# EHC UNIT-1

## 1) Define Data, Information & knowledge.

Data (Raw)

↓ Processing

Informatⁿ (Meaning)

↓ Experience + Rules

knowledge Decision

It combines info+clinical experience+guideli-nes.

### Data

* Raw facts collected from different health-care activities.
* They have no meaning until processed.

Ex: - 98.7°F
- BP : 140/90 mmHg
- Sugar : 180mg/dL
- Heart rate : 110 bpm.

These values alone do not tell whether patient is healthy or sick.

### Information

* When data is organized, processed or interpreted to give meaning.

Ex: 140/90 (Patient has Hypertension stage1)
- Blood sugar = 180mg/dL (Patient is hyperglycemic)

Raw Data → Processed → Info (Compared with normal)

* helps understand patient's conditⁿ.

### knowledge

* Experience based understanding used to make decisions.

Ex: If patient has
- BP - 140/90
- Family history of heart disease
- Obesity

The doctor knows from experience that patient has high risk of heart attack, so medicalⁿ + lifestyle change are recommended.

2) How vocabularies convert data to information

A: In health care, computers store only raw data(number, code, text). By themselves, these data have no meaning.

* To make data useful, we need vocabularies, also called controlled terminologies
  Ex: ICD, SNOMED, LOINC)

* These vocabularies assign meaning to data that converting them into information.

Role of Vocabularies in conversion

Healthcare vocabularies work like a dictionary that maps data to meaning.

Function

1) standard naming
   - Converts data into understandable concepts
     Ex: ICD gives meaning to diagnosis codes

2) Consistency
   - All systems interpret data the same way

3) Interoperability
   - Sharing info b/w EHRs becomes accurate

4) Context addition
   Ex: lab code + unit + reference = meaningful result.

5) Error reduction

- Reduces misinterpretation

**How conversion Happens**

**Step 1** - Raw Data capture

Ex: * 786.05

* HGB = 9.0
* BP = 180/120

**Needed**

- Ensure all hospitals understand data the same way
- Reduce medical errors
- Enable interoperability
- Help convert meaningless data → useful informat? → clinical knowledge

**Step 2** - Apply vocabulary

* ICD-9 code 786.05 (shortness of breath)
* LOINC code → matches Hemoglobin test
* SNOMED CT → maps BP terms

**Step 3** - Add context

* Patient's age, time, symptoms

Ex: HGB = 9.0 g/dL in 32 yr old femal

RAW Data
(numbers, code)

↓ Apply controlled vocabulary

Standardized Meaning
(ICD, SNOMED, LOINC, CPT)

↓ Add context

Information
(Meaningful, understandable)

# 3) Methods that convert info - knowledge.

A: In healthcare, info refers to meaningful, organized data.

knowledge, however, is justified, true belief that helps in decision-making.

* Since raw data cannot directly become knowledge, we first convert data → info → know
* To convert info into know, several scientific and computational methods are

## 1) Data Analytics Techniques

Analytis helps identify patterns, trends and insights from clinical information.

3 main types

1) Descriptive Analytics
2) Predictive Analytis
3) Prescriptive Analytics

## 2) Machine Learning (ML)

Is a core method to transform info → know

* Learns patterns from structed & unstructured clinical data
* Identifies unknown relationships.
* Improves with more data.

Ex: ML model learns from EHR information → predicts which patients likely develop kidney failure

3) Data Mining

* A sub method of ML
* Discovers hidden pattern in large datasets.
* Finds correlations that humans cannot easily use.

4) Text Mining

* Healthcare has large amount of unstructured txt (doctor notes, reports)
* Text mining extracts important information & turns it into knowledge

5) Clinical Data Warehouses (CDWs)

* CDWs store large-scale clinical info and allow researchers to analyze groups of patients.

How
* Combine EHR, lab, radiology data
* Allow population-based studies
* Support clinical research & discovery of new insights.

6) Statistical Interference

Statistical tests help convert clinical info into scientifically valid knowledge.

Methods used

- Hypothesis testing
- Regression analysis
- Probability models

7) Natural Language Processing (NLP)

* NLP helps convert textual infor into computable concepts

Work:
* Extract medical concepts
* Normalizes terminology
* Helps identify diagnosis pattern

8) Knowledge Representat$^{n}$ Models

* Clinical info is transformed into knowledg using Ontologies (SNOMED CT), Decision tree, Clinical guidelines, Rule-based systems

9) Research & Evidence Based Methods

* Clinical trials
* Observational studies
* Meta-analysis
* Evidence-based practice.

