

Introduction:

• Data management includes all aspects of data planning, handling, analysis, documentation and storage and takes place during all stage of a study.

• The objective is to create a reliable database containing high quality data.

• Data management is a too often neglected part of study design, and includes:

- Planning the data needs of the study
- Data collection
- Data entry
- Data validation and checking
- Data manipulation
- Data files backup
- Data documentation

• Each of these processes requires ^{فكر: تفكير} thought and time, each requires pains taking attention to detail.

• The main element of data management are

database files.

Database files contain text, numerical, images and other data in machine readable form. Such files should be viewed as part of a database management systems.

• DBMS^{System} which allows for a broad range of data functions, including data entry, checking, updating, documentation, and analysis.

1. Data collection:

• Before analyzing data must be collected, reviewed, coded, computerized, verified, checked and converted to forms suited for the analyses to be conducted.

• The process must be adequately documented to provide the foundation for analyses and interpretation.

• Data collection import baseline data from:

→ Existing systems

→ Import lab results

- Scan results
- And other digital data

2. Data Management Software (Data analysis):

- Many DBMSs are available for personal computers.

Options include:

- i. Spread sheet (e.g. Excel, SPSS data sheet)
- ii. Commercial database program (e.g. Access)
- iii. Specialty data entry program (e.g. Epi Data)

i. Spreads sheet are to be avoided for all but the smallest data systems, since they are unreliable and easily corrupted.

ii. Commercially database programs are expensive, tend to be large and slow and often lack controlled data-entry facilities.

iii. Specialty data entry programs are ideal for data entry and storage. For example using

Epi Data because it is fast, reliable, allows for controlled data-entry, and is open-source.

◦ Epi Data refers to a group of applications used in combination for creating documented data structures and analysis of quantitative data.

The Epi Data Association, which created the software was created in 1999 and is based in Denmark.

3. Data Entry and Validation:

◦ Data processing errors are errors that occur after data have been collected.

◦ This is distinct from measurement errors, which are differences b/w the true state of affairs and what appears on the data collection form.

◦ Examples of data processing errors

◦ ^{تبادل} Transpositions (e.g. 19 becomes 91 during data entry).

- Copying errors (e.g. 0 (zero) becomes "O" during data entry)

- Coding errors (e.g. a racial group gets improperly coded because of changes in the coding scheme)
- Routing errors (e.g. the interviewer asks the wrong question or asks questions in the wrong order)
- Consistency errors (contradictory responses, such as the reporting of a hysterectomy after the respondent has identified himself as a male)
- Range errors (responses outside of the range of plausible answers, such as a reported age of 290)
- To prevent such errors, you must identify the stage at which they occur and correct the problem.

Methods to prevent data entry errors include:

1. Manual checks during data collection (e.g. checks for completeness, handwriting legibility)
2. Range and consistency checking during data entry (e.g. preventing impossible results, such as ages greater than 110).

3. Double entry and validation following data entry.

4. Data backup and storage:

There are two kinds of computer users.

- i- Those that have lost a major chunk of data, and
- ii- those who are going to lose a major chunk of data.

- Data loss can be due to natural disasters, human error and computer failure.

- You've worked too hard to collect and enter data, and you must now take care of it.

- The most common loss of data among students is due to loss of data somewhere on the computer.

The best way to prevent such loss is to know the physical location of your data (local drive, removable media, network) and to use logical filenames.

All too often students save files to unknown locations (usually the default set up by the program).

- But never find saved files or have the saved

Files deleted by the local area network as a part of routine data clean up.

- Always Be Aware Of The Location And Path "Folder" To Which Files Are Being Written.

- In addition, it is essential to back-up all data (e.g. data files, code books, software setting, computer programs, word processing documents).

- Back up system sentail manual or automated copying of files to removable media (e.g. floppy disks, Zip disks, tape) or to network storage.

- Back up procedures should be thoroughly tested to ensure archived files remain uncorrupted and can be restored.

- Procedures should be written up so that personnel unfamiliar with back up and restore methods could follow them from scratch.

- Health researcher must be aware of confidentiality and ethical requirements when working with research files.

- This is especially important when data contain

personal identifiers and medical information.

• It is each researcher's duty to make him or herself aware of local, national, and international laws governing use of health data.

5. Documentation:

Documentation should include:

- i- The objectives and methods of the study.
- ii- Detailed of events and activities
- iii- For each data collection activity, a record of the detailed procedures used and an accounting of the number of subjects in scope, the number selected for data collection, and the disposition for all subjects (by category, e.g. could not contact, refusal). This material should have cross references to original sources (e.g. computer runs) tenable verification when necessary.
- iv- Lists and descriptions of all interventions applied and materials used (e.g. intervention materials, training materials)

V. Documentation on all computerized data and final analyses (information on data sets, variables, computer runs).