RUT: 79.510.270-1
Av. Encomenderos 260 Of. 83
Santiago-Chile
www.serpro.com

DATA ENTRY PROGRAM REPORT

KOSOVO HOUSEHOLD SURVEY

October 2000

TABLE OF CONTENTS

Data Entry General Considerations. 3
The Household Survey DE Program. 5
Module 1: Household Identification and Control Information. 6
Module 2: Demographic Characteristics and Migration. 8
Module 3: Education. 11
Module 4: Labor. 12
Module 5: Non-Farm Enterprises. 16
Module 6: Networks. 18
Module 7: Agriculture. 18
Module 8: Consumption. 21
Module 9: Health. 21
Other Modules. 21
The Community Questionnaire. 22
Other Tasks in SERPRO's TORs. 24
Conclusions. 27

- Attachment 1 - LSMS Survey DE Program
- Attachment 2 - Community Survey Data Entry Program

•

The software developed to capture the Kosovo Household Survey data is using the Census and Survey Processing System (CSPro) as platform, running under WINDOWS 95/98 or 2000.

Several modules have been designed to capture and monitor the survey data, aiming to facilitate the data entry (DE) operation, minimizing the work required from the DE clerks and increasing as much as possible the detection of errors coming either from the DE clerk or from the data capture in the field. In addition to the data entry, two batch modules have been designed and addressed to the supervisors to insure that the DE operation has been carried out properly.

The data entry program has been developed keeping in mind the following general considerations:

Data Entry General Considerations.

1. The data entry screens have been designed as similarly to the questionnaire as possible. This consideration is important for the DE operator to quickly relate the information on the questionnaire and the screen, minimizing the time required to fix the inconsistencies detected by the program. This factor is important in an interactive DE environment, where there is a constant communication between the program and the operator.

2. The screens have been designed to avoid both horizontal and vertical scrolling whenever possible. This is also important since, at the time of reviewing the information in potential conflict, it is useful to have the whole picture on the screen.

3. The flow followed by the data entry screens is not always the same as the one followed by the interviewer. It has been designed to complete one page of the questionnaire before passing on to the next one. This consideration is also important since it allows the operator to concentrate in just one page at the time rather than going back and forth between consecutive pages.

4. To a large extent, the household (HH) questionnaire has been organized in rosters where each line represents a unit of the specific universe being investigated. The number of lines that each roster should store is given by the universe that the specific module should have. Many times the universe of a specific module is a subset of a larger universe of a predecessor module (i.e. “all household members 10 years and older” is a subset of “all household members”). When a specific universe is known by the system, the appropriate element/line will be identified and brought in without the operator’s intervention; in this case, no blank lines are required to make both rosters parallel (the whole universe and the subset). Both

rosters are linked together by a pointer, usually the id code that is present in almost every single roster. Under this scheme, the operator concentrates in checking that the id retrieved by the system matches the one found in the questionnaire. When the number of lines that have to be filled in a roster is unknown by the program, a common identifier (9, 99, 999, etc.) has been used to determine the end of the list throughout the different rosters. However, if the universe is known, as soon as the last line is entered by the DE operator, the flow continues with the next roster or question in sequence.

5. The interview flow/path followed by the interviewer has been duplicated in the data entry program. This means that whenever the answer to a given question requires that the interviewer skip one or more questions or even a whole section or module, the DE program performs exactly the same skip and the focus is set to the appropriate field. The DE operator doesn't need to locate the next field since this is done automatically by the system. When one or more fields have been skipped by the flow logic, they cannot be accessed unless the answer to the question that originated the skip is modified. This constraint is an important consideration to guarantee that the universe that has to answer a given question is the one defined by its filter question.

6. All answers are passed through a range test that makes the first basic check to ascertain that it is a valid answer. If the answer/code entered is an out of range code, the system will issue an error message asking if the value is accepted or not. CSPro has the option of setting the out of range error as not acceptable, in which case the operator will be forced to change it to a valid code. In this case, when an out of range code/value is entered, the system will issue an error message; to close the message window, the operator needs to move the cursor to the acknowledge check box. Thus, it is impossible to enter an out of range code/value without the operator's acceptance. The decision of accepting out of range values (with the operator's acknowledgement) was taken mainly to admit the inclusion of new codes in the questionnaire without creating a problem at the DE time.

7. Some modules include a long list of items with a filter question (Yes/No) specifying whether it applies or not. In all these cases, the list is a solid template where only those items that show a *Yes* answer are asked the following, predefined, questions. To avoid entering the whole list of items, the *Yes/No* answer has been omitted and replaced by code presence. That is, if a code is entered, the presence of the code means that the answer was *yes* and therefore, the subsequent answers are expected to be entered. When the code is entered, the corresponding item description is immediately displayed. The description display is just a help to the DE operator to make a visual examination that the code entered does correspond with the description on the questionnaire. Again, this approach aims to minimize the amount of information entered and, at the same time, it minimizes the possibilities of originating new errors.

8. The DE program, besides checking the code/value ranges, includes a large number of logical checks that aims to the error detection. Since those errors could have been originated not by the DE operator but by the interviewer/respondent in the field, some of them just issue a warning message and others, considered critical for the rest of the questionnaire, force the operator to fix the conflict before moving ahead. The error messages clearly describes what the conflict is, what variables/answers are originating the conflict, and in many cases the values/codes of each variable. The correction of the detected errors in many cases requires the intervention of a demographer or survey specialist, particularly when the error was originated in the field. Thus, at my request, the DE team was complemented with the presence of Edis Agani, who will advice directly the DE team in this important matter. She will also have to make decisions about what errors are serious enough as to require that the whole Primary Sampling Unit (PSU) questionnaires be sent back to the field.

