002)

- Exploitation of a DW via Internet/Intranet
  - (we have all necessary SAS modules licensed)

- Development of custom SAS/AF, SAS/EIS, Web applications
State Statistical Office of Macedonia

choosed SAS for implementation of this project

because,

SAS:

- offers the ideal mix of products;
- has tools for each level of implementation.
Questions?
Data warehouse implementation in financial statistics using SAS® in the Statistical Office of Macedonia

Lidija Petkovska, State Statistical Office of Macedonia, Skopje, Republic of Macedonia
Valentina Cacorovska–Jankovic, State Statistical Office of Macedonia, R. of Macedonia

Abstract

This paper discusses the Project of automation of the National Accounts System at State Statistical Office of Macedonia (SSOM). As first main functions and tasks of the State Statistical Office of Macedonia and main functions of the National Accounts System are introduced. Next the general principles of data warehousing (as data organisation and as a process) are given; as well as the practical tips and aspects on the organisation and implementation of the Financial Statistics DW by using SAS®. At the end some concluding remarks on our expectations of this project and our future plans are outlined.

1. Introduction

Main tasks of the State Statistical Office of the Republic Macedonia are statistical data gathering, processing, analysis, publishing, estimations, 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) and Censuses.

Besides these tasks, SSOM has also legal obligations for establishing and maintaining Statistical Registers in the Republic of Macedonia.

The main aim is providing objective, neutral, actual and relevant statistical data for all users and also providing quality statistical information for decision support processes of the Government and other business subjects.

National Accounts are important macro-economic indicator showing the level of a development of a country (gross domestic product). System of National Accounts integrates a lot of data in a consistent system providing comparable data for the national economy on the international level.

National Accounts Sector, which is responsible for the implementation of the System of National Accounts, uses data from numerous administrative sources, as well as statistical data produced by the Business Statistics Sector and Social Statistics Sector of SSOM.

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 led to the definition and implementation of “Project of the automation of the System of National Accounts” for establishing and implementing DW on financial statistics (in order SSOM to continue providing data on financial statistics which were provided by Payment Operational Office before the payment’s system transformation).
2. About the Project of the Automation of the System of National Accounts

This project implementation should be realized in two phases:
1. Automation of financial statistics for which data are mainly gathered from administrative data sources
   • (end - May 2001);
2. Integration of basic statistics in the SNA (in the mean time the used classifications in the basic surveys have to be harmonized and standardized with the EU and international ones)
   • (May 2001 – April 2002).

Data Warehouse will contain data from a numerous administrative sources, special statistical surveys and basic statistical data produced by SSOM (Figure 1: Data Warehouse Content).

2.1. Project Goals

In the framework of the project the following should be realised/implemented:
• Establishment of Data Warehouse on financial statistics;
• Standardized data transfer from State Government Organizations – providers of administrative data to Data Warehouse;
• Automation of the processes on financial data transformation, processing, analysis and dissemination;
• Decision-support system for business analysts at phases of data adjustments and analysis;
• Automation of the reporting phase i.e. enabling the end-users to easily prepare and access reports;
• Establishment of the consistent system of input-output tables, supply-use tables and Integrated Economic Accounts;
• Harmonization of the used methodology with ESA ‘95 and Eurostat recommendations.

2.2. Project Team
Project is the result of the joint work of:
• State Statistical Office of Macedonia;
• SAS Institute d.o.o. Ljubljana;
• Eureka Informatika – private informatic firm, Skopje, Republic of Macedonia.

Roles and tasks of the project members were divided as follows:

**State Statistical Office of Macedonia**
- Definition of the business rules for extraction and transformation of data from source systems to warehouse entities;
- Providing of knowledge of the structure and organisation of the in-house systems and data, as well as of the external data sources;
- Design and programming of the algorithms for part of the processes;
- Data Warehouse implementation;
- Validation of the data, processes, and objects in the Warehouse, in fact check if it was worked according previously defined business rules;
- Joint project management with external consultants.

**SAS Institute d.o.o. Ljubljana:**
- Design, build and help at the implementation of the Warehouse;
- Design automated processes in the Warehouse;
- Design and building of the applications for data correction, analysis and exploitation;
- Providing in-house SAS training;
- Providing SAS support.

2.3. Methodology

Rapid Warehouse Methodology is used, series of steps were defined and followed during the work on the project (Figure 2: Rapid Warehouse Methodology). In the frames of one of the phases SAS training was provided for the business analysts from the National Accounts Sector, aiming at providing skills for more adequate definition of business needs and more efficient participation in data analysis processes.

![Figure 2: Rapid Warehouse Methodology](image)

3. Design and implementation of SDW at Financial Statistics

In the framework of activities at Statistical Offices, 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 optimized for decision support applications.

In that way is established a mechanism to bridge all differences between Transactional (Operational) Systems (handle and administer business’s daily activities) and Decision Support Systems (provide a framework to use and analyze data to make effective, informed business decisions – support managerial and executive needs) as are: type of the data, hardware, purpose and users. In fact, from process-oriented systems based on well-defined business requirements, a step is made towards subject-oriented systems designed to provide flexibility for what-if analyses.

3.1. Statistical Information System in SSOM

In order to understand better the organization of this Data Warehouse, this chapter describes state-of-the-art of the Statistical Information System in SSOM.

**Hardware equipment** in SSOM can be divided in two environments:

1. **Mainframe Unisys environment** (purchased 1990/91) – planned to be put out of production as soon as possible (almost achieved);

2. **Client/Server environment** in the Central SSOM building and 8 Regional Offices.

   - Client/Server environment in the Central SSOM building with the following configuration:
     - IBM RISC/6000 Servers with AIX operating system;
     - Windows NT Servers;
     - Windows NT work stations;
     - Network printers;

   The network is 100 Mb LAN.

   - Client/server environment in 8 Regional Offices:
     - 1 Windows NT server;
     - Windows NT workstations.

   The network is also 100 Mb LAN.

   Regional Offices and Central Office are all connected in 64Mb/s WAN.

**Software** installed and used:

- SAS;
- MS Access;
- MS Visual Basic 6.0;
- COBOL;
- DB2/6000 (RDBMS).
3.2. SAS System and SAS/Warehouse Administrator® as a solution

Up to now employees from the National Accounts Sector have made these complex calculations by transferring .txt files from Payment Operations Office (POO) to MS Access, while tools used for data processing and analysis were MS Excel and MS Access. Taking in consideration the fact the amount of data and complexity of calculations, it is clear that the used software was not suitable for the needed tasks.

Therefore, in the frames of the preparations for the above mentioned project, but especially for the needs of the Annual accounts '99 data processing, basic SAS training of the employees from the National Accounts Sector was conducted, and SAS® was used for data processing.

Some of the National Accounts Sector employees started to use Base SAS, SAS/STAT®, SAS/INSIGHT® and Enterprise Guide®. For those having previous experience in MS Access, easy acceptable solution was usage of SAS Query window for “ad-hoc” queries. Having in mind planes for training in the near future, reorganization of data and processing and Data Warehouse implementation, we consider that very soon they’ll fill all the benefits and improvements that SAS® brings.

For building of the Financial Statistics Data Warehouse SAS® is used (a lot of his modules licensed by SSOM).

3.3. Physical Model of the Data Warehouse

The Figure 3 captures the Structure of the Data Warehouse, which consists of:

1. Operational data sources (ODD);
   - Administrative data sources (POO data);
   - Dimension (classification) tables;
   - Business definitions – rules.
2. Detail Tables organized by:
   - Business subjects form POO: administrative data from POO, merged with dimension tables, on the lowest level of detail (company);
   - Classification of enterprises by sectors (sector dimension table ): data on the lowest detail level grouped by sectors (S11 – Non financial; S12 – Financial; S13 – General Government; S14 – Household Sector; S15 – Non profit institutions servicing households).
3. Exploitation (summary) tables (structured for analysis and reporting);
4. Data Marts (Reporting Application).

3.4. Transformation, Aggregation and Adjustments processes
In this paragraph, according Figure 4 (DW Process Model), is described the general flow of the processes for processing of the administrative data sources - process of transformation of data from micro level (level of enterprises) to macro level (macro economic categories calculated on level of institutional sectors).

![Warehouse Environment](image)

**Figure 4: DW Process Model**

Annual accounts data are taken from the Payment Operations Office – POO (Central Register of annual accounts after the POO transformation). There are 8 types of business subjects, which fill in different balance sheets (1. Non-financial organizations – big enterprises; 2. Non-financial organizations – small and medium enterprises; 3. Users of Budget sources and Social Funds; 4. Budgets of local units and Local Funds; 5. Other units (non-profit institutions); 6. Banks and other financial institutions; 7. Insurance companies; 8. Unincorporated enterprises (self-employed-workers)).

Main data processes between operational systems and Data Warehouse, for each of the POO business subjects are:

1. **Extraction, transformation and loading processes on yearly basis (ETL Process)** *(Figure 5)*

![Figure 5: Extraction, Transformation and Loading Process](image)
1. Process of **extraction** of input data sources into SAS environment (DB2 and .txt files into SAS tables);

2. **Data transformation** (derivation of categories, data corrections and adjustments);
   - **P1** (Figure 6: **Process Business Units in detail**);
     - One subject data from POO are *merged with dimension (classification) tables* (Administrative Register of Enterprises – ROE; Public Revenue Office data; Classifications of activities -the old one, the National version of NACE classification - NKD, the link between these classifications, Territorial Classification – OPS, Classification of enterprises in sectors – SEK; Dimension tables used only at SNA calculations – SKOI_SKOP etc.);

   ![Figure 6: P1 - Process Business Units in detail](image)

   - **P2** (Figure 7: **Process Categories Correction in detail**);
     - Basic *categories of gross domestic product are calculated* and added as new variables in the data set;
     - This detail data set (with not corrected data from POO) is used as a data source for reporting, preparation of multidimensional reports, and address lists for the needs of business statistics;
     - Data are *aggregated* on different levels and by different variables - after that data analysis are performed;
     - **Correction of the categories** - outliers are found, additional variables - corrective factors are determined and corrected categories values are calculated (corrections are made by comparing current and previous year’s data, also by the analyzing of the Labour Force Survey and some other administrative and statistical surveys data);
     - In the next process corrective factors are being put back on the level of the Firm and corrected values of the categories are calculated.

   ![Figure 7: P2 - Process Categories Correction in detail](image)

3. **Loading** of detail tables in the Data Warehouse
   - Detail tables organized by *business subject* are stored as permanent tables in DW (also these data organized and grouped by *Classification of enterprises in sectors* (5) are stored as detail tables in the Data Warehouse).
II. Data summarization and aggregation - data storing in summary tables for facilitating analysis, reporting and exploitation.

- Data sets of all POO Business subjects are merged (a new variable indicating the data origin is added), new ESA variables are derived, multidimensional databases (which cover a lot of class variables, hierarchies) and summary tables (having a structure optimized for generation of pre-defined reports) are loaded.

III. Exploitation of Warehouse through Dissemination Application.

IV. Data Archiving – (at the moment data for years 1997 to 1999 are loaded). Based on the business requirements the tables on detail level (1 row = 1 company) will be archived. Tables must include all dimensions for data analysis and source and derived analytical variables.

Dimension tables are used at data imputations where possible, but also special algorithms are designed according the business rules defined by business analysts. DW users will be included in the processes that can not fully automate (the reason is the nature of calculations which require subjective data analysis and data corrections to be performed by business analysts):

- Imputation of missing data;
- Correction of analytical variables (categories of gross domestic product).

3.5. Naming conventions and business rules

Based on the business requirements (inputs, outputs and work with DW) data definition standards have been introduced. Naming conventions and business definitions are explained briefly in the following text.

Some examples of naming conventions are:

- Short names are defined for POO business subjects and used in the DW (BAN - Banks, BUX - Budget, BUK – Users of the budget, DRU - Others, OSI - Insurance, SAM - Unincorporated, SGO – Non-financial big enterprises, SMA – Non-financial small and medium enterprises) – acronyms are derived from the macedonian name of the subject;
- Following naming conventions are used as part of the names of data sets and processes: ODD – operational data; TT – temporary; DT – detail tables; OUT – send for correction in ETL process; Cor – corrected by business analysts and returned etc.).

We have 3 types of business definitions:

- Definitions for extracting data;
- Definitions for calculating the categories;
- Definitions for building the reports.

These definitions are in the MS Excel file, which in different columns contains formulas for:

- Calculation of basic categories of gross domestic product;
- Calculation of corrected categories on the basis of data analysis and comparisons with other administrative and statistical survey’s data (depends from the business subject) – on each inclusion of some data source in corrections, process interrupt is made and business analysts are included by using application for data corrections and analysis;
- Calculation of ESA variables needed for reporting.

This kind of information is stored in separate worksheets for each of the POO business subjects.

A program was designed which reads data from the worksheets and columns of this MS Excel file and generates SAS programs which perform the following calculations: calculation of the categories, calculation of corrective factors and corrected categories values, calculation of ESA variables needed for reporting.

