Software EpiData – Applications for Needs of Biotechnological Processes

Ljakova K.

Centre of Biomedical Engineering "Prof. Ivan Daskalov"
Bulgarian Academy of Sciences
105 Acad. G. Bonchev St., 1113 Sofia, Bulgaria
E-mail: krasil@clbme.bas.bg

Summary: EpiData (free software for entering and documenting data) is presented. Some aspect of this software is shown for needs of database system (DB) and information systems (IS) that can be used in bioprocess system.

Keywords: Information Based System, Dataentry, Database System, Knowledge Based System, EpiData as Freeware.

1. ABOUT **EpiData**

EpiData is a Windows 95/98/NT/2000 based program (32 bit) for DataEntry. EpiData is developed by Jens M. Lauritsen, MD. PhD, Denmark. Program released as freeware by: The EpiData Association, Odense Denmark. Program design: Jens M.Lauritsen & Michael Bruus.

EpiData can be used when we have collected data on paper and we want to do statistical analyses or tabulation of data. Basic frequency tables and lists of data can be made, but other than that EpiData is focused on dataentry and documentation of data [1, 2].

During dataentry calculation of summary scales or restrictions to values can be defined. We can choose an item from a list and save the corresponding numerical code (1 = No 2= Yes), the text lists are exported as "value labels" for statistical programs. Dates are easily entered, e.g. 1909 will be formatted as 19/09/2007 if entered in year 2007 in a "dd/mm/yyyy" field.

EpiData is suitable for simple datasets where it has one source of data (e.g. one questionnaire or one laboratory registration form) as well as datasets with many or branching data forms; only the simple situation is described here.

The idea is that in EpiDate must write simple text lines and the program converts this to a dataentry form for which we can add further control of entry, conditional jumping to other fields or calculations.

Downloading and installing EpiData is free of charge from http://www.epidata.dk. Must be followed the instructions when run the installation file. EpiData will not interfere with the setup of the computer. EpiData consists of one program file and help files. (In technical terms: EpiData comes as a few files and does not depend on, install or replace any DLL files in the system directory. Options are saved in an ini-file).

No limit on number of observations (tested with >100.000). Search with index in 80.000 records < 1 sec on Pentium I 200 MHz). Specification of the data file structure must fit within 999 lines of text.

2. HOW TO WORK WITH **EpiData**

The EpiData screen has a "standard" windows layout with one menu line and two toolbars.

The "Work Process toolbar" guides you from "1. Define data" to "6. Export data" for analysis is shown on Fig. 1.
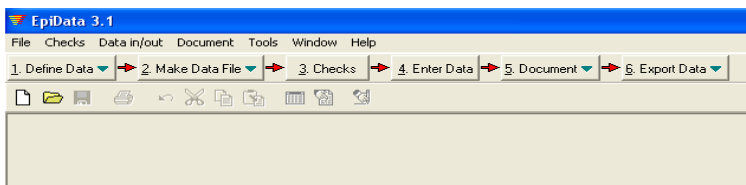

Fig. 1

2.1. Define data

Data is defined by writing three types of information for each variable:

A. Name of input field (variable, e.g. v1 or exposure).

B. Text describing the variable (e.g. mode of something or "day of beginning").

C. A field definition, e.g. ## for two digit numerical.

Other options: e.g. colour of background and fields, line height etc. Users of e.g. Stata or SPSS should use the "first word" principle; field names will be variable names. Users of Stata should choose lowercase field names.

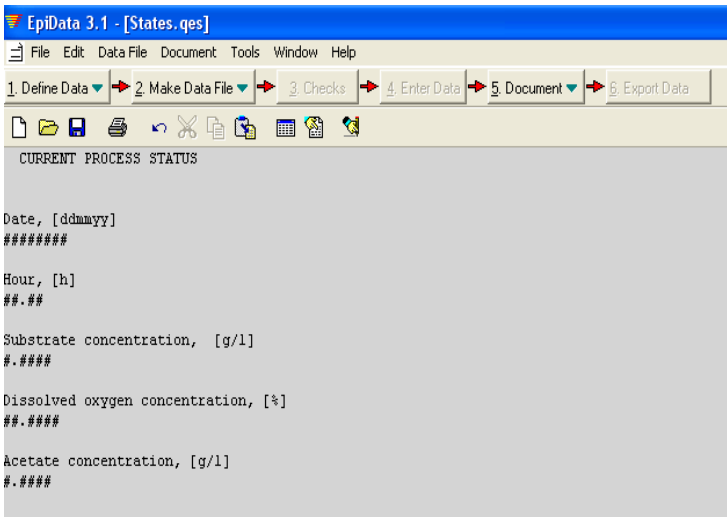The procedure is shown on Fig. 2. The more details are given in [4].

Fig. 2

2.2. Dataentry form:

id (automatic id number) <idnum >
Field1 Date ######
Field2 Hour [h] ##.##
Field3 Substrate concentration [g/l] ###.#
Field4 Dissolved oxygen concentration [%] ##.##
Field5 Acetate concentration [g/l]

The option **range** shows possible value of variables (Fig. 3).

