

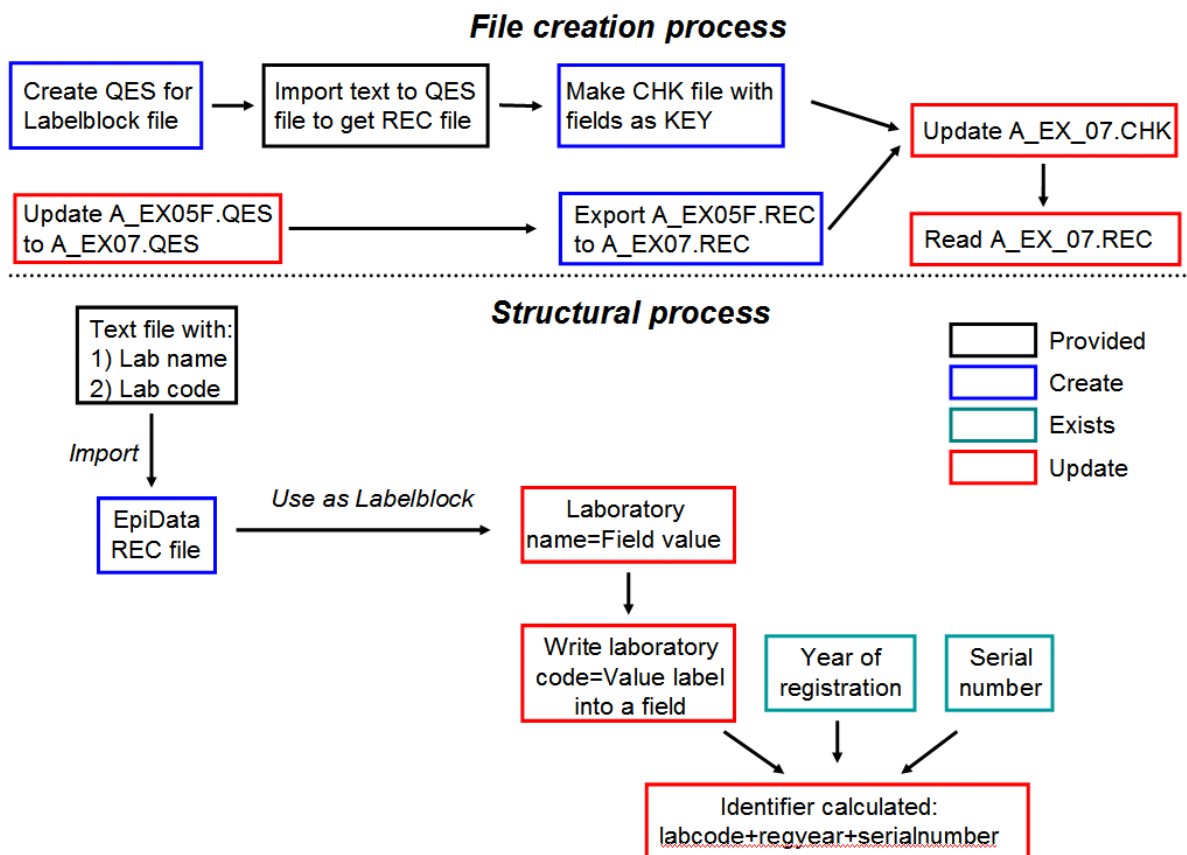
## Exercise 7: Using an external file for Labelblocks

At the end of this exercise you should be able to:

- a. Import a text file into EpiData
- b. Use an external file for Labelblocks
- c. Edit and make calculations in the check file
- d. Create a unique identifier from more than one variable

The WHO/Union-recommended Tuberculosis Laboratory Register uses a unique sequential serial number beginning with 1 each year for each examinee. If one were to record data from more than one year, the serial number would no more be unique but a unique identifier could be constructed using a combination of serial number and year of examination. If more than one laboratory is included, then additionally a laboratory code would be needed to unambiguously identify each examinee from several years in several laboratories.

You will accomplish this in this exercise. It will require you to work both in parallel and in sequence. The following graph summarizes the process.



Because the procedure is a bit complex we will explain the sequence of the steps required to arrive at the solution.

