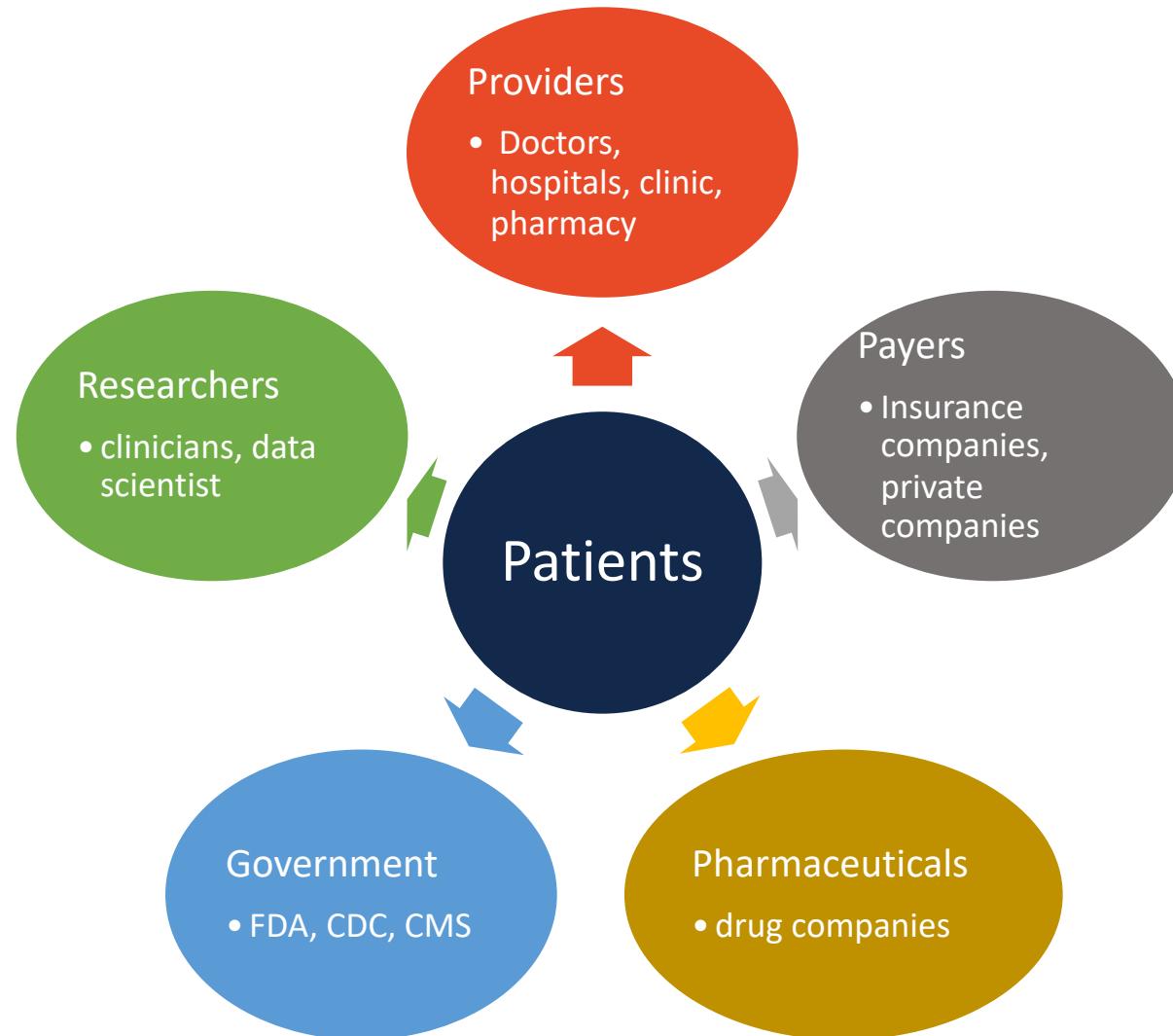


Deep Learning for Healthcare



2.a Health data

Players in healthcare



Life cycle of healthcare data



Pharmacy



Hospital



Insurance co.



