Contributions to Methodology and Statistics
A. Ferligoj and A. Kramberger (Editors)
Metodološki zvezki, 10, Ljubljana: FDV 1995

CASIP - A Complete Automated System for Information Processing in Family Budget Research

Willem E. Saris[1], Poulicos Prastacos[2], and Manuel Marti Recober[3]

Abstract

Specialists on household surveys are convinced that the existing procedures for production of statistics are too costly, too time consuming (three years production time), not harmonized enough within the EC and too troublesome for the respondents. In this paper we report about the develoment of a new procedure for this research, which can be a solution to the above mentioned problems. The amount of work for the respondents can in some parts of the research be reduced by a factor 10 using an expert system for data collection. Similar reductions of efforts for the researchers of the statistical offices are the purpose of this new system. The combination of three subsystems should allow the automatic production of the statistical tables within a few hours after the data collection is finished while the production of the same tables nowadays requires two years or more.

The system will be designed and tested by representatives of different EC countries. Therefore, we can ensure the possibility of use in all countries. Portability across computers and countries has been supported by use of computer programs which can be used on a wide variety of computers, while the systems have been designed in such a way that each country can use its own classification of income sources and classification of expenditures while any kind of summary statistics can be produced.

[1]Department of statistics and methodology, University of Amsterdam, Oudezijds Achterburgwal 237, 1012 DL Amsterdam, The Netherlands
[2] Research Center of Crete, Heraklion, Greece
[3] Universitat Politecnica de Catalunya, Barcelona, Spain

1 Introduction

Family Budget research is done in all modern Western countries and many other countries, but it is one of the most time-consuming research projects undertaken by statistical offices. It requires a great deal of effort on the paert of respondents who have to provide a lot of information about their purchases. It also requires a lot of work on the part of the statistical agebcies as paper diaries are normally used to register all the purchases, and these diaries have to be printed and mailed, the written information has to be coded and entered in the computer and checked for errors, and missing data have to be input. As a result the research process can easily take more than three years. This means that the results from 1993 may not be published until 1996, which is too late to be useful in making many policy decisions.

To improve this situation requires a quite different system in which information processing is completely automated. The researchers working on the project described here have proposed a design for such a system, called the Complete Automated System for Information Processing (CASIP). EUROSTAT provided a grant from the DOSES program to develop this system.

The objectives of this project are:

(i) To design a set of expert systems for family budget research which, combined in an integrated system, provide a complete system for automated information processing. This system will contain expert systems for: (1) data collection, including facilities for data validation, automatic coding, (2) data storage and retrieval of collected micro data and associated meta data, (3) production of standard statistics including facilities for imputation, (4) analysis and presentation of summary statistics.
(ii) The system will be designed in such a way that the following time and money consuming activities are avoided: (1) printing, (2) mailing, (3) coding, (4) punching, (5) creation of system files, (6) data cleaning.
(iii) A system should be designed which can be used in all EC countries from the second half of the 1990's. For this purpose special attention should be payed to the portability of the systems on different computer systems and the usefulness of the system with different classification systems in different countries.
(iv) The system should be able to produce summary statistics which are in agreement with the harmonization proposals of EUROSTAT.

In order to satisfy these goals a system has been designed with four different subsystems which can transfer data and metadata with each other according to the scheme presented in Figure 1. Figure 1 indicates that the process starts with the

Expert system for data collection (EDC). Several systems have been designed for the data collection in budget research. Most systems are based on printed questionnaires. As a consequence, the processing of the data can only be started after coding and punching. Computer assisted interviewing is far more efficient. By these procedures the printing, mailing, coding and punching can be avoided. The data cleaning and the automatic coding can be done during the interview by the expert system which is guiding the data collection. This means that the corrections in the data are made by the respondent, who knows the right answers, and not by the coder. Given these advantages, it is clear that in the late 90's most statistical data will be collected by computer assisted procedures. This development has already started in Great Britain, the Netherlands and the USA. Therefore a system for computer assisted data collection has been developed. It can be used for computer assisted personal interviewing (CAPI) and for computer assisted panel research (CAPAR). This system will produce data and meta data which have to be provided to the next subsystem (MIDAS) in the CASIP.

