Week 2: Clinical Data and Big Data

Dr. Jwaher Almulhem

Assistant Professor in Health Informatics

Medical Education Department
College Of Medicine
King Saud University

email: jalmulhem@ksu.edu.sa

Acknowledgment:

Book: Health Informatics Practical Guide

Chapter 2: Healthcare Data, Information and Knowledge

Elmer Bernstam MD

Todd Johnson PhD

Trevor Cohen MD PhD

Big data : Senthilkumar, S. A., Rai, B. K., Meshram, A. A., Gunasekaran, A., & Chandrakumarmangalam, S. (2018). Big data in healthcare management: A review of literature. *American Journal of Theoretical and Applied Business*, *4*(2), 57-69.

Outlines:

- Define data, information, and knowledge
- Types of clinical data
- Informatics vs. Information Technology and Computer Science
- Data to information
- Information to Knowledge
- Clinical Data Warehouse
- Use of Aggregated Clinical Data
- What Makes Informatics Difficult?
- Big data.

Data, Information, and Knowledge

Data are symbols or observations reflecting differences in the world.

Information is data with meaning.

Knowledge is information that is justifiably believed to be true.

Wisdom is the critical use of knowledge to make intelligent decisions.

Data

- Each zero or one is a bit.
- Byte : ; a series of 8 bits .
- **Bits can occur as various data types**
  - Integers
  - Floating point numbers
  - Characters
  - Character strings
- **File format:**
  - image files (JPG, GIG, PNG),
  - text files,
  - sound files (WAV, MP3)
  - video files (WMV, MP4)

it is important to recognize that neither data types nor file formats define the meaning of the data

Types of clinical data

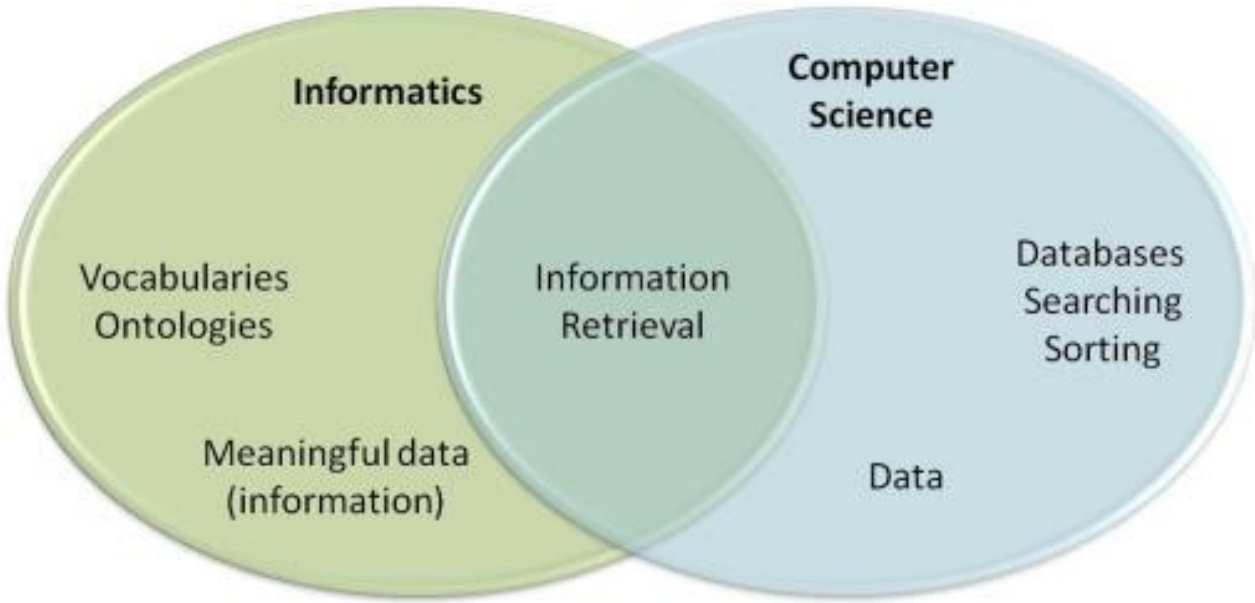
- Narrative: recording by clinician- maternity history
- Numerical measurements: blood pressure, temperature
- Coded data: selection from a controlled terminology system
- Textual data: other results reported as text
- Recorded signals: EKG, EEG
- Pictures: radiographs, photographs, and other images

