WP 19

27 January 2010

ENGLISH ONLY

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 (METIS)
(Geneva, Switzerland, 10-12 March 2010)

THE VALUE OF ADOPTING AND IMPLEMENTATION OF ESMS STRUCTURE IN MACEDONIAN STATE STATISTICAL OFFICE

Submitted by the State Statistical Office of the Republic of Macedonia[1]

I. INTRODUCTION

Importance and role of reference metadata

1. The vision of the State Statistical Office of the Republic of Macedonia (SSO) is to be recognized as an institution that provides quality, timely and internationally comparable statistical data. The values that the SSO upholds regarding the institutional environment, statistical processes and statistical outputs are the 15 principles from the European Statistics Code of Practice. But not to be only declarative, we need means to measure and compare improvements. In this regard, reference metadata are used for better recording and comparison of the data quality reached, by each survey and over time.

2. In particular, there is a bi-directional linkage between statistical metadata and quality. Defined as integral and essential part of completeness of statistical metadata, reference metadata describe the quality of statistics and vice versa, statistical metadata are defined as a quality component that aims at improving the availability and accessibility of statistical data. According to various reference metadata frameworks (ESMS, SDMX, OECD, IMF, SDDS) reference metadata should include conceptual metadata, methodological metadata and quality metadata. Most of these metadata already exist at the SSO, but in separate metadata repositories. The purpose of this paper is to describe the design and implementation of a central database for reference metadata, based on ESMS structure.

[1] Prepared by: Helena Papazoska (helena.papazoska@stat.gov.mk)
Biljana Ristevska Karajovanovikj (biljana.ristevska@stat.gov.mk)
Slavica Lipikj (slavica.lipic@stat.gov.mk)

II. IMPLEMENTATION OF EURO-SDMX METADATA STRUCTURE FOR INCREASING THE AVAILABILITY AND REUSABILITY OF REFERENCE METADATA

Adoption of Euro-SDMX Metadata Structure in the SSO

3. In 2007, Eurostat launched a new standard for reference metadata called "Euro-SDMX Metadata Standard" (ESMS) containing a list of 21 metadata cross-domain concepts, accompanied by a limited number of statistical sub-concepts. Commission Recommendation of 23 June 2009 on reference metadata for the European Statistical System was a step forward in advancing a mission of providing international standardization and harmonization in exchanging and sharing of statistical data and metadata, in particular for reference metadata.

4. Reference metadata in the SSO exist in various forms of shapes and storage; in particular there was no officially promulgated and adopted standard at the organizational level. This has simplified the process of adoption of the Eurostat recommendation for implementation of ESMS structure at the national level.

5. Before actual implementation of ESMS has started, in order to achieve better harmonization, SSO's three-level Theme/Statistics structure was converted into five level structure. The old Theme/Statistics structure consisted of: Statistical Area, Statistical Sub-area and Statistics. The revision was done in accordance with the Eurostat Compendium structure: Domain, Theme and Module. Furthermore, a breakdown was made of the module level into two additional levels: group level and level of statistics.

6. The core of the SDMX model of reference metadata is the concept of "Metadata Structure Definition". Reference metadata may be attached to different object types (a data set, a time series, an observation…) and Metadata Structure Definition also identifies the object the metadata are attached to (Concept 21). For the time being they are attached to a level of statistics, with an intention in the future to make it at the highest level - group level or even at a module level where appropriate.

7. Methodological clarification of ESMS structure was done and the impact on existing reference metadata was analyzed as well. Although the existence of description and ESS guidelines in the ESMS structure is very helpful, at a conceptual and sub-conceptual level we attached an additional explanation available from another source in order to provide enough information for correct interpretation by subject-matter staff responsible for different statistics in the Office. All documents were translated into/from Macedonian language.

Development of Euro SDMX Database and Application

8. The practical side of the work done in the SSO since June 2009 resulted in the development of, as we call it, ESMS Database and ESMS Application. The paper briefly outlines the experiences and the work carried out to develop the system.

9. The Figure shows the overall conceptual model of the system. It is divided into three sectors inter-linked among each other: Organizational Structure, Module-Group/Statistics Structure and ESMS Reference Metadata Framework. The system is organized as a comprehensive structured collection of ESMS concepts for the Module/ Group/ Statistics structure, and in particular for the level of statistics aiming at documenting methodologies, quality and the statistical production processes in general.

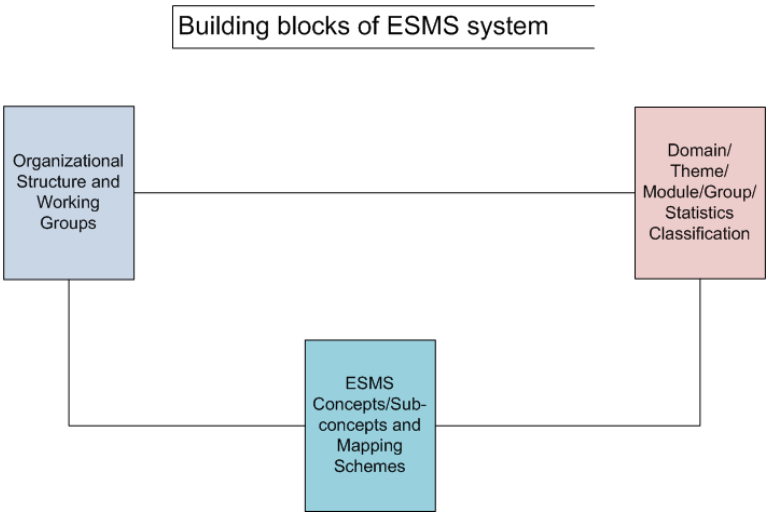

Figure 1.1 – ESMS DB Basic scheme

10. Figure 1.2 is a composite view of the ICT infrastructure of the ESMS system.

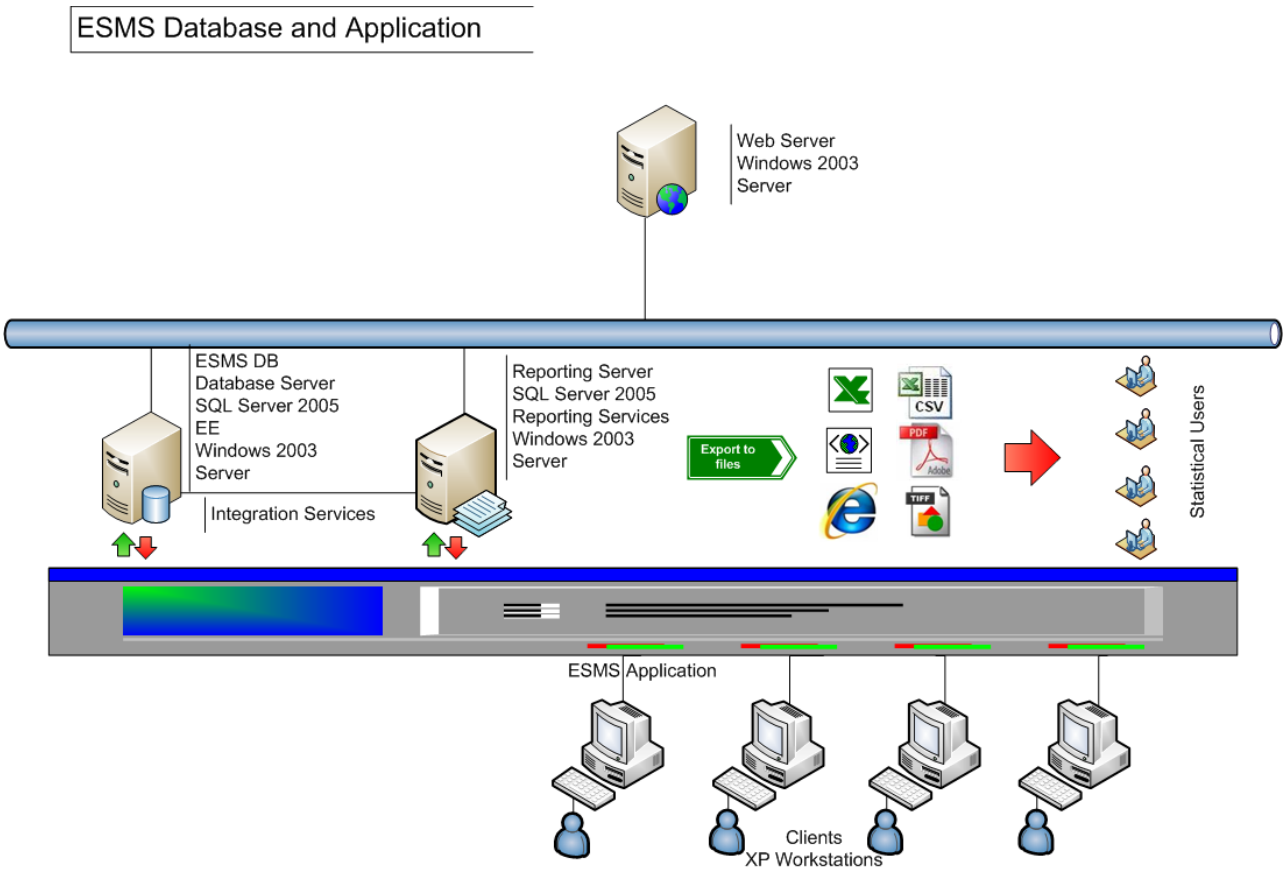

Figure 1.2 – ESMS IT System

11. The database is organized in a relational manner. It is created on SQL Server 2005 platform. The EE of this Microsoft Data Platform is installed on Microsoft Windows 2003 Server, and it is the vision of the organization to manage any data concerning metadata and output databases, at any place and any time. Microsoft SQL Server 2005 Reporting Services provides a complete, server-based platform designed to support a wide variety of reporting needs, enabling delivery of relevant information for statistics where needed across the entire Office. Security enhancements implemented

at the database level are: user schema separation and database roles and programming capabilities as well - stored procedures and user-defined functions.

12. History and version handling mechanism is implemented. History management makes it possible to keep consistency and quality among different versions of reference metadata. It is very important to distinguish how many versions we want to keep: either all or milestone only, in order to reduce uncertainty when doing comparisons between different versions. Only significant changes are denoted as new versions and only one version is current and active. The previous one, which is the base for the new one, is not active any more. Besides activity information each version is accompanied with additional information: user information, date and time of modification, type of modification (initial data entry, modification due to technical correction, modification due to approved and significant correction and deletion). In reference to this, modifying "shared" objects must be carefully managed since it affects the components, the model itself and any previous data being shared.

13. Referring to the application, two-tiered client-server architecture is implemented. Microsoft .NET and Microsoft Visual Studio 2005 are used as a development platform. It is a desktop multi-user application (VB 2005 App) with interface for data entry, data editing and data review.

14. The application has multi-lingual concept and automatically switches keyboard language and layout as needed from a user perspective in order to support Cyrillic code page. Regarding the fact that reference metadata is mainly free text, it has been discussed about using formatted text (styled text, rich text), as opposed to plain text. The advantages of the first one are that it has styling information beyond the minimum of semantic elements: colors, styles (boldface, italic), sizes and special features (such as images, tables, hyperlinks). As enhanced features, we plan to introduce formatted text capabilities and to develop advanced searching capabilities.

15. The application is user-friendly and flexible and it provides user control via intuitive structure of information; however, a user manual for application usage is also available. First usage of the ESMS application was done by five users and five pilot statistics from different statistical domains were documented: Foreign Trade Statistics, Household Budget Statistics, Marriages Statistics and from Industry Statistics: Monthly Industry Survey and Turnover and New Orders in Industry. Figure 1.3 displays ESMS Application user interface.

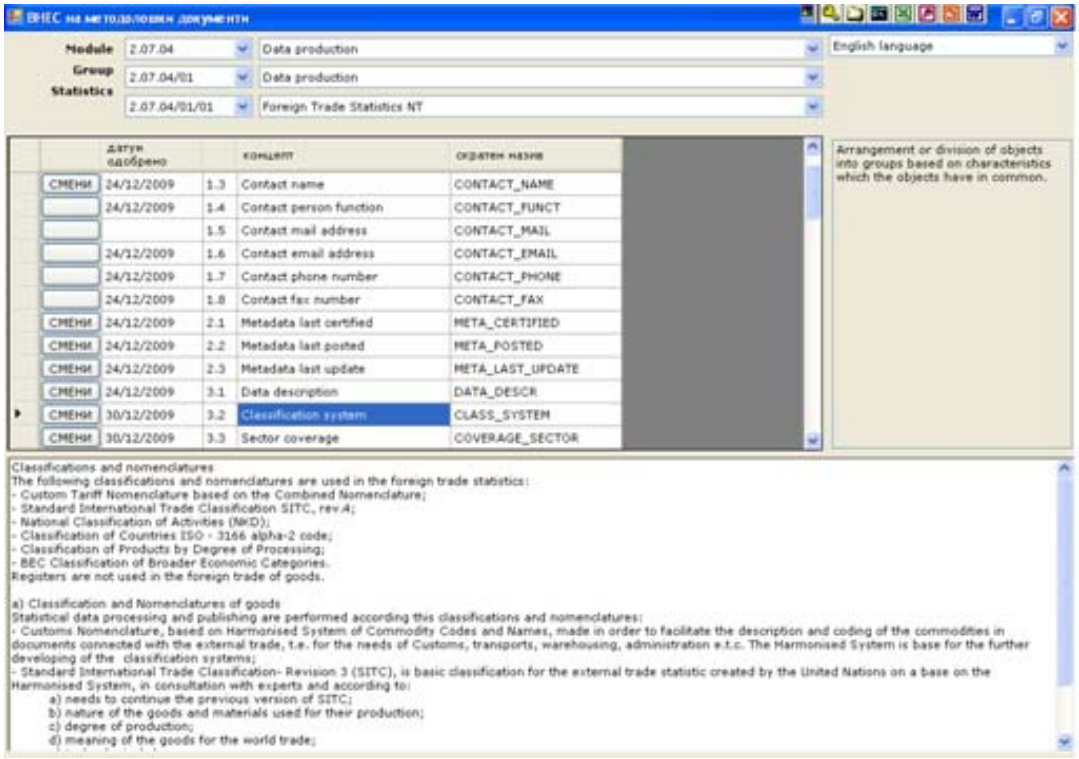

Figure 1.3 – ESMS Application Interface

III. MEETING VARIOUS QUALITY REPORTING REQUIREMENTS

Adoption of ESS Quality Framework in the SSO

16. In early 2009 new versions of the ESS Handbook on quality reporting (ESHR) and ESS Standards for Quality Reports (ESQR) were published in order to promote harmonized quality reporting across statistical processes and their outputs. The SSO immediately recognized the value of this handbook for supporting internal self-assessment, for reporting to Eurostat and also for user-oriented quality reporting because it puts considerable emphasis on output quality.

17. Many of the quality and performance indicators from the ESQR are clearly identified in the ESMS structure. The Euro-SDMX Metadata Structure (ESMS) consists of 21 high-level concepts with appropriate sub-concepts. It incorporates 21 SDMX cross-domain concepts and also reuses a significant part of the previous SDDS standard. More than one-third of ESMS concepts are directly related to data quality, such as: relevance, accuracy, timeliness, punctuality, accessibility and clarity, comparability, coherence.

18. Actually, all of the above-mentioned quality criteria are enumerated in the EP/EC Regulation No 223/2009 of 11 March 2009 on European statistics. By integrating the quality criteria from Article 12, ESMS guarantees certain quality of results because with the ESMS implementation, statistics will be developed, produced and disseminated on the basis of uniform standards and harmonized methods. Of course, the compilation of ESMS based user-oriented quality reports should be extended with producer-oriented (i.e. aimed at internal assessment of process and output quality) in order to fulfil quality reporting standards in accordance with ESQR.

Increased efficiency in monitoring the quality of statistics and preparing reports

19. The SSO devotes a lot of time and energy to reporting to international organizations. The ESMS structure includes data quality information and once statistics are documented, one can easily build predefined quality report just by using concepts and sub-concepts related to quality. The most important benefit of having a central database of reference metadata (ESMS Database) is its exploitation for reporting.

20. Reports could be produced in different shapes, depending on the topic (for internal assessment, for submission to corresponding Eurostat units or other international organizations, for dissemination purposes, etc.). They can also contain all or just part of the concepts and appropriate sub-concepts (Figure 2.1). The compatibility and possibility of mapping between ESMS and SDMX cross-domain concepts, IMF-DQAF, and OECD-Metastore metadata frameworks, allow users to quickly and accurately build different reports for different organizations. The concept of re-usability means an increased efficiency over a period of time. Report Manager is included in Microsoft SQL Server Reporting Services and from a user perspective it is used as a report viewer and provides user interface to a report server where reports are deployed. The report layout design is very similar to Eurostat ESMS reports. We implemented, or rather tried to implement, the proposed mapping documents of ESMS with SDMX Cross-domain Concepts, IMF-DQAF, and OECD-Metastore. It was quite a challenging task because different types of mapping (1:1, 1:n and n:1) existed and at different levels as well. For the time being, the European Standard Quality Report (ESQR) has not yet been implemented because of not clarified mapping. Figure 2.1 and Figure 2.2 display different reports produced on the report server.

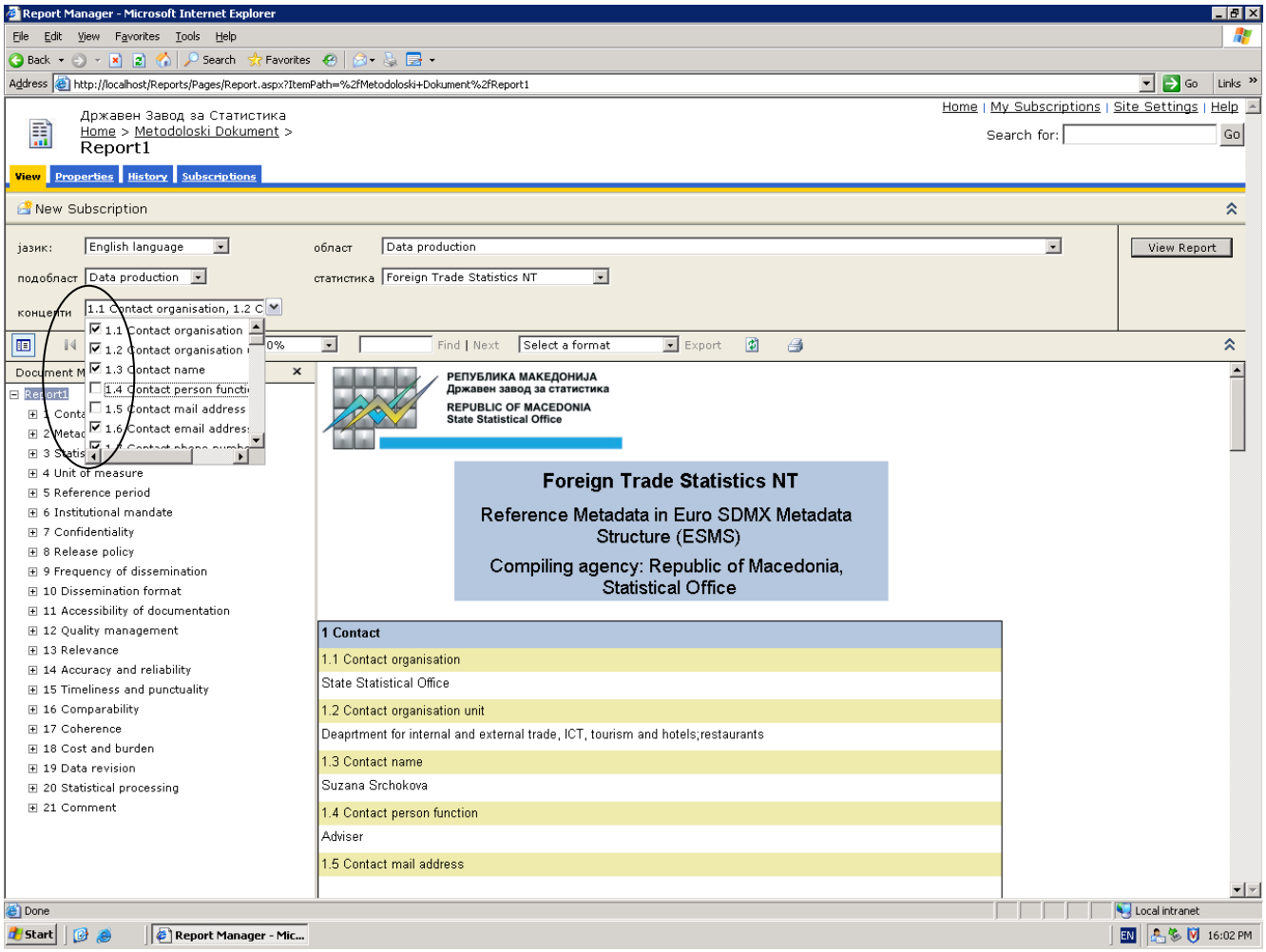

Figure 2.1 – Production of ESMS metadata report for Foreign Trade Statistics

21. The statistical users have the option to export reports in various printable formats (xls, csv, xml, pdf, tiff, web archive) for internal assessments, documentation and dissemination as it is shown in Figure 2.2.

