

Topic (i): Infrastructure issues for statistical metadata

METADATA SYSTEM IN STATISTICS LITHUANIA

Submitted by Statistics Lithuania¹

Contributed paper

I. INTRODUCTION

1. During the last ten years, statistical offices of the eastern and central European countries faced the challenge of integrating their national statistics into the European Statistical System. The rapidly growing domestic and international demand for reliable and internationally comparable statistical information made it necessary to develop a statistics system in compliance with the European requirements and needs. In order to ensure better statistical information management and monitor the ongoing processes, Statistics Lithuania decided to develop a statistical information system, covering the whole process of production of statistical information. In 1995, under the terms of the PHARE project, Statistics Lithuania, in cooperation with "World systems (Europe) Limited" (Luxembourg) started to develop a metadatabase as an integrated part of the statistical information system. This metadatabase was foreseen to cover a comprehensive set of metainformation on the statistics production process and its outputs. The main features attributed to the metadatabase are: a user-oriented retrieval and dissemination system, flexibility, compatibility, integrity and security. The development of the metadatabase is ongoing. Currently, the metadatabase is a metainformation system containing references to data, but not the data itself. Nevertheless, when the output database is established, the metadatabase will be integrated into it as one of the subsystems.

2. This paper is dedicated to the metainformation system, which is being developed in Statistics Lithuania, and the peculiarities of the statistical metadata contents in the transition period. A quite important part of the metainformation system was dedicated to metadata determining national statistics compliance to the EU requirements.

II. GENERAL OBJECTIVES OF THE METADATABASE AND USER NEEDS

3. Statistics Lithuania established the metadatabase in order to support a consistent interpretation of statistical data and to improve the quality of use and processing of statistical information.

4. The requirements for the metadatabase were predefined mostly by user needs. It was agreed that a comprehensive and thorough understanding of different user needs is the basis for the better satisfaction of their expectations concerning metadata. The establishment of the concept of the metadata system predefined two main types of users of the statistical metadata, which are:

¹ Prepared by Daiva Jureleviciene.

- producers of statistical data,
- users of statistical data.

5. Obviously, these categories of users have very specific requirements and demands for the metadata. So the decision was taken to develop the metadatabase capable of satisfying a wide range of possible interests (starting with queries on data availability and finishing with comprehensive methodological details).

6. Therefore, general objectives and requirements for the metadata system were moved into two directions:

- i) Establishment of a metadata system, satisfying internal needs of statistics producers;
- ii) Application of internally developed metadata system for the needs of different categories of external users.

7. Currently, access to the metadatabase is established only for internal users. Metadata internally could be accessed in two different ways - using Windows application (some screenshots of this application are presented in the annex I) and via Intranet.

8. In the near future, access for external users must be established. As external users have very diverse interests concerning metadata, two main groups and some sub-groups of them were distinguished:

- i) Domestic users: governmental institutions, companies, researchers, public, politicians, journalists,
- ii) Foreign users: international organizations, companies, and public.

9. Each of the above-mentioned user categories has very specific needs in terms of contents and scope of metadata. So the metadata presentation for external users requires to have strictly determined main metainformation points, which each category of users could access. The metainformation set should be well-structured and differentiated according to different user skills and interests.

10. Clear and transparent metadata structure would enable to satisfy requests for metainformation in the most comprehensive way, but not to overload the less experienced user.

