Usage of Data Warehouse for Data Preparation for the Needs of the State Statistical Office of the Republic of Macedonia and How Was Transparent Data Dissemination Achieved?

S.Savoska , V.Manevska, S.Neshkovska[1]

ABSTRACT

The State Statistical Office (SSO) is an institution which is in charge of collecting, processing and analyzing data concerning relevant entities in the Republic of Macedonia, as well as, data dissemination in adequate modes for gaining transparent insight.

The main objective of the paper will be:
- Detection of the necessary organizational changes for implementation of Data Warehouse (DW), organizational culture, collection modes, transformation and utilization of data to obtain the required transparent data for public usage;

As the main objective of the paper is rather complex, the following research questions should be looked into with considerable deliberation:
- What organizational changes have been introduced in the SBS to achieve successful implementation of DW?
- What kind of organizational culture is needed to implement such a complex project; how was it achieved; what barriers had to be overcome?
- What were the expected benefits of the project implementation?
- What are the sources of DW of the SBS, the formats and the coded systems in which they were prepared and how is the preparation for filling DW carried out?
- To what extent are these processes automatic?
- In what way has the business logic applied in data processing and preparation for obtaining sector aggregated data been defined?

In order to find the answers to the abovementioned questions, a detailed analysis of the organizational, financial and other requirements concerning the introduction of such a complex data structure called data warehouses , as well as the sources of data and the mechanisms needed for their filling are needed. The support needed in this process both of organizational and technical nature is a topic of interest as well.

The methods used in the research imply usage of:
- Interview with the employees in the management structures and the IT sector aimed at detecting information for implementation in the data warehouses;
- Analytical method for researching sources of data in the SSO and their attributes, used formats and coded systems, mechanisms for collection, filling, cleansing, transformation and preparation of data for further usage;
- Analysis of business processes and software tools used as a part of the overall process of preparation and analysis of ready-made data which are subject of dissemination by the SSO and their quality;

Results and conclusions of the research:
- Implementation of data warehouses is a complex process which is profoundly concerned with the organizational culture and it requires core organizational changes;
- Successful implementation requires unconditional support on the part of top management, i.e. strict sponsorship;
- The benefits of the implementation come after the implementation and the filling of DW and they depend on the extent to which DW architecture has been successfully set.

---

[1] Faculty of Administration and Information Systems Management, "St.Kliment Ohridski" University-Bitola, the Republic of Macedonia

1. Introduction

The State Statistical Office (SSO) is an institution which is responsible for collecting, processing and analyzing of data which refer to relevant entities in the Republic of Macedonia, as well as their dissemination in a suitable form for obtaining transparent insight into data. This task is not simple at all as it touches upon many institutions, organizations, companies and administrative workers who use data from diverse sources. This intricate task most definitely implies a complex organizational structure, as well as mechanisms and software support which should ensure that. The software support, among the other things, includes using a professional database, SAS DW, which serves as a foundation for the integration of all versatile data obtained from many institutions, organizations and agencies, collected with distinct applications, under different operating systems and collected in diverse databases, code sides and coded systems.

Because of the complexity of the issue, we will look into several research questions:

1. What are the sources of data which are used for obtaining these data?
2. Where do those data come from, what data structure is needed and in what way has it been implemented in the SSO?
3. What organizational changes were necessary in the SSO to support such a complex data structure?
4. What obstacles and problems had to be overcome in the course of the implementation, as well as what organizational culture was necessary to implement the project?

As the answers to these questions are not simple at all, it is necessary to do a detailed analysis of all data, organizational, financial and other requests pertaining to the functioning of such a complex structure. Hence, our topic of interest will also be the support which is a need in the process of implementation and usage of DW both of organizational and technical nature. The research questions will assist us in reaching the aim of the paper:

- Detecting the necessary organizational changes for DW implementation, organizational culture, collection modes, transformation and usage of data in order to obtain the necessary transparent data.

We commence by answering the following questions:

- What organizational changes have been introduced in the SSO to reach a successful implementation and usage of DW in the SSO?
- What kind of an organizational culture is needed to implement such a complex project and how was it achieved and what barriers had to be overcome?
- What were the expected benefits of the implementation of the project?
- What are the sources of data for DW of the SSO, the formats and coded systems in which they have been prepared and how is the preparation for filling DW carried out?
- To what extent have these processes been automatised?
- In what way has the business logic been defined during data processing and the preparation for obtaining sector aggregated data?

To this aim, the employees of the SSO have been interviewed, in particular, those working in the Sector for National Accounts (SNA) and the IT sector. The necessary changes of the organizational culture which were introduced in the process of the implementation of DW, as well as the consequences of the implementation have been researched. An analysis of the sources of data, their attributes, formats used and coded systems, mechanisms for collecting, filling, cleansing, transforming and preparing the data for further usage has been done. Furthermore, we have analyzed the business processes and the used software tools. Finally, we analyzed the processes as a part of the entire process for preparation and analysis of data which are subject to dissemination by the SSO and their quality as well.

