BLAISE in Macedonia-Census of agriculture 2007[1]

Liljana Taseva, Mira Deleva, State Statistical Office of the Republic of Macedonia

This paper is dedicated to the expert from Slovenian Statistical Office, M.Sc Pavle Kozjek, to whom we are deeply grateful for the mutual collaboration.

1.Introduction

Census of agriculture was one of the main actions conducted in State Statistical Office of Republic of Macedonia (SSO) in 2007. It was a process of collecting, processing, assessing, analyzing and disseminating of data related to agricultural situation in Republic of Macedonia. In the period after the Second World War, on the territory of Republic of Macedonia the Census of Agriculture was conducted twice, the first time in 1690 and the second in 1969. Data collected through the census are of great importance for creating strategies for the future national and local development.Census of agriculture in 2007, on the territory of Republic of Macedonia was conducted in the period from 01-15.June.2007.

SSO prepared following census instruments:

- Survey form for individual farmers, as well as form for farms registered as business subjects;
- Control form for registering the census units;
- Form for post census survey;
- Methodological guides.

Data collection was conducted in traditional way, by interviewers on the field and using paper forms.

According to planed activities and deadlines, applications for data processing (data entry, control, aggregation, derivation of some variables, tabulation) were created. Considering long experience in using Blaise at SSO, it was decided to use it for entering, checking and editing of data collected via Census of agriculture in 2007.

Presently, the main tool used for data entry at SSO is Blaise .The first application has been developed in 1999 when this software was used only for opinion poll based surveys, but latter SSO has started using Blaise for the rest of surveys as well, including Census of agriculture. Experience from using the Blaise increased with its continuous use, especially for creating of data entry forms as well as for establishing of some uniformed procedures.

2. Brief overview of data entering and processing

The Census of agriculture data entry began one month after the end of the fieldwork, i.e. after publishing the first results.

Before data processing took place in SSO, some preparatory work was to be done, such as: enhancing the IT environment; developing an applicative software for data processing; providing working space for data processing; choosing staff for visual control, coding, entering and correction of the paper material; choosing staff for surveillance of the working process; providing staff training; preparing manuals for data processing and applicative software.

According to the plan of activities and deadlines, SSO was equipped with server and 46 working stations dedicated for this activity. Workstations were placed in the premises provided for data processing. Local area network in Windows environment and software tool Blaise 4, were used for completing the process of entering

---

[1] Views expressed in the paper do not necessarily reflect the views of the State Statistical Office

data of Census of Agriculture. Records affected during the process of data entry counted 197500 farms, and for each and every one of them were entered approximately 980 variables.

Applications for data entry, data control and data editing, developed in Blaise, were placed in special server folder **(POPZEM07).** Access to data and applications on server was controlled via user interface, developed in VB, by typing account name and password.

Several applications were developed in Blaise 4 :

- entrance of identification data for the carrier of the farm;
- entrance of data for the farm related with online control;
- Blaise data models for meta data files, such as: municipalities, settlements, census districts, streets, classification of activities (NACE REV. 1);
- manipula for transfer of meta data files from .txt to . bdb format;
- manipula for batch control;
- manipula for coping of the program for data entry along with data, from one model to another;
- manipula for coping Blaise file to ASCII files;
- manipula for data editing in DB2 database;
- manipila for daily reporting for data entry on the data entry person level and summary report for data entry on the data entry person level for given period of time as well as total summary reports.