**Step 1: Make a new QES file and import the supplementary text file to obtain a REC file that will be used as external Labelblock**

You need to download the supplementary text file A\_EX07\_NAMECODE.TXT from the course folder. You must look at this file in your text editor (e.g. with NotePad™ which comes with the Windows operating system. Note: NotePad™ is a rather basic text editor. We recommend using an alternative such as Crimson Editor®, a free and extraordinarily powerful text editor which you may obtain free from <http://www.crimsoneditor.com>). Shown here are just the first 18 lines among the total of 95:

```
1 Awuna;ML_J
2 Beitbridge;MS_D
3 Bindura;MC_A
4 Binga;MW_G
5 Birchenough;ML_M
6 Bonda;ML_I
7 Brunapeg;MS_G
8 Chegutu;MW_J
9 Chikombedzi;MV_I
10 Chimhanda;MC_H
11 Chinhoyi;MW_A
12 Chipinge;ML_B
13 Chiredzi;MV_H
14 Chitamoyo;MW_G
15 Chitsungo;MC_K
16 Chivi;MW_N
17 Chivu;ME_B
18 Collin Saunders;MV_J
```

Note the following here:

- o This is a semicolon-delimited file
- o It has two fields per row, separated by the delimiter
- o The first field is the Name of the laboratory, the second the Laboratory code
- o The Name of the laboratory has various lengths. We can assure you that none has a name exceeding a length of 20
- o The Laboratory code has a field length of 4

This file was created during an operations research project in Zimbabwe (kindly provided by Biggie Mabaera, University of Zimbabwe). At the time of the study, 95 laboratories performed sputum smear microscopy and utilized the standard Tuberculosis Laboratory Register. Each of these laboratories was assigned a unique code of length 4.

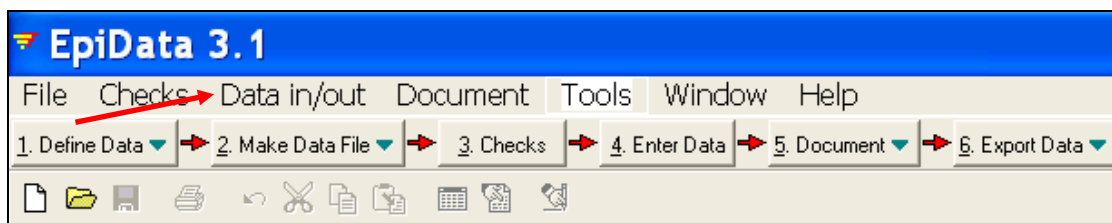
First, a QES file must be prepared into which the two fields will fit. This QES file will be given the name A\_EX07\_NAMECODE.QES and must have two text fields (*not* Upper case text as it has to accommodate whatever was chosen in the A\_EX07\_NAMECODE.TXT file). It is a simple file to make:

Questionnaire to accommodate data from the Text file

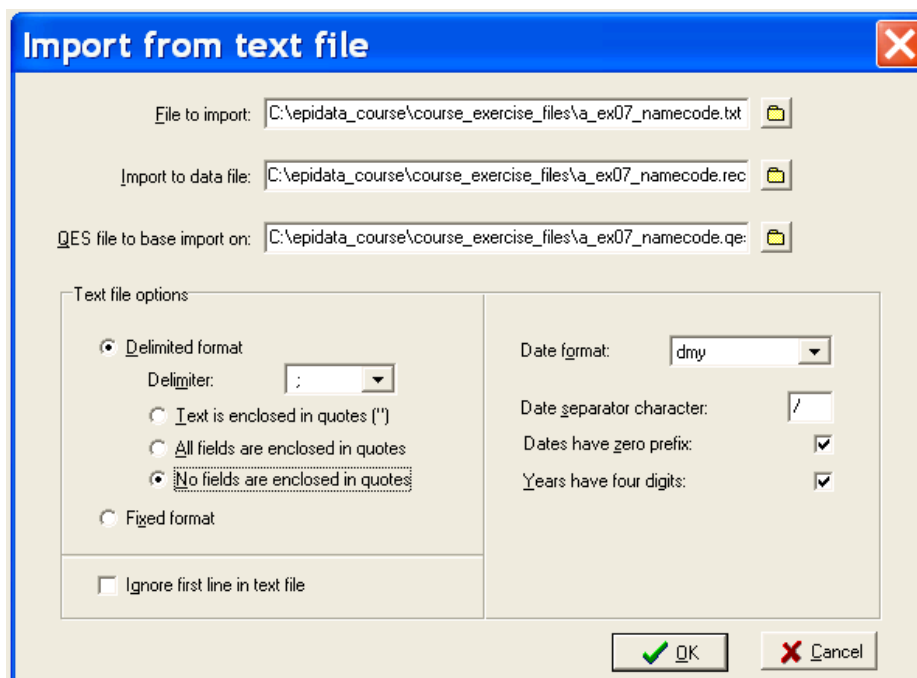
name Name of the laboratory \_\_\_\_\_  
code Code of the laboratory \_\_\_\_\_

**Note:** The sequence of the Fields is important. The first field should be the field that will serve as the Field value, the second the one that will serve as the Value label.