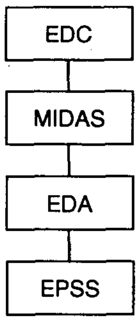

Figure 1: The structure of the CASIP system

Expert systems for storage and retrieval of micro data also exist such as Oracle (tm), Ingress (tm) or dBase (tm). The major points in the design of a new Microdata Database System (MIDAS) were the portability of the system, the applicability to many different data collection systems, the connection with the software for the data collection and the management of the large amount of data collected. Additionally, it was felt that since all these DBMS's are command driven the effort needed to develop a user friendly interface with any of these will be comparable with the effort needed to design a database system specifically for family budget research.

The portability has been taken into account by using as language C while the applicability to many different data collection systems has been facilitated by design of an open system where the user can introduce any kind of classification system. This classification is not a part of the progam. For the connection between the different subsystems procedures to handle metadata have been developed. The size problem required special attention. To reduce data storage requirements, data compression techniques are studied. The systems produces export files being any rectangular data file of households by variables which can be used by the next system to produce standard statistical tables. In its final form the system can be used for fast, fully automated retrieval of the stored information on many different computers and in many different countries.

The third expert system in the sequence is the expert system for data analysis (EDA). This system can produce the standard summary statistics required by the national statistical offices and Eurostat. This procedure does not require that the user is familiar with the database system and with the statistical package used. A simple interface takes care of the communication simple. The combination of MIDAS and EDA allows the production of any type of tables and combination of tables. It is even possible to produce tables for the national statistical office as well as EUROSTAT which might be different by minor changes in the selections and the specification of the metadata.

Furthermore, an information system is developed to store, retrieve, analyze and present summary statistics (EPSS). This system is different than the one storing micro-data information in that the emphasis here is not so much on efficient storage techniques but on providing a user an interface that facilitates the use of this dataset by users without extensive background in computers and/or statistical data.

By combining the different systems a complete automated system for information processing in family budget research will be obtained.

Such a complex system can not be recommended for use in practise without careful testing of the parts and the combination of all parts. These tests are done in two real life data collection processes. One is done in a technologically highly developed country (The Netherlands) using the automatic data collection system of the University of Amsterdam and the other will take place in a technologically less developed country (Portugal) using facilities partially developed within this project.

The work on this project has been started in 1990 and is sponsored by the DOSES program of EUROSTAT. In this presentation we will concentrate on the design aspects of the different components of the system.

2 EDC: Expert system for data collection

Expert systems for data collection have several advantages compared with the usual procedures using interviewers and paper questionnaires. The computer assisted data collection procedures are more efficient due to: automatic routing, automatic coding, automatic consistency checks (validation), reduction of the efforts for the respondents and avoiding costly routine activities like printing, mailing coding, punching and data cleaning (Nichols and Groves, 1986) .

All together the use of expert systems for data collection means that the coding of answers and the consistency checking occur during the interview. At that point in time the respondent can still provide correct information. Normally the coder in his office has to guess what the correct answer should have been. If there is doubt the information given is ignored and a missing value is created. In the new procedures the respondent corrects the mistakes personally so no missing values are created. This change in procedure means that clean data are available immediately after the data collection has been finished and that the data are in computer readable form. The last point means that the next phase, the analysis, can be started immediately (Van Bastelaer et al., 1988).

Given the advantages of the expert systems for data collection as described above, we chose not to design a system which combines the different procedures developed for coding, validation and imputation separately on main frame or mini computers (Knaus, 1987; Lorigny, 1989; Wenzowski, 1989; Rubin, 1987). Such systems can not be used during an interview. Therefore they do not provide the possibilities which we have mentioned above.

