

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

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

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

Work Session on Statistical Metadata

(Geneva, Switzerland, 6-8 May 2013)

Topic (ii): Case studies and tools

**IMPLEMENTATION OF GSBPM AND METADATA STUDIES IN THE
TURKISH STATISTICAL INSTITUTE**

Working Paper

Prepared by Deniz Özkan, Nilgün Dorsan, Gülhan Eminkahyagil, Turkish Statistical Institute

I. Introduction

1. In this age of information, it is a known fact that effective management and development of states could only be possible with sound data. Without reliable statistics, it would be difficult to make short or long term plans and the policy makers would not be able to make sound decisions.

2. Nowadays, accurate, reliable and timely statistical data is very important to assess the current situation of communities, to make the right decisions based on the events taking place and to make informed plans about the future. The information provided to the public serves the communities as a strategic resource: providing necessary information for accurate and quick decision-making, for innovation and development required for information age, creating a strong statistical and information system infrastructure which is important for the present and the future of the countries. Without reliable statistics and a properly functioning statistical system it would not be possible to talk about a trustworthy future.

II. Institutional Background

3. Although statistics collection has a long history dating back to the 14th century in Ottoman Empire, a central organization to collect statistics, Turkish Statistical Institute (TurkStat), was founded nearly a century ago in Turkey in 1926. As the producer and coordinator of official statistics, Turkish Statistical Institute passed through important stages since 1926. Since its establishment, TurkStat is acting with a mission to produce and present high-quality, timely, reliable and impartial statistics by taking into account the needs and priorities of national and international users based on the international standards.

4. TurkStat consists of a central office and 26 provincial offices. It was structured as the Presidency of Turkish Statistical Institute and The Statistics Council, based on the Statistics Law of Turkey No 5429. The Statistics Council is established to advice on the development and implementation of the Official Statistics Programme (OSP).

5. TurkStat is responsible for coordinating the OSP. The OSP is prepared for a 5-year period in order to determine the basic principles and standards for the production and dissemination of official statistics and to produce reliable, timely, transparent and impartial data required at national and international level. A programmed period for the Turkish statistical system has begun with the 2007-2011 programme, and starting 2012 a new five year programmed period began.

6. This programme prevents repetitions in the production and dissemination of official statistics, decreases the burden on respondents, saves on labour force and resources and enhances the confidence in official statistics. Furthermore, the official statistics are standardised, responsible and related institutions are defined, data compilation methodology and the publication periodicity/schedule of official statistics are specified through OSP.

III. Statistical Metadata and the Statistical Business Process

A. Changes in the societies drive institutional changes

7. A new vision and new objectives for European Statistical System were laid out in the Communication Report on the production method of EU statistics. A vision for the next decade¹ report offers a vision for reforming the production method of European statistics. The main intention and strategic direction of this vision is to improve the efficiency of statistical production. It was mentioned in the report that the efficiency policy would enable the ESS to cope with the conflict between increasing information needs on the one hand and resource constraints on the other.

8. By knowing the statistics' central role in the information era, and in line with the developments in the international arena; TurkStat's Strategic Plan (2012-2016) was prepared based on the principles of quality. The vision stated in the Strategic Plan is "to establish a user-centered, sustainable statistical system based on the international standards". The vision is not only focused on statistical system of TurkStat itself but it also focused on the Turkish statistical system as a whole.

9. The progress can be observed and development can be made possible with effective use of information technologies; through production and dissemination of sound and reliable data. Statistical organisations are the main data providers and the society demands more and relevant data much sooner than before. In this context Turkish Statistical Law came into force in 2005 to design the Turkish statistical system by incorporating the quality principles accepted by the international organisations. In accordance with the Statistical Law and the OSP, 63 organizations/institutions were included in the statistical system to make the production of statistics more effective and efficient than before.

10. In 2012, a new organisational structure formed within the context of the Generic Statistical Business Process Model (GSBPM) was implemented in TurkStat. The GSBPM analysis showed how the organization should be structured around the processes to function efficiently. Structuring of organization in accordance with the production flow will optimize the production processes and reduce the cost. New departments were established and the functions of some of the existing ones changed after the reorganisation.

