

CONFERENCE OF EUROPEAN STATISTICIANS

Work Session on Statistical Data Editing
(Prague, Czech Republic, 14-17 October 1997)

Item 4 of the provisional agenda

DATA EDITING IN THE MULTIVARIANT QUESTIONNAIRE
(Application of the DataMan package)

Submitted by the Czech Statistical Office¹

¹ Prepared by Dušan Loutocký.

Abstract

This report briefly provides information on the possibilities of the DataMan system in an environment of heterogeneous surveying, where it is necessary to work not only with one type of questionnaire, but with a set of questionnaires, each of a different structure defined by the needs of structured surveying.

I. INTRODUCTION

1. In 1995 the requirement to take surveys in establishments was raised in the CSO-CZ. In the Czech Republic, there are relatively large production enterprises (economic entities) and each of them has a defined main economic activity. They are involved in enterprise surveys.

2. An enterprise can be divided into production units (establishments). The economic activity of such an enterprise does not have to correspond with the economic activity of its head office. The survey, in which the enterprise reports data according to their main economic activity does not respect this reality. This is why the requirement to take establishment surveys was made. The survey was based on sending not only one questionnaire to enterprises according to their main activities, but a whole set of reporting forms for each establishment in order for them to find the one that would correspond to its economic activity.

3. The basic handling requirement for the reporting forms was to keep them together and not to make any changes in them. No disassembling of the sets and rearranging them by, e.g. economic branch was permitted. No sorting of the forms even within the set was permitted either. It was assumed that all the reporting forms in the set were completed in the order provided in the set. This makes some demands on those responsible for entering data in the form as well as on the system used for doing that. It was also assumed that captured data (for all enterprises and establishments) would be placed in one file in the framework from which they would be transferred into the process of processing.

II. THE DATAMAN SYSTEM IN MULTI-FORM MODE

4. To implement the requirements, which means simultaneous acquisition of different structures of reporting forms, the DataMan system was modified by the introduction of a multi-reporting-form (multi-form) mode of work. This enlargement has no influence on the current single-form mode of work, so that tasks programmed in the single-form version can be run without any change.

5. When the work is started, the DataMan asks the user to select a mode to be run. If the multi-form mode is chosen, tasks for all reporting forms are begun at the same time. The tasks reside in the computer memory from where they are activated by the user. The activated task becomes a control one for the whole family of tasks, its corresponding reporting form is displayed, and its specific checks are activated. The remaining tasks are running in background, and there is a navigation system used which makes them available. The system controls access to sources of all tasks and makes it possible to transfer data between the individual tasks of the multiquestionnaire (set of reporting forms), run interactive and batch check routines (for instance, it is possible to check whether the value of an indicator for the

whole enterprise is equal to the sum of the values of indicators of the establishments the enterprise consists of). The data capture of a selected reporting form is then the same as in the single-form version of DataMan.

6. A security locks system established for the multiquestionnaire was introduced to protect both the reporting form and individual fields. The protection is defined for different groups of users, each group can have specific rights to display, modify, import and export values.

7. The multi-form access in the DataMan system expanded the possibility for single-form tasks, where its use provided several unexpected effects. In programming new single-form tasks it is useful to use the multi-form feature to capture data by means of complex reporting forms. Such a reporting form can be split into parts and these can be taken for different reporting forms when data capture is carried out. Different initial database values can be defined (that can be useful for database function, e.g. imputation of missing values). Another interesting effect, which resulted from the way of implementation, is the possibility to refer to data between individual tasks of the multi-reporting form.

III. OUTPUT FROM THE SYSTEM

8. The problem of the output file was solved by using a new data format - i.e. OTF. It is a non-standard format, which is a disadvantage, developed for data transfer between the DataMan and the database. The OTF format is an ASCII format derived from a standard CSV format; in fact, it is a text file, which may contain records of different structures. A descriptive section is defined for each data structure. The section has its own identifier and describes the data record relevant to a particular reporting form. The file also contains data sections which may be common to several different structures or reserved for one structure only. In these sections each data record corresponds to one statistical reporting form. The first item of the data record is the identifier of the descriptive section used to recognize its corresponding record. The file thus contains not only data, but also format information which can be used to construct a database. Data not only from one multi-reporting form, but also from a number of independent multi-reporting forms can be placed in one OTF file.

9. Tools have been developed to work with the OTF format. They make it possible to export the format to and from the database. The tools also include a transport one -a programme permitting to combine, split, transform and check the files outside the DataMan system, using universal criteria common to all surveys.