It is clear that within a short time most statistical data will be collected by computer based systems. This development has already started in the technically most developed countries like Great Brittain, The Netherlands and USA (Saris, 1989). The expert system is written for the interview program INTERV (De Pijper and Saris, 1986a) which is very fast and requires little space (De Bie et al., 1989). The program can work on any MS.Dos computer.

2.1 The features of the EDC system

It is impossible to collect complete data of expenditures for the households because the complete collection of daily consumption data alone would require 63 KB per household and consequently for a moderate sample of 2500 people one would get a file of 160 Mb (Olivier, 1987). Adding all the other information to it would lead to too much information which can not be handled any more. Therefore, not only sampling of households but also sampling in time will be used to reduce the amount of data and the effort of the respondents.

The designed expert system consist of three components:

(1) one system for income and regular expenditures to be used once per month

(2) one system for all expenses collected by a kind of electronic diary to be used during some weeks

(3) one system for expensive purchases to be used once per month
This is the minimum set of information required by Eurostat with respect to family budget data.

In some countries the data collection is more elaborate (Luykx Draelant, 1981) than this but extensions of this system can be left to the countries themselves to organize. For such extensions the same system can be used.

Two different approaches have been developed as an alternative for the verbal reporting by the respondents and coding afterwards. These two principles are quite different but they can complement each other. The first has been called tree structured coding (De Pijper and Saris, 1986b). This is a procedure by which the respondent has to answer a set of tree structured questions which automatically lead to a code for the specific product. Although answering such questions goes very fast, they cost, nevertheless, a considerable amount of time especially if the products bought are rather unusual. An efficient procedure requires that the most frequently cocurring products are mentioned on the top of the list of categories and that the lists are not too long. However the number of products is so large that such tree structures could become very large and time consuming in use. Therefore, a second principle is also attractive which is string matching. Using this procedure the respondent has to type the name of the product and the program compares the answer with a set of strings given to the program to code the answer. If a string of the program and the answer match the answer is automatically coded. It will be clear that this procedure will require more work from the respondents, therefore it should only be used for unusual products which would make the lists of categories very long. In practice this procedure is used in tree structured coding by specifying a category "other". If this category is chosen, the next question is "What is the name of the product?" The respondents type the name of the product and the computer program automatically codes the answers which are expected in advance, allowing for spelling errors in various ways (De Pijper and Saris, 1986a).

These principles present only two of the possibilities used to make the data collection more efficient and better. Other procedures are: upwards movement through the tree in order to reduce the amount of effort, dynamic consistency checks, substitution of memory by best guesses, summary screens and scheduling to reduce the work for the respondents. For more details of these procedures we refer to a general introduction (Saris, 1991) and to the publications about this system (SRF, 1990; Hartman et al., 1991; Leeuwin, 1991).

2.2 Metadata files

We have mentioned the tree structure because for all sets of variables such hierarchical structures have been used for coding the income and expenses. These classifications do not match the classifications used by any statistical office but all incomes and expendituress can be covered in such a tree and the structure can be changed relatively easily afterward. These trees are automatically provided as part of the metadata by the EDC system to the MIDAS. The MIDAS can make its aggregations automatically knowing the hierarchy and given the required level of aggregation.

2.3 Portability

The expert system has been designed in such a way that the syntacs can easily be separated from the semantics. Due to this possibility it is easy to move from one language to the other. First, one has to translate the verbal text of the system in a different langage. Next, the text has to be adjusted to the typical characteristics of the income sources and expenditures in the specific country. Such adjustments are in general minor, but necessary. After these changes the new EDC can again be made.

If an agency also wants to change the syntax specifying for instance the hierarchical structure, this can also be done in two ways: one is the adjustment of the tree structure. It will be clear that such changes have to be made with care. The other possibility is the correction of the classification through a transformation of the specified categories to the the categories of the new classification. In that case the adjustments are less dangerous and better controlable. In that case the EDC remains the same and only the classification can be changed whenever necessary.

After these changes have been made, the new EDC will also produce new metadata which will be used automatically in the next subsystems without complications. At this moment a Dutch and an English version of EDC is available. In the near future a Portuguese EDC will also be made for the test of the procedure in Portugal.