¹ . Communication From The Commission To The European Parliament And The Council On "The Production Method of EU Statistics: A Vision For The Next Decade", Brussels, 10.8.2009, COM(2009)

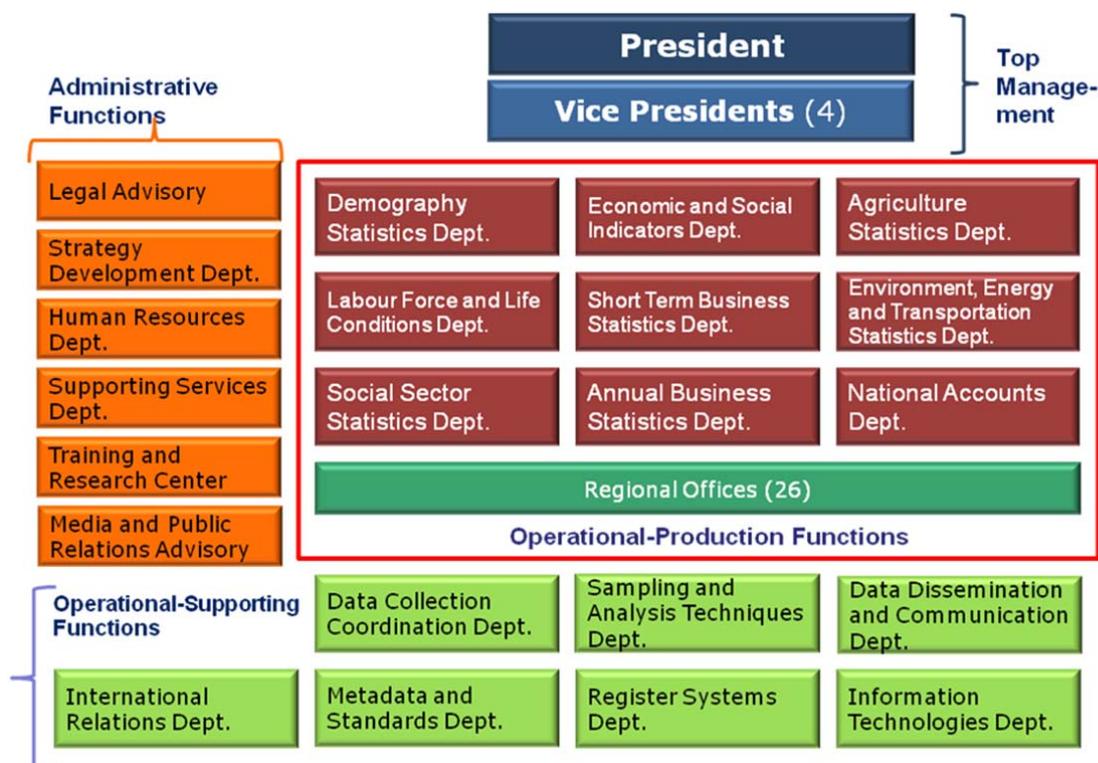


Figure 1: Organizational structure of TurkStat

B. Metadata strategy

11. Statistical production process as a whole is currently being discussed in many platforms. The main aim is the same for all statistical offices: the necessities of being in the information age, the need to produce consistent and relevant statistics, establishment of strong metadata system within statistical business processes are the main areas of concern when building and redesigning the infrastructure of the statistical system. A metadata infrastructure which is comparable with the international standards and the reusability of these standards were important starting points for TurkStat.

12. TurkStat aims at standardising the processes and the data infrastructure, which will serve sustainability and flexibility of the organization. TurkStat had a paradigm change in recent years by replacing the stovepipe production model of statistics by an integrated model. The reason for this change is quite clear. That is, the need for information continues to grow and the resources to meet this increasing demand are limited. The challenge for NSIs lies in meeting the increase in demand for quality and timely information with less cost; without increasing the costs for producing statistics, and without putting too much burden on respondents as well.

13. An action plan was prepared for the standardization of data and the information by making an analysis of the needs of TurkStat. The aims of the action plan are;

- To supply the documentation in every processes
- To prepare the infrastructure for the development of metadata system
- To standardize code lists used the production process
- To make the management and follow up easy by centralizing the metadata system,
- To strengthen the institutional memory,
- To prevent reiterations in business processes and increasing the reusability of the standardized processes

14. Metadata studies gained speed in TurkStat since 2009. A work has been actively going on to customize GSBPM according to TurkStat's business processes, and to customize and use the new

metadata editor (using DDI) to eventually be linked to the survey applications. With the recent reorganization of TurkStat, the studies carried out were merged under a new department called Metadata and Standards in 2012. The new department is responsible for coordinating the efforts to establish a comprehensive metadata system in TurkStat.

15. Currently TurkStat has more than 100 applications and each one is separately designed. The terminology used is not consistent. For a sustainable institutional memory it was inevitable to structure the metadata system as a whole.

C. Statistical business process model

16. GSBPM has been adopted as a reference model to bring together the metadata that get created in all phases of production. A project started in 2010 in TurkStat to define the processes for all statistical products produced in TurkStat.

17. In order to specify the processes in TurkStat, four main operational processes were analysed first: collect, process, analyse, disseminate. Five products were selected to be analysed for the pilot study of the GSBPM model. These products were (i) short term industrial statistics, (ii) household budget statistics, (iii) crop production statistics, (iv) producer prices, (v) consumer prices. Process maps and work flows for these products were drawn. It was seen from the pilot study that the other 3 processes (specify needs, design, build) can be clearly identified too. Based on the process flows of the pilot study a national draft model of GSBPM was developed.

18. This draft model was used as a basis for process modelling and standardization of the remaining products. Approximately 250 statistical products were analysed. Within the scope of the process modelling and standardization project, meetings were held with all units producing statistics to obtain information about the activities done to produce these statistics. Procedures/methods that are used, step by step work flows, inputs and outputs of sub processes, owners of the processes, the number of people required to do these tasks, software tools and other relevant metadata were collected from the subject matter departments.

19. All the process information collected from the departments is currently being analysed, grouped and standardised. The information objects are going to be identified. Different procedures and algorithms used in different products are going to be listed. The draft model will be finalized after the process analysis was completed for all products and processes.

Draft Statistical Business Process Model

1.Specify Needs	2.Design	3.Build	4.Collect	5.Process	6.Analyse	7.Disseminate
1.1.Determine need for information	2.1.Design statistical products and outputs	3.1.Build and enhance production system components	4.1.Establish frame and registers, select sample	5.1.Classify and code	6.1.Evaluate the information for its effect	7.1.Update dissemination systems
1.2.Consult and confirm need	2.2.Design frame, register and sample methodology	3.2.Integrate production system with other systems	4.2.Set up collection	5.2.Micro-edit	6.2.Produce statistics	7.2.Produce dissemination product
1.3.Establish output objectives	2.3.Design data collection methodology	3.3.Test production system	4.3.Run collection	5.3.Macro-control	6.3.Quality assure statistics	7.3.Manage publishing for dissemination product
1.4.Check data availability	2.4.Design process and analysis methodology	3.4.Finalise production system	4.4.Finalise collection	5.4.Imputation	6.4.Examine and evaluate statistics	7.4.Manage user demands
1.5.Determine business plan	2.5.Design production system and workflows			5.5.Calculate weights and derive variables	6.5.Prepare statistics for dissemination	
					6.6.Finalise content	

Figure 2: Draft model used in Turkstat

20. The standardisation of the processes needs to be carefully managed and in fact this process takes a significant amount of time. When someone wants to analyse a survey data, the data can be easily put into a software programme (SAS, Oracle, SPSS, etc.) and queries, crosschecks and other analysis could easily be made. However, when it comes to analysing the processes, steps, inputs and outputs of the processes etc. one faces a different structure. After putting all the process steps collected from various departments into one document, we ended up tens of thousands of lines with different expressions and wording. Analysing all these lines of information, putting job steps, inputs and outputs into a sub level in the model is the main point and the main challenge of the standardisation. While examining these, we saw that some work flows were not properly written, some parts were incomplete or missed. After the information provided by departments is completely mapped to the model the gaps would be seen more clearly. Both the issues in the model and the issues in the information collected from the departments will be addressed as a result of these mappings and the analysis.

21. We haven't included evaluate and archive processes in our draft model. However, evaluate and archive levels will be put into the final model to complete the whole production system in the model.

The draft model will be finalised in June 2013 and the final version will be distributed to the departments to be checked over. Then the final version of the model will be put into a software programme to develop the production system architecture.

D. Reference and structure metadata

22. Metadata was captured in various processes of statistical production in TurkStat, however an integrated metadata system or a central metadata repository to hold these metadata did not exist in TurkStat.

23. Reference metadata about the products were published in the Turkstat's web page, mostly in SDDS format. Definitions of variables and concepts used in surveys were published in the web page for each product separately. There is a classification server which contains the national and international classifications.

24. In 2012, a decision was made to keep the metadata in one place. For this purpose Metadata Editor1.2 was chosen as a new tool to catalogue the survey metadata in TurkStat. A working group was set up in 2012 to customize the generic template according to our specific needs. The group modified the template, standardised the procedures about how the metadata is going to be entered to the template and code lists. Institution wide training was provided to all users. This working group also extracted the code lists used in all business surveys, household surveys, agriculture and environment surveys, and tried to standardise the codes within each group of surveys as a first step. Then these code lists were brought together for all surveys. A catalogue of standard internal code lists was added to the web page for internal use only (figure 3). The variable names were also listed and they are tried to be standardised according to the internally defined standard naming conventions. A catalogue of variable names was created for common variables.
25. Standardization in a narrow sense began with the code lists. The studies were begun with the 12 pilot projects and now the codes were standardized for 132 projects in the production database. The harmonization of the codes in the institutional and dissemination databases is a work in progress.
26. The metadata editor consists of document and study description sections where the reference metadata about the study to be entered. There is also datasets and variable groups sections in the editor in which the structural metadata of the surveys can be entered (figures 4 and 5).
27. The metadata collected using the new editor is used only internally at this point, i.e. not used for dissemination purposes.
28. Another functionality of the new metadata editor is the potential usage of the data structure defined in the metadata editor in the collect process (figure 5). The structural metadata (datasets and variable groups) entered by the departments into the editor would be used as an input to create the data entry tools built by the IT department. This way the data structure defined in the metadata editor would be directly used in one of the components of the production system.
29. The only major downside of the metadata editor is it does not contain a section for concepts and definitions. A tool to catalogue the concepts and definitions needs to be created and integrated into the metadata system.
30. For internal users and methodologists, the methodology documents, handbooks etc. were usually kept in the departments by individuals, and it is up to the people working in these departments to save, archive and share these documents. In some cases the procedures are not fully documented, only parts of the processes are documented here and there. The new metadata system brings together most of this documentation (see figure 4).

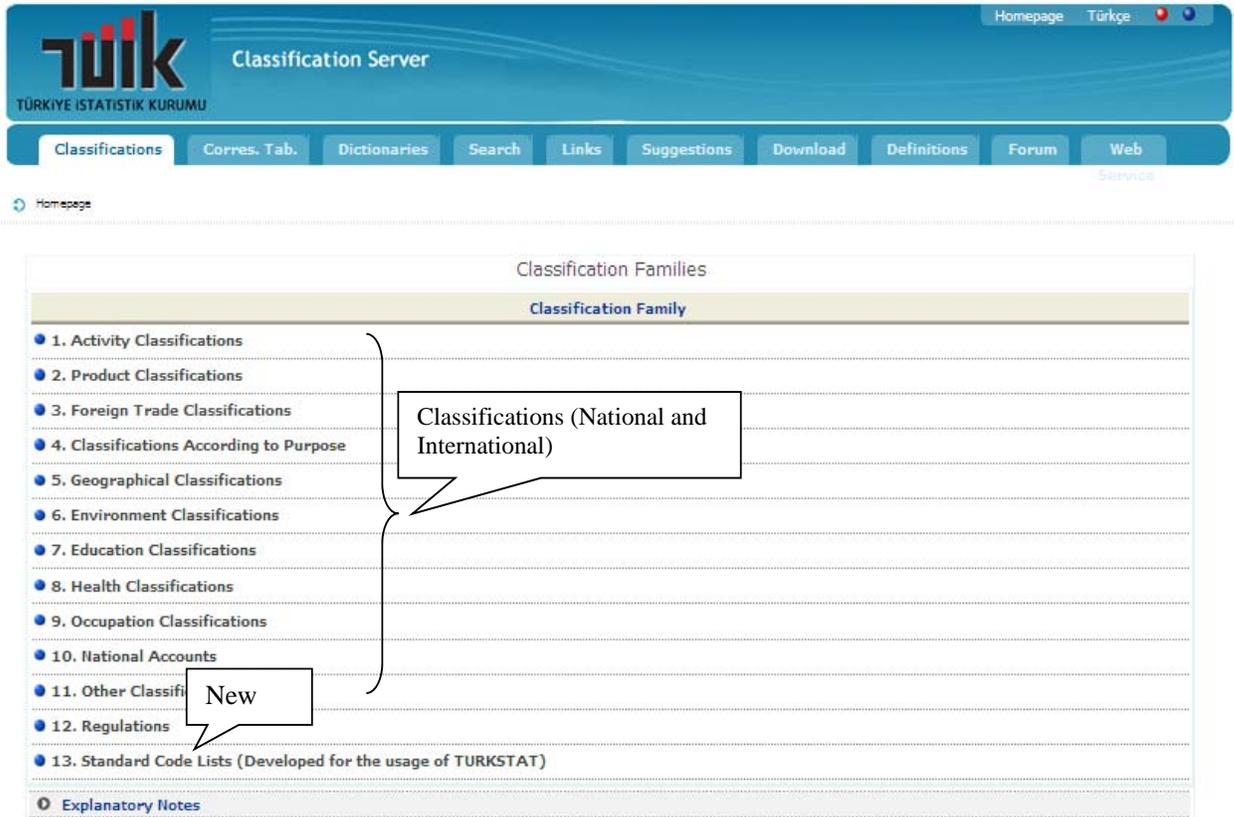


Figure 3: Classification server, including a new section for the standard code lists

