

# Week 2: Clinical Data and Big Data

Dr. Jwahr Almulhem

Assistant Professor in Health Informatics

Medical Education Department

College Of Medicine

King Saud University

email: [jalmulhem@ksu.edu.sa](mailto: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, GIF, 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

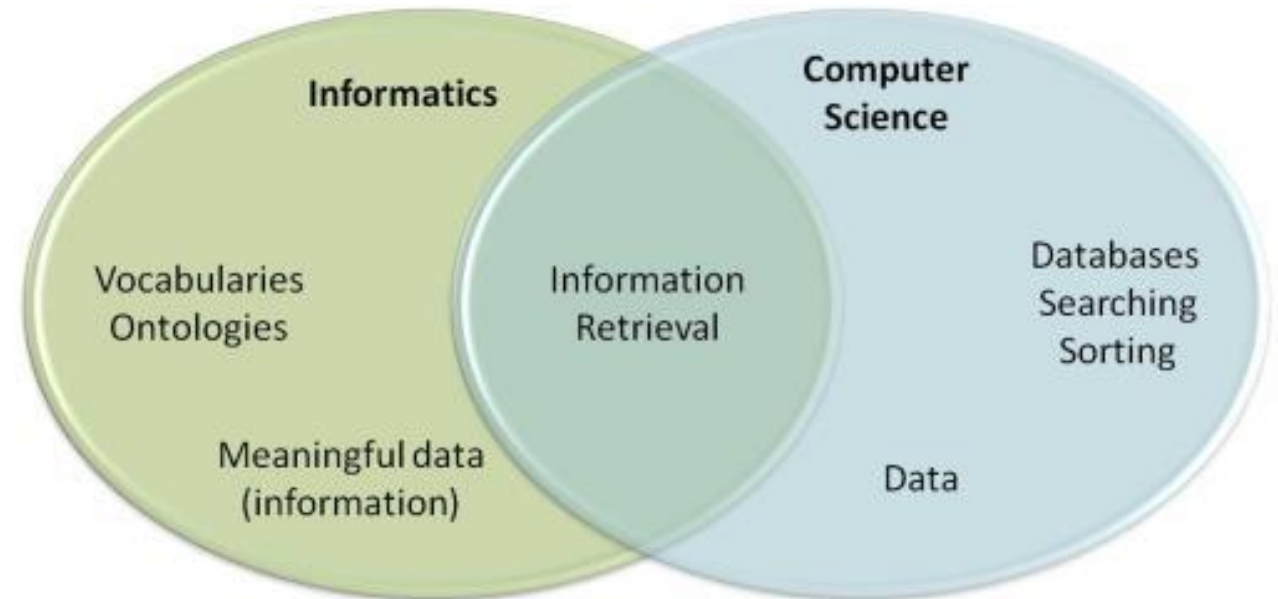
❖ Structured (menu-driven) data entry by mouse or pen

❖ 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)

Liverpool Hospital Neonatal Database - Clinical Data Entry

**Ian TEST DOB 28/2/2008 GA 26+2 BW 1070** Liverpool 2170  
**MRN 123432** Day 33 - Corrected GA 31+0 1250g on 01/04

Log Files (0) Images (2) Calculator

ATTENTION: Brain scan overdue:

**Current Status**  
**Respiratory Support**  
CPAP /5 , FIO2 29

**Fluids / Feeds**  
160 ml/kg/day  
TPN 10% Fat 3g  
14x2 EBM 24cal (134)

**Jaundice**  
09/03 SBr 135 Billblanket  
ceased 08/03

**Other**  
01/03 Mod PDA  
POSSIBLE NEC

**Treatments**  
Pentavite, Folic Acid  
Longline,

**Test Results**  
09/03 Na 136  
09/03 Hb 135  
09/03 Plat 265  
02/03 HUS JVH II  
01/04 Eyes ROP I  
Opened 01 Apr 12:27

Admissions Respiratory Nutrition Other Treatments Test Results

Admission Planning Discharge

Liverpool Hospital Admitted: 28/02/08 at 4 hours

**Admission** Age 0 Corr.GA 26 Weight 1070 76% HC 25.5 71% Length 35 56%

Date & Time 28/02/2008 16:30 Hospital Liverpool Hospital MRN 123432

Bed 01 To NICU Reason(s) for Admission Prematurity

Consultant Ian Callander Insurance Hospital Respiratory Distress

**MATERNAL HISTORY**  
Ann is a 28 year old G2 P1 (now) woman whose blood group is O positive. She was booked to deliver at Campbelltown Hospital under the care of Kaisher however delivered at Liverpool Hospital under the care of Dr Peter Hammill. She had a history of essential hypertension. This pregnancy was complicated by hypertension of pregnancy, fetal growth restriction, Bilateral Renal Pelvis dilatation 5 - 10mm, GBS +ve swab, fever, abnormal Dopplers, prolonged rupture of membranes for 2 days, clinically suspected chorioamnionitis. Ann was treated with antenatal steroids, tocolytics, and antihypertensive drugs. Following the spontaneous onset of labour, she proceeded to a vaginal delivery. Antibiotics were given before delivery.

**PERINATAL HISTORY**  
Ian was born at 13:00 hours with a birth weight of 1070 grams (76th centile). Apgars were 3 at 1 minute and 7 at 5 minutes respectively treated with intubation and ventilation. The arterial cord pH was 7.24 and the base excess -6. Ian was then retrieved to

Added to Worksheet 01/03 Orders on Worksheet 01/03

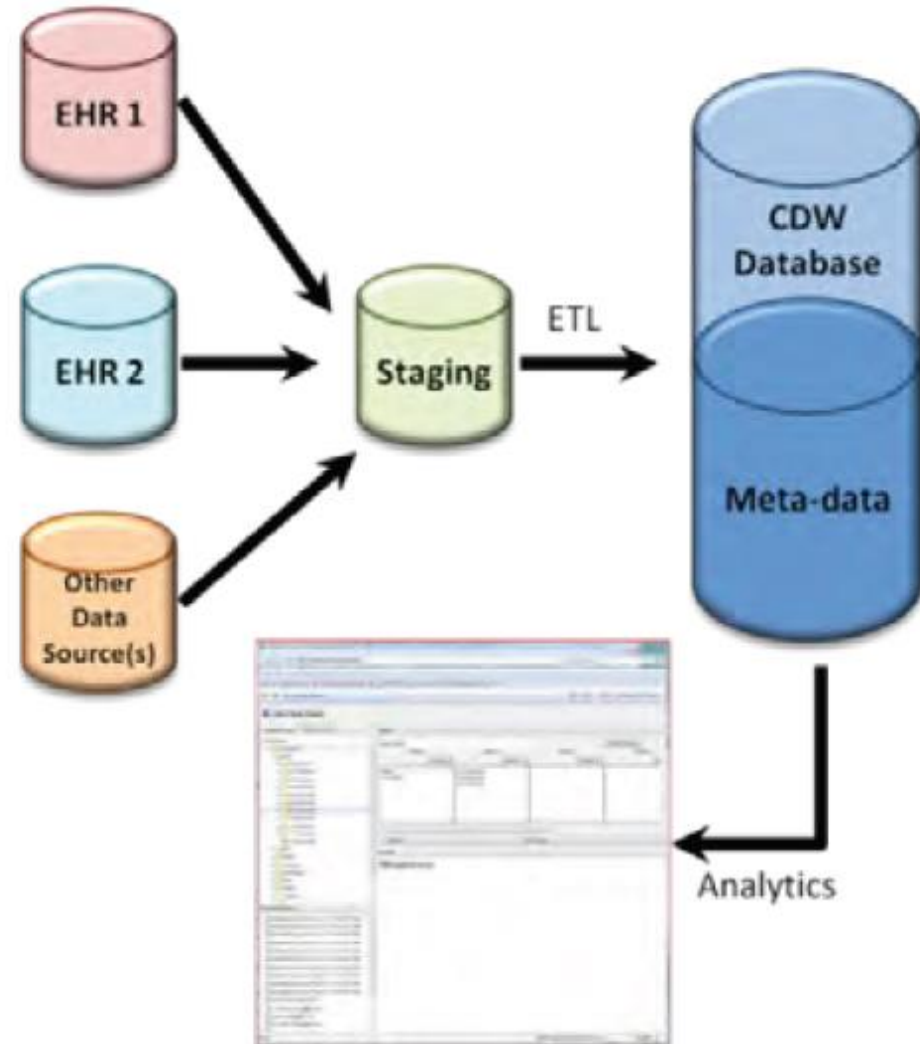
This is freetext Freetext orders (double click on text to delete)