The Household Survey DE Program.

The survey covers a variety of subjects grouped into eleven different modules, some completely independent from each other. However, most of them are closely related to the first module, the household roster. This report will use the same pattern to describe relevant issues about each module. The pattern will take the following structure:

1. General remarks about the DE program for the specific module.
2. Problems found.

Module 1: Household Identification and Control Information.

1. General remarks about the DE program for the specific module.

This module includes all the information to clearly identify each HH, the PSU it belongs to, and the geographic information where the HH is located. It also stores the interviewer and supervisor identification plus the DE operator id.

Since the PSU is a very important piece of information, a special procedure was designed to make sure that it was correctly entered. One DE operator was always responsible for a whole PSU, and each PSU is entered to a different data file. The name of the data file has a fixed structure: "PSU_XXX.dat" where XXX is the PSU number. When a PSU is assigned to a DE operator, the system starts by requesting the name of the data file. Later on, the name of the data file is checked with the PSU number to make sure that the HH information they are about to enter is correct.

An external file (table) was built to store the geographic information of each PSU; the information stored in this external table is as follows: PSU number, Village name and code, Municipality name and code, Urban/Rural characteristic that for the first 150 PSU is always rural, and the Area of responsibility code.

At the time the PSU number is entered, the external table is accessed to ascertain that the PSU exists in that table, and if it does, to recover all the related information that should be stored in the survey data file; that is, the municipality and village code, the Urban/Rural area code and the Area of responsibility code. The program directly copies those fields from the table and the DE operator has no access to them (locked fields).

The key fields to fully identify a HH are the PSU number and the HH number; thus, the system is checking that no HH number can be duplicated or entered twice. However, there is no checking that the same PSU is entered twice in a different computer. This will be a matter of administration and follow up of the whole operation that is important to define. Since the individual PSU data files will need to be concatenated for analysis, a frequency distribution of the PSU number combined with the HH number will show clearly any deficiency in this regard. SERPRO is planning to select important variables of each module to simply analyze the initial part of the DE process. Among them we will produce the frequency distribution mentioned above and I strongly suggest the use of this tool to detect early enough any deficiency in the survey operation.

2. Problems found.

The only problem found regarding the construction of the external table was the lack of information about the urban part of the sample. Therefore, the table currently contains information regarding the rural part and it will need to be updated when the rest

of the information is available. It is important to point out, however, that the urban part cannot be started without updating the external table. The program simply will reject any PSU that is not found at the time it is accessed.

Along with the information already mentioned, the date and time the HH data entry was started and finished is also being stored. This information is being passed directly to the data file without the operation intervention just for future analysis in case of any problem arises.

Module 2: Demographic Characteristics and Migration.

1. General remarks about the DE program for the specific module.

This module has four parts: (i) the Household Roster that gathers the basic demographic characteristics of the HH members plus information that permits to analyze the family composition; (ii) the Displacement part which is the continuation of the HH roster and collects information about the different changes of residence or migration off each HH member; (iii) New HH Members, oriented to gather information of current members of the HH that joined the HH after the conflict (March 1998) and were born before the conflict; (iv) Spouses and Children of HH members living elsewhere, aimed to collect information of those that having a direct link with members of the HH are currently living elsewhere.

1.1 The Household Roster.

Although all the information gathered in any survey is important, the information collected in this module is particularly important since it becomes the reference frame for many of the subsequent modules. Thus, special attention was paid to make sure that the HH roster was structurally correct, and that the HH members passed through a sophisticated consistency check model.

The roster structure was check in two different ways: the program assigned the line number or id code that is later used by many other modules to refer to specific HH members; the number assigned goes from one to the number of HH members. Unfortunately, the field was not locked since the way to leave the roster was a code 99 (end of list) in the id code. However, the program was checking that the operator did not change the assigned code assuring that there will be no gaps between two HH members.

Several parameters were defined to make sure that HH members satisfied at least these minimum consistency rules: *Minimum age at marriage* was set to 15 and later changed to 14; *Minimum age of Head of HH* was set to 15; Minimum age between two generations was also set to 14. Initially, the difference in age between two brothers was also going to be checked but due to the fact that a HH head could have more than one spouse, it was withdrawn from the consistency rules (along with the existence of more than one spouse in the HH).

This part of the survey includes about forty different consistency rules, some of them at the individual level and some to the HH level. To see all the individual consistency rules checked by the program, please refer to appendix one, where each individual rule is listed. At this level, only the broad issues checked by the program will be itemized.

As mentioned before, the consistency rules applied for this module fall into two different categories: those that refer to the individual HH member like "age has to be consistent with the birth date" and those that will guarantee that the family structure is *logically* correct. The later ones are numerous since the questionnaire is requesting information that is precise and error prone. Questions seven, eight and nine collect the spouse id code (line number), the father and the mother id code respectively. The program is checking that (i) the difference in age between two consecutive generations is at least equal or above the parameter defined; (ii) the sex of the related persons are correct (mother's line number belongs to a female HH member, etc.); (iii) the relationship to the head of the HH is consistent with the father and mother's lines numbers, etc.

Generally speaking, an important effort was paid to make sure that any errors in the information gathered in this module were detected and corrected at once. Thus, the family structure should be clean and free of errors that were susceptible of being identified as such.

1.2 The Displacement roster.

As mentioned before, this roster is parallel to the HH roster and thus, it has the same number of lines, one for each HH member. In this occasion, since the total HH members have already defined the number of lines, the id code column was locked and the information was pre-assigned by the program. The DE operator is forced to enter the pre-determined number of lines and the program skips to the next roster as soon as the last one is filled.

The Displacement roster is straightforward and does not admit too many consistency checks. The few consistency checks that are performed refer to the date of birth and question 3, and the number of times the person has changed residence and the main reasons (for having changed residence).

1.3 New HH members.

This roster is subject to a filter question that pre-sets the existence of this part of the questionnaire. In the same manner, the HH data might or might not show this record(s) depending on the value of the filter question. If the value is two, there should not be any record in the data file; if the value is one, there should be one record per new HH member.

1.4 Spouses and Children of HH members living elsewhere.

Once again, the number of lines in this roster is pre-defined by the HH roster. The roster has one line for each HH member older than sixteen years and thus, the program pre-defines also the frame in which the DE operator has to work. The id code for every HH member is copied from the HH roster and the id code column is locked.

Although the marital status of each person is known and therefore, it could as well be copied from the HH roster, the program uses question 2 to double check the marital status.

For HHM that have children living elsewhere there are four places of residence and four fields to store the year they left. However, if there is only one child living elsewhere, the remaining list does not apply. In this case, the operator once again uses the *End of List* code (99 and 9999 for the year) to exit the pre-determined four-field list.

2. Problems found.

There were only two problems that forced to change the scheme of the DE program: One, the number of lines pre-established for the HH roster had been set to twenty but soon it proved to be insufficient. At the beginning of the test period, some large HH were found (over 40 HHM) that forced to accept up to fifty members. Unfortunately, this roster is the one that defines the size of many others that also had to be expanded. As one can easily imagine, the amount of information the system has to store is *above large*. Since this survey was the first large survey ever processed with CSPro, we were wondering if the system was going to be able to accommodate that huge amount of information. The survey has in total about 850 fields that have to be multiplied by the number of occurrences or lines of each roster to which they belong. That is, roughly, about 42,000 different pieces of information stored at any one time. This was without a doubt an excellent “in the field test” for a newly developed system.

The second problem had to do with the Muslim idiosyncrasy in relation with the number of (wives) spouses they might have. Several consistency checks had to be suppressed including the one previously mentioned (inter-births spacing), although a warning message was issued when more than one spouse were found.

Module 3: Education.

1. General remarks about the DE program for the specific module.

This module has a total of thirty-three questions divided into three DE screens that have to be filled in for every HHM five years of age and older. Since the number and order of the HHM that should be answering this module is known, the program creates the working frame and copies the id codes from the HH roster. The id code column has been locked and the DE operator has no access to this information.

The consistency checks are limited mainly to verify that the class and level are consistent, that they are consistent with the age of the HHM, and that the id code of the respondent belongs to an adequate person. At the time this report is being written, I realize that an important consistency check between the highest grade completed and the current grade enrolled was left behind. Unfortunately, at the time this module was prepared, the number of years per level was unknown and the testing was left in the long "to do" list. However, today it has been added and hopefully, tomorrow, it will be in operation.

To avoid wild values entered in answers to questions 18 to 25, the summation of them was check to be less or equal to the total since, in many cases, people only declared the total and in others, particular values were not remembered. The missing value was entered as "99999" to clearly distinguish it from possible amounts.

In questions 30-32, the number of persons that made contributions to the educational expenses was unknown; thus, an "end of list" code (9) was added. If there was only one person making contributions, questions 31-32 were blank, a not acceptable code for those questions. Entering a "9", the program understood that there were no more contributors and a skip to question 32 was performed.

2. Problems found.

The only problem found in this module had to do with the consistency check between the years of education and the age of the HHM attending school. The starting age had been set to 5 but many cases were found to be right on the limit and, after analyzing the questionnaires, we decided to relax the consistency rule. If the years of education plus 4 were greater or equal to the age of the HHM, a warning message was issued. The same rule was applied considering the starting age at 3 rather than 4. If the inconsistency still persisted, the warning was changed to an error and the grade and age had to be revised. In a similar manner, the highest grade completed was also relaxed but considering one more year (5 for the warning and 4 for the error message).

Module 4: Labor.

1. General remarks about the DE program for the specific module.

This module is certainly the longest and more cumbersome (from the DE point of view) of the whole survey and the one that took the greatest effort to transform it into a simple module for the DE operators.

One of the basic premises of this DE program is that there was going to be one DE screen per page and that there would not be turning pages back and forth to fill in the screen. This module however has long skips that sometimes requires skipping several DE screens and others just end prematurely the whole module. Generally speaking, when the target field of any skip is out of the current screen, it has to be carried over to the next or to the appropriate screen in successive stages. The DE operator only sees that the line changes to the next person (without leaving the screen) but the program has to remember the skips line-by-line, screen-by-screen, until the target field is in focus.

1.1 Part A1: Labor force participation.

The universe for the first DE screen is "all HHM 10 years and older" and thus, the program creates automatically the framework to fill in. The column for the id code was locked and the corresponding id codes were copied from the HH roster.

Question 5 was checked as a bi-univocal relationship with questions 2-4 since it is used as the base to define the universe of part B1.

1.2 Part B1: Overview of work in the last 7 days.

Although the universe of part B1 is known by the program, the rigid framework cannot be created since each HHM that had worked in the last seven days can have more than one activity. However, every time a HHM id code is entered, the program checks that the id code belongs to a person that did work in the last seven days (had a 1 in question 5 of part A1). Once the whole roster is finished, the program checks that every HHM with a 1 in question 5 was entered in part B1. If a HHM with 1 in question 5 was omitted, the program issues a warning message that has to be acknowledged by the DE operator. The message includes the id code missing and the operators have instructed to pass the message to the supervisor. However, if the person is missing in the questionnaire, the conflict will have to be solved later. Thus, it is possible that this kind of problems will be present in the data file until the conflict is resolved at a different level.

Once part B1 has been completed, the program checks that question 10 (ranking of different activities of each HHM by hours worked in last 7 days) has been properly done (by the interviewer) as well as properly entered by the DE operator. To do this, the

program takes the different activities and line numbers and constructs its own ranking and compares it with the one that was entered. When a discrepancy is found, the program issues error messages that specify the id code of the HHM in conflict as well as the lines numbers of the different activities and points out the proper ranking. In this context, the program checks that there is one and only one activity ranked "1", and that there is no more than one activity ranked "2". However, if the HHM has more than one activity, there must be one and only one ranked "2".

1.3 Part C1: Main and secondary job in the last 7 days (questions 1-17).

Once again, the universe of this part is known by the program and as usual, it generates the framework based on those HHM that have a code "1" in question 6 of part A or, what it should be the same, based on the id codes of part B1 that have a code "1" in question 10.

Although the program knows the main occupation code declared in question 1 of part B1, the code entered in question 1 of this part is used to double check the original one.

Question 17 is also used to check the number of activities declared by the HHM in part B1. That is, Q17 = 1 $\Leftrightarrow$ N > 1 being N the number of activities declared in part B1.

1.4 Part C1: Main and secondary job in the last 7 days (questions 18-33).

This part applies only to those HHM that declared more than one activity in part B1. However, the universe specified in the questionnaire is for all HHM that have worked in the last 7 days. Thus, the program generates a framework identical to the previous one (1.3) and, for those HHM that declared only one activity, it will skip automatically to the next line. From the DE operator's point of view, this approach will work as if only those with secondary jobs had been selected since there will be no access to the id code; nevertheless, for those HH where there is no HHM with secondary occupation, the data file will show a roster with only the id code filled.

Again, the occupation code entered in question 18 is used to double-check the corresponding secondary occupation (ranked 2) of part B1.

1.5 Part A2: Labor force participation in the last 12 months.

This part applies only to those HHM that being 10 years or older, did not report work in the last 7 days (part A1). This universe is known by the program and accordingly, the rigid framework is generated automatically from the information entered in part A1.

Question 5 is used to double-check the answers to questions 2 to 4 according to the following relation: (Q2 = 1 or Q3 = 1 or Q4 = 1) ⇔ Q5 = 1.

1.6 Part B2: Overview of work in last 12 months.

This part applies to all HHM that either had worked in the last 12 months or in the last 7 days. Just like in part B1, the framework cannot be generated since each HHM might have more than one activity. Nevertheless, the list of HHM that should be included in this part is automatically built based on the information found in part A1 and A2. Thus, when the DE operator enters the HHM id code, it is checked again the list. If the id code is not found, an error message is issued and the operator has to check the relevant information. That is, question 5 of part A1 and question 5 of part A2 for the HHM with the id code just entered.

The checking mentioned above prevents the DE operator to enter a wrong id code but it does not prevent the omission of a HHM id code. This checking is performed once the DE operator has filled in the first screen related to part B2. At this point, the list created by the program is confronted with the id codes entered; if one or more omission is detected, an error message is issued specifying the id code(s) missing.

Question 7 is used to double-check the values entered in questions 5 and 6 and, question 9 is checked with question 7 in the same manner the ranking test was performed in part B1.

1.7 Part C2: Main and secondary job in the last 12 months.

Once again, the program knows the universe participating in this part and, accordingly, the DE operator framework is automatically generated, copying the id code of every HH member that had a 1 in question 9 of part B2.

This part (C2) has been divided into two subparts, one for the primary activity (questions 1 to 16) and one for those that had a secondary activity (questions 17 to 31). In both cases, the program extracts the universe automatically from part B2.

Since the program selects the right id codes from the information provided in B2, the presence of the right HH Members is guaranteed. At this point, the occupation code entered is checked with the one provided in B2 for both, the primary and secondary activities (if any).

1.8 Part D: Activities in 1991.

This part of the labor module is for all HHM older than 20 years of age and, again, the program knows which HHM should be in this roster. However, the rigid framework cannot be automatically generated since some of them might have no activity, one, or more than one activity.

The main consistency checks performed here are carried out once the roster has been entered in order to guarantee that no HHM that should be in this roster is missing. On the other hand, every time an id code is entered, the corresponding age is checked to make sure it belongs to a HHM older than 20 years of age.

In a similar manner to the checks performed in parts B1 and B2, question 5 is checked to certify that there is no more than one activity reported as most important nor more than one reported as next to most important.

1.9 Part E: Main Activity in 1991.

The universe selected in this part is known and defined by all the HHM that have a 1 in question 5 of part D. Thus, the rigid framework is generated to ascertain that this part will contain every HHM that should be in it. At the same time the id code of these HHM is copied from the preceding roster, the corresponding occupation codes –of the most important activity- is also copied.

2. Problems found.

Most of the problems found in this module came from the questionnaires that were either uncompleted, with skips to the wrong question or with erroneous information. Initially, the problems were so severe, that the interviewers had to be retrained.

There are many reasons to select “intelligent data entry” over “heads down data entry”. However, one of the most important arguments rests in the very early detection of problems in the fieldwork that, otherwise, can threaten the quality of the information gathered. The DE program becomes the fieldwork’s analyzer and, likewise, the DE operators become the DE program testers.

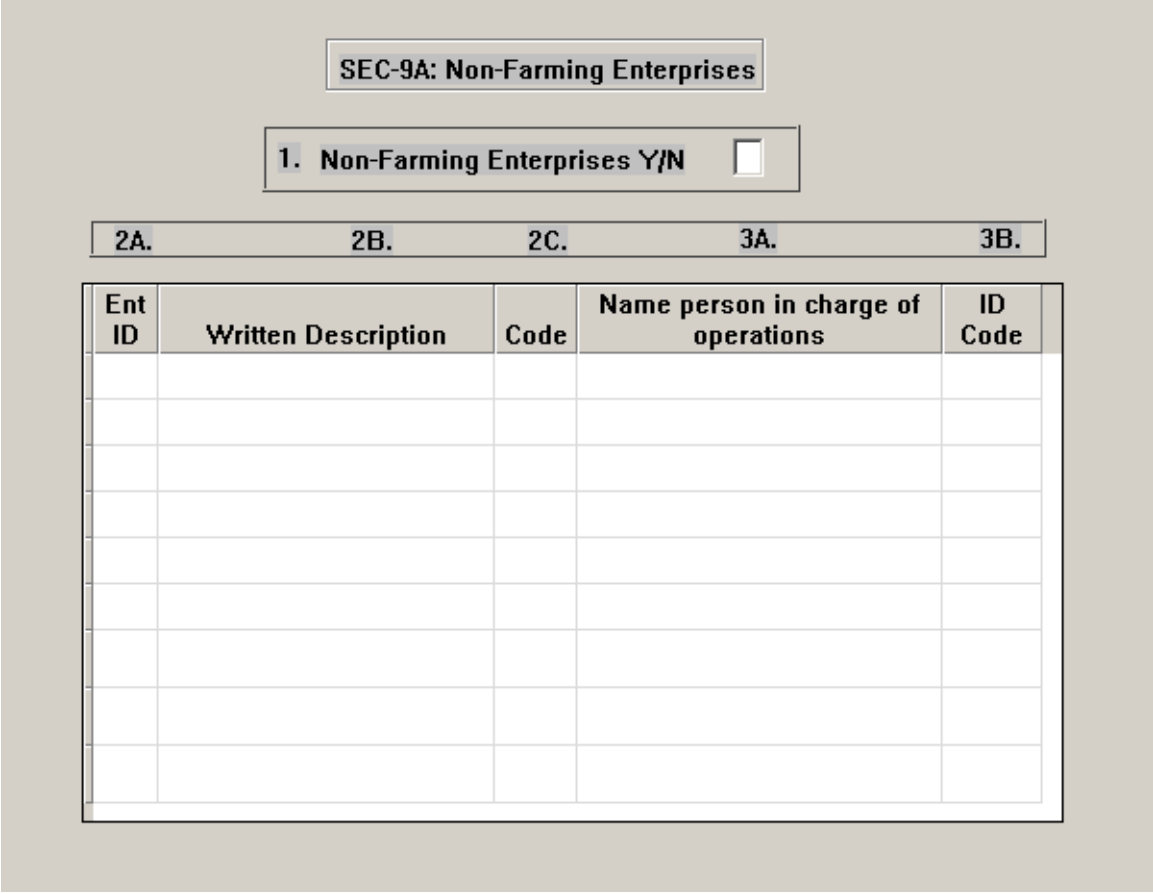
This module has a higher degree of difficulty for the interviewers mainly due to two reasons: one, they have to look to questions –on specific roster lines- that can be way behind the current question, to make decisions that might affect the questionnaire flow or to decide if a given HHM should or should not be part of a given roster; although this is not a difficult task to do, when it is performed many times throughout the interview of a large HH, with many HHM being interviewed, the chances of making mistakes increase accordingly. The second reason is the combination of two different subjects, the duration of the interview and the numerous skips that this module –given its nature- requires. This combination together with the first reason mentioned, makes it easy to do mistakes that end up with questionnaires with missing information.

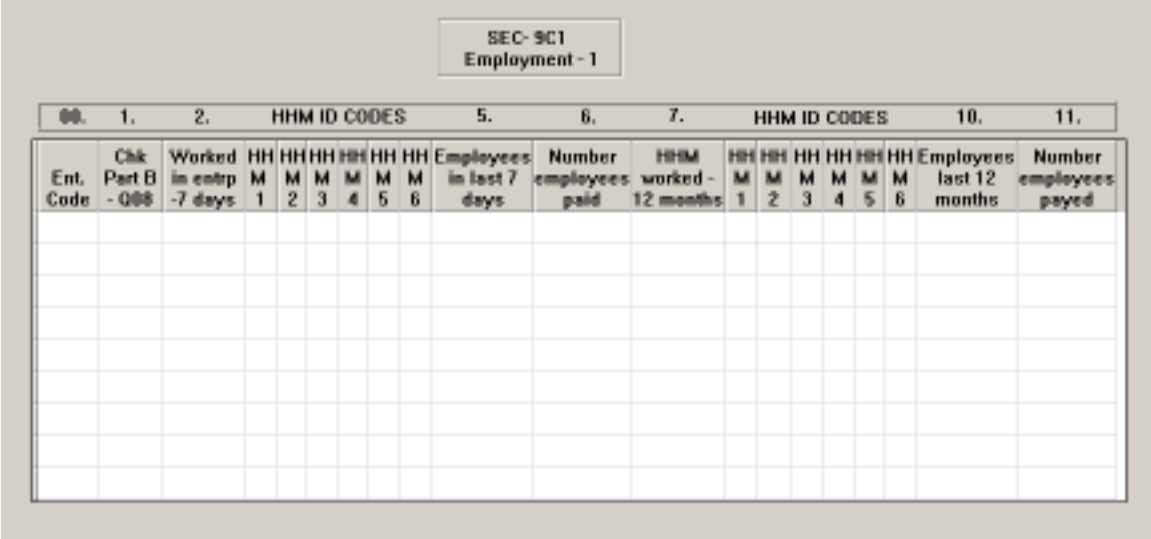
Module 5: Non-Farm Enterprises.