These convert collected info into validated medical knowlege

(4) Distinguish informatics from computer Science and other computational Disciplines.

A: Informatics & CS both deals with computers & data, but they differ in purpose, focus, methods, and application areas.

*

Screenshot

Ex:

Computer Science : Builds a database system to store patient lab results.

Informatics : Analyzes those lab results to identify disease risks and create clinical alerts.

(5)

Screenshot

* A hospital's data-centric system stores patient BP readings (Ex: 180/120)

* An information-centric system analyzes those reading and gives a clinical alert.

Scanned with OKEN Scanner

6 Screenshot

A hospital wants to reduce patient readmission

1) Descriptive Analytics

Shows last yer's data → 12% of patients were readmitted.

2) Predictive Analytics

Predicts who is likely to be readmitted in next 30 days

3) Prescriptive Analytics

Recommends actions → Provide extra home follow-up for high-risk patients.

## ✅ Table Format (Perfect for 10 Marks)

| Point of Difference | Informatics | Computer Science |
|---|---|---|
| **1. Definition** | Study of data, information, and knowledge in a specific domain (e.g., healthcare). | Study of computation, algorithms, hardware, and software systems. |
| **2. Goal** | Convert **data → information → knowledge** to support decision-making. | Design efficient **algorithms, programs, and systems** for computation. |
| **3. Focus** | Focuses on **meaning, context, and interpretation** of data. | Focuses on **processing, structure, and computation** of data. |
| **4. Nature of Work** | Applied, domain-based problem solving (health, biology, business). | Technical, mathematical, and system-level problem solving. |
| **5. Domain Dependency** | Strongly **domain-specific** (e.g., health informatics, bioinformatics). | Mostly **domain-independent**; principles apply to all computing fields. |
| **6. Problems Addressed** | Reducing medical errors, improving EHRs, analyzing clinical data. | Designing OS, networks, compilers, algorithms, database engines. |
| **7. Skills Required** | Domain knowledge + data analysis + communication + statistics. | Programming, mathematics, algorithm design, system development. |
| **8. Data Meaning** | Concerned with **meaning** (semantics) of data. | Concerned with **structure** and **processing** of data. |
| **9. Human Interaction** | Strong focus on **user needs, workflow, usability**. | Limited focus on human factors; more technical. |
| **10. Output** | Produces **knowledge**, insights, and decision support. | Produces **software, algorithms, and computing systems**. |

| Point of Difference | Data-Centric Technology | Information-Centric Technology |
|---|---|---|
| **1. Focus** | Focuses on **storing, managing, and processing raw data.** | Focuses on **understanding, meaning, and use of information.** |
| **2. Purpose** | Ensures data is **collected, saved, and retrieved** efficiently. | Ensures information is **interpreted, shared, and used** effectively for decision-making. |
| **3. Meaning** | Does **not consider meaning** of data (no context). | Considers **context, meaning, and relevance** of information. |
| **4. Main Activities** | Data entry, storage, formats, databases. | Information analysis, interpretation, representation, knowledge creation. |
| **5. User Interaction** | Used mostly by **technical staff** (IT, database admins). | Used by **domain experts** (doctors, managers, analysts). |
| **6. Output** | Produces **raw data** (numbers, codes, files). | Produces **useful information** (reports, insights, alerts). |
| **7. Technologies Used** | Databases, data warehouses, storage systems. | Decision support systems, analytics, dashboards. |
| **8. Data Flow** | "Data moves between systems." Raw, unprocessed. | "Information moves between people and systems." Processed and meaningful. |
| **9. Goal** | Efficiency in **data handling** (speed, storage). | Improvement in **decisions, workflow, knowledge.** |
| **10. Example** | A lab database storing "HGB = 9.0". | The system showing "HGB = 9.0 → Anemia risk (High Priority)". |