2. History and a need for DW implementation in the SSO

As an integral part of the former administration of SFRJ, the Federal Statistical Bureau (FSB) at the time was working on computer systems compatible with the Public Revenue Office (PRO), i.e. the UNISYS systems, A series. One of the most powerful hardware solutions at that time, UNISYS A12 belonged to the Federal Bureau of Statistics of the Republic of Macedonia. Via this platform (MCP, DB2, Cobol tool, UNIX systems and the like) most of the tasks entrusted to the Bureau were being solved.

Even though the SSO as a successor of the PRO has already have a long tradition of collecting, preparing and dissemination of data, yet the implementation of SAS DW has brought about many organizational changes in the organizational structure and organizational culture of the Office itself. The need for introducing DW stemmed from the need for a higher level of

integration of all relevant information in the state. Consequently, the formal initiative for the implementation of DW was directly connected to the growing needs for information and the influx of data from various sources, inconsistent, not well-studied formats and different coding and metadata. Furthermore, the need stemmed from the introduction of new 'business' processes which were supposed to generate a greater number of information in a shorter time span and with a greater degree of automation. At the very outset, in order to create this information, detailed knowledge was needed, which was to be transformed into procedures, which were supposed to be generated by means of some automatic CASE tool, which could be controlled by business managers and which are not available to informaticians. The defined phases, transformed into procedures are presented in Figure 1. What is characteristic about this diagram is that the preparation processes themselves include three types of processing, i.e. first, second and final run, which are preconditions for creating a consistent data warehouse where one can obtain the necessary analyses and reports.

3. Organizational preparedness for DW implementation vs. theoretical suppositions

The organizational preparedness for development and DW implementation could be defined as creating a plan for DW implementation by using five key factors [2], which are necessary for successful introduction of DW. Also, it was necessary to detect the possible weaknesses and to foresee the potential issues by means of constant documenting. The weaknesses and the problems were supposed to be resolved as they emerge or to be resolved even prior to the implementation itself. The five key factors which influence the preparadness for creating and implementing DW of a company are the following:

1. *Strong managerial inclusion*. This is an important and perhaps the most instrumental factor for success. The sponsors of the project should set a clear vision concerning the role of DW and its impact, i.e. its future implications. Managers must have a strong criterion to realize the vision, the plan and the activities of the implementation plan. Having at least one sponsor could be very convenient, but it is much better if there are sponsors from various business areas. Within the SSO, during the implementation period, the Sector for National Accounts (SNA) coupled with the SSO's management at that time displayed enormous maturity and literary 'carried the project upon their shoulders'. In the interviews, the employees of the SSO of that time stressed the role of the current Governor of the Republic of Macedonia as a very strict sponsor together with the SSO's management at that time. According to the interviews, the personal example and the engagement of the managing staff of the SSO, especially of the SNA, were crucial for the strong motivation of the employees as well as the successfulness of the implementation of the project to a large extent.

2. *Business Motivation Support.* DW is not a means to an end. Its value rests upon the role it performs as a realizer of specific business processes and strategic initiatives. It should move the organization forward in the direction of standardization of the processes and business activities. Only then can one expect business motivations with satisfactory results to happen. In the SSO, the SNA alongside with the SSO's managers at that time, have mostly set the business rules and have had all the procedures led by a complete support of the business motivations of DW. The then set procedures and business rules are still functioning nowadays and belong to the SNA's sphere of competence. The business rules which have been set in the SSO have been transformed into software components and formulas placed into Excel cells, which are easily addressed and which are used as software business objects. The SNA is responsible for setting the business rules and the outcomes of the transformations of data and the obtained outputs.

3. *Quality partnership among the information system and business demands.* The degree of their integration and the elimination of the internal redundant processes can make a positive contribution (synergy) to DW implementation. The entire organisation, i.e. its personnel in the IT sector, as well as its business partners must be included in the project so that they can integrate all the business functions in a more coherent higher level of culture in the organisation. According to the interviews with the former employees in the SSO, the achievement of a coherent integration of the working processes of the IT sector and the employees in the SNA, was essential. The outcome was significant whereas the organisational culture was raised on a higher level. This coherence has been maintained even to date. The mere usage of SAS software on the part of the SNA and its employees contributes to the increased comprehension of problems and IT solutions and the constant creation of symbiosis between the two sectors in the SSO which yields enhanced productivity.

4. *Organisation's analytical culture.* Organisation's analytical culture is a key element of DW implementation simply due to the fact that the organisation has to set high values to the information and analyses. The most efficient DW can be found in organisations which tend to review all the facts and take decisions on the basis of

those facts or in other words the culture of the organisation is heuristic. Our research of the analytical culture of the SSO shows that it is highly positive and heuristic as well. Namely, the data which are diverse, taken from various sources (Table 1), under different formats, have been successfully "collected" and prepared in a format required for filling the data in DW, transforming them and preparing them on several format levels. The complexity of this process has been supported by a coherent team for analysis which consists of employees who work for the IT sector and the SNA.The analyses themselves present an integration of the management and are being prepared in a mainly automatic manner by using SAS DW tool with many different types of software end-user tools (Figure 2). The information tools enable the creation of detailed user's reports via interactive menus and request analysis, which are user-friendly and which could be created by non-informatics staff. All the tables from the warehouse could be subjected to analysis, according to years or other selected parametres.

5. *Economic payoffs of data warehouses.* This is a very specific factor which can neither be measured easily nor be assessed directly, but it should be performed indirectly via assessing other measurable parameters. In the case of the SSO, the costs consist of maintenace costs of the entire system of the SSO, employees, material costs and the part referring to the software tools SAS DW's licences and other applications employed in the working processes. Although the costs of the licenses could be very high (at the time of the supply their price was even 10 times higher than it is today), they could be justified as thus important information of state's interest could be obtained. It should also be pointed out that the costs for the training of the employees at the SSO have been largely reduced, despite the permant fluctuatuion of staff in both the SNA and the IT sector. Even though the number of employees at the SSO has not been significantly reduced since the time of the implementation, yet their structure is different – there is a smaller number of employees working in the IT sector and the SNA but they perform a much larger range of activities, doing analyses and writing reports and a greater degree of automatization of the business processes at the SSO. The problem with the outflow of the IT professionals and the highly skilled analytic experts is still present owing to the low salaries of the civil servants at the SSO (as well as in the other institutions of the Republic of Macedonia), which may result in a lack of trained staff needed to support DW.

An important factor which should influence the implementation of DW is also the preparedness for handling errors in the working processes. The factor requires both managerial support and trained IT and analytical personnel. In the interviews, the employees stated various problems and errors which followed the operational implementation of SAS DW in the SSO [2,3]. Yet, the strict managerial support and the high level of determination for successful implementation, as well as the commitment to the vision and the mission, led to a fast and efficient resolution of the problems and the errors which occurred.
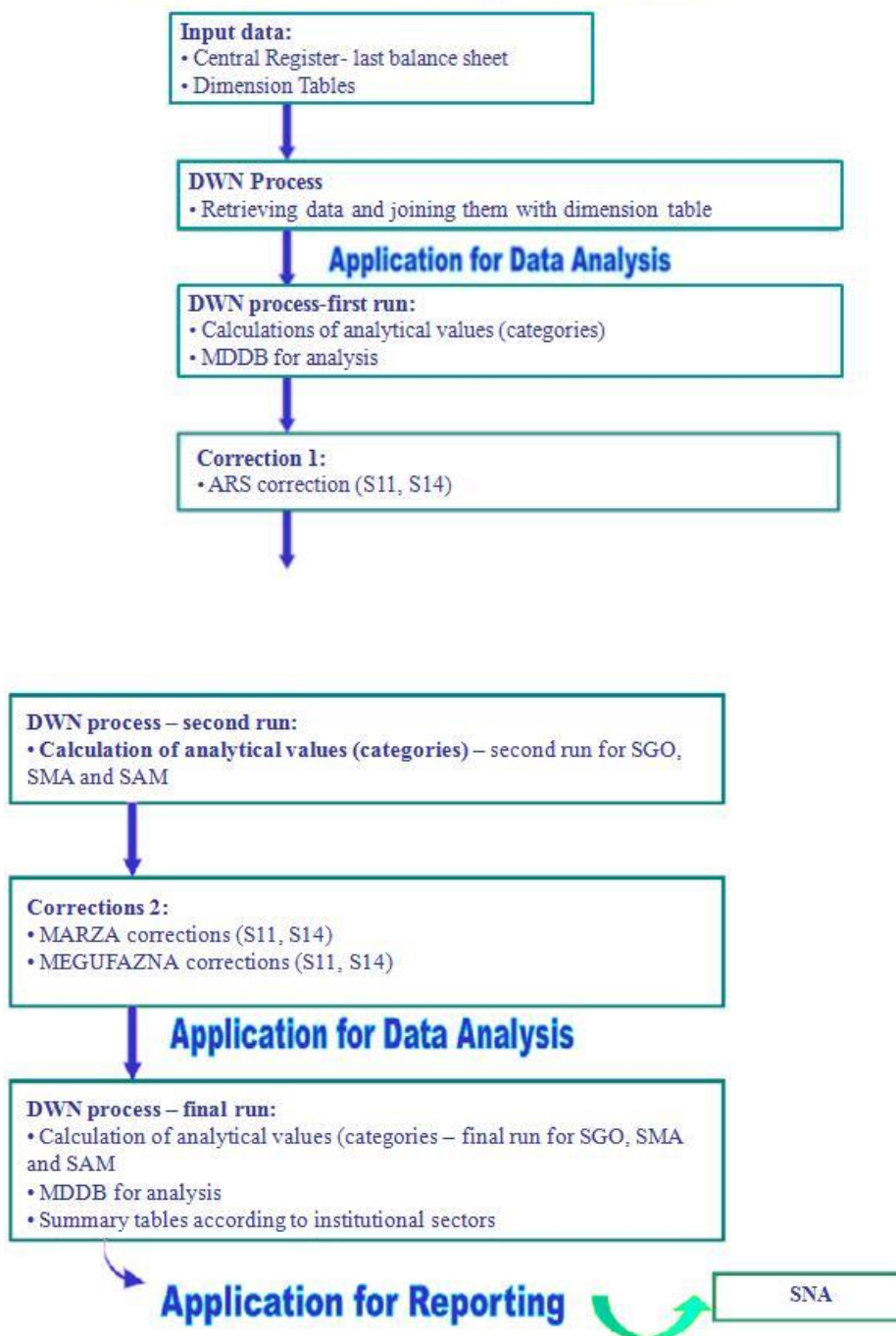
Schematic presentation of the processing phases

