

Each datum is a **single observation** of a patient and has **5** elements:

- the patient (Amr Jamal)
- the attribute (heart rate)
- the value of the attribute (52 beats per minute)
- the time of the observation (1:00 pm on 1/1/2015)
- the method by which the attribute was obtained (heart monitor)

Types of data: Narrative (Stories and paragraphs), Numerical measurements (BP, Temperature), Coded (MI: Myocardial infarction), Textual (text), Recorded (EKG, EEG), Pictures (Radiographs, Photographs)

Types of clinical data document:

- 1- History and physical examination (by clinician)
- 2- Progress notes (by primary, consulting and ancillary)
- 3- reports (by specialist and ancillary providers)

General categories of data entry:

- 1- Free form (writing, dictation and typing)
- 2- **Structured (menu-driven) by mouse or pen**
- 3- speech recognition

Benefits of structured data entry (menu-driven)

- Data codified for easier retrieval and analysis
- Reduces ambiguity if language used consistently

Drawbacks of structured data entry (menu-driven)

- In general, more time-consuming
- Requires exhaustive vocabulary
- Requires dedication to use by clinicians

“**only** a human can prioritize and determine what the chief complaint really is”

Issues with coded data:

- “pick from a list” allows wrong selection
- compliance concerns
- over documentation for care
- cloning

File Organization concepts

- Database: A set of related files
- File: Collection of records of same type
- Record: A set of related field
- Field: Words and numbers

- **Database < File < Record < Field**

- **Relational model** links records to **tables**
- Most query capabilities are based on **Structured Query Language (SQL)**- special language in **relational database**

Big data: Science of Data Management & analysis

What is BIG/VAST ? **Zettabytes** (10^{21} gigabytes) to **Yottabytes** (10^{24} gigabytes)

Example of Big Data is Bioinformatics

The four V's of big data:

Volume (Scale of data), **Variety** (Different forms of data), **Velocity** (Analysis of streaming data), **Veracity** (Uncertainty of data)

“80% of medical data is **unstructured** and is clinically relevant”

Sources of big data:

- **Clinical Data from CPOE**
- **Clinical decision support systems** (Written notes & prescriptions)
- **Imaging systems: PACS, Radiology Information systems**
- **Sensor data** (monitoring vital signs)
- **Social media data-** Tweets from Twitter, wall and status updates on Facebook
- **Emergency care data**
- **Literature from medical journal**