Figure 1: user interface, developed in VB6, used for accessing applications in Blaise for data entry and control.
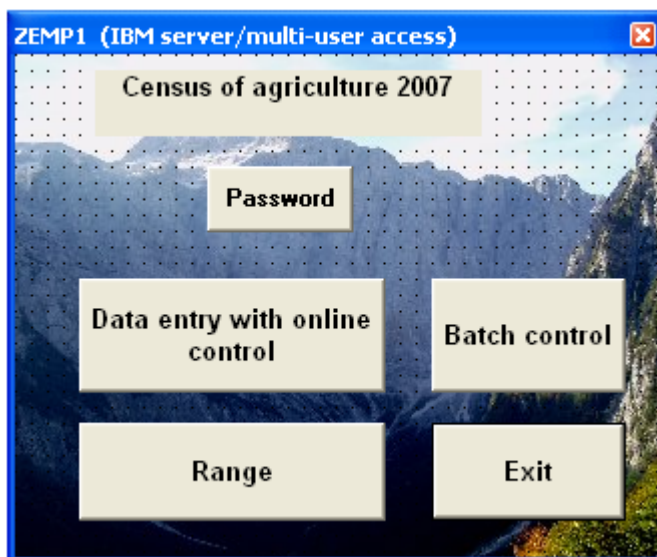
2.1 Application for data entry

User interface for data entry was designed according to the questionnaire design. It was accomplished by standardized code of blocks and tables, fields were textual and numerical.

Every farm had its own identification, recognized as primary key, which was a complex key composed of municipality, census district and ordinal number of the farm. There was an online control of the existence of such municipality and census district regarding meta data files. In the case of incorrect choice of municipality or census district, an error dialog appeared.

Application design was made to follow the flow of the questions. Data were entered in the questionnaire order, from question to question, from table to table. In the table where data of household members were placed, all jumps were programmed, so data entry person had to follow only filled fields of the questionnaire. Positioning on the proper field was automatically guided.

The same application was used for data entry and for data correction. Hence, in the application were implemented all logical-mathematical controls. Controls were very complex, so part of them were implemented

during the data entry process, and the other part after the data was entered. Part of the logical-mathematical controls were implemented as on-line control, so the entering of illogical data was reduced. In the case where illogical data were entered, error dialog appeared with error specification. Automatic positioning on the error fields was done and data entry person had to correct the value of positioned fields.

2.1.1.Example of on-line control in the table for household members

RULES

```
IF lice[I].cl_5k=child  THEN
        (lice[i].cl_4k < lice[1].cl_4k) AND (((lice[1].cl_4k - lice[i].cl_4k) < 50)  and
                                ((lice[1].cl_4k - lice[i].cl_4k) > 14))
        "child  ^lice[I].cl_4k difference with carrier
has to be less than 50 years and greater than 15 years   ^lice[1].cl_4k          Error 106_tab40
ENDIF
```

2.1.2 Example of on-line control on table 4 – land in household disposal

```
AUXFIELDS
c4_1:  0..99999999
c4_2:  0..99999999
.

FIELDS
greska : ARRAY [1..100] of STRING[1]
.
RULES

c4_1:=((tabela4.P4[1].k4_h*10000+ tabela4.P4[1].k4_m) +
      (tabela4.P4[2].k4_h*10000+ tabela4.P4[2].k4_m) -
      (tabela4.P4[3].k4_h*10000+ tabela4.P4[3].k4_m))
  c4_2:= tabela4.P4[4].k4_h*10000+ tabela4.P4[4].k4_m

IF (c4_1)=(c4_2) THEN
          greska[1]:='1'
ELSE
          error involving (tabela4.P4[4].k4_m) "Error 1 – sum error in table4"
ENDIF
```

In this case in the field for certain error, the correct combination is signed as in the case above with greska[1]:='1' and for the incorrect combination appears error dialog for the combination that has to be corrected at the moment of data entrance (on-line).

The rest of the controls were executed during the data entrance by marking incorrect and correct combinations in the following way:

```
RULES
.
.
IF tabela16.P16[19].k16_h*10000 + tabela16.P16[19].k16_m =
    Tabela17.P17[14].k17_1h*10000 + Tabela17.P17[14].k17_1m THEN
          greska[11]:='1'
ELSE
          greska[11]:='0'
ENDIF
```

Once data entry for each municipality was accomplished, an application (manipula) for batch control was executed, in order to create a list of erroneous records. Each erroneous record consisted of identification data (primary key) and type of the error. Since data entering created one .bdb file, in application (manipula) for batch-control existed option for its executing for one or more chosen municipalities. Data control and correction were executed until database was cleaned of errors.