Patient

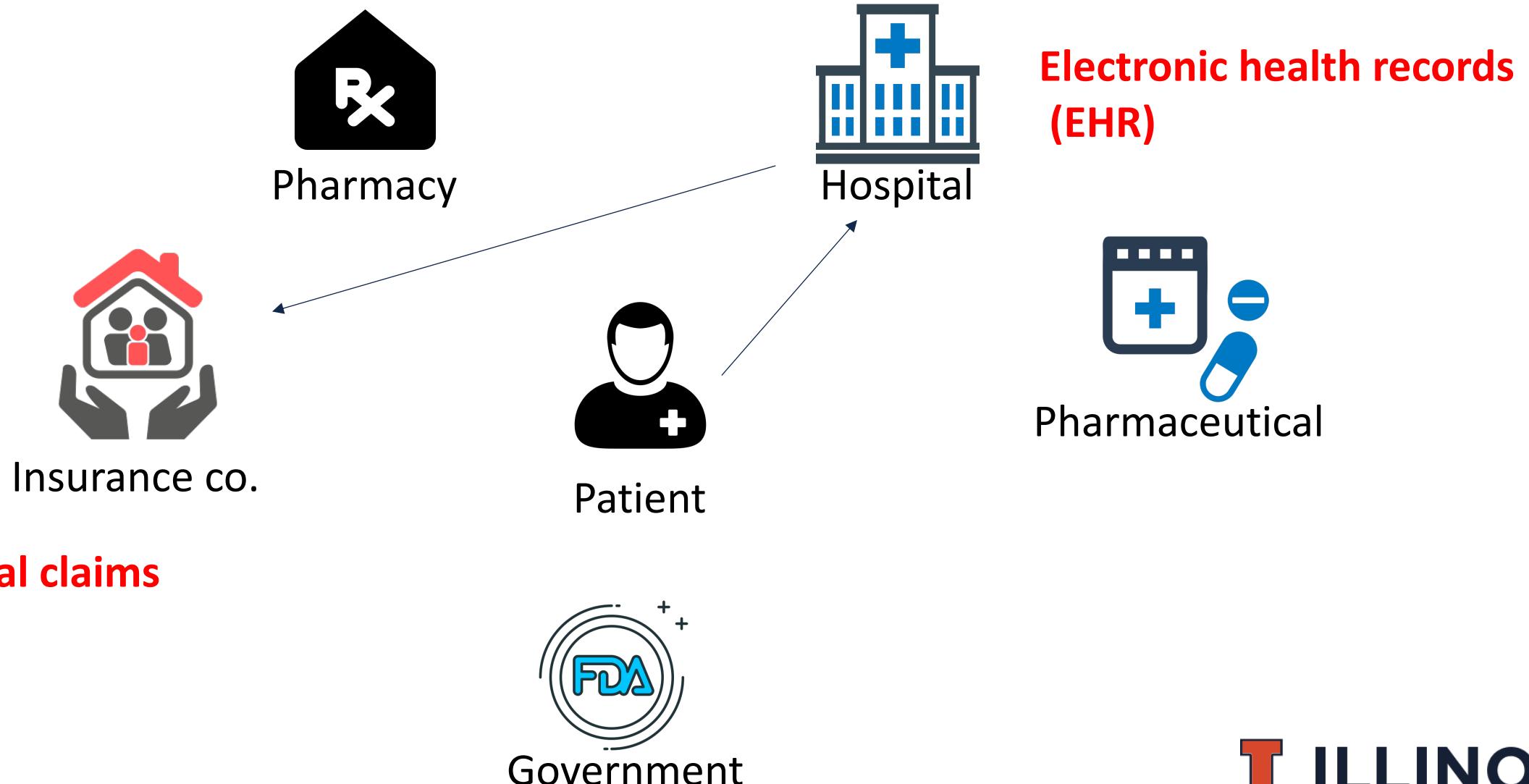


Pharmaceutical



Government

Life cycle of healthcare data



Life cycle of healthcare data



Pharmacy claims



Pharmacy



Insurance co.



Patient



Hospital



Pharmaceutical



Government

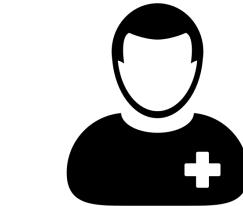
Life cycle of healthcare data



Pharmacy



Insurance co.