No REC file is made as this is created when importing the data from the EpiData Menu Data in/out menu, that is do not go to “2. Make Data File”, but to “Data in/out”:



This allows importing a text file into a pre-existing QES file, resulting in a REC file. Read the menu and options carefully to choose the correct options:

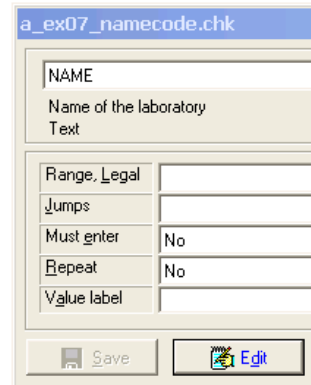


After completing the import, you have to make a CHK file. Use Edit to make each of the fields a KEY field, numbering the keys:

Questionnaire to accommodate data from the Text file

name Name of the laboratory

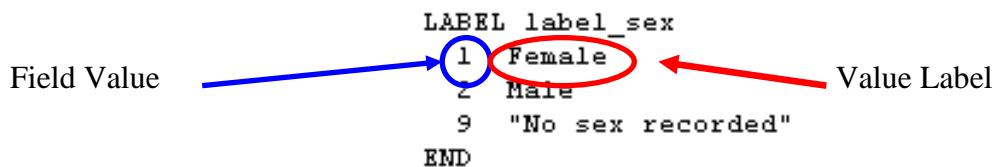
code Code of the laboratory



The field CODE will then be KEY 2.

### ***Role of the KEYS***

The command KEY in a field instructs EpiData to create an index for this field (see earlier exercise). This file will be used as a Labelblock, and you remember that a Labelblock has two components the Field Value and the Field Label:



Slightly different, but in analogy, we will now use the field NAME in the A\_EX07\_NAMECODE.REC file as the Field Value for a field LABNAME in the file A\_EX07.REC (see below) and the field CODE as the Value Label for a field LABCODE in that A\_EX07.REC file.

When you are done with the CHK file, you have completed this component of the exercise and you have a “A\_EX07\_NAMECODE” QES-REC-CHK triplet.

### **Step 2: Making the main QES file A\_EX07.QES**

To create a unique identifier that consists of three components, use the A\_EX05A.QES questionnaire as the starting point, save it as A\_EX07.QES and make the necessary amendments.

The unique identifier will have the field name IDCODE and will consist of three components, the code that unambiguously identifies the laboratory (Field length 4), the year of registration (Field length 4), and the laboratory serial number (Field length 4), the components separated by a hyphen (for better visualization of the components only), e.g.:

AA\_J-2006-1234

While the calculation for the field IDCODE can obviously be made only after the information on the three fields making it up has been entered, the actual position of the IDCODE in the data entry form is not important. We will arbitrarily place it at the top:

This is the questionnaire for the laboratory register

idcode	Laboratory identifier	<input type="text" value=""/>	
labcode	Laboratory code	<input type="text" value=""/>	
labname	Laboratory name	<input type="text" value=""/>	
serno	Laboratory serial number	<input type="text" value=""/>	Write note (F5) if alternate required
regdate	Registration date	<input type="text" value=""/>	Enter 01/01/1800 if not recorded
sex	Examinee's sex	<input type="text" value=""/>	
age	Examinee's age in years	<input type="text" value=""/>	Enter 999 if not recorded
reason	Examination reason	<input type="text" value=""/>	
res1	Result of specimen 1	<input type="text" value=""/>	
res2	Result of specimen 2	<input type="text" value=""/>	
res3	Result of specimen 3	<input type="text" value=""/>	

**Suggestion:** it can be useful to place any field that contains information that is calculated by the CHK file and not entered by the data entry person physically separated from the fields which have to be entered (above or below). As the IDCODE will be “calculated” to be the result for 3 fields, it is proposed here to place it at the top.

In previous exercises we had the CHK file writing the Field Value into the field and get the Value Label just typed next to its right during data entry, as for example for the field SEX:


```
sex
  COMMENT LEGAL USE label_sex SHOW
  MUSTENTER
  TYPE COMMENT
END
```

resulting during data entry in:

sex                      Sex of examinee  Female

The Value label Female does not become part of the REC file (it belongs to the CHK file), only the Field Value 1. As the intention in this exercise is to create an IDCODE that uses the Value label (i.e., the Laboratory code) as one of the three components, this information must be written into a field, and thus be foreseen with a Field length of 4 in the QES file:

This is the questionnaire for the laboratory register

idcode	Laboratory identifier	_____	
labcode	Laboratory code	_____	
labname	Laboratory name	_____	
serno	Laboratory serial number	####	Write note (F5) if alternate required
regdate	Registration date	<dd/mm/yyyy>	Enter 01/01/1800 if not recorded
sex	Examinee's sex	#	
age	Examinee's age in years	###	Enter 999 if not recorded
reason	Examination reason	#	
res1	Result of specimen 1	##	
res2	Result of specimen 2	##	
res3	Result of specimen 3	##	

With this addition the QES file is complete: it has three new fields: LABNAME that will be entered, and LABCODE that will be the written Value label when the Field value for the new field LABNAME is entered.

**Note:** After completing the A\_EX07.QES file do not make a Data file as this will create an empty data file. Instead, we are going to preserve the data already entered into A\_EX05F.REC.

**Step 3. Preserving the existing validated dataset**

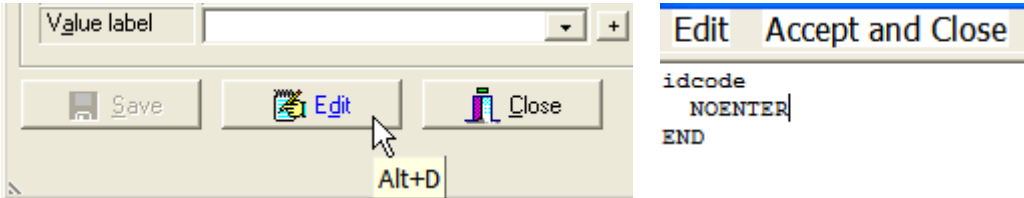
You have already a final, validated A\_EX05F.REC file. After you updated the A\_EX05F.QES file to the A\_EX07.QES file, you must export the A\_EX05F.REC file to A\_EX07.REC file. This will also give you the A\_EX05F.CHK file, copied to a A\_EX07.CHK file.

**Step 4. Editing the main CHK file A\_EX07.CHK**

We start by opening the A\_EX07.REC file (created by the export) and you will be prompted to update it to the A\_EX07.QES file. After affirming, close the file, and start editing the A\_EX07.QES file.

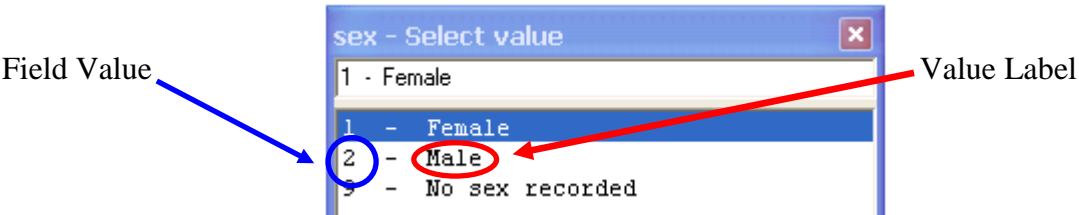
The first step might be editing all three new fields:

idcode  
labcode  
labname

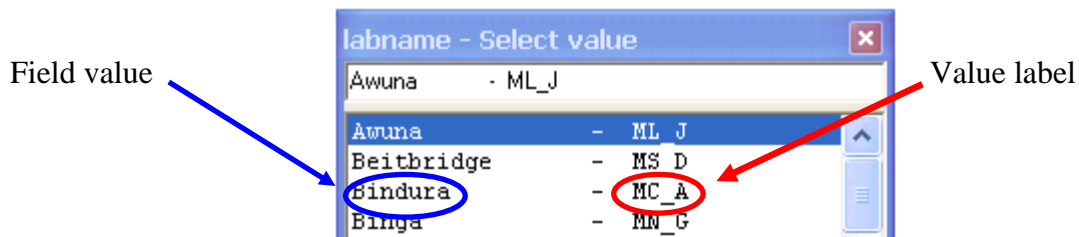


making them NOENTER fields.