Before making *.rec file (the main file of dataentry), we can see an appearance of the form (in option preview) and if it is necessary we can make correction.

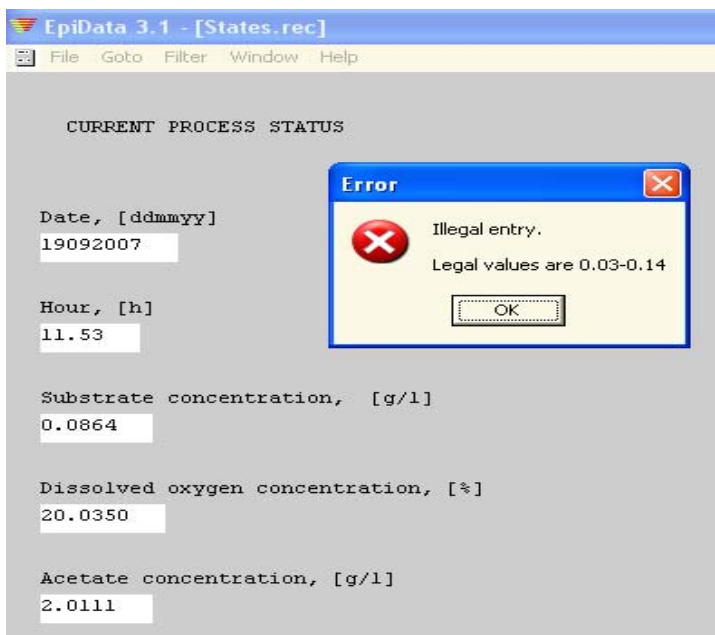The dataentry is taking only in ready datafile – *.rec file.

Fig. 3

2.3. Making datafile

After writing the definition the user can preview dataform and create the actual datafile (Fig. 4).

2.4. Add/Revise checks – at dataentry

A strong part of EpiData is the possibility to **specify rules and calculations during** dataentry.

- Restrict dataentry to certain values and give text descriptions to the numerical codes entered.
- Specify sequence of dataentry e.g. fill out certain questions for males only, **(jumps)**.
- Apply calculations during dataentry. E.g. age at visit based on date of visit and date of birth. But typically most calculations are done at the analysis stage.
- Help messages and other extended definitions of computations, "if… then ...endif" structures.
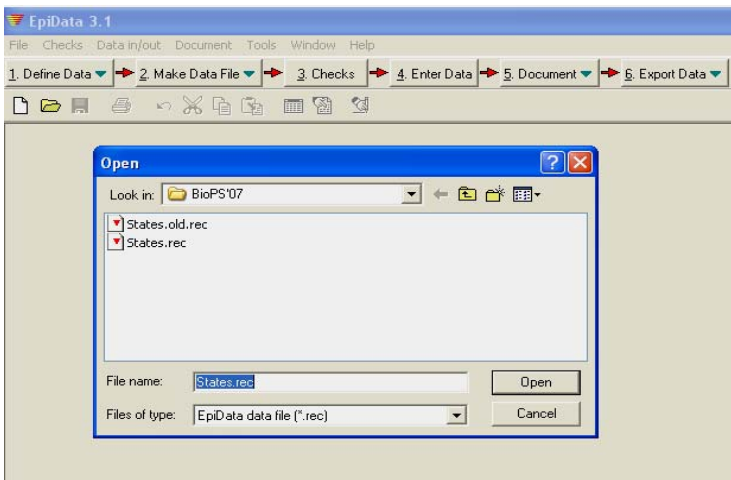

Fig. 4

When the user start the "add/revise" part a new screen appears: At the top the variable name (FIELD) is shown. Below that the label (Rigidfix) and the variable type (number). Following this the definition blocks:

      Range, Legal defines which data can be entered.

      Jumps specifies where to go to after entry (here on value 1 jump to field v10)

      Must enter: If set to Yes a value must be given. (Otherwise leaving the field blank is accepted).

Repeat: Repeat value from previous record, e.g. if data are from groups a value will repeat until next group. The value can still be changed.

Value label: For categorical data this defines the meaning of the values. E.g. 1=aerobic 2=anaerobic. The values are edited via the ”+” . The drop-down list lets you pick an existing label definition.

Edit: Many other aspects can be defined here as “free hand” editing for this field. See the collection of commands in the help file.

Save: Save current definitions.

In the example **Range**, **Legal** and **Value labels** are defined. In a typical dataentry this would not be the case. Only one of them would be used (Fig. 5).

2.5. Enter data

Open the file and enter, add or search data. The blue explanatory text to the right of the input fields is added by EpiData after entry of data based on labels in check file. Files saved:

A. Dataform definition file. E.g. **states.qes**

B. Actual datafile containing the data. E.g. **states.rec**

C. A file with the defined checks. E.g. **states.chk**

D. Supplementary files, e.g. **states.not** with notes taken during dataentry or **states.log** with documentation.

The current data in states.rec is real experimental data from *E. coli* fed-batch cultivation process [3].

