

# **BLAISE IN EUROPEAN COMMUNITY HOUSEHOLD PANEL OF STATISTICS ITALY**

*Alessandra Sorrentino, ISTAT, ITALY*

## **USE OF BLAISE IN ISTAT**

The presence of Blaise in Istat dates back to year 1994.

Some CADI and CAPI application, even dealing with complex questionnaires, have been developed.

Salary and Wage Structure Survey, Small Enterprises Survey, Italian Household Budget Survey and European Community Household Panel made use – and some of them still use - Blaise System.

Interesting studies regarding, for example, the Italian Household Budget Survey - an important monthly survey - have been conducted on comparison between Blaise controlled data-entry and traditional data-entry (almost without check). Some consideration about this matter:

- 25% time increases during Blaise controlled data-entry. On the other hand this is a natural consequence of this new methodology concentrating on data quality rather than on acquisition speed;
- however this increase in time for acquisition, results in a decrease in time for manual revision of questionnaires;
- moreover Blaise user interacts with this new way of working and seems to be more interested in his job.

Blaise version 2.5 has been used until 1996. Between 1996 and 1997 Istat bought Blaise III licence that simplified the realization of cross-sectional and longitudinal check programs.

A quite user-friendly language represents one of the most interesting Blaise features even if the user is not a skilled software developer. This feature is one of the reasons for its diffusion.

Many surveys are carried out in Istat, but the data-entry is usually performed externally by private companies, with the help of other software. Certainly this is one of the reasons of restricted diffusion of Blaise system in Istat.

Best way for exploiting Blaise system features is to design a controlled data-entry application; in Istat, however, we managed to obtain many other advantages, by means of computer aided revision and edit.

## **EUROPEAN COMMUNITY HOUSEHOLD PANEL - ECHP**

The most demanding Blaise application in Istat is the one concerning the European Household Panel (ECHP).

The ECHP is a yearly survey carried out in States members of European Community. Eurostat had planned this survey and coordinates the activity of different European Countries.

The first Wave started in 1994 and the surveys are going on till 2002.

The models collect questions about work, income, properties, residence, training and education, health, migration and other social indicators concerning household and his members.

The feature of the panel technique is that the interviews are carried out on the same sample of persons and households every year. This allows, after the first year, to analyse the evolution of the sample, from different points of view.

Through a longitudinal comparison of two or more waves or a cross-sectional study between data collected in the 14 countries involved, it is possible to single out social and economic characteristics of Countries, becoming a part of a European context.

## **CAPI EXPERIMENTATION IN ECHP**

The department of methodological studies in Istat has carried out a little CAPI experimentation regarding ECHP survey with Blaise programs. Some interviewers together with Istat employees have interviewed about a hundred of the households using a notebook.

It has been the first CAPI experimentation in Italian Statistical Institute and the outcome was successful both for the positive reaction of households and the good quality of data collected.

At least 50% of households have showed interest in using a notebook.

However this mode of conducting the survey results too expensive without a good network of interviewers (depending on Istat and equipped with notebook) covering all the national territory.

## **BLAISE IN ECHP SURVEY**

On the basis of four standard models sent by Eurostat to Community Countries Members, three Italian models have been prepared (each one with its one record type), keeping the same information in a different format. This adjustment for the Italian social-economical reality has been essential to allow the interviewed households to understand and answer to questions.

The three models are:

1. Household Register is a basic instrument of operational control in the ECHP, collecting information about number of household members, the coding of interview result, the relationships between the individuals and demographic information.
2. The Household Questionnaire contains information about migration, income and economic situation of household.
3. The third and more complex model is the Personal Questionnaire. Each member in the household aged 16 years or more is interviewed. The individual model collects detailed information on each person's economic activity and income, instruction, health and a large number of other variables.

Blaise 2.5 has been used for the first and second waves (1994 and 1995). This version has allowed simple management and automatic counting of dirty and clean records, but this has involved restrictions in datamodels linkages.