Data entry

- 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

- Data are the domain of computer scientists, but information is the domain of informatics and informaticians
- Information retrieval involves both computer science (data) and informatics (information).



Artificial Intelligence (AI)

AI is concerned with the development of systems that can do something that previously required human intelligence.

Dermatologic (skin) lesion categorization

Data to Information

Example: “C34.9”

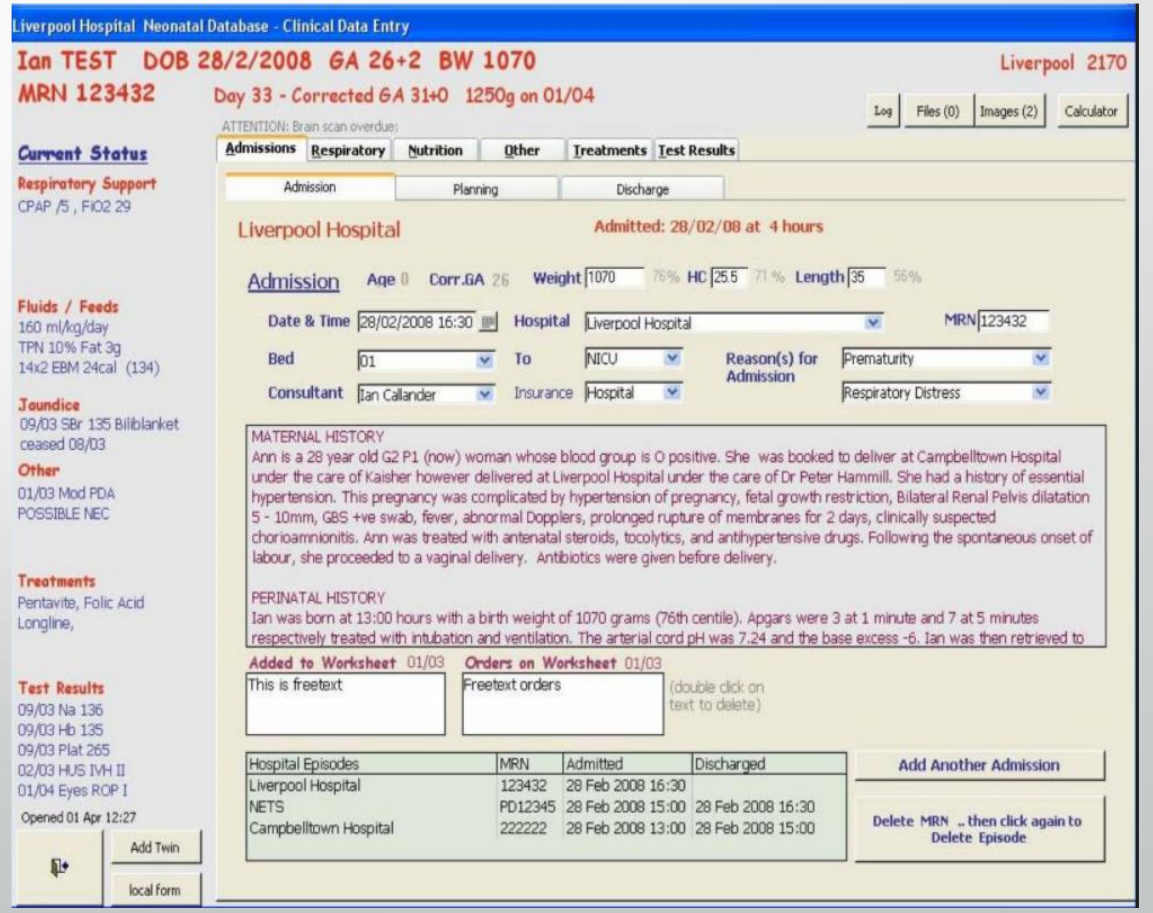
Vocabulary (ICD-10-CM)

