

# 19: Notes on Data Management

## Introduction

Data management entails:\*

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

Each of these processes requires thought and time; each requires painstaking attention to detail.

The goal is to create a **database management system (DBMS)** that allows for a broad range of data functions. Many such systems are available. Example of DBMSs include:

- Spreadsheets (e.g., Excel, SPSS)
- Commercial DBMS programs (e.g., Oracle, Access)
- Public domain systems (e.g., EpiData, Epi Info)

Spreadsheets do not make good DBMSs. They are prone to error, corruption, and mismanagement. They lack data controls and have limited programmability. Thus, spreadsheet based DBMSs are suitable for small and trivial projects only.

Commercial DBMSs are designed for data management. They are powerful and often have good programmability and data flow controls, but tend to be slow & expensive when used on PC systems.

Public domain programs often offer controlled data entry and good programmability. Several such systems are particularly well suited for research and field use.

In this class, We will use a two platform DBMS: EpiData for controlled data entry and documentation, and SPSS for analysis and reporting.

---

\* Bennett, S., Myatt, M., Jolley, D., & Radalowicz, A. (2001). *Data Management for Surveys and Trials. A Practical Primer Using EpiData*. The EpiData Documentation Project. Available: [www.epidata.dk/downloads/dmepidata.pdf](http://www.epidata.dk/downloads/dmepidata.pdf).