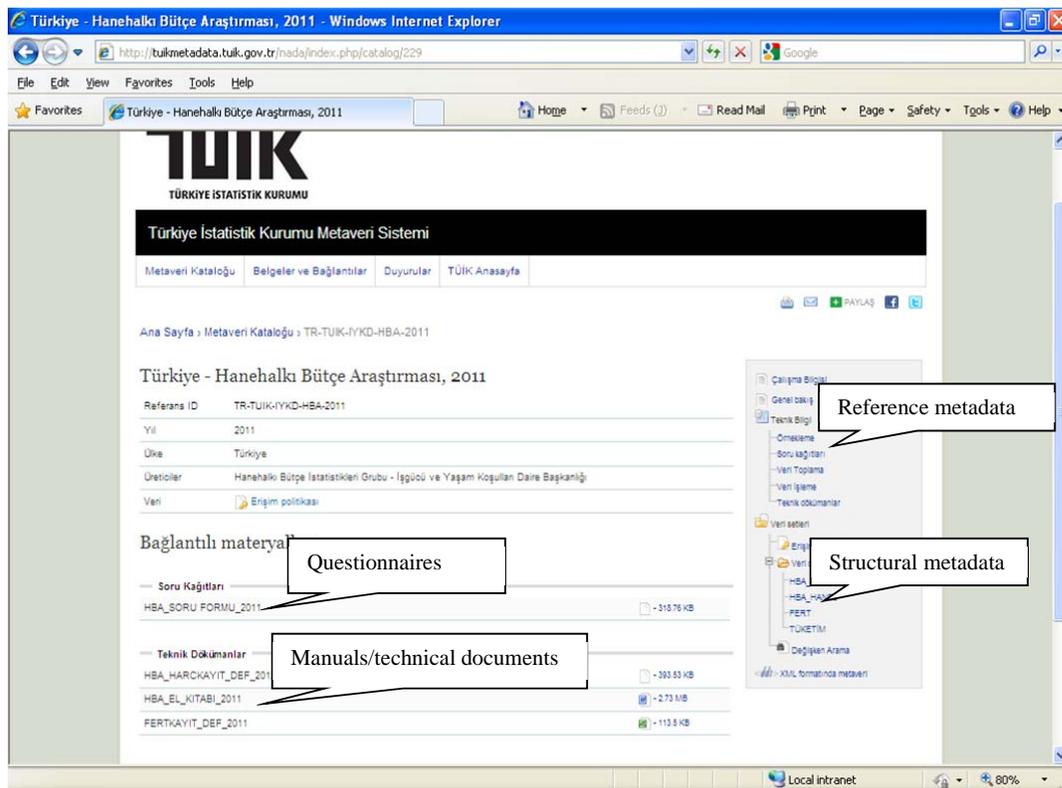


Figure 4: TurkStat's new metadata catalogue

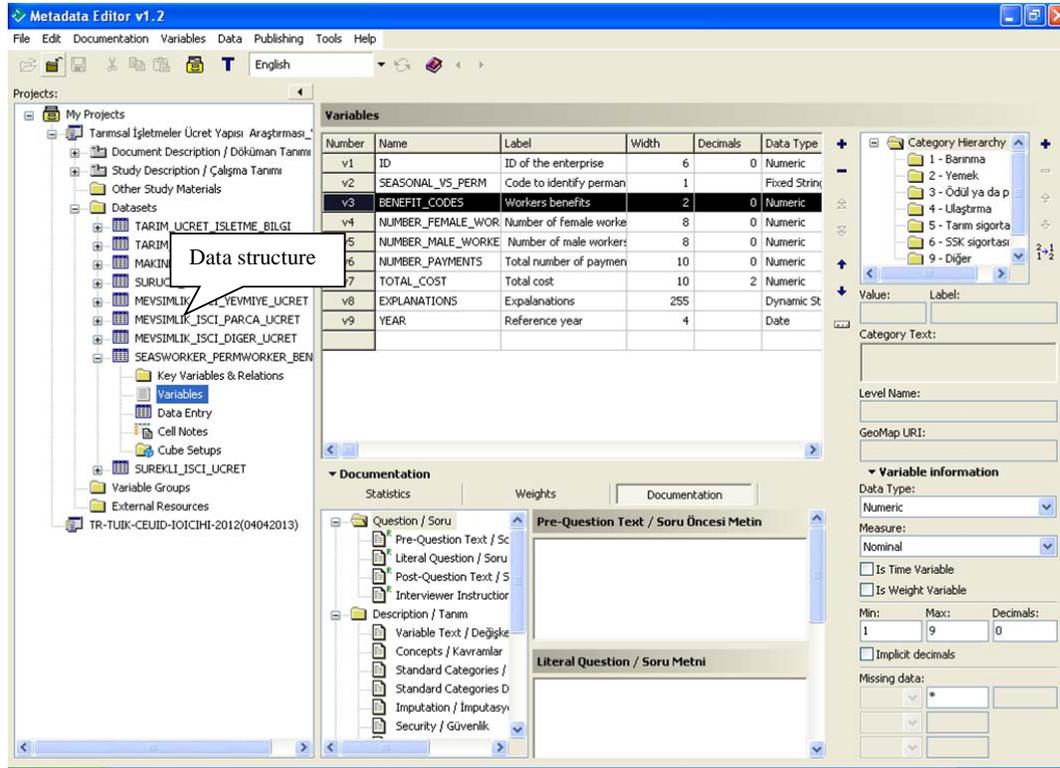


Figure 5: A snapshot of the dataset from the metadata editor

IV. Conclusion

31. The national statistical offices and international organizations are constantly developing the standards and guidelines for statistical metadata. As the statistics production is becoming more industrialized; the metadata systems are becoming important tools for NSIs to ensure the establishment of common processes, to create common systems and a common terminology. In that sense, GSBPM has become a very useful to create a common terminology within the statistical offices, as well as between the NSIs and international organizations.

32. Turkish Statistical Institute adopted GSBPM to define the statistical production processes and the metadata. The objective is to design a new metadata system around GSBPM for the efficient functioning of the business processes. A draft statistical business process model was prepared by TurkStat and the metadata for the processes were collected from the departments. This process metadata is going to be organised and standardised and the objects are going to be identified and the draft model will be finalized based on TurkStat's internal processes. A new metadata editor is being used to collect statistical metadata in a central location. Although this editor is used only internally right now, it will soon be used for dissemination of metadata to the external users as well.

33. A plan was put into action in order to:

- establish the infrastructure for the development of a comprehensive metadata system,
- centralize the metadata system, create a central metadata repository,
- define the production systems,
- supply the documentation for every process,
- strengthen the institutional memory,
- standardise classifications, internal code lists, variables,

- prevent reiterations in business processes and increase the reusability of the standardized processes,
- specify KPIs for each process for monitoring and evaluation,
- review processes and optimize the process workflows.

V. References

Turkish Statistical Institute (2012) “Official Statistics Programme, 2012-2016, (General Information and Tables)”, January 2012, <http://www.turkstat.gov.tr/rip/rip.pdf>

Turkish Statistical Institute (2012) “Strategic Plan, 2012-2016”, August 2012
http://www.turkstat.gov.tr/stratejik_plan/str2012.pdf

UNECE Secretariat (2009) “Generic Statistical Business Process Model: Version 4.0 – April 2009”, Joint UNECE/Eurostat/OECD Work Session on Statistical Metadata (METIS)

COM(2009) “Communication From The Commission To The European Parliament And The Council On The Production Method of EU Statistics: A Vision For The Next Decade”, Brussels, 10.8.2009
<http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2009:0404:FIN:EN:PDF>