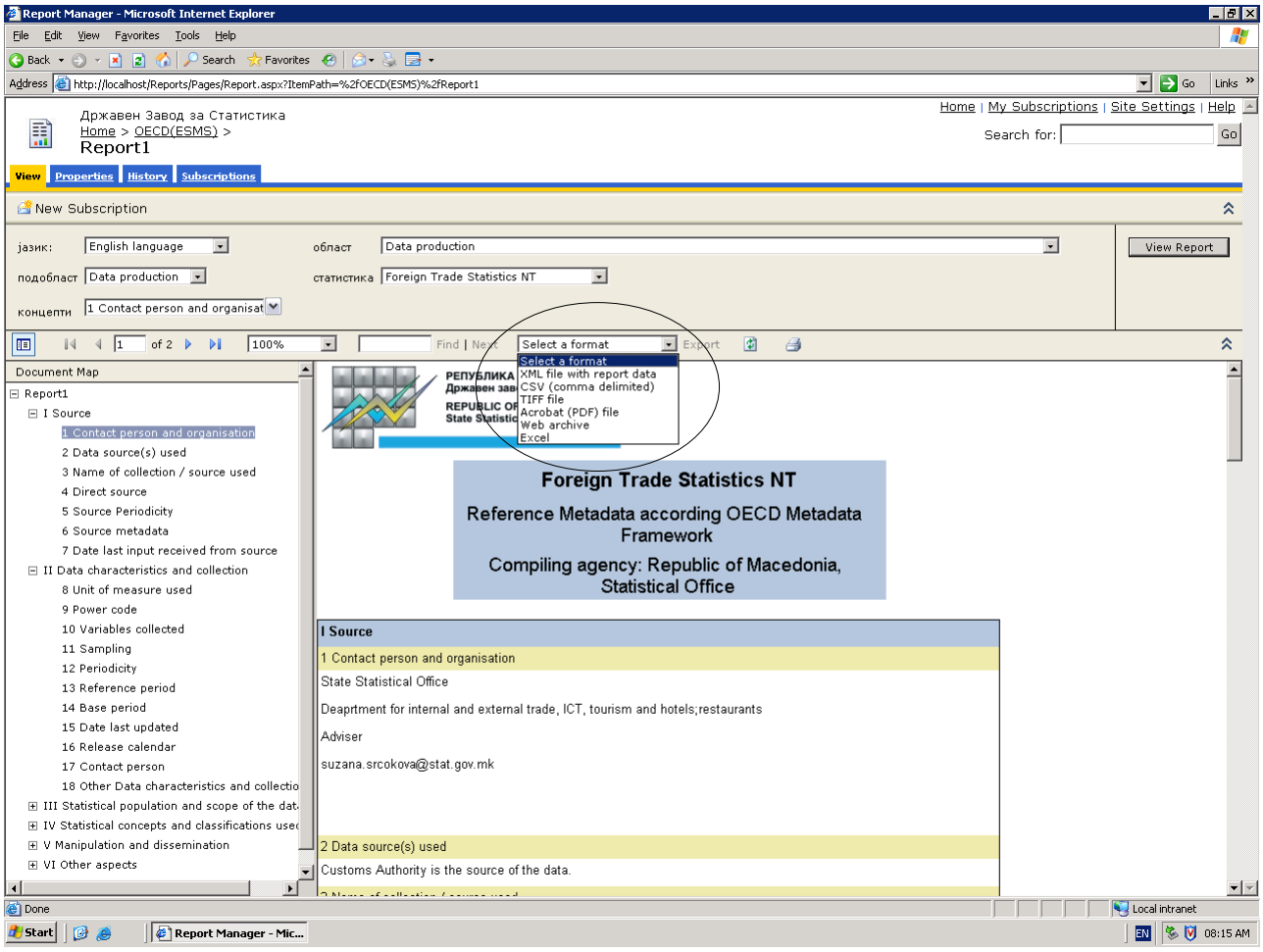

Figure 2.2 – Production of OECD metadata report for Foreign Trade Statistics

IV. MANAGEMENT AND ORGANIZATION FOR ESMS IMPLEMENTATION

Teamwork approach in SSO

22. In general, teamwork as a high level of collective performance is a desirable way of work and the SSO is committed towards building a teamwork culture in the Office. In order to provide relevant information on all the concepts and sub-concepts from the ESMS structure, different types of teams and Working Groups (WG) are established: Metadata Working Group (MDWG), Statistical Confidentiality Committee, WG for analyzing respondents' burden, WG for implementation of the cost calculation system, WG for statistical outputs quality. High-level management commitment and support is a key factor that contributes to the success of the cross-domain projects at the SSO.

23. MDWG Group was established under the auspices of the Managerial Board to promote and coordinate statistical metadata activities in the SSO. Dealing with metadata is a multidisciplinary issue, so it has been decided that a cross-domain working group should be established as an "umbrella" over all the activities. The Director General is the leader of the working group, which means that there is support from the highest authority in the SSO, and the development of a central metadata repository has been defined as one of the highest priorities at the SSO.

24. Members of the MDWG are from different job positions and different organizational units: from subject-matter departments, IT staff and staff dealing with quality issues at the SSO. MDWG is responsible for:

- Development and implementation of centralized storage of reference metadata for statistics;
- Dissemination of reference metadata;
- Documentation, standardization and harmonization of SSO's Business Process Model;
- Implementation of the new BPM in the SSO environment;
- Standardization, documentation and variables inventory;
- Establishment of harmonized variables repository.

V. INFORMATION SYSTEM ARCHITECTURE FOR ESMS IMPLEMENTATION

Information system architecture for reference metadata

25. The Statistical Information System architecture for reference metadata is shown in Figure 3.1. The architecture is modular and composed of relatively self-contained mutually linked subsystems as presented in the figure. Unified approach is assured with a specific main principle, which is consistency of metadata inside the subsystems and between them, and with unified technological tools for implementation.