Input data:
• Central Register- last balance sheet
• Dimension Tables

DWN Process
• Retrieving data and joining them with dimension table

Application for Data Analysis

DWN process-first run:
• Calculations of analytical values (categories)
• MDDB for analysis

Correction 1:
• ARS correction (S11, S14)

DWN process – second run:
• **Calculation of analytical values (categories) – second run for SGO, SMA and SAM**

Corrections 2:
• MARZA corrections (S11, S14)
• MEGUFAZNA corrections (S11, S14)

Application for Data Analysis

DWN process – final run:
• Calculation of analytical values (categories – final run for SGO, SMA and SAM
• MDDB for analysis
• Summary tables according to institutional sectors

Application for Reporting SNA

Figure 1 – Phases in the process of data preparation for DW

Table1. Sources of data relevant for the SSO's DW

Administrative sources	Data about:	It refers to:
Central Register	subjects' final balance sheet	non- financial organizations (small, medium-sized and large companies), budget beneficiaries, the budget and the funds, other non-profit organizations, banks and other financial institutions, insurance companies, independent performers of economic activities and individual merchants
Public Revenues Office	tax- informative base	VAT, income tax, profit tax, estate tax
Ministry of Finances	public revenues and expenditures – treasury accounts	Employees' accounts, subjects' accounts
National Bank of the Republic of Macedonia	monetary statistics	
Health Insurance Fund	health contributions	employed, unemployed
Pension and Disabilities Insurance Fund	pension and disabilities contribution	employed, unemployed
Statistical research	industry, foreign trade, domestic trade, construction and catering, agriculture, polls on households' consumption, statistics on prices and polls on workforce and 'grey' economy	Citizens' and other subjects' data
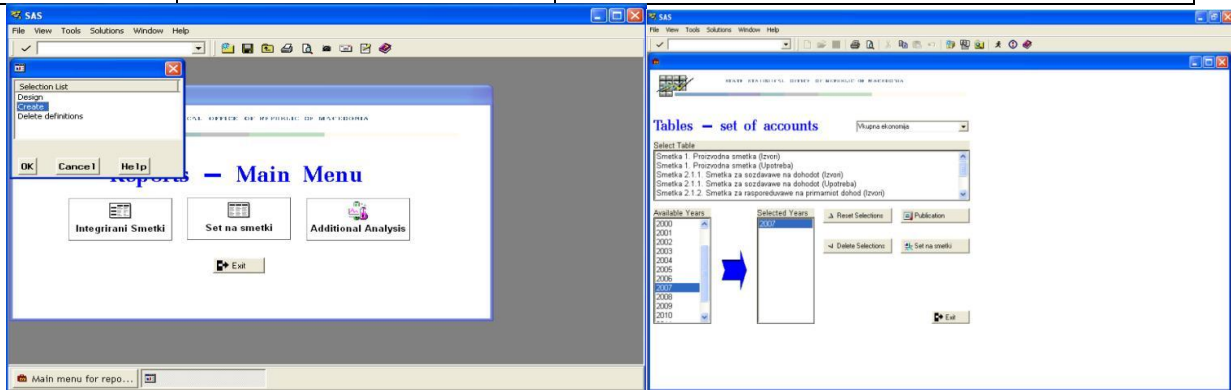
Figure 2   SAS Report menu tool which is used for creating reports by non - IT staff

5. Requests for business analyses - one of the main reasons for using DW

Requests for high level business analysis are certainly one of the basic reasons for development and implementation of DW. Hence, the efforts for a multi-level business analysis, should provide the answers to several key questions [1], which are essential for the success of the construction and operation of the company's DW. We present those key questions as follows: 1. What are the key performances of success (KPI) or success indicators of a business?; 2. What are the crucial business processes which the management normally perceives and wishes to influence via business initiatives?; 3. What is the potential impact of the approach to quality information for each of the existing KPI?