These facilities indicate that the system is designed in such a way that it can very easily be adjusted to the wishes of the different users and that is suitable to the multilangual situation of Europe.

3 MIDAS: The Microdata Database System

One of the problems faced by statistical agencies when performing family budget research is the shear volume of the collected information. Individual persons are surveyed to provide information on incomes received and money spent on more then 3000 products. Once collected, this information has to be analyzed to estimate means of the survey sample and subgroups of the samples. The massive amount of collected data - usually more than 5000 persons participate in national surveys- necessitates the organization of all information in a database.

The microdatabase (MIDAS) fulfills this function. It creates the databases that store information on individual persons and households in an efficient manner and makes them available on demand for estimating standard type summary statistics or performing other statistical analysis. It provides an information system that permits users to access, search and retrieve data as needed. As shown in Figure 1 MIDAS receives input from the EDC subsystem and provides output files to the EDA subsystem. A complete description of MIDAS can be found in Prastacos and Diamandakis (1990, 1991).

3.1 Features of the system

Data, coming from the EDC system are stored for persons and households. Income and expenditure information for individual persons is collected during the survey process and imported in MIDAS using ASCII files. MIDAS stores the data and metadata in such a way that information can be easily searched with the system. Household data are estimated in MIDAS by aggregating income/expenditure of all persons in a household and then stored in the databases.

The key feature of MIDAS is its capability to produce rectangular, so called, export files. The data items to be included in an export file can be completely predefined or selected by the users. The former approach is used when the same type of analysis is performed repeatedly on the same type of data set. An example of this is the production of standard statistics by statistical officies. MIDAS can automatically generate export files that contain all data items needed to perform a standard type analysis.

The latter approach is useful when conducting a nonstandard type of analysis. MIDAS can produce data files for any type of panel survey analyses. Subgroups of the sample and/or of the variables can be selected on the basis of various criteria. Selection criteria can be specified on demographic, income or expenditure variables. This is a powerful feature since it provides users with the capability to search the database and retrieve only the data needed.

MIDAS also has other features designed to facilitate the user. Users can browse through the complete database and selectively edit parts of it. The contents of the export file can be displayed on the screen before actually creating the file and various statistics (e.g., mean, count, standard deviation) can be estimated and displayed on the screen for any variable in the export file. Utilities are included to delete or rename databases or to temporarily suspend the execution of the system.

To facilitate the use of the system and therefore increase its acceptance, special consideration has been given to the design of the user interface, the information system through which users interact with the database. The system is completely menu based with users specifying their selection through a "point - and - shoot" interface. Users do not have to be familiar with the structure of the database or learn query/programming languages. The menu system translates user's requests to the appropriate commands for accessing and retrieving the data.

Another feature of the system is availability of on-line data dictionaries that contain the names of all data items, their description, value labels etc. Users can specify variables to be included in the export file by browsing through the data dictionary.

The information systems consists of two major components: the databases and the software for accessing the databases and providing the interface with the user. The database is of the relational type and has a structure optimized for family budget research. Data for every period are organized in separate files. For every time period there are several files; incomes, expenditures, demographics for persons and households, metadata files that describe the data, history files that have information on the various databases and others. Users can acces all these databases through the information system.

Income and expenditure data are stored using a special data structure that takes advantage of the characteristics of the database and stores only the non-zero entries. The data matrices for family budget research are very sparse since each person/household does not receive income from all sources and does not spend on all items. The experience of the Central Bureau of Statistics in the Netherlands shows that the expenditures matrix is about 80% sparse. Hence, substantial storage savings can be obtained by storing only the non-zero entries.

All data are stored as a matrix with rows representing households/persons and columns representing attributes with respect to income sources or expenditures. The data for every record (household or person) are stored in a manner that minimizes storage requirements. Every record consists of two parts, part A and part B. Part A consists of bits that take the value 1 or 0. There is one bit for each income source or expenditures item. The bit takes the value of 1 if the household is receiving income from that source or consuming that product and 0 otherwise. Part B of the record contains the values of the nonzero responses. Major savings in data storage are accomplished because 8 bits in Part A can be represented by one byte. Experience with actual data has shown that this data structure compresses the original data by an avarage of 75 percent. Additionally, this data structure speeds up significantly the data retrieval process.