Patient



Hospital

Clinical trials

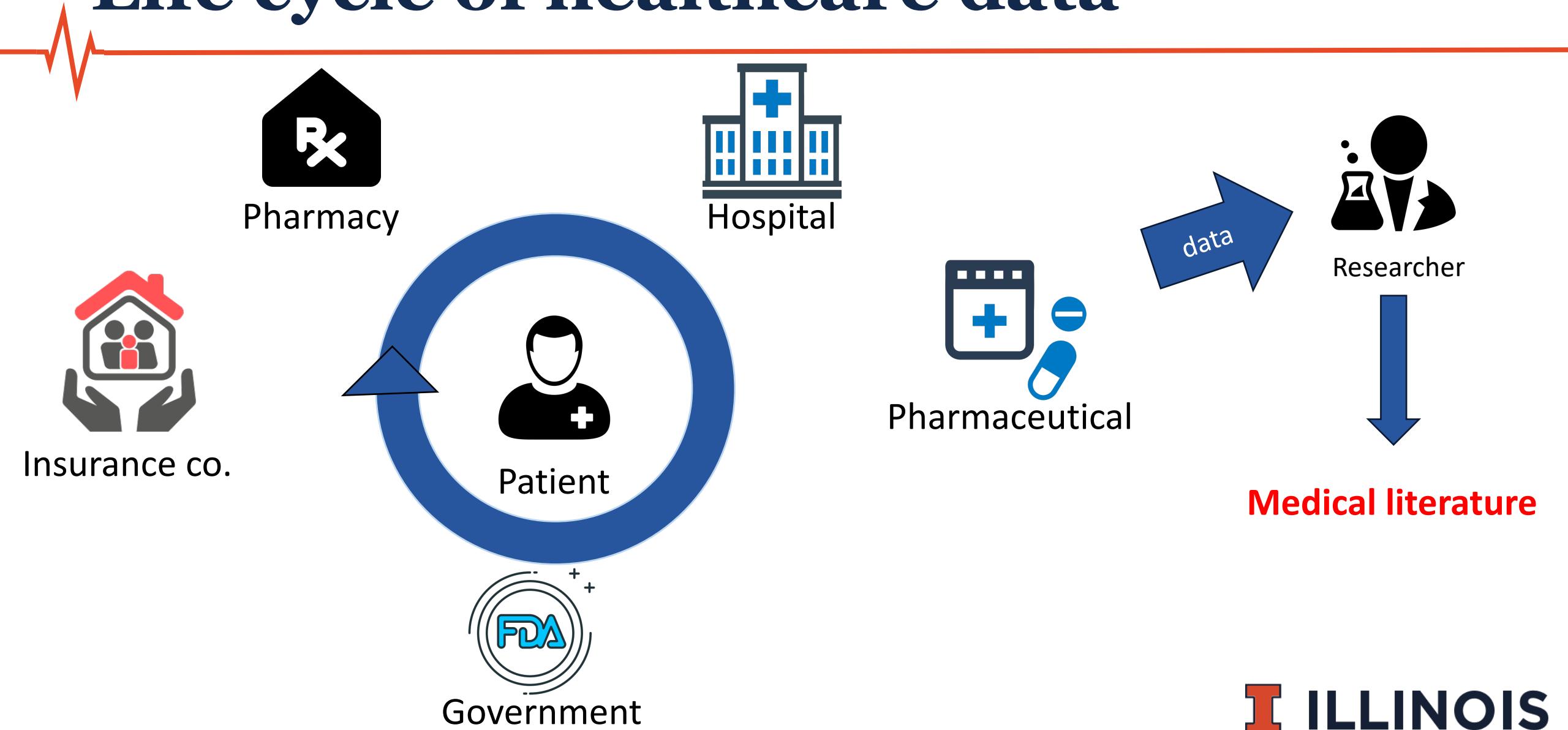


Pharmaceutical



Government

Life cycle of healthcare data

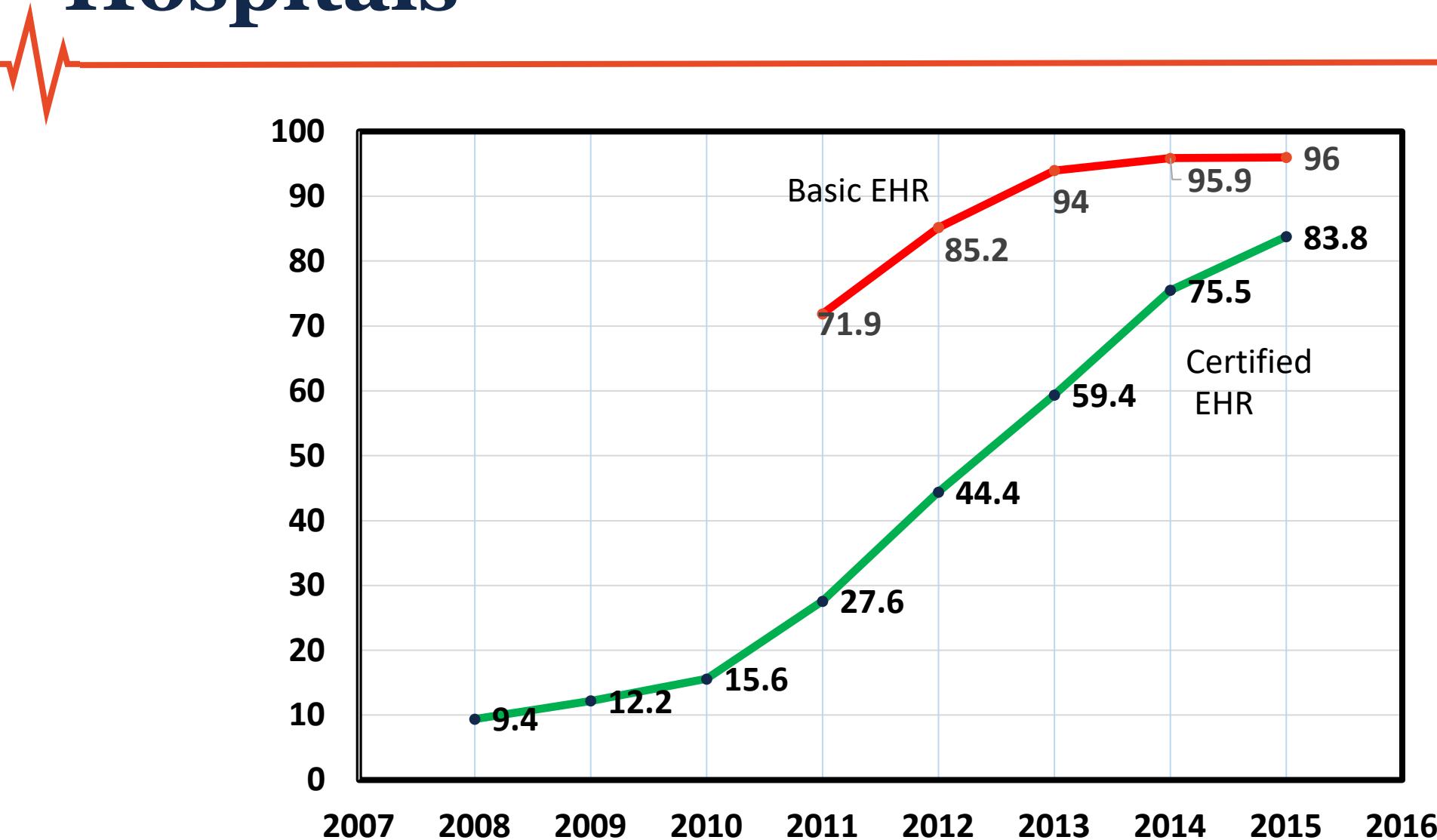


Health data

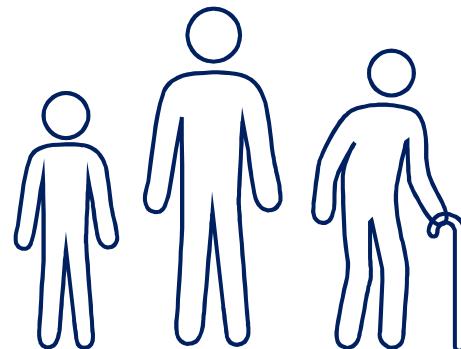


- Electronic health records
- Medical claims
- Clinical notes
- Medical literature
- Continuous signals
- Imaging data
- Medical ontology
- Clinical trial data
- Drug discovery data

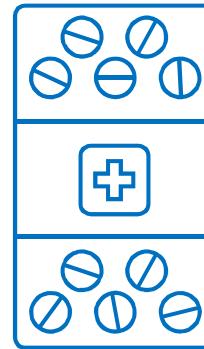
Adoption of EHR Systems among U.S. Hospitals



What is in Electronic health record (EHR)



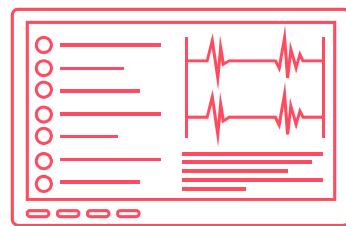
Demographics



Medications



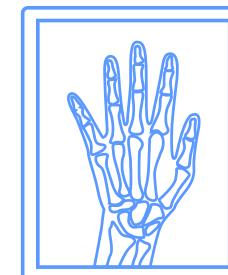
Clinical Notes
and Reports



Continuous
Monitoring Data

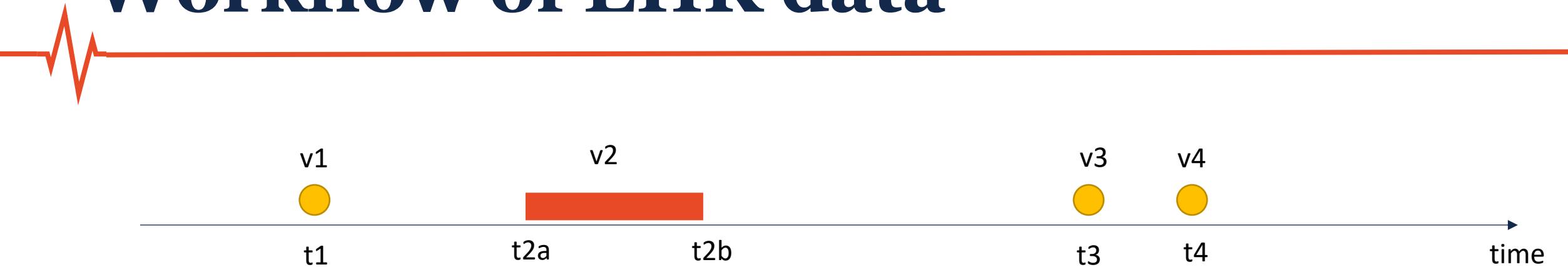


Multi-typed
Medical Codes



Medical
Images

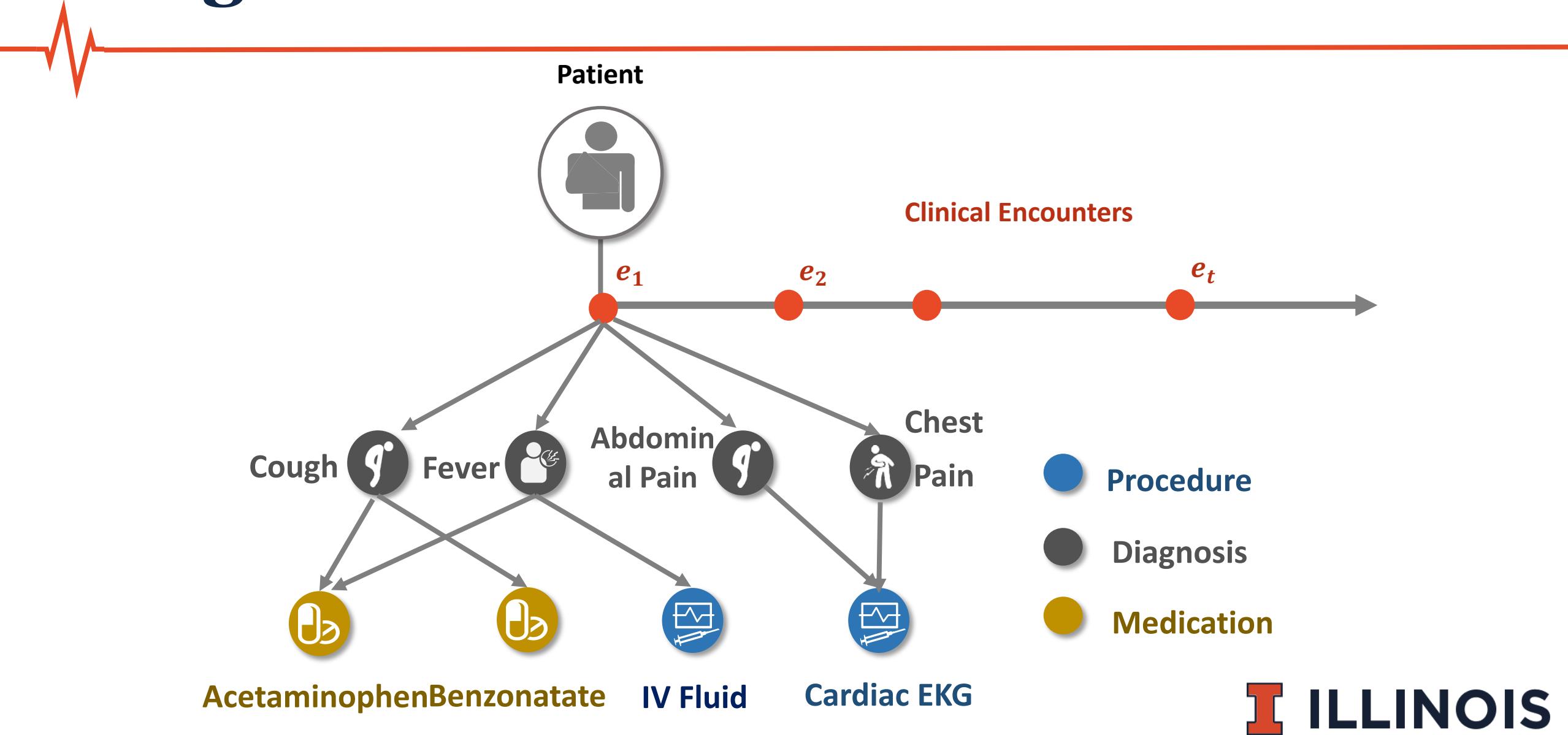
Workflow of EHR data



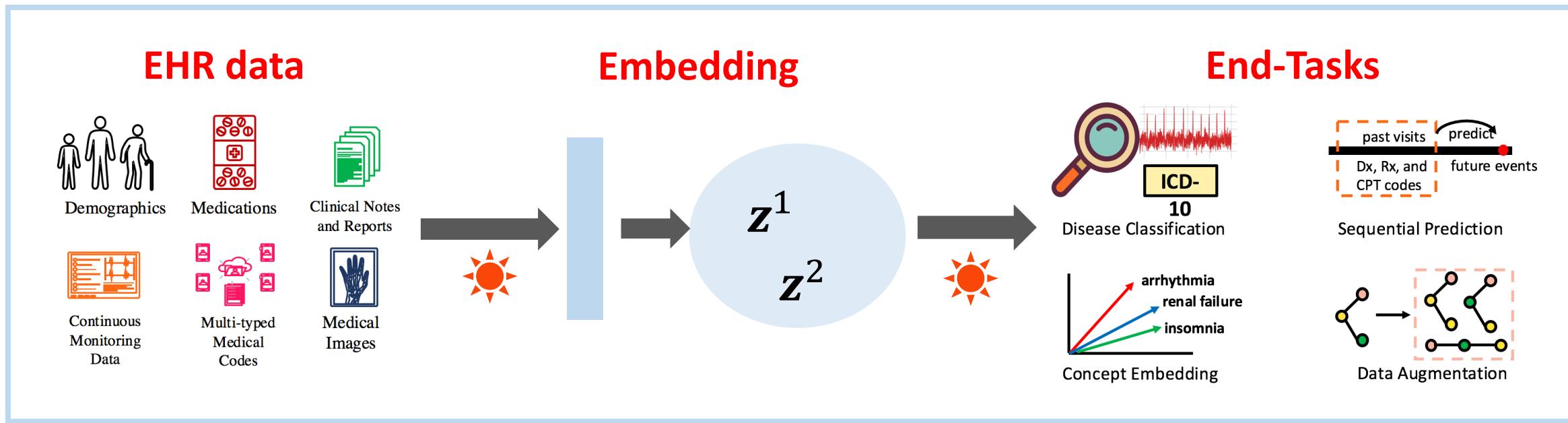
- Outpatient visit/encounter 
 - Symptom, diagnosis, lab tests, medication prescriptions
- Inpatient visit/encounter 
 - Admission: chief complaints
 - Continuous monitoring data, imaging data
 - Discharge: discharge summary

(c)

Longitudinal EHR data



Deep Learning as Effective Tools



 Deep learning models are effective tools for all phases of health data modeling.