As an example we could mention the calculation of the gross production value which adds up Gross_k and trade_margin columns and accommodates them in P1 at ESA position under the title Gross in the final_run procedure.

Figure 3 – Excel file Categories which have the role of metadata for business rules of DW

The answers to these questions require identification of the business area of interest, provision of choice and opportunities for error correction, as well as protection of DW from breakdowns or unreadiness to meet those business requests. In the SSO, these business analyses typically followed the strict statistical and mathematical standards which are applicable in this area of operation to a great extent. However, the tool which has been implemented, assissted significantly in defining the necessary parameters, both from the aspect of built-in knowledge and functions in the software application itself, and its user-friendly interface, the easy way of defining the business rules (with an excel file with multiple tables which are easily maintained and upgraded and it has the role of metadata – Figure 3) [1,6]. The pleasure derived from the usage of this tool contributed to the smooth overcoming of many business problems which ranged from invalid and incorrect data to incoherent structures of data, insufficiently defined areas and inconsistent demands for input analysis.

6. Benefits of using DW

The successful implementation and usage of DW can introduce numerous visible and invisible benefits. The visible benefits include reduction of the needed time and the costs for overtime work, number of employees, less paper and media, less engaged management structures, a higher quality of information, consolidation of the roles of the employees, as well as increased control, efficiency and productivity [8]. All pieces of information which can be obtained in this way, and which was not the case in the past, are also considered a benefit. The invisible benefits (Table 2) could be the improvement of the organisational image and service quality; the improved decision-making and employees' morale; more correct and more timely information; an improved public and joint relation and enhanced organisational flexibility, leverage and inventiveness [2]. From the interviews and the analysis of the benefits of DW implementation in the SSO it could be stated

that the benefits of the implementation were huge, for both the employees and users of data as well. At the very outset, an integration, which previously was considered impossible with any other tool, was achieved, whereas the updating was dependant on the sources of data.

The process starting from the admitting up to the filling of data in DW was greatly automatized and it was consequently much faster. The number of employees who had an access to the original data was minimized, whereas the time needed for processing data and the quality of data was drastically increased. The entire office 'grew' in its organisational structure and different organisations of the data-marts were performed (sector's data warehouses) which were filled in an automatic way. The public service was much improved by the implementation of flexible structures of databases which were drawn directly from SAS DW and were transparently displayed in the desired shape – visually, in tables or in any another shape. The insights of the citizens were widened which led to an increased inventiveness while creating analyses and reports. On the whole, the SSO obtained a powerful tool which enhanced its information capacity and the timeliness, accuracy and representation of data. However, due to the limitations of licensing funds, a further enhancement of the facilities by supplying other SAS modules for data representation was not allowed.

Table 2. The most important benefits of a new DW

Improvement of employee's morale
Improvement of the corporative and public image
Improvement of products and services' quality
Notable decrease in the marketing time by buyers
Improved decision-making
More timely information
Improvement of the organizational flexibility
Improved allocation of resources and leverage
Improvement of strategic and comparative advantages
Improved public and joint relations
Improvement of environment-friendly handling of activities
Reduction of employees' redeployment
Increased employees' work quality
Proactive attention to ethical topics
Proactive addressing of inherited topics (outcomes)
Increased work space and/or safety of communities

7. Business rules and DW

We have already mentioned that one of the key factors for successful implementation and usage of DW is the defining of mutual business rules which make the identification of algorithms and the logic for creation of mutual dimensions, measures, divisions, classification and structures, possible. If the designer of DW cannot reach consensus concerning these business rules, individual models should be identified of each calculation method and for each department separately. If this is the case, then what follows is integration via data consistency tables which should be included in the warehouse. The business rules of DW of the SSO have been defined by Business Performance Management (BPM) process which follows the previously precisely defined procedures, which are related on multiple levels (Figure 4) [4]. Behind each business process object there is a program code which is executed by starting the object. Figure 4 presents objects behind which there are program codes of the First run procedure which takes and places data from databases according to arrows (first) and the procedure which takes annual data from the warehouse and by means of SQL join, SQL target I SAS splitter procedures prepares them to accommodate them in data marts (sector data warehouses).
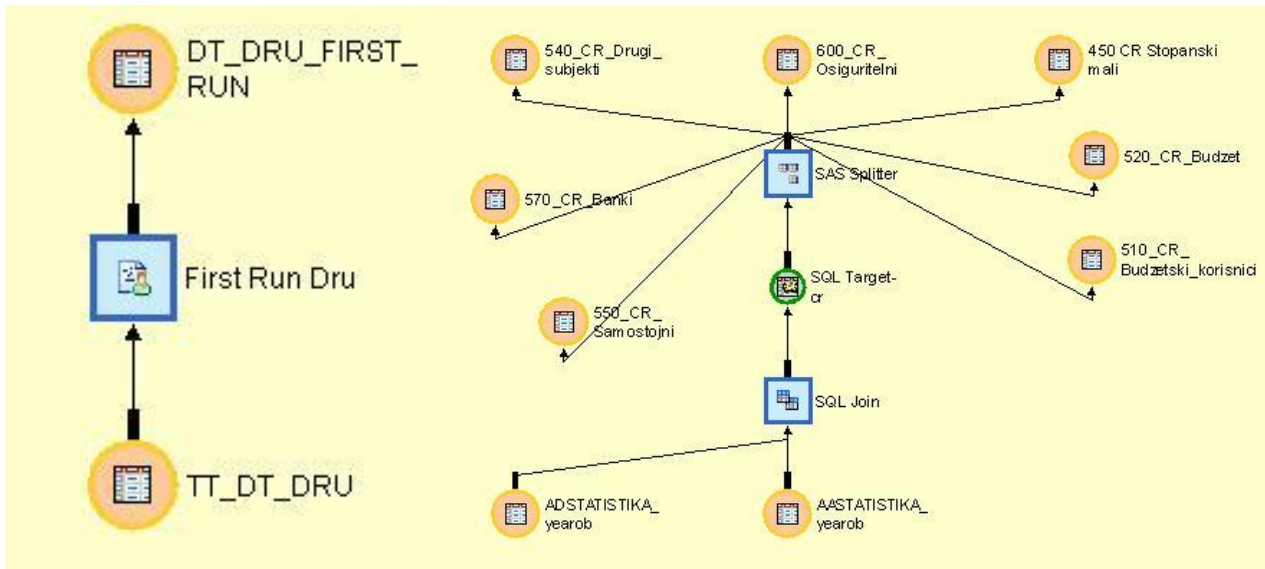
Figure 4   The business processes are defined objects behind which there is a program code
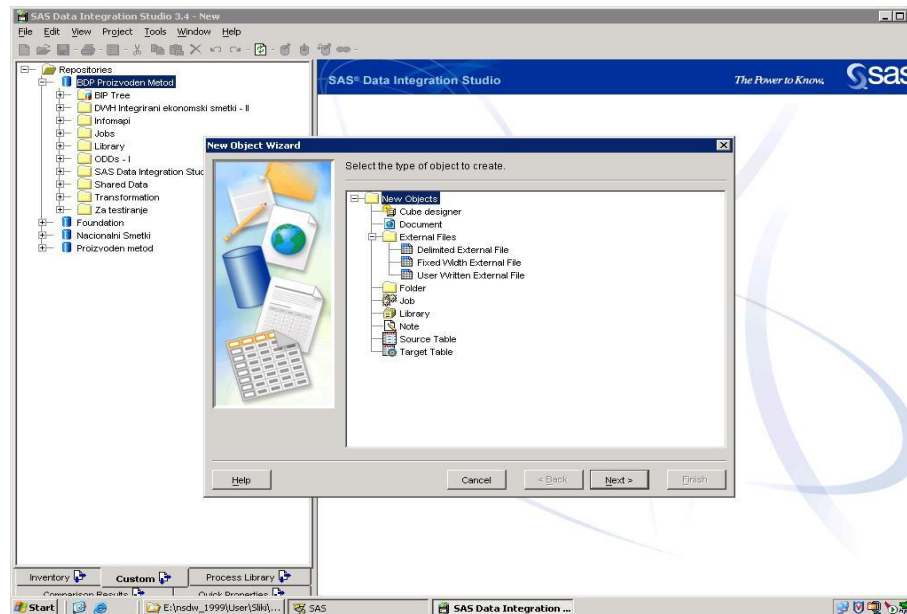

Figure 5 – Display of the screen of the SAS Data integration studio

As the data which are filled in DW of the SSO are to a great extent inconsistent with the business rules, even at its admission, the data are subject to control and necessary recoding into a format which has previously been defined with the strict business rules of the SSO (Figure 5) [6]. SAS Data Integration studio is the tool which makes it possible for data to be selected and placed at various locations and then to be classed according to dimensions. Such data are initially accommodated in temporary files, grouped according to time intervals and then are retrieved into the offical databases of the warehouse, being checked twice beforehand (according to the procedures) and recoded into a format previously defined for retrieving of DW. After this phase, two other procedures called First run (Figure 6) and Second run with which the data are accommodated in the parts of the warehouse according to certain algorithms for aggregation or reprocessing are being started. The possibilities for visual display of data and for connecting them in the first run process, make it possible to link certain data with parts of other data and place them in the second processing phase. If a need arises, certain corrections could be made by using Adjustment Analysis Tool within these phases (Figure 7) and finally, after starting the Final run

procedure, they are put in the specified form in the sector DW, which process them further for transparent dissemination to users via web or company Intranet.
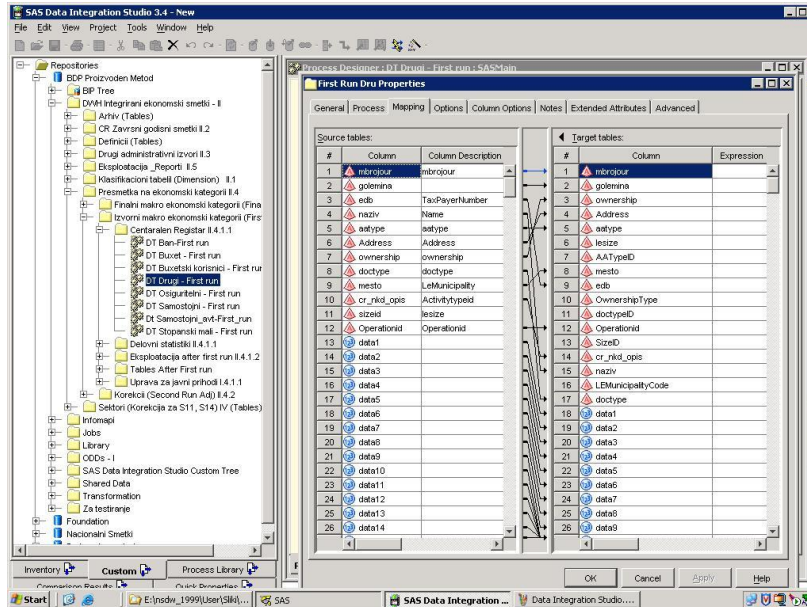

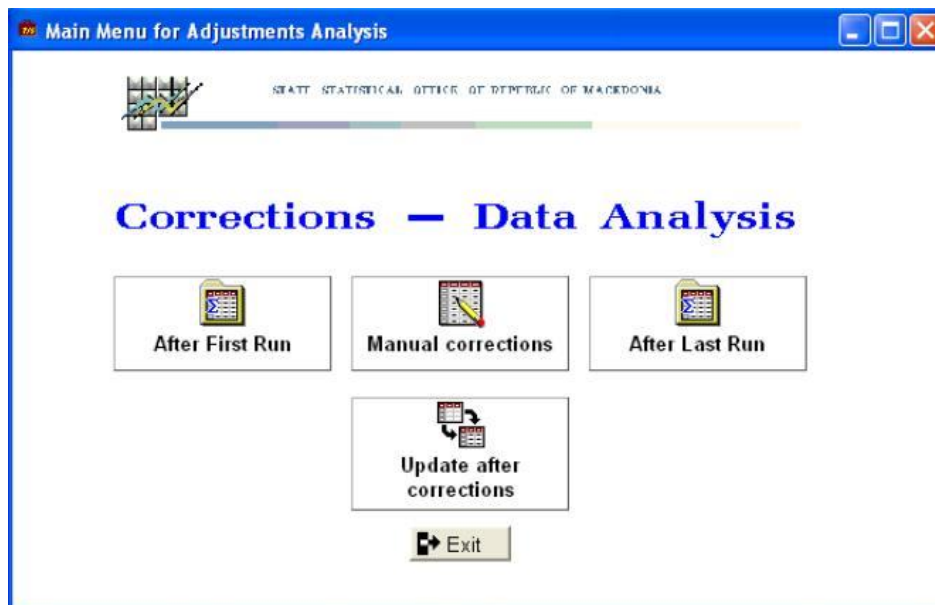Figure 6 – SAS Data Integration services – First run screen

Figure 7 - Screenshot Adjustment Analysis – Corrections – Data Analysis

8. Dissemination of data from DW

Although the dissemination of data is not carried out directly from DW to the Web, we should mention what types of information could be obtained via the automaticized mechanisms for retrieving data from DW. However, this task is not the administrators' task; it is a task of a special department within the SSO. Apart from publishing special annual publications of the SSO, the data are being disseminated via the SSO's website, stat.gov.mk. The website has been created upon Microsoft ASP.NET platform, and its content is quite elaborate and it offers opportunities for improvement. It offers a range of data which can be obtained by selecting multiple attributes – parameters for statistical analysis. A modern HCI, which is very user-friendly, has been provided. Data visualization and visual representations with geographic benchmarks have also been provided. By selecting a database, the needed data can be extracted mainly from the agregated tables according to the categorisation which has been carried out peviously with the procedures for data preparation and creation of agregated tables of DW.