1. General remarks about the DE program for the specific module.

This module is the only one in the whole survey where some of the DE screens are different than the corresponding ones in the questionnaire, the reason being some temporary limitations in the CSPro software.

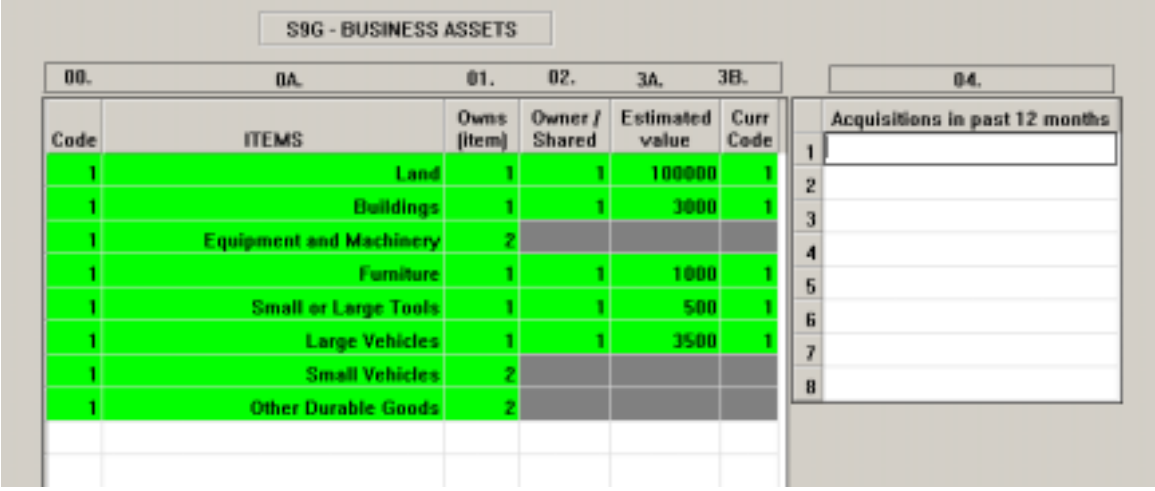
The following images show the screens as they are presented in the DE program:





SEC- 9C1
Employment - 1

00.	1.	2.	HHM ID CODES						5.	6.	7.	HHM ID CODES						10.	11.
Ent. Code	Chk Part B - 008	Worked in entrp -7 days	HH M 1	HH M 2	HH M 3	HH M 4	HH M 5	HH M 6	Employees in last 7 days	Number employees paid	HHM worked - 12 months	HH M 1	HH M 2	HH M 3	HH M 4	HH M 5	HH M 6	Employees last 12 months	Number employees payed



S9G - BUSINESS ASSETS

00.	0A.	01.	02.	3A.	3B.	04.
Code	ITEMS	Owns (item)	Owner / Shared	Estimated value	Curr Code	Acquisitions in past 12 months
1	Land	1	1	100000	1	1
1	Buildings	1	1	3000	1	2
1	Equipment and Machinery	2				3
1	Furniture	1	1	1000	1	4
1	Small or Large Tools	1	1	500	1	5
1	Large Vehicles	1	1	3500	1	6
1	Small Vehicles	2				7
1	Other Durable Goods	2				8

In the last screen, the code 1 refers to the Enterprise id code and, according to the number of enterprises, an identical block of 8 items is preset as a rigid framework.

The DE program in any other way is very similar to the labor module and the same concepts are used.

Module 6: Networks.

1. General remarks about the DE program for the specific module.

This module is straightforward and there are no remarks about it.

Module 7: Agriculture.

1. General remarks about the DE program for the specific module.

This is probably the longest module of the whole survey and, at the same time, with a higher degree of difficulty given the existent relationships between different parts in which it is divided.

1.1 Parts A1 and A2: Operated Land.

These parts are straightforward with the system providing the plot number or plot code and the operator using the code 99 to signal the end of the plot list.

1.2 Part B1: Crops and Production.

The most distinctive characteristic of this part is the list of 47 different crops that can be potentially planted. Instead of following the approach used in the questionnaire, the DE program was designed to accept only those crops that were really planted rather than generating a rigid framework where there is a fix line per crop, independently if a specific crop was planted or not. The advantage of this approach is that is faster for the DE operator than the alternative option. As a visual help and confirmation, immediately after the crop code is entered, the crop name is display on the screen. Consequently, the yes/no answer is omitted since only *Yes* answers are registered.

As the crop codes are entered, the following consistency checks are performed: (i) the code has not been previously entered; (ii) if it is a secondary crop and the main crop has a code lower than the current crop code, it should have been previously entered as a main crop (codes 1 or 2 in question 3).

Once all the crop codes have been entered, the following tests are performed: (i) all main crops of all secondary crops should have been entered; (ii) the summation of all the surfaces of the different crops planted –converted to a common unit- should not be higher than the total surface of all the plots entered in part A1.

1.3 Part C: Inputs.

There are seven different types of inputs for all crops that, from the DE program point of view, are very similar. In fact, the editing checks performed are also the same. The only important difference among them is that some of them can have none, one or more than one line per crop, and others can have none or maximum one line per crop.

The difference is important since, for those that have only one line per crop, a rigid framework is created with one line per crop entered in part B1. For inputs that can have more than one line per crop, the DE program can't generate it and has to accept the crop codes from the operator.

The inputs that can have only one line per crop are seeds and seedlings; the one that can have more than one line per crop are the remaining inputs: fertilizers, manure, pesticides, labor and machinery that, all of them can have more than one type of the same class of input (i.e. fertilizer: compound and super phosphate).