Because of the complexity of this surveys, the high number of questions, the longitudinal follow-up year by year both the mass of data and the possibility of mistakes increases and, consequently, the number of checks to implement.

The passage to Blaise III, between 1996 and 1997 gave way to an easier and exhaustive development of checks, carried out either in respect of previous years or current year data.

Thanks to Statistics Netherlands the passage to new release was not a problem.

Our Statistic Institute did not buy Maniplus licence, that had allowed a personalization in procedure. In order to make a user-friendly application, we prepared simple Dos screen in third and fourth Waves (1996 and 1997), and a Visual Basic application for two last waves (1998 and 1999).

As described before, in ECHP survey Blaise is not used as a data-entry system – except the CAPI experimentation - but for the following phase regarding computer aided revision and edit.

Procedure steps in ECHP are listed below:

## **1. DATA-ENTRY BY EXTERNAL COMPANY**

Until 1998 recording of compiled models is carried out by a specialized data-entry company, free to use any software for data-entry. This company also carried out a minimum number of checks.

Data-entry concerns about 7000 households for a total of 30000 questionnaires (7000 Registers, 6500 Household questionnaires, 15000 personal questionnaires).

## **2. COUNTING RECORDS AND KEY CHECKS USING SAS**

Records are stored on a PC Pentium II Hard-disk.

First step on rough data is the count of questionnaires ordered by geographic region.

By means of Sas language double keys are checked out.

## **3. LOADING DATA AND CHECKRULES USING MANIPULA**

Six PC Pentium II are available for this work.

By means of a LAN different users can access data at the same time.

After storage 'Check rules' step starts.

Manipula provides the conversion from Ascii to Blaise format and marks the incorrect records.

We perform batch data processing at evening or nighttime, because data checking in this step could take a large amount of time.

Manipula processing performs the following actions:

- pointing out the range errors;
- pointing out the route errors;
- pointing out the incompatible values among variables of current wave (cross-sectional checks);
- pointing out the incompatible values between current wave variables and previous wave ones (longitudinal checks).

Check programs contain two types of incompatible errors (hard and soft errors):

- If the user doesn't correct the error the record remains dirty and data-entry could be inhibited: this is a hard error.
- if the user decides that the variable is correct he can close the message box relevant to the error and retain the original variable value (but the record hereafter will be considered as clean): this is a soft error.

To improve data query performances during checking, different 'datamodels' containing 'NACE', 'ISCO' and countries code are linked to current wave questionnaires.

'Datamodels' containing previous years variables are also prepared.

## **4. COMPUTER AIDED REVISION AND EDIT USING BLAISE**

Computer aided data correction is a successive step after Manipula 'checkrules'.

Check plan considers incompatibilities between different variables of current year (cross-sectional check).

Moreover, some variables from previous years data files, are extracted and linked to current year data (longitudinal check); links between current year variables and code files has been made as well.

To avoid system overloading during this phase of correction, only selected variables from previous year have been extracted.

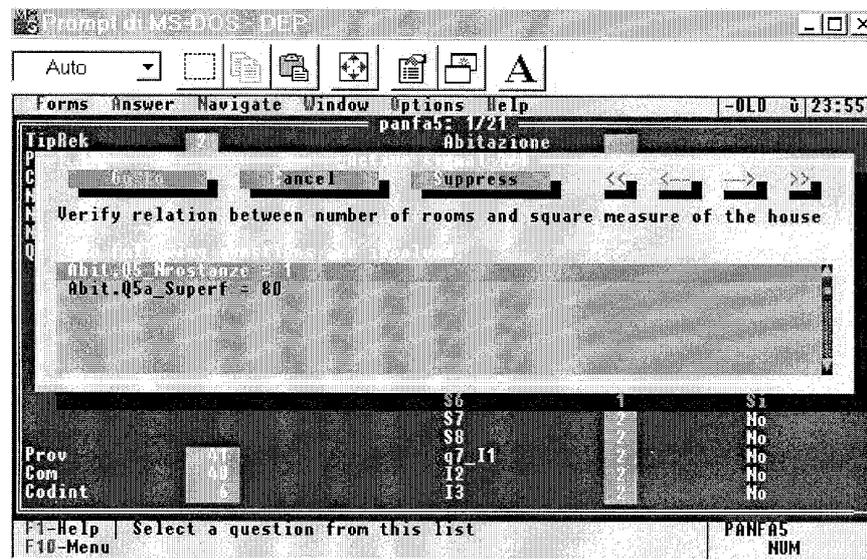
Comparison between data belonging to different waves is necessary in a panel survey.

