

**STATISTICAL COMMISSION and
ECONOMIC COMMISSION FOR EUROPE**

**COMMISSION OF THE
EUROPEAN COMMUNITIES**

CONFERENCE OF EUROPEAN STATISTICIANS

EUROSTAT

**Joint UNECE/Eurostat Work Session
on Statistical Metadata**
(6 - 8 March 2002, Luxembourg)

Working Paper No. 16
English only

Topic (ii): Users and metadata, statistical information portals

**PROGRESS ON STATISTICAL METAINFORMATION SYSTEM IN THE
NATIONAL STATISTICAL INSTITUTE**

Submitted by the National Statistical Institute, Bulgaria¹

Contributed paper

SUMMARY

The Integrated Statistical Information System (ISIS) in Bulgaria will be based upon the Statistical Metadata System (SMS) and other systems as well. The SMS operates in all stages of the statistical data life cycle: survey preparation, data collection and preparation, statistical modelling, estimation, presentation and dissemination. For the time being, the SMS includes administrative elements, classifications, nomenclatures, registers, databases, methods and indicators.

The aim of SMS is to create a standard integrated environment to support ISIS design and implementation. The current status, problems and tasks are presented in this paper. The National Statistical Institute of Bulgaria (NSI) is conducting research on collection, content, storage and delivery of metadata. Integrating these components of metadata management requires careful planning. Standards are a means to facilitate this work.

Special attention is given to the use of metadata to explain statistics, assets of data and metadata. The two basic groups of metadata users are divided into producers of statistics and users of information. The market research on data and metadata users are indicated. The main methods of and approaches to metadata dissemination are described below. NSI's priority goal is to design a unified approach, using standard models and international recommendations.

Some aspects of future development and conclusions are outlined.

I. INTRODUCTION

1. NSI has started work on improving the statistical metainformation system (SMS). These efforts are being made for several important reasons:

- ? implementation of international recommendations;
- ? improvement of service to users;
- ? development of an integrated statistical information system (ISIS).

2. SMS vision and concept are in accordance with today's information technologies. Following the recommendations and instructions coming from EU, UNECE, IMF and OECD, NSI has recently begun a research of SMS organization and design, as a tool for ensuring integrity of standards and comparability

¹ Prepared by Delcheslava Delchava.

of data with the international requirements. NSI has responded to the need to be more customer oriented by establishing an effective dissemination policy. Metadata and SMS must play a key role in the process of ISIS elaboration. SMS has always been a tool for ensuring harmonization. Following the international process of integration of official statistics, NSI has also initiated some activities to solve related problems.

II. CURRENT SITUATION AND PROBLEMS

3. During the last years, NSI has approved several important papers that have changed the course of NSI action. The Statistical Act, as of 25 June 1999, amended 27 April 2001, provides the legal basis for the Bulgarian official statistics and NSI. The Act provides the general framework for “realization of the statistical activities by Bodies of statistics, registers and information system they maintain, as well as data exchange between them.”

4. The Act provides the legislative basis of SMS. The content and principles of the Statistical Act define the NSI tasks and designates NSI as a responsible body for the coordination of statistical activities in the country. The tools for the administrative and infrastructural coordination and elements of SMS are as follows:

- ? National Programme for Statistical Surveys (NPSS);
- ? Statistical Infrastructure (classifications, nomenclatures, registers, methods and database);
- ? Methodological Descriptions of Indicators.

5. In line with the Statistical Act, NSI has prepared a long-term Strategy for Development of Bulgarian Statistics 2000-2006. The Strategy identifies the main objectives, i.e:

- ? Improvement of quality of statistical information;
- ? Raising the cost effectiveness of statistical activity;
- ? Heightening the trust in the statistical information;
- ? Improvement of cooperation between all participants in statistical activity – users and respondents;
- ? Improvement of international cooperation.

6. According to the Strategy there is no adequate system to keep and ensure coherency between data and metadata, including a system of indicators and specific characteristics to assess the updated processes.

II.1 National Programme For Statistical Surveys (NPSS)

7. According to the Statistical Act, NPSS will be adopted annually by the Council of Ministers and it includes the following:

- ? permanently conducted statistical surveys with established public significance;
- ? other statistical surveys for the respective year or period;
- ? surveys related to international obligations undertaken by the Republic of Bulgaria.

For each statistical survey the following will be indicated in the Programme:

- ? theme of the survey;
- ? content and object of the survey;
- ? way of conducting the survey;
- ? sources and form of data collection;
- ? body/bodies conducting the survey;
- ? obligation for data provision;
- ? budget of the survey.

8. While elaborating the NPSS, NSI must accomplish several different duties, substantially concerned with:

- ? control and formulation of the statistical systems (elaboration of concepts, measuring the phenomenon object of study);
- ? control of statistical precondition on which the systems are based (definition of samples, evaluation of errors);
- ? control of the process to achieve final products.

The NSI President also reports on the programme implementation for the preceding year to the Council of Ministers.

9. The Statistical Act requires the results of the statistical surveys, included in the NPSS, to be accessible to all. NSI is committed to preparing each year:

- ? an operating plan;
- ? performance indicators;
- ? schedules.

II.2 Classifications, nomenclatures and registers

10. The Statistical Act also provides a suitable legal basis for coordination, including mandatory use of the national statistical and information systems and registers, classifications, nomenclatures and methods, approved by the NSI President.

11. Concerning harmonization of the statistical infrastructure with the requirements of the EU legislation, the following European and international classifications and nomenclatures have been implemented:

- ? Bulgarian economic activity classification in conformity with the NACE Rev.1;
- ? National product classification by activity in compliance with CPA,
- ? National PRODCOM nomenclatures, compatible with PRODCOM list;
- ? Bulgarian classification of types on construction;
- ? Combined Nomenclature adopted by the customers;
- ? Unified classification of administrative and territorial entities;
- ? Classification of territorial entities for statistical purposes (NUTS);
- ? COICOP;
- ? COFOG;
- ? Classification of education;
- ? Classification of occupation ISCO'99;
- ? Classification of accident at work;
- ? International Classification of diseases.

12. According to the Statistical Act, NSI has to establish and maintain the following registers:

- ? Unified Register for identification of economic and other subjects carrying out activity on the territory of the country, the so called BULSTAT Register;
- ? Unified Register for identification of farmers, self-employed individuals and other natural persons;
- ? Statistical Business Register- based primary on the BULSTAT Register.
- ? National Registers of Settlements.

II.3 Data base

13. NSI has the following databases:

- ? Population;
- ? Economy;
- ? Regional data base.

Updating the databases has continued. Data and metadata loaded into the database have been used in survey planning, descriptions, sets of questionnaires, micro-data, macro-data, time series. The aim of the survey description is to:

- ? present information for statisticians;
- ? provide metadata for user;
- ? help to programme and conduct the survey;
- ? enable the survey' control.

II.4 Indicators

14. At present, a full methodological description is available for almost all indicators (input and output) and it has been disseminated. A methodological description of results and impact indicators, required under the Statistical Act, has not been completed yet.

15. The main objective is to develop a user-oriented SMS, which will support the planning and design of statistical surveys, collection, processing, storage and dissemination of data.

16. As is well known, metadata can be divided into two groups, according to the function of the system:

- ? end-user oriented;
- ? production oriented.

The SMS architecture in NSI is structured on these two levels. Each of them has its specific metadata to:

- ? satisfy users' needs regarding identification, location, processing, interpreting and analyzing of statistical data;
- ? satisfy production requirements on planning, design, implementation, processing and evaluating activities.

17. There are a number of data users, who require different kinds of metadata:

- ? managers – about budget, schedules and processes;
- ? analysts and researchers, who have more interest in sample design, questionnaires, standard errors;
- ? programmers – who are interested in record layouts, data dictionaries, file storage medium and other information needed to process data. [Creecy, 1995]

II.5 Users and dissemination

18. NSI has many users, both internal and external. The needs of these users are different. Main SMS user groups are: government, business, research institutions, non-profit entities and citizens.

19. Market research is very important in order to meet user needs. The first study on information supply of management purposes during transition to a market economy was carried out by NSI and the University of National and World Economy in 1996. Last year, and through this year, NSI conducted a similar study of the main users of business and government sectors. A number of positive and negative aspects of NSI performance have emerged. Marketing provides a better means of dissemination of statistical information, and effective statistical data dissemination is the best kind of marketing.