3.2 Metadata

Traditionally, information systems for statistical operations have been designed with the objective of storing only numerical data. Information on the non-numerical data, that is the metadata, were not given appropriate attention since their management has always been a thorny issue (Cubitt, 1983). In the CASIP system the automatic manipulation of the metadata is an essential component and has therefore been given special attention.

MIDAS obtains data and metadata from the EDC system and has to transfer the information about the variables and the value labels etc to the EDA system. In the next section we will indicate that the system even allows the substitution of the metadata of one research agency by the one of another agency without effecting the functioning of the system. This possibility allows the portability of the system from one statistical office to the next without much problems, each using its own classification system or metadata.

3.3 Portability

The MIDAS was developed to be portable since it was recognized that the list of potential users include national statistical agencies and other government and private organizations that use a variety of computers, operating evironments and procedures. Although complete portability is difficult to obtain, MIDAS contains features which facilitate the implementation of the system in:

- various computer platforms
- different national statistical agencies.

Portability across various computer platforms is achieved through the use of standard ANSI C and has been developed on PCs under DOS. The standard C code has been used to facilitate the transfer of the complete system to other computing environments. Commercially available C libraries - C Tree Plus (tm) and C-scape (tm) - are used for index creation and screen management. Presently, the system has been implemented on IBM-type personal computers under DOS and SUN workstations under SunOS.

Designing a system that can be easily transferred across various statistical offices which might organize their family budget research in a different way was a more difficult task. This is accomplished by storing all information about the metadata files which can be easily altered. As discussed earlier, MIDAS users define the structure, codes, labels of the various data items in metadata files which the system "reads" to produce a customized version of the microdata system. The alternative approach, hardcoding meta information in the system, would have made the transfer of MIDAS across statistical offices a difficult and time consuming process since extensive changes would have to be made in the source code.

For example, by storing the description, codes and lables for the various expenditures in the metadata files rather than coding them directly in the code, MIDAS can be implemented by just changing the contents of the metadata files without altering the computer code and recompiling. Transferring the system in a different statistical office requires the creation of a new metadata file that has the codes and descriptions of the expenditure items used in this office. This information comes automatically from the EDC system, if that has been used.

4 EDA: An Expert system for Data Analaysis

The goal of the EDA (Expert system for Data Analysis) is to produce a series of tables which we will call a report. Although it is an open system, EDA should also automatically produce standard reports for statistical offices.

The inputs are the export files with the corresponding meta information provided by the MIDAS. The user can specify a list of tables (report) which should be produced. The output will be a set of tables. The major advantage of the EDA system above alternative programs like ABACUS, is that the system can produce complex tables with different types of variables and therefore allowing the calculation of different statistics in one table which is not the case in other programs.

4.1 Features of the system

Due to the complexity of the reports that can be generated by EDA , it is necessary to use a relational model to specify the reports. A report is seen as a set of tables while a table is defined by the crossing of a categorical variable, each category in one column, by a list of different variables (socio-demographic variables, income sources and expenditures etc.) in the rows. There may be as many variables in the rows as desired. The last set of variables can also contain newly derived variables like cost of living, buying power. The system provides also possibilities to create aggregated variables by use of the hierarchical structure of the variables in the export file.

If the variable is continuous the variable will occupy only one row presenting for each column category a mean or percentage of the total (income or expenditure). It is possible to take into account nesting of the categorical variables. If the variable is categorical each row will be represented in a row and the cell will give the abolute frequencies or the percentages on the total. For every row variables EDA may calculate a different statistic.

The set of individuals in the tabels will be defined by the Domain field of the table. The column variables partition the population of households and for these partitions EDA provides statistics for a large number of variables specified as row variables.