The main feature of a panel consists on the longitudinal dimension. This means that some variables could be in contrast with the same variables in the previous years of the survey. While what is needed is that outcome be coherent with the past waves

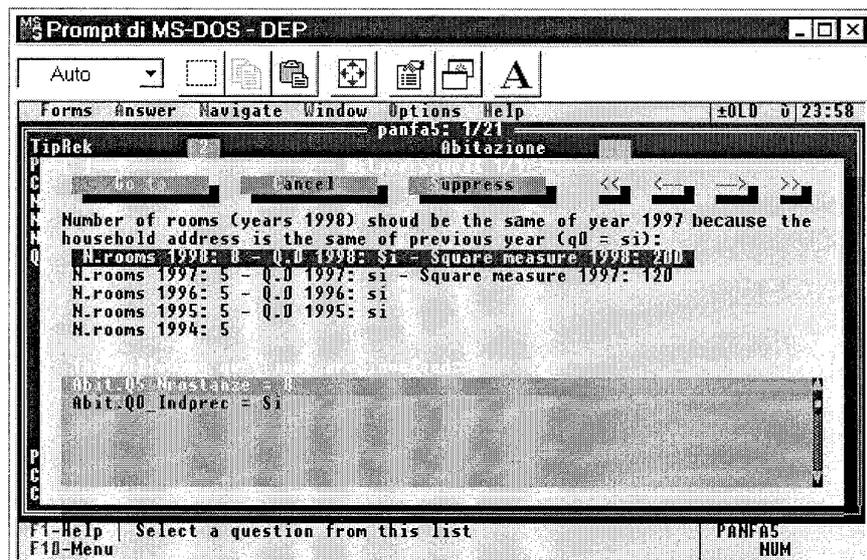
Each data resulting in contrast with assigned constraints is enlightened (to enlighten) as a mistake and a relevant message box appears on the screen. On this message box, even the variables involved in the error, appear.

Some examples of checks here below:

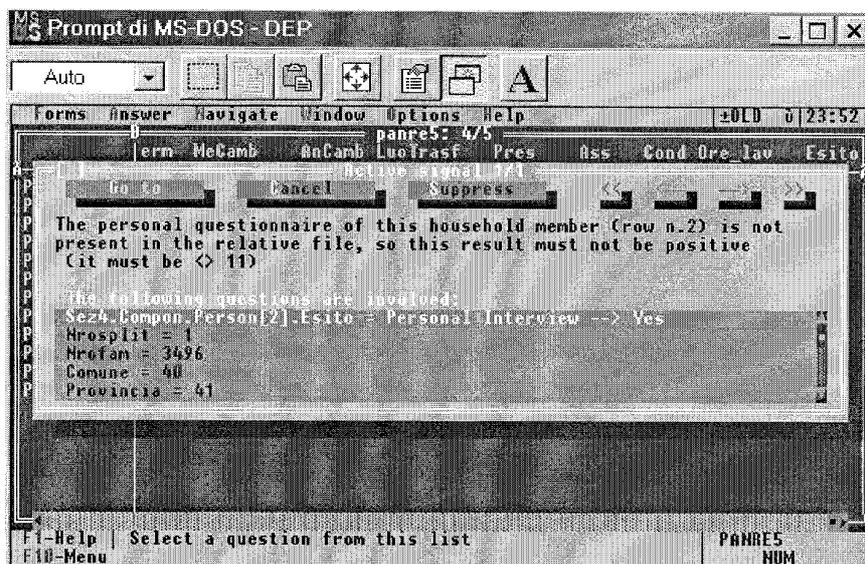
- ❖ Household Questionnaire: relation between the number of rooms in the house and its square measure is incompatible:



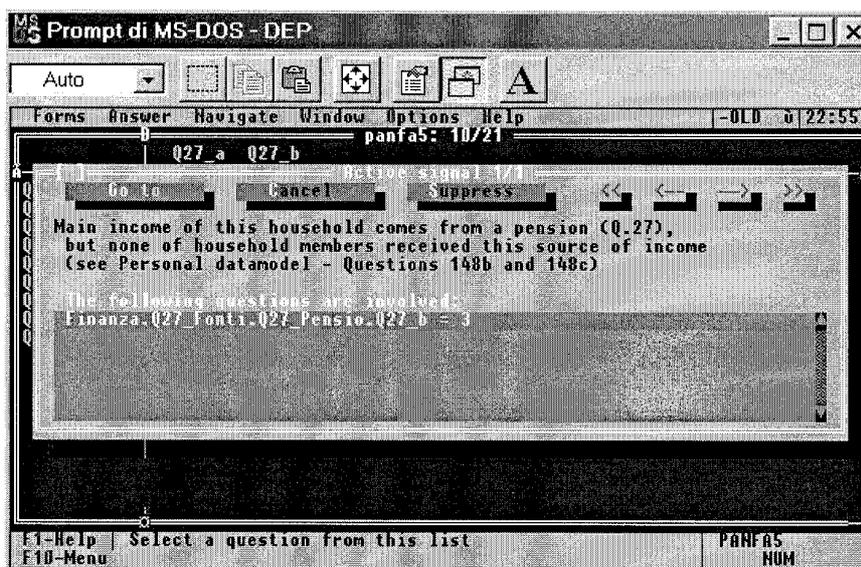
- ❖ Household Questionnaire: the number of rooms in the house is different from the number of rooms in previous year. The household, however, lives at the same address. An explicative message shows the answers for all previous waves: the user can decide which is the best answer.



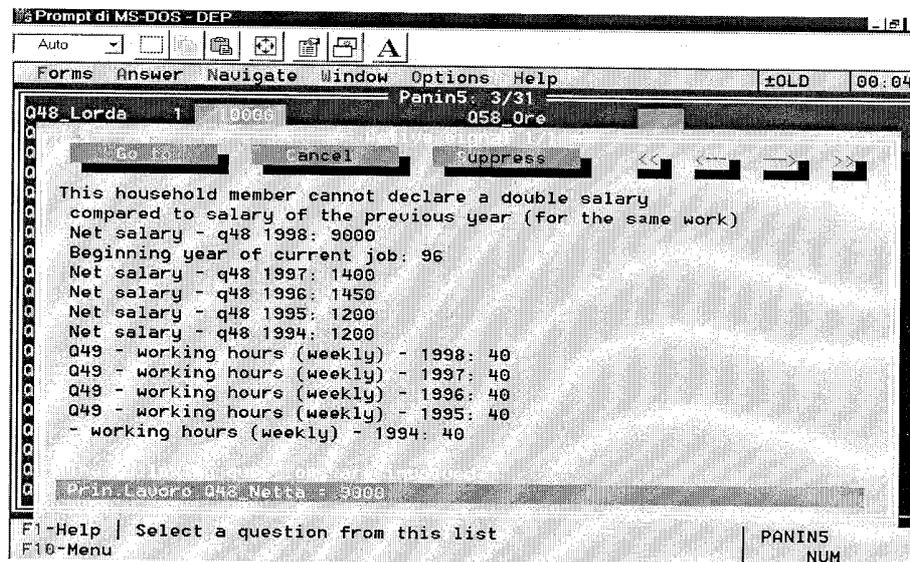
- ❖ Household Register: even though in household register is declared that two personal questionnaires are filled in, there is only one personal datamodel. Blaise requires to verify the number of existing paper model:



- ❖ Household Questionnaire: the application links some information from household questionnaire (main source of income coming from pension) with some other information from the personal questionnaires (source of income). Error arises because none in the household receives a pension indeed.