That is described in the code below:

```
Settings
  DESCRIPTION = 'List of errors - zemp1'

 USES
          InputMeta 'zemp1'

DATAMODEL  outf
          FIELDS
                    oneline:STRING[100]
ENDMODEL

INPUTFILE InFile: InputMeta ('zemp1',BLAISE)
OUTPUTFILE OutFile: outf ('zemp1_greski.txt', ASCII)
SETTINGS
MAKENEWFILE=yes

 AUXFIELDS(global)
  i:INTEGER
  k:INTEGER

MANIPULATE
IF recordnumber (infile)=1 THEN
oneline:='----------------------------------------------------------------------------'  outfile.WRITE

oneline:='ZEMP-1  Pogresni podatoci  na den:'+ datetostr(sysdate)+'      vo:'+timetostr(systime)
outfile. WRITE
oneline:='                                                        '
outfile. WRITE
oneline:='opstina popisen krug stopantvo  greska '                                 outfile. WRITE
          oneline:='                                                 '
          outfile. WRITE
ENDIF

IF (ops = '0035')  or (ops = '0116') or (ops = '1414') or (ops = '1228')
    or (ops = '1643')  or (ops = '1872')  THEN

          FOR i:= 1 to 12 DO
                    IF greska[i]<'1' THEN  k:=i
oneline:= ops + ' ' + krug + ' ' + stop + '  Greska ' + FORMAT(str(k),2,right)
                    OutFile.WRITE
                    ENDIF
       ENDDO

       FOR i:= 14 to 17 DO
              IF greska[i]<'1' then k:=i
 oneline:= ops + ' ' + krug + ' ' + stop + '  Greska ' +       FORMAT(str(k),2,right)
                    OutFile.WRITE
                ENDIF
       ENDO
ENDIF
```

Example of created report from batch control

municipality	settlement	Census district	farm	Error
1040	BEROVO	059	013	Error 2
1040	BEROVO	059	034	Error 88
1040	BEROVO	059	034	Error 89
1040	BEROVO	060	004	Error 27
1066	BITOLA	060	012	Error 66

1066	BITOLA	075	002	Error 12
1066	BITOLA	068	004	Error 6
1678	KUMANOVO	261	021	Error 5
1678	KUMANOVO	261	021	Error 45

Identification data regarding the farm holder such as: name, surname, personal ID, municipality, settlement and address (street and house number), phone, were placed in separate file apart from the rest of data regarding the farm, hence the necessity for checking the range of the first file against the range of the second file and vise versa. This control was executed continuously during the data entry.

3.Following carried out procedures

Backup of the .bdb file of data-track, was created at the end of each working day. Also, the contents of the folder POPZEM07, where applications and data resided,

were copied into another server folder named ARHIVA, on the daily basis. It was done in order to provide data entry to flow continuously and simultaneously with the rest of data processing procedures, such as batch-control of data, creating necessary reports for the process of data entry etc. This was of the great importance for the batch control because in the folder named ARHIVA resided data of data entry from the previous day, enabling the possibility for data entry and batch control to run simultaneously and control of data to end up at the same time with data entry.

After data entry and control were accomplished, followed the process of migrating data to DB2 database. Because of the lack of the modules for direct communication between Blaise files and DB2, manipula for transferring data from Blaise file to ASCII file was developed. With this manipula data entry file is transformed into several .txt files, mutually connected by primary key, where data were separated by delimiters (>).