When the rigid framework can't be generated, crop codes are checked –as they are entered- with the list of cultivated crops to ascertain that the specific input was used in one of these cultivated crops. At this time, the input code entered is checked to make sure that it is not duplicated (the same line is not entered twice). Once the individual input roster is finished (with a 99 as crop code), the program checks the list entered in part B1 and the list of the specific input to make sure that there are no crops missing.

1.4 Part D: Disposition of Crops.

Given its nature, this part is the most error prone section of the agricultural module. The disposition of crops contemplates eight different distribution categories for the volume harvested, each one of them potentially using a different unit code. Furthermore, some of the units (cartload and other) inhibit the arithmetic check since there is no equivalence between them and a weight or volume unit (i.e. cartload to Kg).

The universe included in this roster is each one of the crops for which a harvest volume greater than zero was reported in part B1. Thus, the program automatically generates the rigid framework, guaranteeing that there are no missing or duplicated crops in this roster.

Not withstanding the potential use of inexact volume/weight unit codes, the following arithmetic checks are performed for each harvested crop: (i) the individual disposition category is converted to Kg –whenever possible- and then compared to the total harvested volume (checking for individual wild amounts); (ii) whenever the unit is

not exact, the amount is assumed to be zero; (iii) the summation of the partial categories has to be less or equal to the total harvested crop volume; (iv) the summation of all disposition categories has to be equal to the total harvested (if not inexact unit was used). Given the inexact calculations, whenever a conflict is unveiled, a warning message is issued displaying the harvested volume and the individual category volume or the summation total. Therefore, it is likely to find discrepancies between the total harvested and the summation total of all disposition categories.

1.5 Part E: Farm Capital Inventory.

This is a list of a maximum of 17 items where only those having a *Yes* answer in question 1 are entered. This decision was made thinking that the average of HH would have a low number of this items (less than 7) and thus, the DE process could be speeded up with this approach. As usual, a code 99 was entered as item code when the list was completed. When the item code is entered, the item description is display as a way of visual verification.

1.6 Part F: Livestock.

This is a list of 16 different livestock that has been treated following the same approach as specified in part E.

1.7 Part F2: Animal Feed.

This is a short list of six different animal feed that, given its size, is generated as a rigid framework to fill in.

Module 8: Consumption.

1. General remarks about the DE program for the specific module.

This module is characterized by a long list of items for which, common questions are asked. The filter question is whether the item was consumed or purchased during the last 12 months. Once again, instead of generating a large rigid framework that would slow down the DE process, the code of all items having a *Yes* answer is entered; again, as a visual verification, the name or item description is displayed as soon as the item code is entered.

Since all the different parts of this module follow the same pattern, we will not give specific explanations about them.

The amount of online editing performed is limited -given the nature and relationships of the information gathered- to duplication of the item code and range checks.

Module 9: Health.

1. General remarks about the DE program for the specific module.

This module is divided into two parts: the Health Care use and the Self Reported Health Status. The first part is a roster identical to the HH roster in which, all HHM are included. The second part is only for HHM older than fourteen years of age. In both cases, the number of lines per roster is known and thus, the program automatically generates the rigid framework with the proper id codes for each line.

As in the previous module, other than the range checks and automatic skips, there is almost no online editing.

Other Modules.

There are two other minor modules, the Miscellaneous Income module and the Dwelling module. Both are small and straightforward and do not require special explanations.

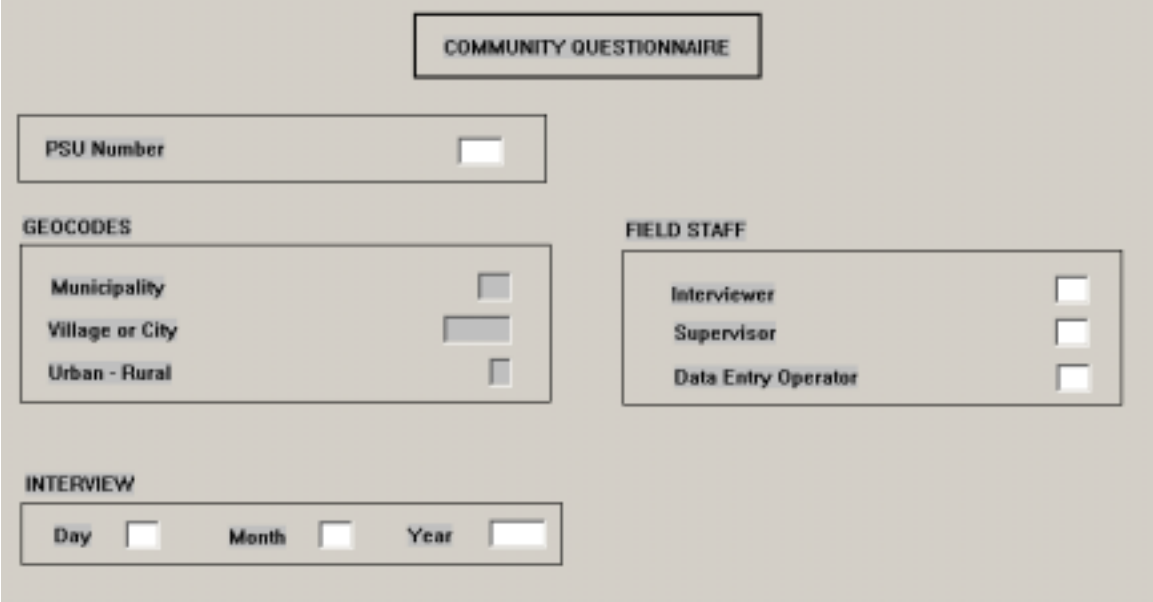
The Community Questionnaire.

The Community survey DE program is a completely independent program that follows exactly the same pattern as the HH survey DE program.

The questionnaire –and also the DE program- has been divided into several small modules as follows:

1. Identification.

The following image shows the information gathered in this section of the Community DE program (CDE).



The most important piece of information captured in this section is the PSU number that is linked to the Municipality, Village and Urban/Rural area characteristic. Just like in the HH survey, this module uses the same PSU external file to check that the PSU code is a valid one. Thus, the same observations made for the HH id are valid here.

The dark fields indicate that they are locked fields and used by the DE program with no DE operator access to them. These fields are directly copied from the control file once the PSU number has been validated.

2. Other Sections of the Community Questionnaire.

The remaining sections of this questionnaire follows the same pattern and behavior as the HH DE program and thus, they will not be analyzed with the same detail.

The following are the remaining sections of the Community survey addressed by the DE program:

- Respondent Characteristics
- Basic Characteristics of the community
- Institutions and Infrastructure
- Economy
- Agriculture
- Community Organization and Displacement
- Health Facilities
- School Facilities

Other Tasks in SERPRO's TORs.

1. Training Country Staff.

Training of the DE operators and the DE supervisor in the use of the program is a very important task since during the experts mission period, they have to get acquainted with all the peculiarities of the DE program and gain as much experience as possible in solving the numerous conflicts they will have to face during the duration of the DE stage.

Normally, a one-week period for the DE training is considered to be sufficient; however, considering the length and the complexity of the survey, eight working days would have been desirable. However, given the time constrains, only a three-day training period was provided. The lack of training was compensated with "on the job training" that, apparently, did work fairly well.

Selection of the DE operators was somewhat difficult for reasons that are well known of the WB survey's director. The main problem was that, with the operators available at the Statistics Office (KSO), the success of the DE process would have been jeopardized. Knowing that an interactive DE process is completely different from a "heads down" DE, we tried to identify university students that are more likely to adapt to a more analytic type of work than what the KSO operators were used to do. Our effort was compensated and seven or eight students plus all the KSO DE (10) operators attended the training sessions. At the end, the eight survey DE operators would be selected by a final test consisting in one real interview, the same one for all the potential DE operators.

During the training, it was apparent that the KSO operators were not going to be suitable for the survey purposes. Furthermore, they decided not to take the test and instead, to exercise pressure through the KSO director to be accepted directly without any screening process. Pierella Paci, the WB survey director, solved the political problem generated and we ended up having six DE operators that fully satisfied our expectations.

2. Testing of the DE program.

Given the extent of the questionnaire and, consequently, of the DE program, a large number of HH questionnaires would have been desirable to test the program. Once again, the time constrains and other circumstances didn't allow us to give the program a fair test. In fact, when we left Kosovo, only 4 complete interviews had been passed through the DE program. This is by no means an acceptable number of questionnaires for a test, especially when the extent of the survey is considered.

Taking in consideration that the Internet communication was adequate, we thought that any error in the DE program could be promptly fixed and submitted to the DE supervisor. The Kosovo survey was fortunate to have an excellent DE supervisor as

Besfort Azizi ended up being. His high degree of responsibility and technical capabilities are indeed a great asset for the survey, allowing a fluent communication of the errors detected, and a rapid implementation of the corrections and adjustments to the program that were carried out. The number of errors in the DE program –detected after our departure from Kosovo- was not more than 10, but at least one of them was very important since a whole page of the labor module was being skipped systematically.

In the future, I think that enough time should be allocated to the testing of the DE program since we might not be so lucky as to have good Internet connection or a DE supervisor as diligent and efficient as Besfort Azizi.

3. Installation of a WINDOWS Network.

This should be a fundamental task in the data processing TOR since it greatly facilitates the day-to-day activities of the DE supervisor. These are: (i) software dissemination and updates through the different PCs; (ii) sharing different devices such as printers and other expensive peripherals; (iii) backup and concatenation of the data files.

The importance of the Network was obvious during the training period when the DE program was first submitted to a testing and debugging process. At that time, the software had to be updated several times during the day and, not having the net installed, it was a time consuming effort. Having this experience in mind, SERPRO together with the KSO proceeded to install a Windows Network for eight PCs.

We strongly suggest that in the future, Network should be budgeted together with the hardware cost.

4. Data Entry Follow-up.

As a principle, SERPRO's services don't finish with the software development and the link between the SO and us is maintained until we feel the software has been sufficiently tested and a high degree of confidence in its performance gained. In this case, knowing that the program test was insufficient, that follow up has been more intense. Especially in the first three weeks following our departure, a deep and thorough data analysis has been performed. Thanks to this task and the coordination between SERPRO and the local survey executives, several errors have been unveiled.

5. Data Quality Control Software.

Two pieces of software are being submitted to the survey supervisors for checking the quality of the data entered. The first one is a batch-editing program that will perform the same online editing carried out by the DE program. The purpose of this program is to identify the cases where the DE operators could not fix conflicts in the information entered and the error messages issued were ignored.

The second software aims to the production and observation of the frequency distribution of selected –or all- variables being investigated.

Conclusions.

This has been a long and complex survey, characterized by a very short time for the design and implementation of the DE program, and the selection and preparation of the DE operators.

The DE program is by large the biggest one SERPRO has been involved in more than 15 years of experience in all types of surveys. The program contemplates the capture of 862 different variables, 72 DE screens of which 48 were rosters' having a maximum number of occurrences varying between 15 and 75 with an average time to enter a full questionnaire of about 45 minutes. In addition, the Community Survey has 220 variables and 20 DE screens.

SERPRO would like to emphasize the great collaboration and support of the local professional staff in Kosovo, in particular of Valerie Evans, Survey Director in Kosovo, Edis Agani and Besfort Azizi.