- ❖ Personal Questionnaire: the interviewed person has declared too large amount of salary compared to the amount for the previous years (obviously for the same work). This is an example of longitudinal check:



The following table summarizes some information concerning datamodels and error types for Wave 5 – 1998:

Datamodel	Program rows	Questions Number <sup>(1)</sup>	Fields	Error Type A <sup>(2)</sup>	Error Type B <sup>(3)</sup>	Error Type C <sup>(4)</sup>	Total Error A+B+C	Automatic Correction
Register	1755	20 + 13	400	4	12	59	75	19
Household	1261	45	120	11	3	25	39	47
Personal	3420	202	400	18	16	177	211	75

<sup>(1)</sup> - Question number for Register: 20 questions concerning household + 13 questions for each household member.

<sup>(2)</sup> - Error type A: pointing out of the errors arising from linking variable values contained in other models for current wave or code files.

<sup>(3)</sup> - Error type B: pointing out of the errors arising from linking variable values from previous waves.

<sup>(4)</sup> - Error type C: pointing out of the errors arising from inconsistency between variable values within the same datamodel.

## 5. WORKING IN DOS AND VISUAL BASIC

To make user-friendly the Blaise III application we prepared simple Dos form for Wave 3 - 1996. This Dos menu is designed by means of Dos batch programs.

In short the Dos menu allows:

- ◆ direct access to selected datamodel, without passing through the main Blaise menu;
- ◆ automatic activation of certain data-entry mode (Data Editing Mode);
- ◆ clean and dirty records counting;
- ◆ conversion from Blaise to Ascii format;
- ◆ automatic data saving on floppy-disk.

Moreover, a very useful feature that we have used, is the possibility to change easily data-entry mode. For example, Data Editing Mode facilitates navigation in datamodels, even if there are hard type errors.

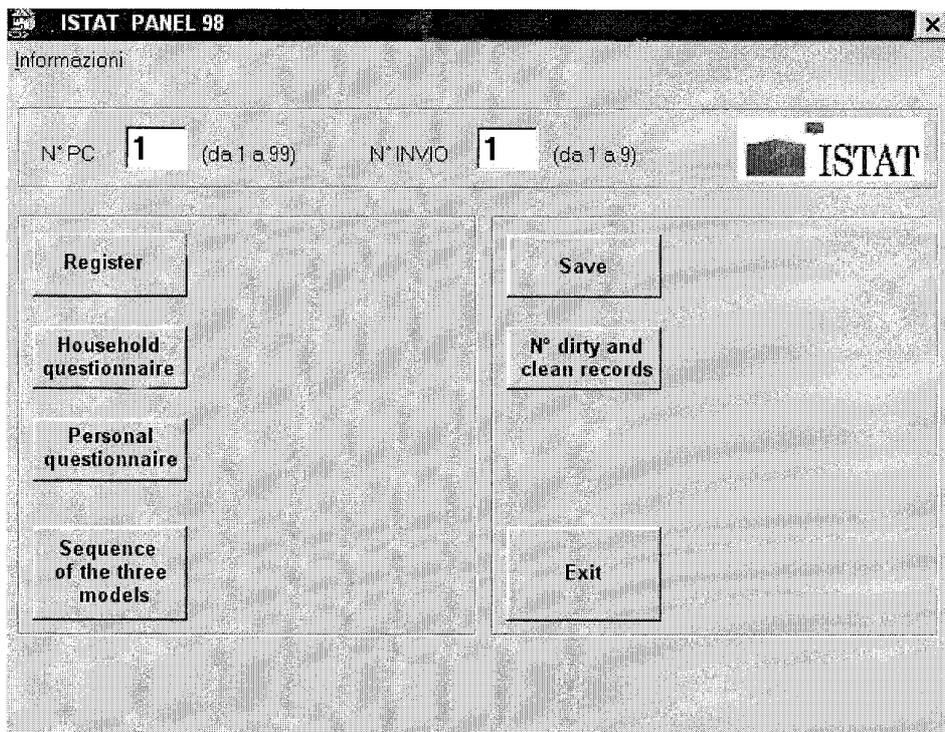
In our department revision and edit is organized in geographical areas and each employee works on a particular area. This means that a global counting of dirty records was a nonsense. That's why we prepared other Dos Bat files, managed by a Visual Basic application.