Basic structure of this application is given in the example below:

```
SETTINGS
DESCRIPTION = 'BLAISE to ASCII'

USES
InputMeta 'zemp1'

DATAMODEL zemp1_d1
        FIELDS
                OPS  "Opstina": STRING[4]
                KRUG "Popisen krug": STRING[3]
                STOP "Stopanstvo": STRING[3]
                TABELAID :BID
                tabela4 :tab4
            .

                Tabela18: tab18
.
.
ENDMODEL

DATAMODEL zemp1_d2
        FIELDS
.               OPS  "Opstina": STRING[4]
                KRUG "Popisen krug": STRING[3]
                STOP "Stopanstvo": STRING[3]
                TABELAID :BID
                Tabela19: tab19
                .
                .
                Tabela21: B21
```

```
.
ENDMODEL

DATAMODEL zemp1_d3
        FIELDS
        OPS  "Opstina": STRING[4]
                KRUG "Popisen krug": STRING[3]
                STOP "Stopanstvo": STRING[3]
                TABELAID :BID
                Tabela22: tab22
                .
                .
                Tabela32: B32


.
ENDMODEL

DATAMODEL zemp1_d4
        FIELDS
        OPS  "Opstina": STRING[4]
                KRUG "Popisen krug": STRING[3]
                STOP "Stopanstvo": STRING[3]
                TABELAID :BID
                Tabela33: tab33
                .
                .
                Tabela46: tab46


.
ENDMODEL

INPUTFILE InputFile1:InputMeta ('zemp1', BLAISE)

OUTPUTFILE OutputFile1: zemp1_d1 ('zemp1_file1.txt', ASCII)
SETTINGS
SEPARATOR = '>'
OUTPUTFILE OutputFile2: zemp1_d2 ('zemp1_file2.txt', ASCII)
SETTINGS
SEPARATOR = '>'
OUTPUTFILE OutputFile3: zemp1_d3 ('zemp1_file3.txt', ASCII)
SETTINGS
SEPARATOR = '>'
OUTPUTFILE OutputFile4: zemp1_d4 ('zemp1_file4.txt', ASCII)

MANIPULATE
  OutputFile1.WRITE
  OutputFile2.WRITE
  OutputFile3.WRITE
  OutputFile4.WRITE
```

Scripts for creating tables in DB2 database and import of .txt files into DB2, were made by using CAMELEON. From DB2, data were exported into SAS, which was used as a tabulation tool. During the preparation of data for tabulation, additional variables were derived in order to easy the process of tabulation in technical terms and to improve efficiency in finalizing output results. Applications for aggregation of data and tabulation of the output results were developed. Afterwards results were analyzed and published.

For the first time it was achieved that conducting, processing and publishing of the Census data was be accomplished within the same calendar year.

WEB dissemination of Census of agriculture 2007 data was realized through software tool PC AXIS in which was offered a greater part of published data.

4. Post Census survey – data entry

In order to check and confirm the range and quality of data collected from Census of agriculture in 2007, statistical survey based on sample of 50 census districts over the territory of Republic of Macedonia was conducted independently, right after ending the field work from 16-20/06/2007 and included 2% of total farms. The survey was conducted in traditional way by repeated visits and collection of data. In Blaise was developed application, with all necessary on-line data controls implemented. Copy of application was created with different name of .bdb file, in order to have the same application twice, so the double data entrance could be performed. Manipula was developed for comparing data from the two files. After two files were matched, data from post Census survey were processed in SAS.

5.Conclusion

Conducting of the Census of agriculture in 2007 was successfully accomplished task and for the first time it was achieved that conducting, processing and publishing of data of Census to be carried out within the same calendar year. Great contribution to this task gave Blaise 4.7, as a software tool in which were developed applicative solutions that provided overcoming of some deficiencies in the previous versions of Blaise (for instance hospitalizing of data etc)

Because of the successful use of this software, SSO is tending to extend the use of Blaise from opinion poll based surveys to the rest of statistical surveys. Experience of using Blaise for Census of agriculture confirms that this software can be used as a tool for data processing for other mass actions.