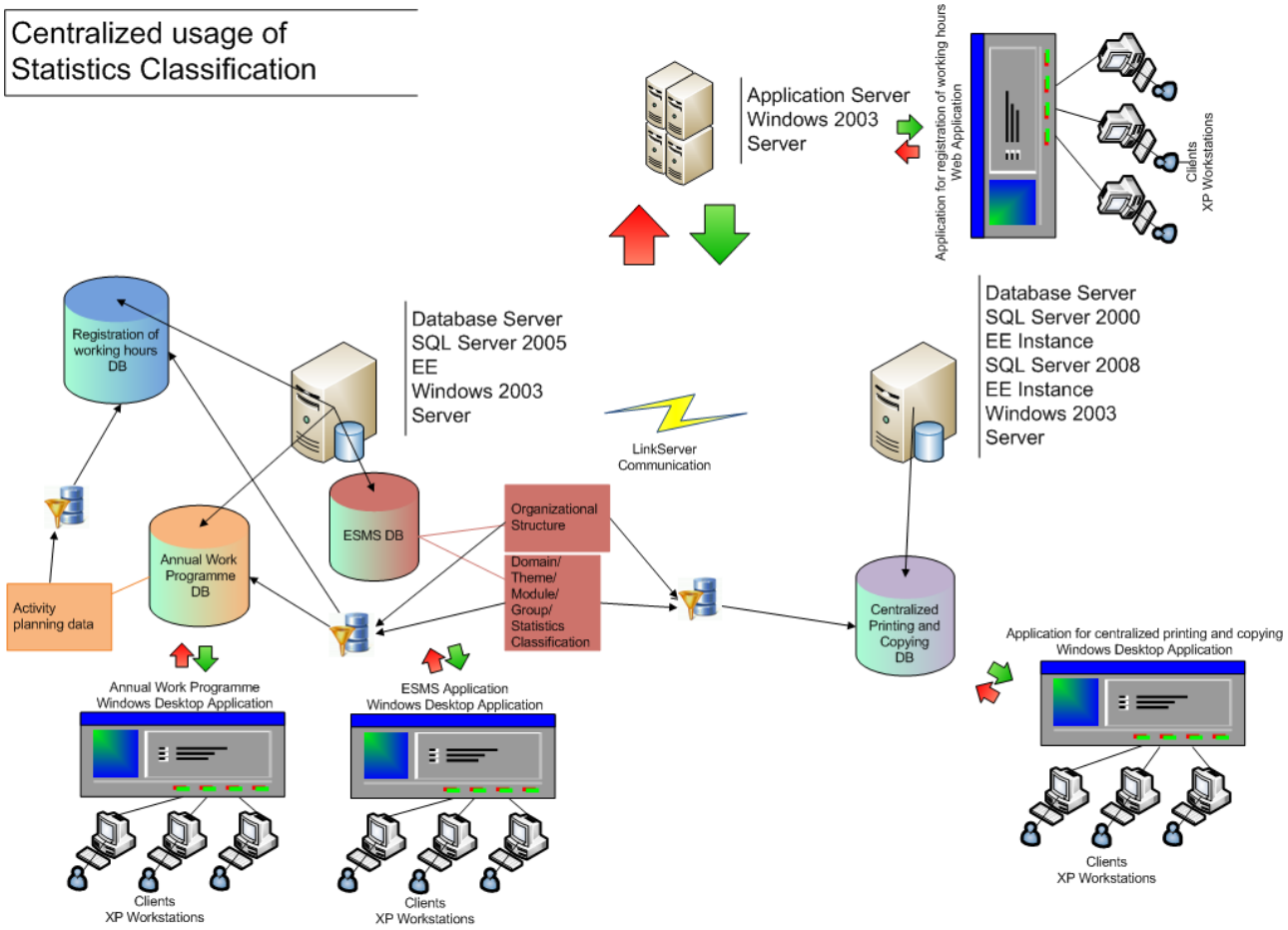

Figure 3.1 – Centralized usage of Classification of Statistics in the production system of SSO

26. The central element in the system is Module/Group/ Statistics classification. Being a basic source of information on statistical surveys it is used for preparation of an annual and long-term Statistical Work Programme, development of cost-calculation system and development of the application for monitoring the realization of the planned activities for each employee. This emphasizes the constitution of links to other supporting processes: cost controlling, work efficiency and respondents' burden evaluation, phases from the existing BPM implemented in the annual working plan.

VI. CONCLUSIONS

Future plans

27. As future steps from this perspective we have a vision and challenges to implement enhanced features of the ESMS Database and Application. In order to produce Quality Reports we must implement Quality Indicators related to ESMS Metadata Sub-concepts and linkage with ESQR Framework as well.

28. The main aim of the usage of the application is to document all existing statistics in accordance with the defined time frame in order to fulfill the already mentioned reporting needs, but also for dissemination puposes i.e. to release reference metadata on the internet in a standardized manner, and to develop a Reporting Database as well.

29. For the time being some of the concepts and sub-concepts from the ESMS structure are still not fully available or officially adopted, such as: Microdata policy and access to microdata, Statistical disclosure control, Expenses and reporting burden, Policy on data revision. Each of these issues needs adequate treatment in the relevant WG and will be included in the ESMS concepts as soon as available.

30. A future challenge is to enable subject-matter staff to produce SDMX-ML files with reference metadata directly from the database in order to send them to Eurostat via the "Single Entry Point" in SDMX format.

Summary

31. A short summary as a conclusion could be just repeating the main purpose of this paper - to present the practical side of the in-house work done at the SSO, but strongly relied on clear recommendations and standards provided by Eurostat. We are very optimistic about the further improvements and exploitation of the Euro SDMX Database, that it will have an impact on the efficiency of statistical production, improvement of the quality of statistical information and reduction of burden. And finally, it is a step forward to better and more standardized documentation of statistics.