The relational model for this system is implemented in a prototype using microsoft C as language, which includes thirteen relationships and their corresponding files. The user will interact with a friendly interface using Microsoft Windows, which will translate all the specifications in relationships. To implement these report definitions the nine relationships of Table 1 are used.

Table 1: The relationships used to define the tables in EDA

RELATIONSHIP	DOS FILE NAME	DESCRIPTION
REPORT	REPORT.DCT	Documental part of the report
REPTABLE	REPTABLE.DCT	Set of pairs report-table
TABLES	TABLES.DCT	Info about tables
GROUPS	GROUPS.DCT	Infor about groups of variables
GROUPVAR	GROUPVAR.DCT	Set of pairs group-variables
COLUMN	COLUMN.DCT	Info about columns of each table
ROWS	ROWS.DCT	Info about rows of each table and the statistics to be calculated
OPERATIO	OPERATIO.DCT	Formula for statistics
DERIVE	DERIVE.DCT	Formula for derived variables.

The specified relationship should have the following functions:

- *Report relationship* defines the list of different documents (for the moment reports) which can be produced in a standard way.
- *Reptable relationship* defines per report the list of different tables formed.
- *Tables relationship* enumerates the list of tables defined in EDA in a given moment. The combination of such tables will form a report.
- *Groups relationship* serves to define a group as a set of digits of the codevar field.
- *Groupvar relationship* enumerates the list of variables per group.
- *Colum relationship* serves to specify for each table the variable(s) which will appear in the columns.
- *Rows relationship* serves to specify for each table the variables which will appear in the rows.
- *Operatio relationship* enumerates the list of statistics that will be supported by EDA.

- *Derive relationship* contains the algebraic expressions used for the creation of new variables.

Using these nine relationship EDA can produce any table on the basis of the export file generated by the MIDAS.

4.2 Metadata

In order to simplify the work for the user, not only the export file has to be made available for use in EDA but also the metadata about the contents of this export file. In order to make this transfer of information as smoothly as possible the designers of the CASIP have agreed on a set of rules for transfer of information from the MIDAS to EDA. It would lead too far to discuss them here. But the consequence of these rules is that the user does not have to do anything to obtain the information about the variables and the category labels etc. They are automatically produced by the system and immediately ready for use. Knowledge about these rules is only necessary if one uses a different DBMS than the one discussed here (MIDAS). For more details about these rules we refer to Marti Recober et al. (1991).

4.3 Portability

Due to the fact that Microsoft C is used for the program which works under Microsoft Windows the system is quite portable since the use of Windows is growing rather quickly.

The input of the system is a set of ASCII files with data and metadata. This part of the system allows the use of the system for most database management systems as long as the rules for the specification of the metadata are taken into account.

5 EPSS: Expert system for storing and presenting summary statistics

To complete the fully automated system for the data collection and analysis of the family budget survey there is a need for an information system to store, retrieve, present and analyze the summary statistics estimated by the expert system outlined above.

The potential users of the summary data set include not only the staff of the National Statistical Offices but also other researchers, EUROSTAT and public and private agencies involved in economic planning, marketing etc. Since the users are not necessarily familiar with the contents of the family budget dataset, the information system should provide an intelligent way to access the information. The information system should be designed to answer the five key questions users of this database will be faced with:

- what kind of data are available,
- how to access the stored information,
- what are the various metadata associated with the different variables,
- how to analyze the data,
- how to present the information.

This sytem is still under development but it has been decided that the metadata is again automatically transferred from EDA to EPSS and that the user can select specific rows and columns from the tables specified by the EDA system. Such a selection will lead to a new smaller table and then the user can specify in what form he/she would like to get the results. This means that the user is offered a choice between numeric values and graphical representations like bar charts etc.

It will be clear that this program is completely dependent on the EDA system for its information and can not be used for any other set of tables.

6 Conclusions

The four information systems were discussed independently of each other. To produce an automated system, however, the four systems should be closely interconnected. This interaction should be transparant to the user.