When we entered the field SEX during data entry, the Labelblock showed:



Similarly, we want for LABNAME:



The A\_EX07\_NAMECODE.REC file will be invoked to serve as a Labelblock. Of course, we could make Labelblocks in the CHK file as introduced in an earlier exercise. If the list of values is very large, this becomes tedious work and may make for a very long check file, and at some point may even exceed the legal limits. In such an instance it may become more efficient to use another EpiData REC file that is invoked.

The notation in the example for the field SEX in the CHK file was:

```
sex
  COMMENT LEGAL USE label_sex SHOW
  MUSTENTER
  TYPE COMMENT
END
```

As you learned, the command TYPE COMMENT has the effect that the Value label is written to the right of the field SEX. Alternatively you can tell EpiData Entry to write the Value label into another field:

```
varx
  COMMENT LEGAL USE label_x SHOW
  MUSTENTER
  TYPE COMMENT otherfield
END
```

The grammar to accomplish the same thing with an external REC file rather than a Labelblock is very similar:

```
labname
  MUSTENTER
  COMMENT LEGAL a_ex07_namecode.rec SHOW
  TYPE COMMENT labcode
END
```

Note that USE is dropped from the command.

Once you have the relation made to these external files, you will have the Field value for the field LABCODE which you need for creating the identifier. Creating the identifier is like making a calculation. If you have two numeric fields, where the third field summarizes the information from the two numeric fields, we basically have:

```
field1
* some commands
END

field2
* some commands
```

```

AFTER ENTRY
  field3=field1+field2
END
END

field3
  NOENTER
END

```

If FIELD1 and FIELD2 are numeric fields with values 2 and 6 respectively then FIELD3 will get the value 8. If FIELD1 and FIELD2 are text fields with values “AB” and “XYZ” respectively, then FIELD3 will have the value “ABXYZ”. The identifier can obviously be created only after all the three values required are available, and this is placed in an AFTER ENTRY block:

```

regdate
  RANGE 01/01/2000 31/12/2005
  LEGAL
    01/01/1800
  END
  MUSTENTER
  AFTER ENTRY
  idcode=labcode+"-"+year(regdate)+"-"+serno
  END
END

```

Note above the function to extract the year from a date field.

Finally, we want to prevent a data entry person from ever even getting into the field IDCODE. This is accomplished by changing the MUSTENTER into:

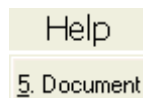
```

idcode
  NOENTER
END

```

**Note:** Do not make the field that takes the laboratory name (the Field value) a REPEAT field. In the current version of EpiData Entry, the Check commands in a field that uses an external file as a Labelblock will not be properly executed. Here the Value label will not be written into the field made for it and no IDCODE will thus be created.

Finally, before you complete this exercise, we encourage you to try the help file:



which you can access by clicking of with the shortcut **Alt+H**. **Please use it, that’s what it is for.**

**Important note:** Although later IDCODE will become the unique identifier, for the time being you must keep SERNO as the unique identifier. Why is that so? If you were to make IDCODE as the unique identifier, you could not update your A\_EX07.REC file, because all 15 records are missing it (thus non-unique).

## Step 5. Updating the A\_EX07.REC file

Open now the A\_EX07.REC file, go to the first record, add the LABNAME field, and keep pressing Enter to automatically update the three new fields:

IDCODE  
LABCODE  
LABNAME

Then go to the next record, and so on, until all 15 records are updated.

### **Step 6. Finalize the A\_EX07.CHK file**

As the last step, change the A\_EX07.CHK file to make IDCODE a KEY UNIQUE field and remove that instruction from the SERNO field.

#### **Tasks:**

- o Make the QES file A\_EX07\_NAMECODE.QES*
- o Import the text file A\_EX07\_NAMECODE.TXT which creates the A\_EX07\_NAMECODE.REC file*
- o Make the two fields in the Check file A\_EX07\_NAMECODE.CHK Key fields*
- o Edit the A\_EX07.QES file*
- o Export the A\_EX05F.REC file to A\_EX07.REC*
- o Open the A\_EX07.REC file to adjust its structure from the newer A\_EX07.QES file, then close it*
- o Edit the A\_EX07.CHK file to:*
  - Use the file A\_EX07\_NAMECODE.REC instead of a Labelblock for the Fields identifying the laboratory*
  - Create an identifier IDCODE from the three fields LABCODE, the year of REGDATE, and SERNO*
- o Update the A\_EX07.REC file with LABNAME (and the three fields calculated) for all 15 records*
- o Finalize the A\_EX07.CHK file to make IDCODE instead of SERNO the KEY UNIQUE*