| Type of Analytics | Descriptive Analytics | Predictive Analytics | Prescriptive Analytics |
|---|---|---|---|
| **1. Purpose** | Describes **what has happened** in the past. | Predicts **what is likely to happen** in the future. | Recommends **what should be done** to get the best outcome. |
| **2. Focus** | Summaries, trends, reports. | Forecasting, risk prediction. | Decision-making, optimization. |
| **3. Time Orientation** | Past and present. | Future. | Future actions + outcomes. |
| **4. Techniques Used** | Reporting, dashboards, data aggregation. | Machine learning, statistics, regression models. | Optimization models, simulations, "what-if" analysis. |
| **5. Question Answered** | **"What happened?"** | **"What will happen?"** | **"What should we do?"** |
| **6. Data Used** | Historical data. | Historical + real-time data. | Predictions + constraints + business rules. |
| **7. Output** | Charts, tables, summary reports. | Risk scores, probability values, forecasts. | Action plans, recommendations, optimized decisions. |
| **8. User** | Managers and analysts who need reports. | Data scientists predicting outcomes. | Decision-makers choosing best actions. |
| **9. Complexity** | Low | Medium | High |
| **10. Example** | Showing last year's hospital infection rates. | Predicting which patients may be readmitted in 30 days. | Recommending the best treatment plan to reduce readmission risk. |

## 7) Characteristics of Big Data

Big Data refers to extremely large & complex datasets that cannot be handled by traditional data-processing tools.

Char

1) Volume

* Refers to huge amount of data generated every second.

Ex: Hospitals generate large volumes of EHRs, lab reports, images, and sensor data.

2) Velocity

* Speed at which data is generated, collected and processed.

Ex: ICU monitors & wearable devices send patient data in real time

3) Variety

* Data comes in many different types & formats
  - Structured (lab values, BP readings)
  - Unstructured (doctor notes, images, videos)
  - Semi-structed (JSON, XML)

4) Veracity

* Refers to trustworthiness, accuracy & quality of data

Ex: Incomplete patient records / incorrect entires affect analysis.

5) Value

* The usefulness of data in providing meaningful insights

6) Variabitity

* Data meaning & format frequently chang

7) Visualization

* Presenting complex data as charts, dashboards, & graphs for decision-making.

8) Validity

* Ensuring data is correct, releavant, & appropriate for analysis.

Scanned with OKEN Scanner

## 1. Programming Skills

- Knowledge of data-oriented programming languages such as Python, R, SQL.
- Used for data cleaning, analysis, visualization, and model building.

## 2. Statistical Knowledge

- Understanding of probability, statistical tests, regression, hypothesis testing.
- Essential for making valid conclusions from data.

## 3. Data Management Skills

- Ability to handle large datasets.
- Knowledge of databases, data warehouses, ETL (Extract-Transform-Load) processes.

## 4. Machine Learning and Modeling Skills

- Ability to apply ML algorithms such as classification, clustering, and prediction models.
- Helps in building predictive and prescriptive analytics solutions.

## 5. Data Visualization Skills

- Using tools like Tableau, Power BI, Matplotlib to present insights clearly.
- Helps decision-makers understand complex results easily.

## 6. Domain Knowledge

- Understanding of the specific field (e.g., healthcare, finance, marketing).
- Helps interpret data correctly and solve real-world problems.

## 7. Critical Thinking and Problem-Solving

- Ability to identify patterns, detect errors, and make logical decisions.
- Crucial for converting information into knowledge.

8

## 8. Communication Skills

- Ability to explain findings to non-technical people (managers, doctors, clients).
- Includes report writing and presentation skills.

## 9. Data Cleaning and Preprocessing Skills

- Ability to handle missing data, inconsistencies, and errors.
- Ensures high-quality data for meaningful analysis.

## 10. Understanding of Privacy and Ethics

- Knowledge of data security, confidentiality, and responsible data use.
- Especially important in healthcare and finance.

## Discuss the Critical Role Electronic Health Records (EHRs) Play in Healthcare Data Analytics

Electronic Health Records (EHRs) play a foundational role in healthcare data analytics because they are the **main source of clinical data.**

According to the PDF, **clinical data are collected via electronic health records (EHRs)** and then used for further analysis and knowledge generation. unit-1-18.6.25HEALTH PPT-IT (1)

---

### 1. Primary Source of Clinical Data

The PDF states that **clinical data are collected via EHRs** and these records contain both **structured data (lab results, billing codes)** and **unstructured data (free-text doctor notes).** unit-1-18.6.25HEALTH PPT-IT (1)

This makes EHRs the main input for analytics systems.