“*Lung neoplasm, not otherwise specified.*”

Interoperability

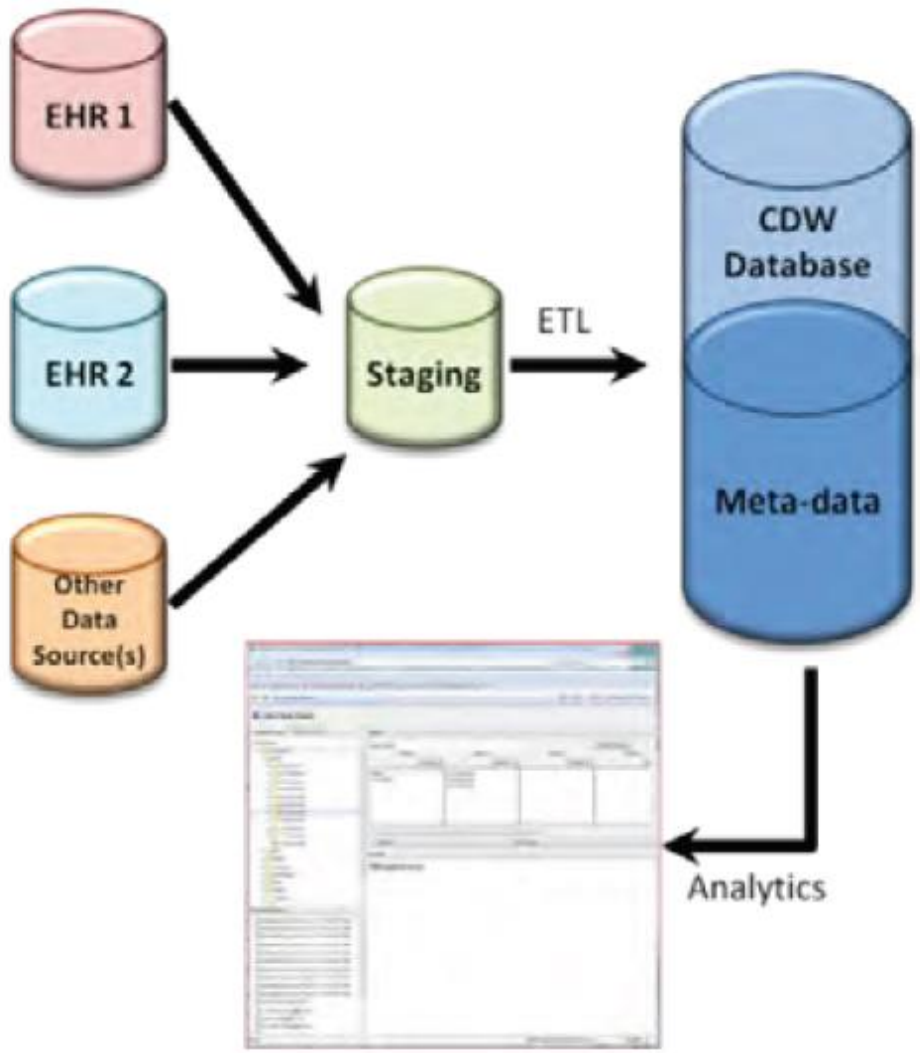
Information to Knowledge

- Two types of data:
  - Structured data
  - Unstructured (free text)
- Advantages and disadvantages
- Natural language processing (NLP)



Clinical Data Warehouse

- A modern way to convert medical information to knowledge is to use a clinical data warehouse (CDW).
- A clinical data warehouse is a database system that collects, integrates and stores clinical data from a variety of sources including electronic health records, radiology and other information systems.
- Differences between EHR and CDW.



Clinical Data Warehouse

- **Uses**:

1- Monitor quality by allowing users to query for specific quality measures

2- Identify trends.

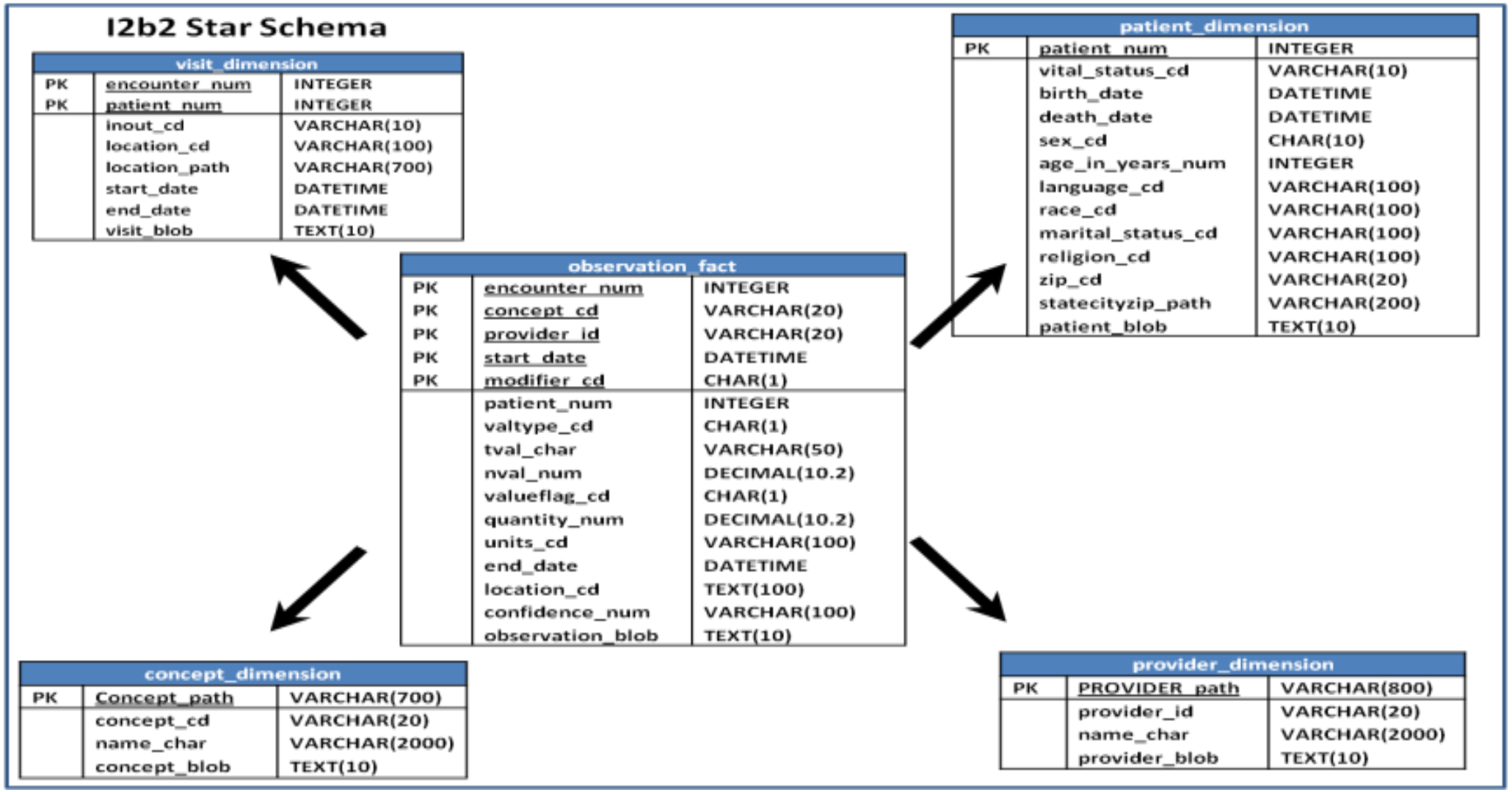
3- Comparative effectiveness research (CER)

i2b2 platform

https://www.i2b2.org

* Informatics for Integrating Biology and the Bedside (i2b2) is a Harvard project used by many other academic institutions in the US
* The program is open source and modular and incorporates genomic and clinical information for research purposes
* Database consists of **facts** (diagnoses, lab results, etc.) queried by users and **dimensions** that describe the facts