Hospital Episodes	MRN	Admitted	Discharged		
Liverpool Hospital	123432	28 Feb 2008 16:30		Add Another Admission	
NETS	PD12345	28 Feb 2008 15:00	28 Feb 2008 16:30		Delete MRN .. then click again to Delete Episode
Campbelltown Hospital	222222	28 Feb 2008 13:00	28 Feb 2008 15:00		

Add Twin local form

# 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

visit_dimension		
PK	<u>encounter_num</u>	INTEGER
PK	<u>patient_num</u>	INTEGER
	inout_cd	VARCHAR(10)
	location_cd	VARCHAR(100)
	location_path	VARCHAR(700)
	start_date	DATETIME
	end_date	DATETIME
	visit_blob	TEXT(10)

observation_fact		
PK	<u>encounter_num</u>	INTEGER
PK	<u>concept_cd</u>	VARCHAR(20)
PK	<u>provider_id</u>	VARCHAR(20)
PK	<u>start_date</u>	DATETIME
PK	<u>modifier_cd</u>	CHAR(1)
	patient_num	INTEGER
	valtype_cd	CHAR(1)
	tval_char	VARCHAR(50)
	nval_num	DECIMAL(10.2)
	valueflag_cd	CHAR(1)
	quantity_num	DECIMAL(10.2)
	units_cd	VARCHAR(100)
	end_date	DATETIME
	location_cd	TEXT(100)
	confidence_num	VARCHAR(100)
	observation_blob	TEXT(10)

patient_dimension		
PK	<u>patient_num</u>	INTEGER
	vital_status_cd	VARCHAR(10)
	birth_date	DATETIME
	death_date	DATETIME
	sex_cd	CHAR(10)
	age_in_years_num	INTEGER
	language_cd	VARCHAR(100)
	race_cd	VARCHAR(100)
	marital_status_cd	VARCHAR(100)
	religion_cd	VARCHAR(100)
	zip_cd	VARCHAR(20)
	statecityzip_path	VARCHAR(200)
	patient_blob	TEXT(10)

concept_dimension		
PK	<u>Concept_path</u>	VARCHAR(700)
	concept_cd	VARCHAR(20)
	name_char	VARCHAR(2000)
	concept_blob	TEXT(10)

provider_dimension		
PK	<u>PROVIDER_path</u>	VARCHAR(800)
	provider_id	VARCHAR(20)
	name_char	VARCHAR(2000)
	provider_blob	TEXT(10)



# 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.

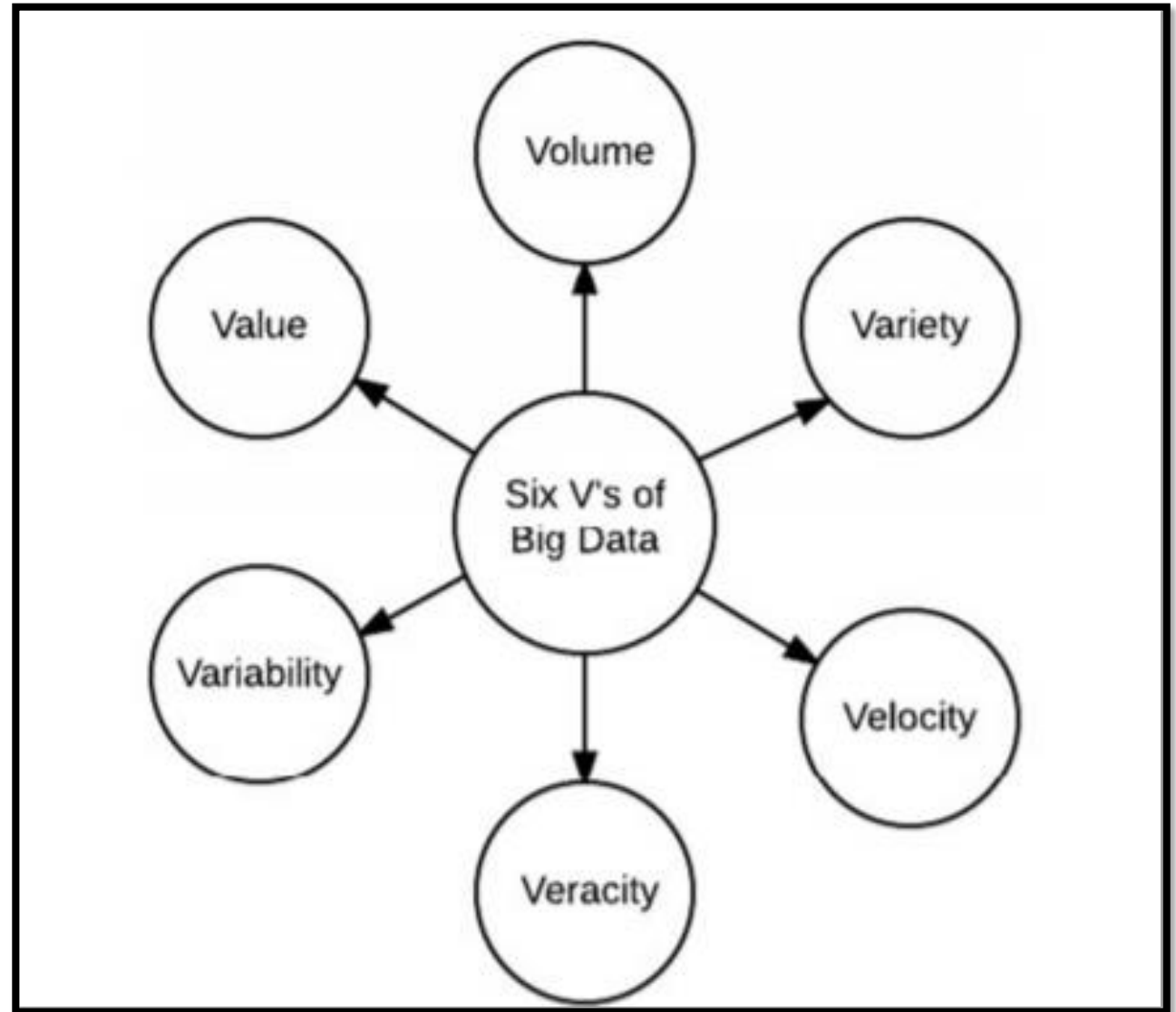
4- Vague information.