Using this way of defining business rules, the user writes them in a familiar environment (MS Excel file) - the other is done the program generator.
4. Exploitation interfaces

Based on the experiences in the past, Business users express some requirements, which are solved by integrating an appropriate GUI in the DW.

<table>
<thead>
<tr>
<th>Requirement</th>
<th>Solution</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Corrections and adjustments-</td>
<td>GUI (Graphical User Interface) – Application which will enable users to change values in the detail DW tables</td>
</tr>
<tr>
<td>To be able to make corrections on the detail level (companies) after first calculation of analytical variables (categories) and to make notes about changes</td>
<td></td>
</tr>
<tr>
<td>2. Analysis (due to corrections) - To be able to display data on summarization level about sectors</td>
<td>Objects in the application – multidimensional reports</td>
</tr>
<tr>
<td>3. Reports from detail data – Design of the reports is defined by the National Accounts Team</td>
<td>GUI (Graphical User Interface) – Application for reports definitions and creation is developed</td>
</tr>
</tbody>
</table>

4.1. Application for corrections

The main menu of the application for adjustments consists of two selections (Figure 8):

- Subject (one of the POO business subjects, 3 characters are entered according previously explained data definition standards); and
- Sector.

**Functionality:**

- **Selection of variables** (all columns of the detail table of the selected subject are available in this list box),
- **Showing variables** in Table View (edit mode and browse mode);
- Editing of data in edit mode and audit trail of all modifications to the SAS tables in a log file;
- **Tracking** of data modifications (shows audit trail file (html) in which all modifications are recorded, date and time of the changes, who has made the changes etc);
- **Summary** by selected variables is available;
- **SAS/INSIGHT®** procedure can be called with the selected observations (subject and sector);
- Showing of all columns that the selected column is computed from or shows message that the selected column is not computed (by **double clicking the column**).

![Figure 8: Application for adjustments](image)

4.2. Application for Data Analysis
The look of multidimensional reports (dimensions and hierarchies) needed for data analysis is defined by business users (Figure 9).

4.3. Dissemination application

A sequence of frames of the SAS/AF® application that helps in definition of reports is shown on the Figure 10.

On the main menu type of the report is chosen:
- Integrated economic accounts;
- Set of accounts;
- Additional analysis.
Further are defined:
- Table title;
- ESA '95 codes.
On the next screen are specified:
- Years;
- Type of the report.
5. Current Status (organization and used SAS® modules) and Future Plans

**SAS/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 **ODD groups** according to the administrative evidence they take the origin from. **SAS/ACCESS®** 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 organized 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 organized in groups (**Data Groups**) depending on data source, purpose, format (detail, OLAP), permanency (permanent/intermediate).

DW data are kept in SAS format, **relational databases** (DB2/6000) and **multidimensional databases** on AIX platform (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/Warehouse Administrator® software for automatic 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 necessitates the provision of very clear documentation of the process involved. This documentation is not only useful when performing operational tasks, but is also important for future system development and planning. Usage of SAS/Warehouse Administrator® software and its metadata infrastructure enables automatic documentation and metadata management, as well as easy management of all DW changes. Metadata can be exported in other formats and we hope that this will enable establishment of integrated metadata system.

5.1. Future Planes

Our planes for the future are as follows:

• To continue with the development in this direction in Module 2 of the Project (May 2001 – April 2002);
• Exploitation of a DW via Internet/Intranet (we have all necessary SAS modules licensed);
• Development of more custom SAS/AF and SAS/EIS applications, and also Web applications.

6. Concluding remarks

Expected benefits of this project are:

• SAS software to satisfy our expectations for successful implementation and exploitation of this DW (SAS was chosen because it offers an ideal mix of products to setup such an environment, and because it has tools for each level of implementation);
• 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.

References

1. Project Documentation;

Trademarks


Contact Information
Authors of this paper worked on DW implementation in State Statistical Office of Macedonia.

**Lidija Petkovska**  
State Statistical Office of Republic of Macedonia,  
Research and Development Sector  
Department for Methodologies, Statistical Analysis, Estimations and Prognosis  
Consultant for software support and SAS Administrator  
*E-mail:* lidijap@stat.gov.mk

**Valentina Jankovik Cacorovska**  
State Statistical Office of Republic of Macedonia  
Research and Development Sector  
Department for Metadata and Management of Metadatabases  
Head of the Department  
*E-mail:* valec@stat.gov.mk