I2b2 Star Schema



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

What Makes Informatics Difficult?

- Biomedical informatics is difficult:

1- Incomplete information.

2- Uncertain information.

3- Imprecise information.

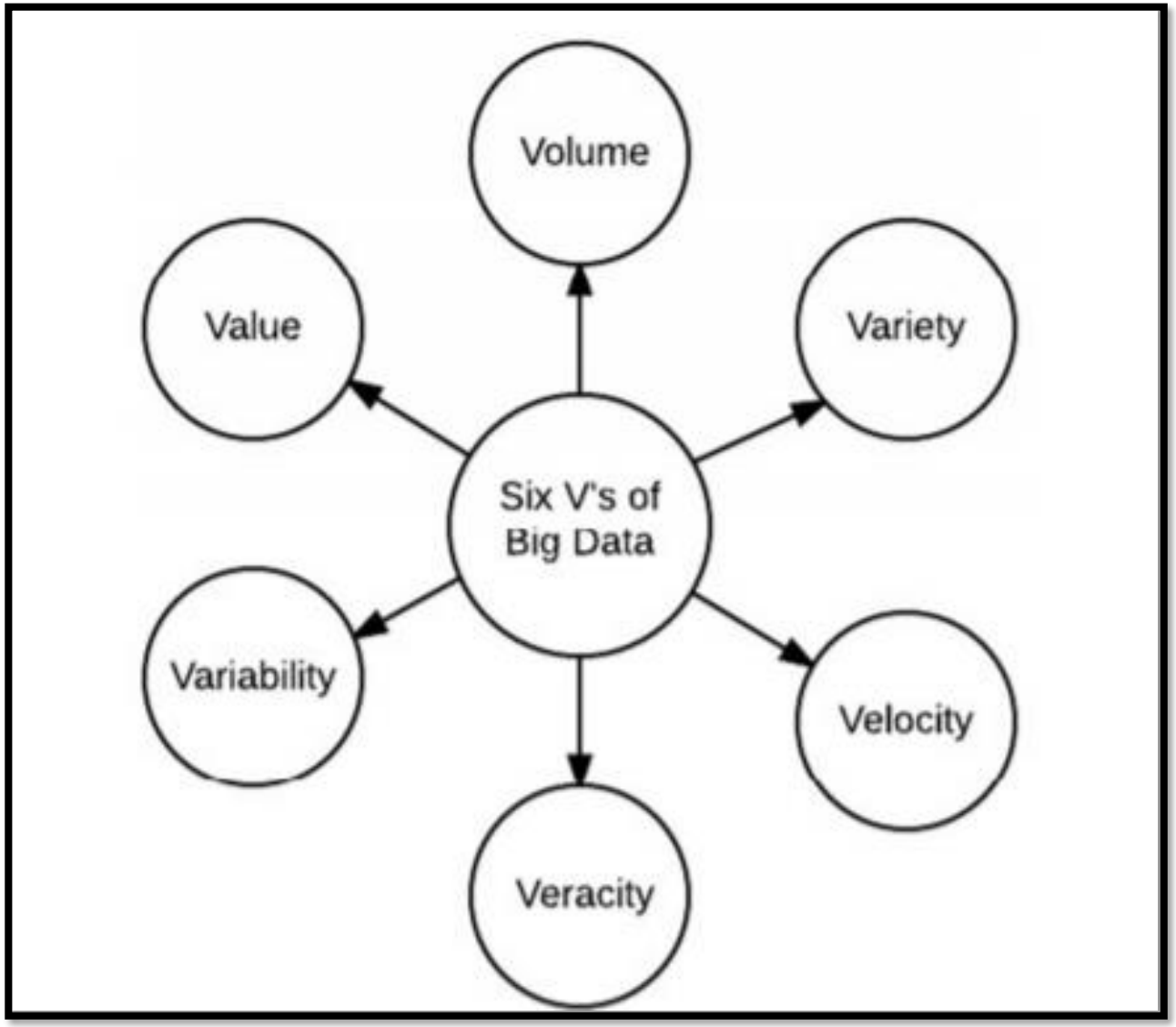
4- Vague information.

