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. 6
English and Russian only

Topic (i): Infrastructure issues for statistical metadata

SOME ASPECTS OF THE ESTABLISHMENT OF THE METADATA SYSTEM AT THE ARMENIAN NATIONAL STATISTICAL SERVICE

Submitted by the Armenian National Statistical Service (NSS)[1]

Contributed paper

Summary

Until now, the automation of statistical processing has been based on the establishment of local relational databases.

In 2001, the Armenian National Statistical Service (NSS) began, in conjunction with representatives from Synergy Corp., to work on the design of a reform of the Armenian statistical information system. In keeping with this design, a system oriented towards user needs was adopted. The plan is to establish a single data warehouse and, within it, databases for direct access over the Internet (datamarts). Execution of this design will be accomplished in three stages. The corporate data warehouse will be the nucleus of the data storage architecture at NSS.

The circle of users of NSS statistical output has expanded significantly since 1999 and systematic descriptions of statistics have become a necessity.

The metadata base will contain descriptions of official terms used in Armenian statistics and comparisons with international concepts, descriptions of the methods by which data are processed, and the resulting statistical products. The metadata must support three languages: Armenian, Russian and English.

NSS has developed a software package called GENER, which constitutes the base of the metadata, to control its information resources and equipment. Information on how regularly the data acquired using a particular reporting form are processed, a list of the organisations submitting that information and instructions on how the form is to be filled in are entered into the metadata base. The document is accompanied by an information sheet describing any particular features of the calculations applied to the reporting form.

When disseminated, data will be accompanied by the corresponding metadata, graphs and charts for ease of visual interpretation.

I. Introduction

1. Until now, the automation of statistical processing has been based on the establishment of local relational databases.

2. In 2001, the Armenian National Statistical Service (NSS) began, in conjunction with representatives from Synergy Corp., to work on the design of a reform of the Armenian statistical information system. In keeping with this design, a system oriented towards user needs was adopted. The plan is to establish a

[1] Prepared by Vanush Davtyan, Natalia Pogosyan.

single data warehouse and, within it, databases for direct access over the Internet (datamarts). Execution of this design will be accomplished in three stages. The corporate data warehouse will be the nucleus of the data storage architecture at NSS.

3. The circle of users of NSS statistical output has expanded significantly since 1999 and systematic descriptions of statistics have become a necessity.

II. General

4. Experience at international organisations shows that the best and most reliable means of storing and exploiting data is the corporate data warehouse. The principles upon which such a warehouse is founded are:

- Very rapid data access
- Data relatedness and data integrity
- Ability to obtain and compare cross-sections of data
- Ease of data manipulation for the user
- Data quality
- Data protection.

5. The corporate data warehouse will be the nucleus of the NSS data storage architecture. It will be constructed around a central registry of items on which data are acquired and processed. It will contain micro- and macrodata reflecting the aggregate outputs of official statistics. Both the micro- and the macrodata will be accompanied by metadata.

6. The macrodata will be stored in "macrodata cubes". Since the generation of macrodata from microdata is to take place automatically, creating a set of classifiers for use in creating the statistical product becomes an issue. If, on the other hand, the data aggregation process involves more complicated functions, the macrodata-generating algorithm will be added to the metadata.

7. The metadata base will contain descriptions of official terms used in Armenian statistics and comparisons with international concepts, a description of the means by which data are processed and the resulting statistical products.

8. The warehouse will also provide the foundations for a database accessible over the Internet for external users. Because of this, the metadata must support three languages: Armenian, Russian and English. Only corrected data will be stored in the warehouse.

9. The data and metadata in the warehouse will serve as the sources for publications put out by NSS every month. The warehouse design must encompass all the indicators used in official statistics, offer compatibility with various data processing platforms, data compatibility and data integrity, and permit information to be added over time.

10. Data will be aggregated in accordance with versions of classification schemes appropriate to the time when a particular statistical product is derived. A description of the data-acquisition process will invariably exist in the metadata even before the definitive microdata are obtained, particularly where data coding is involved. Descriptions to simplify data searches and access will also be included.

11. When disseminated, data will be accompanied by the corresponding metadata, graphs and charts for ease of visual interpretation. Standard periodicity tables will be generated automatically. The data warehouse will give users direct access to a database of non-confidential statistical information by means of queries. Access to microdata will be subject to official decision.

12. Hence we will have a single product instead of a proliferation of diversified databases requiring explanations and descriptions every step of the way. This should improve data quality, speed acquisition of the final product and facilitate access to vital information.

III. Information flows

13. At present NSS handles 239 different statistical return forms; it also conducts periodic sample surveys. For the time being it mostly sends hard-copy forms to the districts where they are filled in and returned to a central office. There they are checked, coded and stored electronically. It is planned to transfer these functions to the regional level.

14. In the initial phase, input will take place at *marz* (province)-level offices; later it will be possible to enter data at the regional level as well. Data transfer will take place over telecommunications circuits, initially by modem but, once the corporate data network has been set up, over an NSS WAN. A standard program will be developed for data input; this will also be distributed electronically. The software will have to be accompanied by classification schemes and a selection of sample units on which observations are to be taken, descriptions of the return forms (instructions) and descriptions of the statistics.