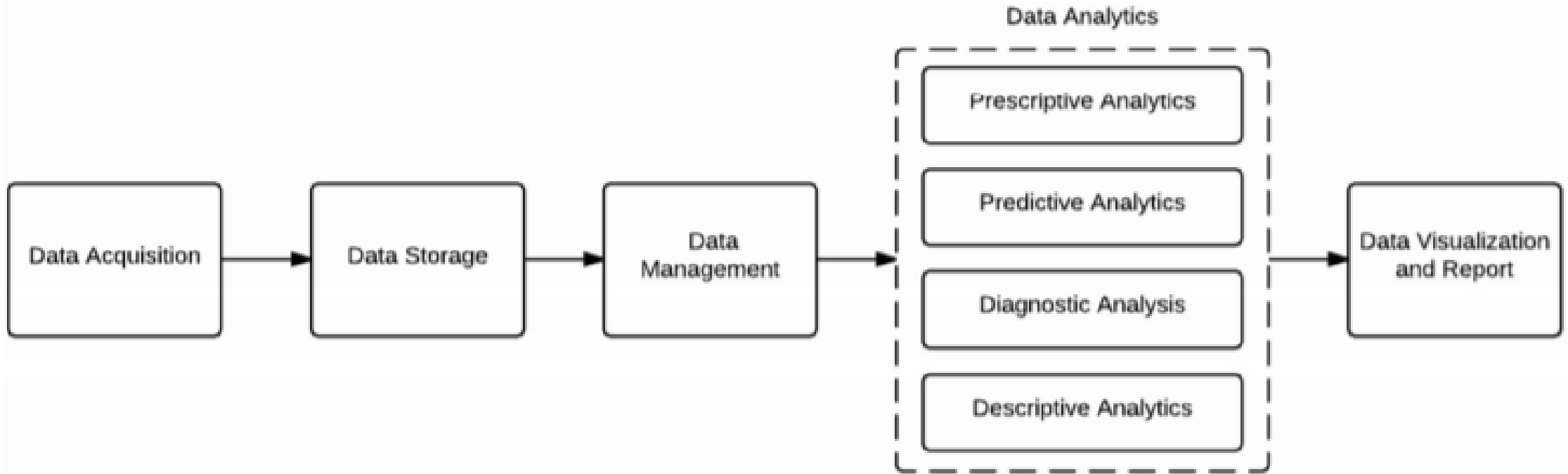
5- Inconsistent information.

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 increasingly 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](mailto:jalmulhem@ksu.edu.sa)