PC-AXIS for Windows is the tool which has been offered for special formats compatible with all the European statistics [7] and which is utilized for data dissemination. The easy user interface makes the obtaining of various data combinations possible, which can be imported to Excel files and be recorded on the disk, where rows and columns could be erased or added, depending on the user's needs. The first selection is according to: population, foreign trade, census of the population, agriculture and statistics according to regions. Depending on the selected topic area, the performed selection of data from several offered categories could result in obtaining data presented in tables and graphic formats which can be exported to many formats (most frequently Excel format) and can be graphically presented as maps or diagrams. Figure 8 presents a diagram concerning the population according to the "step by step" approach.
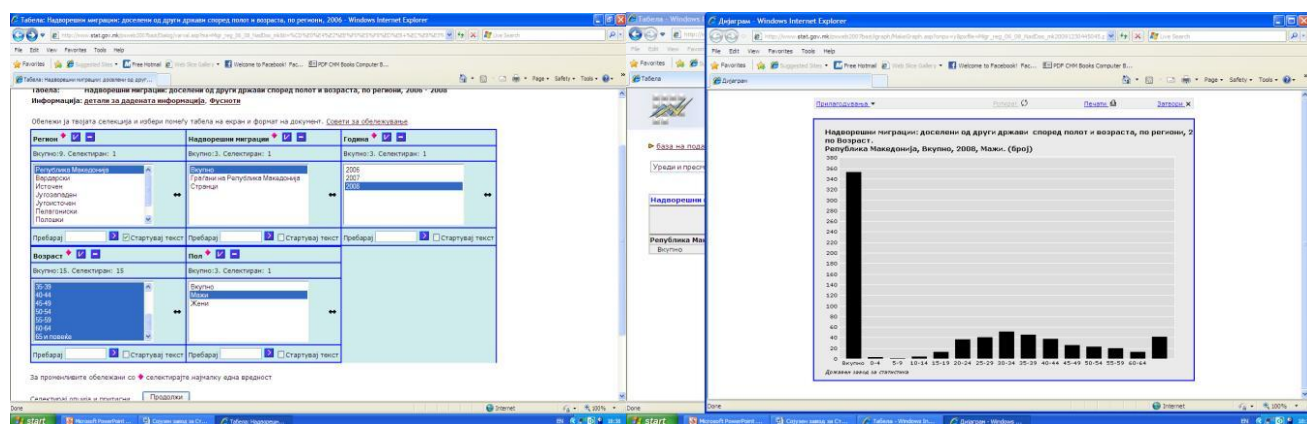

Figure 8 Interactive utilization of databases of the SSO via accessing the web approach to the website stat.gov.mk

 Visual presentations of regions, statistical data referring to particular features, topic areas according to various categorizations, all stem from the selections of databases of DW at the SSO's website, www.stat.gov.mk.

9. Conclusion

The State Statistical Office is an institution which is in charge of collecting, processing and analyzing data concerning relevant entities in the Republic of Macedonia, as well as data dissemination in adequate modes for gaining transparent insight. Because of these reasons the SSO utilizes professional software for DW with the necessary components for extracting, retrieving, cleansing, transforming, filling and processing of data from the data sources. In order to implement and use this complex structure, a high organizational structure of employees; as well as extensive knowledge of the business rules; precisely defined business entities and connections which should be transformed into business procedures, are required. These business procedures have been defined and automatized to a great extent via the Business Process Management tools and the defined objects behind which there is a program executive code.

SAS DW and the other SAS tools which are being used in the SSO are a powerful software tool which incorporates intelligence and a range of data obtained according to specific, detailed and harmonized procedures from dispersed, diverse,

cleansed and integrated data on the state level. Naturally, this concept requires great deal of discipline to collect, prepare and transform data as well as accurate data dissemination. Its capacity for complex statistical usages makes it one of the most desirable solutions for the complex systems of statistical institutes worldwide. However, the extent to which this software is taken advantage of, depends on the employees and the level of statistical knowledge of those who are in charge of those sectors.

Bibliography

[1] S.Savoska, V.Manevska, Business Intelligence tools for statistical data analysis, 32-nd International Conference of Information technology interfaces, ITI, 2010, Cavtat, IEEE, Region 8, [pages 199-204]

[2] G. M. Marakas, Modern Data Warehousing, Mining and Visualization, 2003, Indiana University, New Jersey

[3]Savoska S., Dimeski B., The Future Challenge: Data Warehouse as a Method of Data Integration from Public Administration Legacy Systems for Decision-making in Republic of Macedonia, 5th ISGT'2011 Proceedinngs, 2011, [pages 47-56]

[4] Napoli, J.P, K. Kaloyanova. An Integrated Approach for RUP, EA, SOA and BPM Implementation, Proceedings of the 12th International Conference on Computer Systems and Technologies, Vienna, Austria, 16-17 June 2011pp.63-68

[5] www.stat.gov.mk [2010/2011]

[6] SAS Institute Inc., SAS Data Integration server, USA, 2009; www.sas.com/offices, [2009/2010];

[7] Kozielski S., Wrembel R., New Trends in Data Warehousing and Data Analysis, Springer, 2009, [Pages 93-113];

[8] W.H. Inmon, Building the Data Warehouse, Fourth edition, Wiley Publishing. Inc., 2005