III. CONTENTS AND STRUCTURE OF THE METADATA SYSTEM

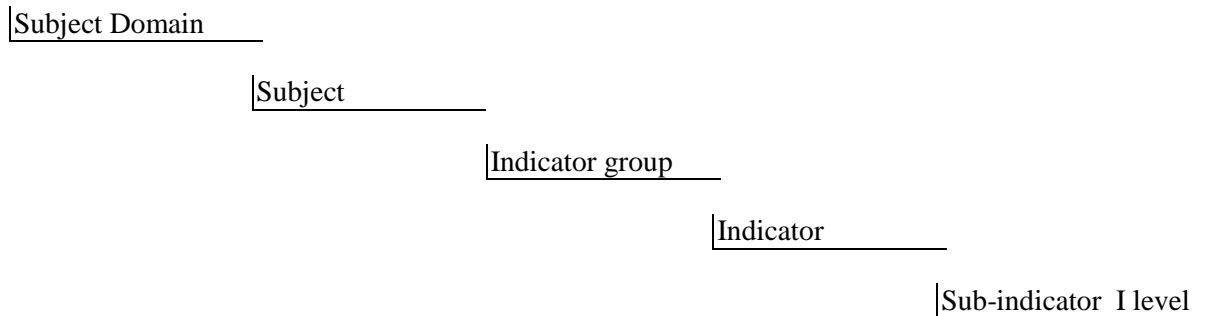
11. The metadatabase is foreseen to provide metainformation on indicators, time series, surveys and questionnaires. Currently the indicators and parts of time series are most comprehensively developed. Once all elements of the metadatabase are developed, the links between Indicator-Time series, Indicator-Survey, and Survey-Questionnaire will be established.

12. The metadatabase contains information on indicators produced by Statistics Lithuania and other governmental institutions. Metainformation is available in the Lithuanian and English languages. At present the database application is developed only in Lithuanian as metainformation is used only internally. The Intranet application is being developed bilingual.

13. There are about 5.5 thousand indicators in the metadatabase, of which approximately 800 are submitted by ministries and governmental institutions.

14. As an easily understandable, well-structured and informative metadatabase predefines the users' willingness to use the information stored, we aimed to structure and present the metadata in the most efficient and user-friendly way. The indicators were structured using up to 9 hierarchical levels: 3 higher levels identify subject domain, subject and indicator group to which indicator is assigned; the next levels

are assigned to the indicators itself and sub-indicators. This way of presentation is visually easily understandable:



15. All the indicators in the metadatabase are structured according to the subject domain they deal with:

Geographical data	Agriculture, forestry and fishing
Environment protection	Industry
Population	Energy and material resources
Education	Construction
Research and development	Domestic trade
Health and social security	Transport, tourism and communication
Recreation, culture and sports	Business and other services
Household statistics	National accounts
Crime and justice	Prices and purchasing power
Dwellings and community economy	Foreign trade
Labour	Investment
Other social statistics	Money and finance

16. The development of the structure of the metadata system predefined assurance of metadata consistence, quality and relevance. The metadatabase contains the following main metainformation set on indicators and time series:

Title of the field	Purpose of the field
Indicator title	Full label of the indicator by which the indicator is known or referenced
Measurement unit	Unit by which the indicator is measured
Territorial level	The lowest territorial level defining the availability of the indicator (national level corresponds to NUTS 1-2 levels, county - to NUTS 3 level, municipality - to NUTS 4 level and neighborhood - to NUTS 5 level)
Periodicity*	Frequency with which data are compiled
Start date of data collection*	The date when the time series starts
End data of data collection*	The date when the time series ends if the related data are not collected or calculated anymore
Breaks in times series*	Interruption in the statistical observations, due to change in methodology, definition and other reasons
Definition of the indicator	Conceptual explanation, meaning and use of the indicator
Calculation method*	Description of the method used to compile the indicator
Seasonal adjustment*	The method used to seasonally adjust a time series
Data source*	Origin of primary data used for the compilation of indicator (survey/ administrative data)

Title of the field	Purpose of the field
Publications*	Paper or electronic product containing the indicator
Responsible division	Statistics Lithuania division responsible for producing the indicator
Responsible governmental institution	Statistics Lithuania, ministry or other governmental institution, responsible for producing the indicator
Main users*	Main users to which indicator or time series is constantly submitted

* Metainformation available on times series.

17. Monitoring of compliance with the EU requirements also could be treated as a separate and very important integrated part of the metadata system for the transition period. Currently about 2000 indicators with predefined EU requirements are stored in the metadatabase. 86.8 per cent of them are produced. About 75 per cent of indicators fully comply with the EU requirements. 13.2 per cent of indicators are not yet produced, but most of them will be started to be produced by 2004 (prospective date of the accession of Lithuania to the EU).

