A short overview of dataentry with

EpiData

EpiData is a program for entering and documenting data.

Use EpiData when you have collected data on paper and you 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 data entry and documentation of data.

During data entry calculation of summary scales or restrictions to values can be defined. You 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. 2301 will be formatted as 23/01/2003 if entered in year 2003 in a "dd/mm/yyyy" field.

EpiData is suitable for simple datasets where you have 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 principle is rooted in the simplicity of the dos program Epi Info v6.
The idea is that you write simple text lines and the program converts this to a dataentry form for which you can add further control of entry, conditional jumping to other fields or calculations.

Downloading and installing EpiData is free of charge. Download from www.epidata.dk and follow the instructions when you run the installation file. EpiData will not interfere with the setup of your 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 your system directory. Options are saved in an ini file).

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

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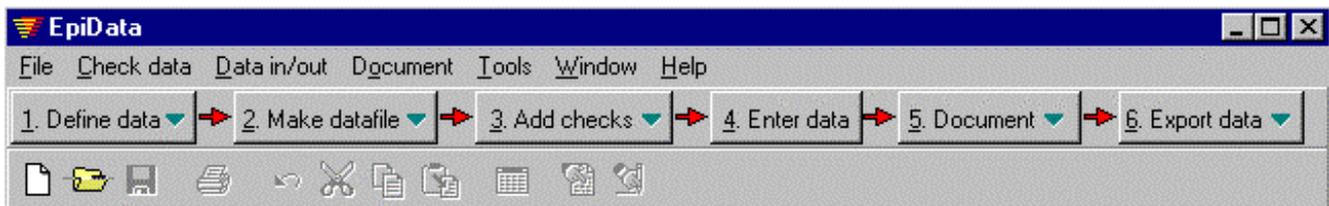
The EpiData screen has a "standard" windows layout with one menu line and two toolbars.

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

1. Define Data
Define data 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. sex or "day of birth")
  C.. A field definition, e.g. ## for two digit numerical.

Other field types are boolean (yes-no), encrypted or soundex fields.

```
My first DataEntry Form
id          (automatic id number)  <idnum >
V1                            sex  #
V2                Height (meter)  #.##
v3                Weight (kilo)  ###.#
bmi              Body Mass Index ##.##
v4                Date of birth  <dd/mm/yyyy>
age                Age today      ###
s1            Country of Residence  _____
s2         City (Current address) <a    >
t1                  Todays Date  <Today-dmy>
```

Setting options (file menu) It is important that the user decides on one of two principles for naming variables. With "first word" the name of the input fields will be taken as first word on the definition line. With "automatic" a combination of the first 10 characters in the line. The example would give:

      a. v1sex (10 first characters in sentence)   "**automatic**" field name is **v1sex**
      b. v1 (first word of sentence).   "**first word**" field name is **v1**

Version: 25th August 2005

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.

2. Make datafile.

After writing the defintion you can preview your dataform or create the actual data file.

3. Add/Revise Checks - at Entry of Data

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.
  (See examples installed with EpiData, or get further examples from www.epidata.dk/examples.php).

When you start the "add/revise" part a new screen appears: At the top the variable name (v8) 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=man 2=women. 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 both **Range,Legal** and **Value labels** are defined. In a typical dataentry this would not be the case. Only one of them would be used.

4. 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. An example of calculated fields is the Body mass index and age.

Files saved:

A. Dataform definition file.
E.g. **first.qes**
B. Actual datafile containing the data. E.g. **first.rec**.
C. A file with the defined checks. E.g. **first.chk**
D. Supplementary files, e.g. **first.not**   with notes taken during data entry or **first.log** with documentation.

5. Document Data

After creating the datafile you can document file structure. An example (part of *first.rec)* is:

```
DATAFILE:  C:\data\first.rec
Filelabel: My first test datafile is an example

Filesize:          612 bytes
Last revision:     28. Jan 2001 12:14
Number of fields:  7
Number of records: 0
Checks applied:    Yes (Last revision 28. Jan 2001 12:02)

Fields in datafile:

No. Name Variable label    Fieldtype       Width  Checks           Value labels
------------------------------------------------------------------------------------
  1 id                     ID-number       6

  2 v1   sex               Integer         1                       sex
                                                                   1: Male
                                                                   2: Female
                                                                   9: Unknown
  3 v2   Height (meter)    Fixed number    4:2    Legal: 0.0-2.30,9

  6 v4   Date of birth     Date (dmy)      10

      (other fields omitted)
```

And after entering data you may list values for some or all records or you could view the data with or without labels in a spreadsheet like window (Try Ctrl+Alt+V).

```
Observation 1

    id               1       v1          Male       v2              1.92
    v4      12/12/1945       s1       denmark       s2        Copenhagen
```

A "codebook" can include raw frequency tables. (example not based on first.rec file)

```
v2 ----------------------------------------------------------------------- Sex
              type:  Integer
       value labels:  sex
        range/legal:  1-2,2

            missing:  0/25
              range:  [1,2]
      unique values:  2

         tabulation:  Freq.   Pct.   Value  Label
                         11    44.0       1  Male
                         14    56.0       2  Female
v3 ----------------------------------------------------------------------- Temp
              type:  Floating point
        range/legal:  36.00-40.00

            missing:  0/25
              range:  [36.00,37.50]
      unique values:  12

               mean:  36,84
           std. dev:  0,37
```

6. Analyse with EpiData Analysis or export for special analysis and securing data.

Export the data for special analysis or backup all files associated with a data set to a selected user defined backup folder. You can analyse data directly with EpiData Analysis (avaialable autumn 2005) or When exporting data to Stata you should select export's lowercase option.

As part of the tool menu you can also create zip file compatible archives with optional encryption. The encryption applies AES/Rijndael strong encryption approved by many data authorities.

Please note that there is **no** way to retrieve a password if you forget it.

Tools and other

Epidata includes other aspects like double entry or comparing two files and listing differences at field level. 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.

Analysis

EpiData includes functions for overall tabulation (codebook). EpiData Analysis will be released in autumn 2005. Follow the information on www.epidata.dk. Export of data for special 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).

Sources for support

The help file and various other supplementary documents are available for free from
http://www.epidata.dk

An e-mail based discussion and support list among users is available at
http://lists.umanitoba.ca/mailman/listinfo/epidata-list

About EpiData

EpiData is a Windows 95/98/NT/2000/XP based program (32 bit) for DataEntry. EpiData works on Mac computers based on emulation and on LINUX versions based on WINE.

EpiData development initiated by Jens M. Lauritsen, MD. PhD, Denmark. Program released as freeware by: The EpiData Association, Odense Denmark. Program design: Jens M.Lauritsen & Michael Bruus.

Suggested citation:

Lauritsen JM & Bruus M. EpiData (version). A comprehensive tool for validated entry and documentation of data. The EpiData Association, Odense, Denmark, 2003-2005.

First version of EpiData released as Lauritsen JM, Bruus M., Myatt MA, EpiData, version 1.0-1.5. A tool for validated entry and documentation of data. County of Funen Denmark and Brixton Health UK. 2001.

Funding and acknowledgements.

An updated list of attained funding is available at Http://www.epidata.dk/funding.htm and further credits and acknowledgements at: Http://www.epidata.dk/credit.htm . International translations made to several languages, see Http://www.epidata.dk

For donations to further development see help file or send an e-mail to info@epidata.dk