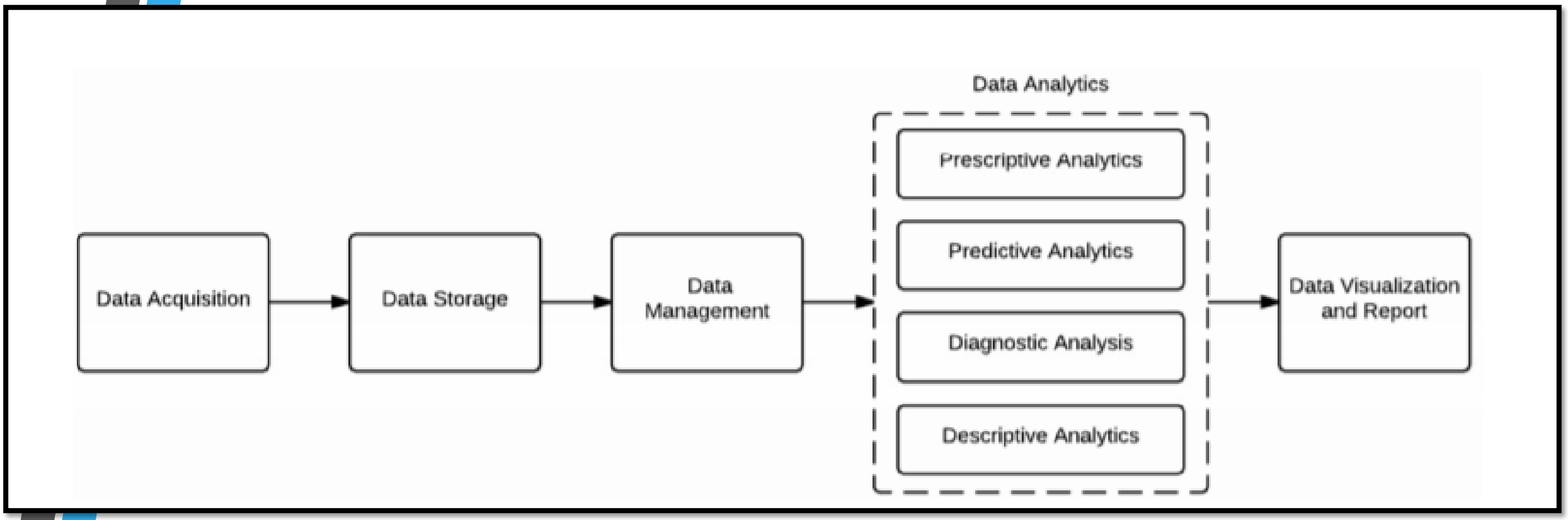
5- Inconsistent information.

e.g Birthdate: 8/29/66 and 9/17/66.

Big data:

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





Big data analysis

Conclusion

- Users must be able to "*make sense*" of clinical data; to make clinical data meaningful (data → information) and then learn from aggregated clinical data (information → knowledge).
- Computer scientists focus on data, while informaticists focus on information
- The transformation of information into knowledge is a primary goal of informaticists
- Clinical data warehouses are increasing used to research clinical questions and generate knowledge from information
- Sources of big data

Thank you for listening

Any questions

Send email : jalmulhem@ksu.edu.sa