This application allows a record counting selecting a given area (provincia). So each user is able to check the number of the record that he has cleaned so programming his work timetable.

Visual Basic application allows the following functions:

- ◆ one or more questionnaires selection;
- ◆ saving of data on other PC's;
- ◆ clean and dirty records counting;
- ◆ recording of a file (log) containing the number of records cleaned and the dates in which this count was required;
- ◆ exit.

Visual Basic application, that manages described query, is showed down:



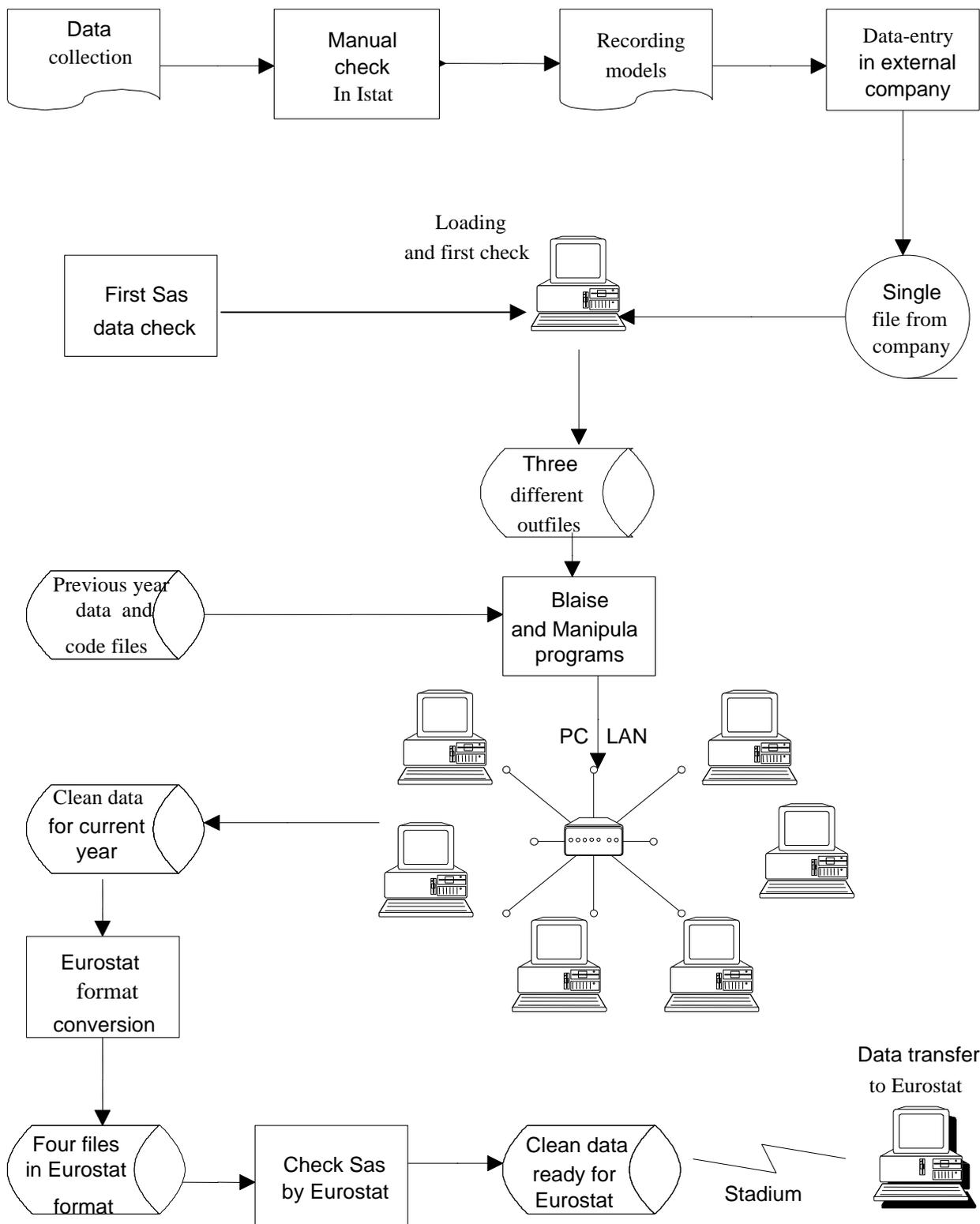
## 6. EUROPEAN DATA FORMAT CONVERSION USING COBOL AND SAS

Once Blaise data correction on Italian format is performed, record types are converted in four Eurostat record types, by means of Cobol and Sas language. Output data are submitted to Sas checks written by Eurostat. Basically these checks account for relationships among individuals, structural errors (i.e. incoherence between two models in the same wave), plausibility (checks on variables – for example income, pensions, mortgage – between two different waves) and so on.

Although the error checklist by Eurostat is large enough, we thought to enlarge this list by adding our own Blaise error codes.

At present we are also thinking to improve automatic editing using appropriate software tool, such as Scia.  
 Scheme of Panel steps survey (from 1994 to 1998) here below:

### FLOW-CHART OF PANEL PROCEDURE



## QUALITY INDICATORS

Since Blaise does not count the total errors for each datamodel, but gives only total number of clean and dirty records in a file, for a better understanding of what type of error it is and its frequency, we completed the records (in Blaise format) by inserting error codes at the end. In this way it is quite simple to get statistics about the errors.

The table shown below – as an example – is calculated on a sample of 500 households. The example regards the household questionnaire only:

Error Code	Error Frequency	%
CBF001	10	2,81
CBF002	10	2,81
CBF006	26	7,30
CBF010	1	0,28
CBF011	81	22,75
CBF013	1	0,28
CBF017	6	1,69
CBF019	3	0,84
CBF020	10	2,81
CBF023	1	0,28
CBF024	3	0,84
DBF001	43	12,08
DBF003	1	0,28
DBF004	2	0,56
DBF005	1	0,28
DBF008	1	0,28
DBF009	75	21,07
DBF011	12	3,37
SBF001	9	2,53
SBF002	6	1,69
SBF003	2	0,56
SBF004	2	0,56
SBF005	11	3,09
SBF008	38	10,67
SBF010	1	0,28
<b>TOTAL</b>	<b>356</b>	<b>100,00</b>

Code C...: cross-section error type.

Code D...: automatic correction.

Code S... : structural error type.

Analysing this table allows the researcher to have a feedback to operate on questionnaire formulation or on more instructions to the interviewers.

## **FUTURE DEVELOPMENT OF BLAISE IN ISTAT**

For the 6th Wave of ECHP (1999) we decided to perform controlled data entry using Blaise III release.

At present we are planning to pass to Blaise IV release for Wave 7 (2000).

We are also examining the possibility to carry out an important survey, such as Italian Labour Force Survey, making use of Blaise IV Capi mode.

## **CONCLUSIONS**

In this paper the experience of Statistics Italy using Blaise has been presented, with a particular attention to European Community Household Panel.

Blaise III has become an integral part of Panel survey processing methodology, helping (by designing the appropriate datamodels) researchers to understand how to improve questionnaire and easing the checking phase too.

This document shows also the good properties of Blaise III, Manipula and Visual Basic as tools for a rapid application development.

Thanks to:

Sara Mastrovita and Pierpaolo Massoli for the active collaboration in paper translation;

Stefania Macchia for studies on Italian Household Budget Survey and CAPI experimentation on ECHP;

Sandro Baldanza for Visual Basic application;

Concetta Pellegrini for Flow Chart of Panel Procedure.

Alessandra Righi, Giuliana Coccia, Stefania Macchia and Giulio Barcaroli for comments on paper.