Reviewed By

Noura Alturki
Jehad Alorainy

# Clinical Data

Click here for: Book

**Color index:**

- Important
- Golden notes
- Doctors notes

Editing File

# Data

- **Data:** are symbols or observations reflecting differences in the world. (ex: 5)
- **Information:** is data with meaning. (ex: 5 fingers)
- **Knowledge:** is information that is justifiably believed to be true. (ex: elevation in FBG level, the likelihood of diabetes is increased, smokers are likely to develop lung cancer more than non-smokers)
- **Wisdom:** is the critical use of knowledge to make intelligent decisions.

Each zero or one is a bit, each 8 bits equal a byte

| Bits can occur as various data types: |
| --- |
| • Integers (numbers ex: 14)<br>• Floating point numbers (decimals)<br>• Characters (A-Z)<br>• Character strings (ex: words) |

| File format: |
| --- |
| • image files (JPG, GIG, PNG),<br>• text files<br>• sound files (WAV, MP3)<br>• video files (WMV, MP4) |

Is important to recognize that neither data types nor file formats define the meaning of the data.

## Types of clinical data

| Narrative | Numerical measurements | Coded data |
| --- | --- | --- |
| recording by clinician- maternity history. | blood pressure, temperature, glucose level. | selection from a controlled terminology system, anything that has a drop down menu (example being the term MI that may mean myocardial infarction or mitral insufficiency). |
| **Textual data** | **Recorded signals** | **Pictures** |
| other results reported as text. | EKG, EEG. | radiographs, photographs, and other images. |

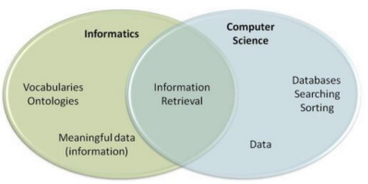
# Clinical Data

**Clinical data:** are a collection of observations about a patient.
**A datum:** is a single observation of a patient.

## General categories of data entry:

- **Free-form** entry by historical methods:
  - Writing
  - Dictation
  - Typing
- **Structured (menu-driven) data** entry by mouse or pen.
- **Speech recognition** for either of above.

## Informatics and Computer Science




- **The domain of:**
  - Computer scientists is **Data**
  - Informaticians is **information**
- Information retrieval involves both computer science (data) and informatics (information)
  - Retrieving all the documents related to Aspirin & heart attacks.

## Artificial Intelligence (AI)

- AI is concerned with the development of systems that can do something that previously required human intelligence. (chess, driving vehicles)
- **Example in medicine:** derma skin lesions categorization
  - A system is given large sets of images which are labeled (benign, malignant,...)
  - The system tries to learn to be able to identify lesions in the features

- Limitations:
  - Require large sets of Data
  - Cannot justify judgment (why did it classify this picture as benign?)

# Clinical Data

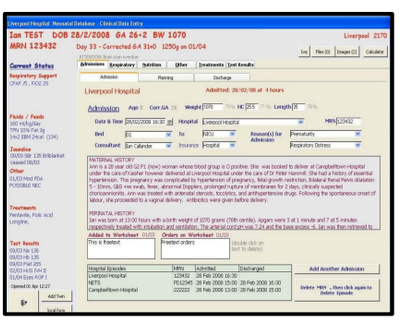
## Data to Information:

**Example:** C34-9[1] (data) → **Vocabulary** (ICD-10-CM[2]) → "Lung neoplasm, not otherwise specified." → **Interoperability**

**ICD:** International Classification of Disease, 10 is the version.

**Interoperability:** Ability of two or more systems or components to exchange the information and use the information that has been exchanged. (ex: medical record at hospital A, sends information to a medical record at hospital B without losing this information).

## Information to Knowledge:



**Two types of data:**

- **Structured data** (from *drop down menus* such as; billing codes, lab results, problem nests and medication nests).
- **Unstructured (free text)** (simple human language ex: history).

Dr. Notes:
Each has its own **advantages** and **disadvantages**, they both are needed and depended on.
The **structured data** is easier to manage and faster in analysis while **unstructured** gives the ability to express thoughts in natural language but it's harder to manage. (ex: retrieval of all patients medical records that are in ICD-10-CM system with the same disease such as breast cancer).

**Natural Language Processing:** studied the analysis of unstructured data. (takes the natural human language as an input, process it (low level processing: assigns the part of speech to words "verb, noun" while high level processing such as Siri).

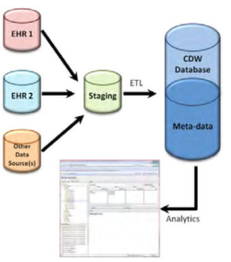
## What Makes Biomedical Informatics Difficult?

- Incomplete information.
- Uncertain information.
- Imprecise information.
- Vague information.
- Inconsistent information.
    - Ex: Birthdate: 8/29/66 and 9/17/66.

1- C34-9 in itself is meaningless
2- however, under the ICD-10-CM system it refers to the code for "Lung neoplasm not otherwise specified"

# Clinical Data Warehouse

- A modern **way to convert medical information to knowledge** is to use a clinical data warehouse (CDW).
- **Function:** a database system that collects, integrates and stores clinical data from a variety of sources including electronic health records **(EHR)**, radiology and other information systems.
- **Staging:** the process of cleaning the data from any incorrect or missing values.
- **Meta-Data:** "data about data" example: ICD-10-CM
- Difference between EHR & CDW:
  - **EHR:** are systems implemented in hospitals that focus on
  - individual data for one patient.
  - Address real time updates regarding patient data
  - **CDW:** support many EHR, focus on group query.
    - Information update depends on how often it is updated (monthly, weekly,...)