20. Provided with metadata, the users could be better informed about what is available and what tool meets their particular requirements. They are therefore better informed about decisions related to what kind of data they wish to use. Thus the users' cost will be lower, as they would be less likely to specify or purchase data of little or no utility.

21. Rules for dissemination of statistical information, products and services have been prepared. The statistical dissemination objectives derived from the NSI's goals are as follows:

- ? statistics should promote democracy by informing citizens about the development in society;

- ? statistics should assist the public and private enterprises in their decision-making, using an effective information service;
- ? statistical data should be accessible to the academic, researchers and non-profit institutions.

22. NSI information activities can be divided into two parts using the content of the information. These are:

- ? dissemination of statistical data within our data system;
- ? marketing of NSI products and services.

These two parts are interrelated. Basic principles of dissemination are: reliability, timeliness, equal access, impartiality, openness, active role of statistics, emphasis on the sources of information, harmonization, documentation.

23. Updating and further improvement of the NSI web-site is an important task in the development of dissemination of statistical data and metadata. The NSI web-site provides the possibility for publishing and updating of:

- ? the yearly calendar for presenting the results of the statistical surveys;
- ? Statistical Act and secondary legal acts;
- ? catalogue of statistical publications.

24. The Bulgarian participation in the IMF GDDS dates back to May 2000 and it is planned to meet the SDDS requirements by 2006 at the latest. Bulgaria has participated in the IMF pilot study on metadata for the GDDS. [IMF 1998]

25. From the content point of view, metadata are data which describe the micro data, macro data or other metadata through the system, application and administrative components. The statistical metadata components can be defined as:

- ? System – physical and logical characteristics needed for computer programming and data base management, include information such as location, storage, medium, records layout and data dictionary;
- ? Application – information needed in order to understand or use the data, information such as methodological notes, definitions, sample design, questionnaires, response rates, standard errors;
- ? Administrative – programmes, costs, schedules, budgets and related information associated with surveys and processes and used by management for analysis. [Creecy, 1995]

II.6 Problems

26. The metadata system that presently exists in NSI is still partially compliant to the two basic purposes outlined above in para. 16 and it is not based on a unified approach. The main problems are:

- ? incomplete methodological documentation of the survey process;
- ? application of non-formalized free text format ;
- ? partially harmonized methodological descriptions of statistical concepts and linking mechanism to the data;
- ? incomplete standardized presentation formats;
- ? incomplete integration tools.

27. Documentation means, on the one hand, a lot of additional work for data producers and, on the other, no guarantee that the documentation is accurate and presents the real status. In order to overcome these problems, we consider the following principles: necessity for semantic description of metadata, terminology integrity, user orientations that help with the correct interpretation of data and play an important role in quality assessment.

28. To address some of the problems that arise from the non-uniform approach, NSI is conducting research in ISIS and SMS development. NSI has developed several organizational structures to solve these problems. For example, NSI has established the Methodological Council and Editing Council,

while the creation of an ISIS Department is foreseen. It is important to elaborate ISIS, which will be planned primarily for production oriented purposes. SMS is to be designed with both these purposes in mind – user-oriented and production-oriented.

II.7 Tasks

29. The main tasks are:
- ? conceptual harmonization;
 - ? ensuring full documentation of statistical survey and products-administrative, survey contents, survey plan, completed data collection, statistical processing and presentation;
 - ? introducing detailed classification and formalized text format;
 - ? implementing common standardized format;
 - ? introducing quality assessment. [Groenenez, 2001]
30. A survey of possibilities for adopting and elaborating statistical standards for metadata description, connected with the methodology of statistical surveys and data processing, is still going on. Based on experience, international recommendations and formats for SMS description, the main elements will be proposed.
31. The setting up and use of a unified concept is an important SMS element which facilitates the uniformity of the statistical data for different categories of statistical information users.
32. The new SMS model will have to take all these needs into account. An important part of SMS management is the engagement of all staff in the continuity process. There is a need to foster the knowledge among the statisticians in the national statistical system, so that they can understand the linkages between various data sets, check data and communicate better on technical issues. Provision of training on standards and methodologies to the staff would contribute to quality improvement.
33. A successful SMS implementation depends on the professional ability of staff and the possibility to enhance their technical skills. The staff need to be trained in new standards and approaches.
34. A working group will be established to deal with the lack of a unified approach in the existing metadata system. The group will work out a proposal for SMS development and consolidated collection of metadata by means of:
- ? identification of users and needs;
 - ? fields of metadata;
 - ? development of electronic template and applied standardized manner.

III. ASPECTS OF THE FUTURE

35. This paper describes current NSI practices. While some progress has been made in using metadata, it is apparent that in order to make significantly greater progress a more unified approach must be undertaken to SMS organizing within NSI. The ISIS is the starting point of such an approach, based on the metadata flows and aimed at ensuring a comprehensive list of metadata, necessary for the:
- ? input - survey preparation, data collection, data preparation;
 - ? aggregation - statistical modelling, estimation;
 - ? output - presentation, dissemination.
36. The recommendation that metadata should be captured when they arise is most important. They should then be automatically transformed whenever the underlying survey data are transferred or transformed and, after that, be automatically updated when the survey process is changed. [Sundgren, 1995]

37. At present, however, the practical arrangements for achieving coordination are not yet fully developed. Three approaches are to be followed simultaneously while developing these arrangements, i.e:

- ? preparation of the National Programme for Statistical Surveys and the rules regulating the activities of the Bodies of statistics;
- ? development of statistical infrastructure, e. g. classifications, nomenclatures, methods, registers, databases;
- ? improvement description of statistical indicators.

38. In future, we hope to make the comprehensive list of the statistical surveys, carried out by the Bodies of statistics, using the Guide of the National statistical system. It puts emphasis on the output side and provides a detailed information about statistical data, products and services. In general, NSI has a key role to play in NPSS development as an effective entity and integrity of the national statistical system.

39. In view of NSI responsibility in implementing standard classifications, NSI is planning to develop a computer system for maintenance of the national classifications and nomenclatures. The main functions of this computer system are to promote a uniform use of classifications and nomenclatures, ensure good access to them and assist development and maintenance. The establishment of the national system for maintenance of classifications and nomenclatures, including a common database and metadata system, is in progress. NSI also plans to put the classifications and explanatory notes on Internet.

40. The work on improving the impact and result of statistical indicators will be carried out by an inter-institutional working group. The characteristics of the statistical indicators will be described using the following elements: subject-matter attributes, observed statistical units, type of observed time period, measurement unit, code, sources of indicators, level of harmonization with international requirements, etc.

41. It is very important for NSI to be aware of the changing public demand for information. The IT will play an important role in statistical data and metadata dissemination. The use of adequate software to disseminate the statistical metadata facilitates the use of this kind of information.

IV. CONCLUSIONS

42. Work on SMS has started. The available parts are under development, the new ones will be built up and integrated in ISIS. The metadata users, i.e producers of statistics and users of statistical information must be investigated. More detailed research is necessary on end metadata users.

43. The procedures concerning dissemination of statistical data and metadata require further development. Metadata are a helpful tool for eliminating the most serious errors and misuses of statistical data. Wide metadata dissemination is an effective tool for clarifying and understanding the responsibilities and rights of statisticians, users and respondents. Research on metadata quality is needed.

44. I wish to stress the need for international cooperation in the field of metadata. Budgetary constraints necessitate intensive sharing of information on common international standards and best countries' practices. That is why we have to organize our creative strength to provide higher visibility to our work and convince our Institute of the importance of our activity.

REFERENCE

Creecy, Robert H., D. W. Gillman, M.V.Appel (1995), Metadata, Statistical Software, and Information Systems at the U.S. Bureau: Current Practice and Future Plans, ISI. Bulletin of the International Statistical Institute (Proceeding of the 50th Session) Book II

Groennoz, Dominique, B. De Norre, S. Paganoni(2001) Metadata Structures in Eurostat, Metadata Workshop 2001, Luxembourg

IMF(1998), The General Data Dissemination System. Standard for the Dissemination by Countries of Economic and Financial Statistics. IMF Washington. D.C.

Sundgren, B (1995), Making Statistical Data More Available, NSI Bulletin of the International Statistical Institute (Proceeding of the 50th Session) Book II

UNECE(1995), Guidelines for the Modelling of Statistical Data and Metadata. UNECE, NewYork and Geneva