Of crucial importance in the system is not only the transfer of the data but also the transfer of metadata. In this project a lot of emphasis is placed on this aspect so that the user will not detect that (s)he is working with four different systems. The data and metadata go automatically from one system to the next and the hierarchical structure which is available in the categories of products and sources of income in the data collection can be used in the different subsystems for aggregation.

These hierarchies can be specified by the user in the data collection or in the step from the data collection to the data storage. This means that each user can use a different classification as is the case in Europa. But a special agency like Eurostat, who would like to harmonize the results obtained from the different agencies, only has to change this hierarchy to obtain its own classification.

The final results can be specified once and can then be used in an automatic way, allowing for very fast data processing. The speed of the process is even more improved by the fact that the data collection system contains most of the validation checks which are normally done after the data collection. If these checks work fine the production of the standard statistics is more a matter of hours after the data collection has been finished, than a matter of years as we see nowadays.

References

[1] Bastelaer A. van, F. Kerssemakers, and D.Sikkel (1988): A test of the Netherlands labour force survey with handheld computers: interviewer behavior and data quality. In D.Sikkel (Ed.): *Quality Aspects of Statistical Data Collection.* Amsterdam: SRF (Chapter 3).

[2] Bie S.E. de, I.A.L. Stoop, and K.I.M. de Vries (1989): *CAI Software: An Evaluation of Software for Computer Assisted Interviewing.* Amsterdam: VOI.

[3] Cubitt R. (1983): Meta data: An experience of its use and management. In Proceedings of the Second International Workshop on Statistical Databases Management, LBL,USA

[4] Hartman H., W.E. Saris, I.N. Gallhofer, W.M. de Pijper, J.Leewin, J.Verwey, N. Lemmens (1991): *Data Collection on Expenditures.* Working paper, Amsterdam: SRF.

[5] Knaus R. (1987): Methods and problems in coding natural lanquage survey data. *Journal of Official Statistics*, **3**, 45-67.

[6] Lorigny J. (1989): Quid, une méthode générale de chiffrement automatique. *Techniques d'enquête*, **14**, 307-316.

[7] Luykx Draelant E. (1981): Geharmoniseerde gezinsbudget onderzoekingen 1979. Report made for Eurostat.

[8] Marti Recober M. (1991): New proposal on meta-information organization. Working paper: DOSES-UPC-EIO-FIB, Barcelona.

[9] Nicholls W.L. and R.M. Groves (1986): The status of computer assisted telephone interviewing. *Journal of Official Statistics,* **2**, 93-115.

[10] Olivier A.J. (1987): Het samenstellen en beheer van gegevens als onderdeel van een beslissings-ondersteunend systeem ten behoeve van het marketing management: een case study. Ph.D. thesis, Unversiteit van Groningen.

[11] Pijper W.M. de and W.E. Saris (1986a): *The formulation of interviews using the program Interv.* Amsterdam: SRF.

[12] Pijper W. M. de and W.E. Saris (1986a): Computer assited interviewing using homecomputers. *European Research*, **14**, 144-152.

[13] Prastacos P. and M. Diamandakis (1990): The design of the information system for storing microdata information in family budget research. Working

paper 90-6. Institute of Applied and Computational Mathematics. Research Center of Crete. Heraklion, Greece.

[14] Prastacos P. and M. Diamandakis (1991): The information system for storing microdata information in family budget research. Working paper 91-7. Institute of Applied and Computational Mathematics. Research Center of Crete. Heraklion, Greece.

[15] Rubin D.B. (1987): *Multiple Imputation for Nonresponse in Surveys*. New York: Wiley.

[16] Saris W.E. (1989): A technological revolution in data collection. *Quality & Quantity*, **23**, 333-349.

[17] Saris W.E (1991): *Computer Assisted Interviewing*. London: Sage.

[18] SRF (1990): Data collection on incomes. Working paper. Amsterdam: SRF.

[19] Wenzowski M.J. (1989): Un système généralisé de codage automatique. *Techniques d'enquête*, **14**, 317-326.