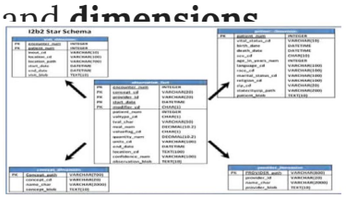



## Uses:

- **Monitor Quality** by allowing users to query for specific quality measures
- Identify trends
- Comparative effective research: practice based research, answers very specific questions (is the treatment A better than treatment B? We can answer this question by looking for the prognosis of people who were treated with both in the system)

## i2b2[1] platform (an example of CDW)

- A harvard project to integrate biology and the bedside used by other institutes
- Its an open source and modular and incorporate genomic and clinical information for research purposes
- Database consists of **facts** (diagnoses, lab results, etc.) queried by users and **dimensions** that describe the fact
- **Design:** star scheme
  - In the middle there is a fact table ( quantitative, information)
  - On the side dimension tables ( more information about these facts.



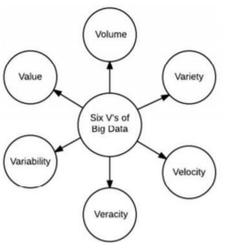

## Use of Aggregated Clinical Data

- **Concept extraction:** the problem of identifying concepts within unstructured data, such as discharge summaries or pathology reports.
  - Usually, these concepts are mapped to a controlled vocabulary.
- **Classification:** the problem of categorizing data into two or more categories.
  - Supervised machine learning.

1- I2B2 stands for "**i**nformatics for **i**ntegrating **B**iology & **B**edside

# Big Data

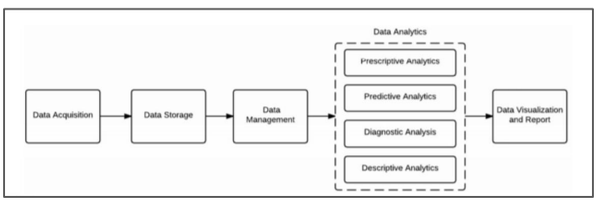
**Definition:** collecting large collections of data from various healthcare foundations followed by storing, managing, analyzing, visualizing, and delivering information for effective decision making.




**The Six <u>V's</u> of big data:**

**Everything after this line is Dr. notes**

- **Volume:** large quantity of data produced by organization.
- **Variety:** a lot of sources in different formats, contribute in big data.
  - **Organized** (structured data: laboratory data, sensors data, data from databases).
  - **Semi-organized** (ex: website data).
  - **Unorganized** "*most common*" (ex: social media data, medical records that doesn't use medical terminology).
- **Velocity:** massive frequency during the current details created, supplied and managed.
- **Veracity:** refers to accuracy of this data if it's accurate or valid (ex: in big data there is usually low veracity compared to other databases).
- **Variability:** data is fluctuating, changeable data, guidelines are changeable as well and might me updated so data might change during life.
- **Value:** methods of extracting valuable information from these data (big data analytics).




**Analysis:** process of transforming raw data into information.

**Stage 1 Data acquisition:** there are different sources (Electronic record, social media, smart phone, apps, websites… etc) to collect the data.

**Stage 2 Storage:** store in cloud chambers.

**Stage 3 Management:** organizing, cleaning and data mining to make sure that the data is valid, no missing data or values.

**Stage 4 Analytics:**

- Descriptive analysis: collects historical data, *what happened in the health care?*
- Diagnostic Analysis: route cause of the problem, *why did it happen?*
- Predictive Analysis: use both real time data and historical data, it has probabilities, *what will happen? what is the future trends?*
- Prescriptive Analysis: number of different outcomes before the decision, *what should we do? And what is the best outcome? And how we can make it happen?*

**Stage 5 Data visualization and reporting:** *how I represent these results?* (ex: in graphs).

# QUIZ

**1. which of the followings is picture data?**
A. Temperature.
B. Blood test.
C. X-ray.
D. EMG.

**2. Which of the following is a major limitation in using abbreviation in narrative data?**
A. Produce ambiguity.
B. Hard to read.
C. Needs skills.
D. Against patient's privacy.

**3. "Pick from a list" allows wrong selection, compliance concerns, cloning and limitations are issues for?**
A. Coded data.
B. Voice recognition.
C.Language processing.
D.Manual data entry.

**4. The veracity of big data is due to which of the following?**
A. Uncertainty of data
B. Velocity of data
C. Multiple sources of data
D. Different forms of data

**5. What of these describe narrative?**
A. Business Documents.
B. Choose from a list.
C.Report as text.

**6. "140/90 mmhg" is considered which one of data elements?**
A. attribute
B. value of the attribute
C. time of the observation
D. method obtained

**7. "Collection of records" is definition of what?**
A. Field
B. File
C. Database
D. Record

**8. The type of data that makes the data (understandable) for the computer:**
A. Coded data
B. Free form data
C. Unstructured data
D. Dictated data

1.C ,
2.A ,
3.A ,
4.A ,
5.C ,
6.B ,
7.B ,
8.A

# Good Luck!

TEAM LEADER
JUDE ALOTAIBI

TEAM LEADER
KHALID ALKHANI

## DONE BY OUR AMAZING MEMBER:

**Omar Aldosari**

## NOTE TAKEN BY OUR SHARP MEMBER: