Distr.
GENERAL

CES/AC.71/2004/20 (Summary)
8 March 2004

Original: ENGLISH

UNITED NATIONS STATISTICAL COMMISSION and ECONOMIC COMMISSION FOR EUROPE (ECE) CONFERENCE OF EUROPEAN STATISTICIANS

EUROPEAN COMMISSION STATISTICAL OFFICE OF THE EUROPEAN COMMUNITIES (EUROSTAT)

ORGANISATION FOR ECONOMIC COOPERATION AND DEVELOPMENT (OECD) STATISTICS DIRECTORATE

Joint ECE/Eurostat/OECD Meeting on the Management of Statistical Information Systems (MSIS)
(Geneva, 17-19 May 2004)

Topic (ii): Development of IT strategies in statistical offices

INTEGRATED META DATABASE AS A BASIS FOR DEVELOPMENT

Supporting Paper

Submitted by Central Statistical Bureau, Croatia [1]

Summary

I. INTRODUCTION

1. The aim of this paper is to describe the general model of the Integrated Statistical Meta database and its use in statistical survey processing in Central Statistical Bureau of Croatia (CROSTAT). The model is under development but partially implemented.

2. Placing metadata in the central role of development illustrates our belief that the development of a statistical system is the development of the metadata. If metadata exists or it is organized in the form of a database, this could define a good basis for development of the whole statistical system. A centralized repository of standards, classifications, methodologies and other subject matter knowledge allows for faster development of the statistical system and it immediate availability for public use.

II. GENERAL ARCHITECTURE

A. Main elements

3. The general architecture comprises several main elements in the integrated meta database. It consists of a description of the organizational structure, annual plan and publishing calendar, survey descriptions, classifications and other elements. Beside the meta database stands a statistical database, in micro data and macro data versions. The third element is the statistical production system, which connects statistical data and statistical metadata.

4. The relationships among all elements are very complicated in practice but generally it is possible to define the data flow and sequence of events, which constitute the production of the statistical data.

[1] Prepared by Zdenko Milonja (zmilonja@dzs.hr)

5. The importance of information technologies in statistical survey processing and dissemination is increasing and consumption of the resources is significant. Through the integrated meta database we open the possibility to increase the productivity in various parts of IT implementation. Separating the data from programs, but also metadata will probably lead to more adaptable programs.

B. Basic relationships

6. The basic relationship shows that statistical metadata come before statistical data. To be able to produce the statistical data the metadata must already exist. They are usually hidden in literature, in the knowledge of subject matter specialists and other places. Without statistical metadata it is impossible to produce statistical data and in that sense the metadata precedes statistical data. The connecting element between the metadata and data is the process of producing the statistical data.

III. BASIC SYSTEM FUNCTIONS

A. Creating and maintenance of organisation structure

7. The organizational structure of the institution is the basic metadata concerning planning and management processes. Because of the complexity of the production process, a lot of people and many activities are included in it. To manage this process effectively we designed the meta database to include data about internal structure of the institution.

B. Registers and respondents lists production

8. An interesting question arises about the position of registers. Do they belong to the statistical data or statistical metadata? We put the registers into the metadata part, together with respondent lists, which are basically derived from the register. The knowledge needed to develop the register (respondent lists) belongs to the metadata although the collected data about statistical units are included in statistical data.

C. Classifications maintenance

9. Classifications are a part of the meta database and processing and maintenance is organized in one place and unique control. We would like to develop a unique and coherent centralized system for the handling of classifications. Each survey is the user of the classification system.

D. Planning

10. Production of the statistical data is a very complex and expensive process. Planning includes assignments of tasks to various parts of the organizational structure of the institution. Therefore, in the metadata, there are detailed records about the organization, tasks and the planning facility for entering the time for planned execution for tasks. There is also a possibility of keeping track of the actual time for of all activities during the production process.

E. Software development

11. IT technologies are significantly involved in collecting, processing and disseminating the statistical data. There are two methods of development: using the cheap general software not produced for specific need or producing the software tailored to specific need but the price is high. Having the metadata collected in one place it is interesting to inspect the idea of automated software production for statistical purposes.

IV. PROJECT STATUS

12. The large and complex projects can’t be developed very quickly and brief information on the project status will be described.