IV. Establishment of the Nss metadata base

15. The metadata system is being developed in close conjunction with the data warehouse. Two kinds of metadata will be used: links, for ease of searching for the information required; and information interpreting the statistical data.

16. NSS has developed a software package called GENER, which constitutes the base of the metadata, to control its information resources and equipment. It began by setting up a repository of all the indicators used in official Armenian statistics:

- Indicator code and title;
- Indicator description
- Reporting form code;
- Keyword for searches;
- Link to classifier;
- Link to aggregation table;
- Link to IMF table;
- Date of completion;
- Operator code.

A related table of reporting forms had been drawn up beforehand, indicating:

- Form code and title;
- Frequency of submission;
- Reporting organization;
- Link to instructions on how to complete form;
- Link to help resource;
- Date of completion;
- Operator code.

17. Details of how often the information acquired using a particular form is processed, a list of the organizations submitting it and instructions on how to fill the form in are entered into the metadata base. The document is accompanied by an information sheet describing the particular features of the calculations associated with the form. Thus when working with a particular form a staff member should already have at his disposal enough information culled from the metadata base.

18. The uploaded metadata are linked by the same keys to the datamarts generating the statistical output. If the return form is modified, the datamarts should be amended automatically.

19. Using the metadata base, GENER can also generate new return forms or be used to make changes to old statistical forms.

20. In the future, as this system develops, the intention is that the output information should also be accompanied by metadata.

V. Hardware and software

21. With the equipment currently available to it, NSS can begin the construction of a data warehouse for storing both microdata and macrodata accompanied by metadata. The tool it has selected for this purpose is Microsoft SQL Server v. 7.0. The software can be developed in MS Visual C++ and MS Visual Basic. ADO (Active Data Object) will be used to link to MS Access/SQL datamarts. The server module in the client-server architecture, which transforms user protocols into SQL.RDS (Remote Data Service) queries, is a simple and effective device supporting the three-tier system. The system allows the client and server to be linked at the query and data-set level with appropriate security features.

22. The Central Data Warehouse will be installed on an SQL server in the NSS central office. The microdata will come from the Business Register, the Administrative Registers and household surveys. Data will be stored for five years. The base will hold one million records. The macrodata are aggregates of the information obtained from classification schemes or supplementary descriptions. Since the information is constantly changing the intention is to keep different versions of the classification schemes and store full details of the changes and the different versions of the schemes in the database. The classification schemes are stored together with tables of transition keys. Since 1 July 2001 it has been a requirement that a national classifier based on NACE must be used to aggregate data on types of occupation, and a geographical classified developed at NSS to aggregate data on geographical areas.

23. The metadata base which is also installed on the server will contain the designations and descriptions of all indicators, and a description of data-collection and -processing methods.

VI. Involvement in international projects on metadata system development

24. Realizing that the use of the IMF General Data Dissemination System (GDDS) would offer increased opportunities for development, especially as regards the supply of timely, reliable statistics to users, NSS informed the IMF Statistics Commission in early November 1999 that Armenia wished to join the System.

25. Since April 1999, NSS statistical publications have been compiled and produced in accordance with GDDS, and can be consulted over the Internet at www.armstat.am .

26. Work began in July 2000 on gathering the information needed to compile tables of metadata for the main statistical indicators in the format proposed by the IMF Statistics Commission. These tables, compiled by the IMF Statistics Department specifically for the main types of data input into GDDS, are intended for a special GDDS site which will, it is planned, contain metadata for all GDDS signatory countries. The intention is that the information in the tables (a total of 22) will be updated every year.

27. The compilation of metadata for Armenia is based on specimen pilot metadata for certain countries, including Kazakhstan, that were compiled with IMF expert support for the express purpose of offering guidance to other countries. Besides the statistics methodology unit, this work is being coordinated by an intradepartmental working group on the data dissemination system whose tasks include coordinating efforts to secure from other governmental departments (the Armenian Central Bank and the Ministry of Finance and Economy) the information needed to compile metadata. The metadata tables are being compiled in Russian and English.

28. Realizing the importance of such data for statistics users within the country, we plan eventually to translate the tables into Armenian for their benefit. It is also our intention to expand the circle of statistics users for whom metadata are to be compiled. We consider that metadata could be made available to the general public either on request or in statistical publications.

VII. Distribution of information

29. The data warehouse needs to support the decision-making process: it will contain stable data, and the only changes will consist in the addition of data to those already stored there. The data in the warehouse must be organized to be responsive to user queries. Some of the information will be displayed, using online analytical processing technology, on the Internet at www.armstat.am .

30. The design of the warehouse, supporting a decision-making system based on the technology of data cubes, will contain category descriptions and values. This kind of multiple-definition design makes it easier for the user to submit complex queries, pull data together for reporting purposes, turn from the general to the particular, and slice the data in various ways. The final design will be built around the metadata base.

VIII. Conclusion

31. Given the experience of the international statistical bodies in developing a blueprint for an integrated statistical space, NSS regards the development of the metadata system as a task of the highest order. Unfortunately, our staff have little experience in this area and would like to be able to study more closely the experience of countries that have had metadata systems for years. We are building the NSS metadata system with a view to its development both by adding to and expanding current information and by developing new attributes. There is room for growth, for example, in the development of meta-information standards. Once we have assimilated these we shall be ready to enter the international statistical information space.