18. The metadatabase provides the following information on statistics compliance with the EU requirements:

Title of the field	Purpose of the field
Is the indicator the subject of the EU requirements?	The purpose of this field is to indicate if specific EU requirements have been defined for the indicator
Source of the EU requirements	EU legal act or other basis defining the production of the indicator
Compliance with the EU requirements	The field indicates the compliance with the EU requirements achieved. Possible following choices: (1) Compliance with the EU requirements fully achieved; (2) Compliance with the EU requirements partially achieved; (3) Compliance with the EU requirements not achieved; (4) Indicator is not yet produced (5) EU requirements are not specified in a detailed way; (6) National indicator
Non-compliance reason	The reason why indicator is not in accordance with the EU requirements. Possible choices: (1) Calculation method, (2) Definition, (3) Statistical observation unit, (4) Classification, (5) Periodicity, (6) Territorial level (7) Other
Date when the compliance will be achieved	The preliminary date when compliance with the EU requirements will be achieved
Is the indicator supplied to Eurostat?	The purpose of the field is to indicate if the indicator is submitted to Eurostat

19. The possibility and relevance of introduction of detailed definitions provided in the EU legal acts, manuals or international agreements simultaneously with national definitions as well as monitoring of compliance with other international organizations' requirements is being considered.

20. The surveys and questionnaires are other very important elements of the metadata system, which defines and characterizes the production process of statistical data. Currently this part of the metadata system is being developed and is undergoing testing. In order to avoid a complicated and overloaded structure, two separate metainformation sets were assigned to the Survey and the Questionnaire.

21. The metadata structure foreseen for the survey and questionnaire is as follows:

Title of the field	Purpose of the field
<i>Survey</i>	
Survey title	Full title of the survey
Statistical observation unit	Observation entity on which data can be obtained
Periodicity	Frequency with which survey is conducted
First conduct	Date when the survey was conducted for the first time
Survey type	Census or sample survey
Period of data collection	Time lag when the data collection starts and the moment when it ends
Data transmission media	Media used to transmit data (post-paper, fax, diskette, e-mail, etc.)
Responsible institution	Institutions responsible for conducting the survey (Statistics Lithuania, ministry or other governmental institution)
Responsible division	Responsible division if the survey is conducted by Statistics Lithuania
Contact person	Contact person for getting queries
<i>Questionnaire</i>	
Questionnaire title	Full title of the questionnaire
Short title	Short title of the questionnaire
Date of approval	The date when questionnaire was officially approved
Registration number	The specific number with which this questionnaire was recorded, once it was approved
Periodicity	Frequency with which the questionnaire is disseminated
Reporting units	The units which report data
Last update	Last update of the questionnaire

IV. FUTURE DEVELOPMENT AND FINAL REMARKS

22. The implementation of the statistical information system is time consuming, requiring know-how, experienced personnel and technical resources. Nevertheless, Statistics Lithuania aims to establish a well-balanced and practical system, satisfying internal and external user needs, ensuring statistical data quality and efficient management.

23. Key tasks related to the further implementation and application of the metadata system:

- i) Full establishment of standardized and comprehensive metadata system and its integration into other databases and information systems of Statistics Lithuania.
- ii) Establishment of the access to the metadata for the different categories of external users, which should support the dissemination and consistent interpretation of statistical data.
- iii) Establishment of coding system for some categories in the metadatabase, which would ensure efficient data flows management and standardization. The following coding criteria were predefined in advance: unique reference, shortness, simplicity, expandability and flexibility.
- iv) Development of the concept of the output database and its implementation. Integration of the metadata system into the output database.

24. The challenge to maximize the benefits for the user with the full establishment of a well-balanced and compatible system with other existing parts of the statistical information system metadatabase is still in force. This has to be done in order to ensure the smooth and consistent integration into the European Statistical System.

References

“Guidelines for the Modeling of Statistical Data and Metadata”. United Nations, Statistical Commission and Economic Commission for Europe, Conference of European Statisticians, Methodological material, 1995.

“Information Systems Architecture for National and International Statistical Offices. Guidelines and Recommendations” United Nations, Statistical Commission and Economic Commission for Europe, Conference of European Statisticians, Statistical Standards and studies No. 51, 1999.

“Guidelines for Statistical Metadata in the Internet”. United Nations, Statistical Commission and Economic Commission for Europe, Conference of European Statisticians, Statistical Standards and studies No. 52, 2000.