---

### 2. Basis for Clinical Data Warehouses (CDWs)

The document explains that data from EHRs are copied into **staging databases**, cleaned, and stored in **Clinical Data Warehouses (CDWs)** for analysis. unit-1-18.6.25HEALTH PPT-IT (1)

Thus, EHRs serve as the **foundation** for all large-scale healthcare analytics.

---

### 3. Support Group-Level Analysis

EHRs focus on **individual patient care**, but analytics requires **group-level** insights (e.g., "average age of lung cancer patients").

CDWs built from EHR data allow these queries. unit-1-18.6.25HEALTH PPT-IT (1)

---

### 4. Enable Knowledge Generation

The PDF states that we need to **transform information into knowledge**, and CDWs (fed by EHR data) are essential for this. unit-1-18.6.25HEALTH PPT-IT (1)

Researchers use this data to find:

- disease patterns
- treatment outcomes
- public health trends

---

### 5. Not Designed for Real-Time Analytics but Vital for Historical Analysis

EHRs update in **real time**, while CDWs are usually **not real-time**, making them suitable for retrospective analytics. unit-1-18.6.25HEALTH PPT-IT (1)

This setup ensures accurate historical data analysis.

# Discuss the Critical Role Electronic Health Records (EHRs) Play in Healthcare Data Analytics

Electronic Health Records (EHRs) play a foundational role in healthcare data analytics because they are the **main source of clinical data.**
According to the PDF, **clinical data are collected via electronic health records (EHRs)** and then used for further analysis and knowledge generation. unit-1-18.6.25HEALTH PPT-IT (1)

---

## 1. Primary Source of Clinical Data

The PDF states that **clinical data are collected via EHRs** and these records contain both **structured data (lab results, billing codes)** and **unstructured data (free-text doctor notes).** unit-1-18.6.25HEALTH PPT-IT (1)
This makes EHRs the main input for analytics systems.

---

## 2. Basis for Clinical Data Warehouses (CDWs)

The document explains that data from EHRs are copied into **staging databases**, cleaned, and stored in **Clinical Data Warehouses (CDWs)** for analysis. unit-1-18.6.25HEALTH PPT-IT (1)
Thus, EHRs serve as the **foundation** for all large-scale healthcare analytics.

---

## 3. Support Group-Level Analysis

EHRs focus on **individual patient care**, but analytics requires **group-level** insights (e.g., "average age of lung cancer patients").
CDWs built from EHR data allow these queries. unit-1-18.6.25HEALTH PPT-IT (1)

---

## 4. Enable Knowledge Generation

The PDF states that we need to **transform information into knowledge**, and CDWs (fed by EHR data) are essential for this. unit-1-18.6.25HEALTH PPT-IT (1)
Researchers use this data to find:

- disease patterns
- treatment outcomes
- public health trends

---

## 5. Not Designed for Real-Time Analytics but Vital for Historical Analysis

EHRs update in **real time**, while CDWs are usually **not real-time**, making them suitable for retrospective analytics. unit-1-18.6.25HEALTH PPT-IT (1)
This setup ensures accurate historical data analysis.

### 6. High-Quality Metadata Support

EHR data are loaded into CDWs along with **metadata**, such as ICD codes, which give meaning to the stored information. unit-1-18.6.25HEALTH PPT-IT (1)
Metadata helps analytics tools interpret data correctly.

---

### 7. Improve Clinical and Translational Research

The PDF states that CDWs built from EHR data are used by **clinical and translational researchers** to identify trends and generate new knowledge in medicine. unit-1-18.6.25HEALTH PPT-IT (1)

---

### 8. Helps in Turning Data → Information → Knowledge

The document emphasizes that a core goal of science is transforming meaningful information into knowledge.
Since EHRs provide the **raw clinical information**, they are critical for achieving this transformation. unit-1-18.6.25HEALTH PPT-IT (1)

---

### 9. Support Analytics Pipeline

The analytics pipeline described in the PDF begins with **input data sources**, including EHRs. unit-1-18.6.25HEALTH PPT-IT (1)

They provide the raw input needed for:

- feature extraction
- statistical processing
- prediction generation

---

### 10. Essential for Modern Healthcare IT Systems

As the PDF highlights, healthcare data is complex and often imperfect (incomplete, uncertain, imprecise).
EHRs help standardize data collection, improving data quality for analytics tasks. unit-1-18.6.25HEALTH PPT-IT (1)