2.6. Document data

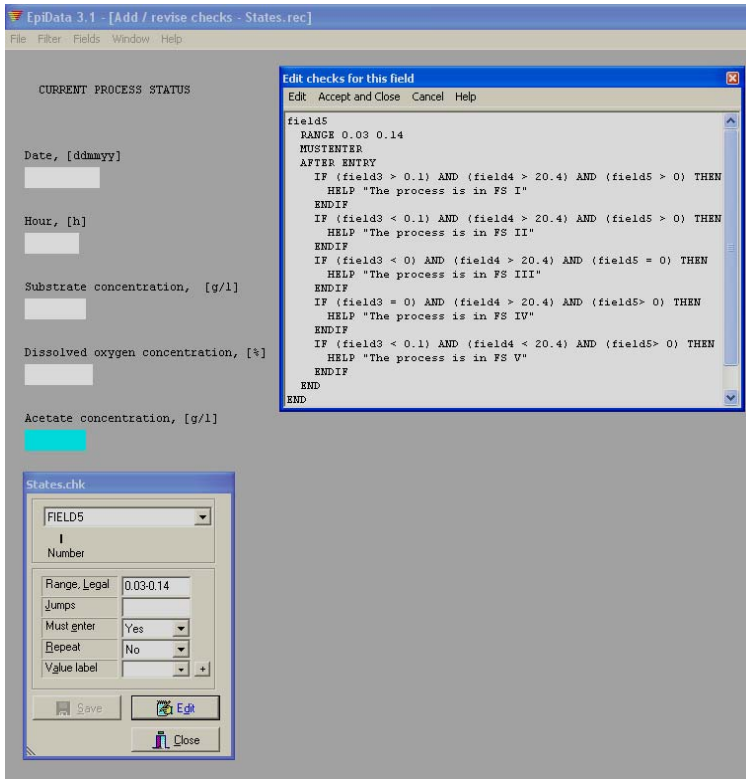After creating the datafile the user can document file structure. An example is shown on Fig. 6.

Fig. 5

2.7. Export for analysis and securing data

Export the data for analyses or backup all files associated with a data set to a selected user defined backup folder (Fig. 7, Fig. 8).
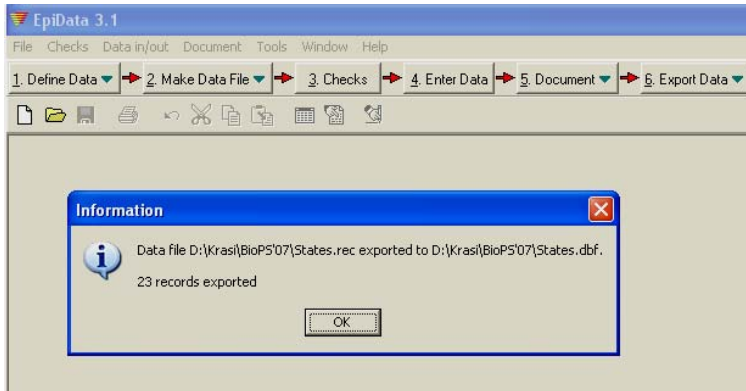
Fig. 6

Fig. 7

Fig. 8

2.8. Tools and other

EpiData includes other aspects like comparing two files and listing differences at field level. That is very important by analysis of the bioprocess.

Revising data file structure without loosing already entered data. Hierarchical coding, relational dataentry, check for logical consistency. Future additions could include functions for adding user defined extensions to menu systems and the check file language.

2.9. Analysis

EpiData includes functions for overall tabulation (codebook). Work has been started to develop a basic analysis package. Export of data for analysis can be made in various data file formats (data and variable names: comma separated ascii file, dbaseIII or Excel) or a complete export (data and variable names, plus labels: Stata, SAS and SPSS).

3. CONCLUSION

EpiData free software is very suitable ware for needs of modelling and analysis of bioprocess systems. We make a first attempt to use this freeware in [4] for a creation of a system for functional states recognition of *Escherichia coli* fed-batch cultivation.

REFERENCES

1. Lauritsen J. M., M. Bruus, EpiData (Vers. 3). A Comprehensive Tool for Validated Entry and Documentation of Data. The EpiData Association, Odense, Denmark, 2003.
2. www.epidata.dk
3. Georgieva O., M. Arndt, B. Hitzmann, Modelling of *Escherichia coli* Fed-batch Fermentation, *International Symposium* "*Bioprocess Systems 2001 - BioPS'01*", Sofia, Bulgaria, October 1-3, I.61-I.64.
4. Ljakova K., O. Roeva, A System for Functional States Recognition of *Escherichia coli* Fed-batch Cultivation, *International Symposium* "*Bioprocess Systems 2007 - BioPS'07*", Sofia, Bulgaria, November 6-7, I.75-I.84.