Properties of EHR



Pros



Cons

- **Rich data types:** structured + unstructured data
- **Longitudinal:** multiple visits over time

- **Complex:** multi-modality data
- **Siloed data:** each hospital has its own EHR
- **Privacy concern:** difficult to share/access



Clinical notes

- Free text documentation of patient status
- Various types of notes
 - Admission
 - ED notes
 - Progress
 - Nursing
 - Radiology
 - ECG, Echocardiogram
 - Physician
 - Discharge summary
 - Respiratory
 - Social work
 - ...

Clinical notes - template



- Progress note (**SOAP**)
- **Subjective:** what patient tells you
- **Objective:** exam findings, labs
- **Assessment:** diagnosis
- **Plan:** treatment

Properties of clinical notes



Pros

- **Detailed:** describes various info about patient status
- **Universal:** clinical notes exist in all clinical encounters
- **Flexible:** various information has been documented – symptoms, lab test



Cons

- **Unstructured:** free text narrative can be difficult to process
- **Noisy:** low documentation quality - acronym, typo
- **Sensitive:** free text often contain protected health info (PHI)

Claims Data

- Electronic transaction data between patients and healthcare providers.
- Differ based on insurance types:
 - Medicare: national insurance program in the US
 - Commercial/private insurance

Claims Data - Medicare



- Part A: Hospital insurance → inpatient visit
- Part B: Medical insurance → physician service, outpatient visit, ambulance visits
- Part C: Medicare advantage plan → capitation plan
- Part D: Prescription drug

Claims Data – Commercial/private



- Medical
 - Inpatient: hospital visit, nursing facility, nursing home, hospice
 - Outpatient: Ambulatory visit, outpatient centers, facility fee
 - Professional: Doctor fee
- Pharmacy
 - Prescription drug

Claims data – data elements

- Member info, date, place
- Diagnosis: ICD
- Procedure: HCPCS (CPT, other)
- Drug: NDC
- Financial information
 - Charge, payment, expenditure (total spending)

Properties of Claims



Pros

- **Large volume:** plenty of claims data are available
- **Holistic view:** all patient's interactions with providers
- **Medication compliance:** Every fill/refill of a prescription

Cons

- **Coding errors:** Billing purpose only not detailed clinical documentations
- **Inaccurate data:** imperfect reflection of actual patient status
- **Time lag:** claims filing delay
- **Temporally limited:** intrinsic point events, hard to track individual over time

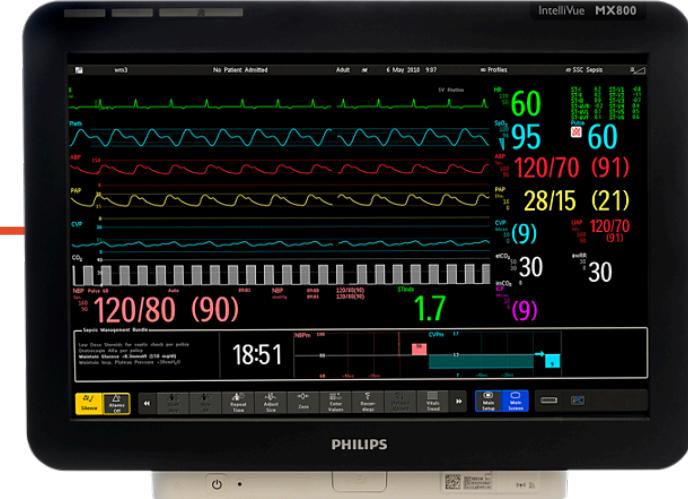
Claims vs. EHR



	Claims data	EHR data
Scope of data	All transactions of patient provider interaction	Capture only the portion of care provided by a specific provider (hospital/clinic)
Medication data	Medication filled	Medication prescription, not whether it is filled
Data richness	Limited: structured codes – ICD, CPT, NDC	Rich: clinical notes, lab, vital, problem list, social history, imaging

Continuous signals

- Continuous signals are commonly collected in hospitals
 - Important data for real-time clinical decision making



ECG monitoring



Oxygen saturation (SpO₂)



Heart rate monitoring



Blood pressure



EEG

Properties of Continuous signals



Pros

Detailed data: continuous monitoring data

Objective measures: no human opinion or variations



Cons

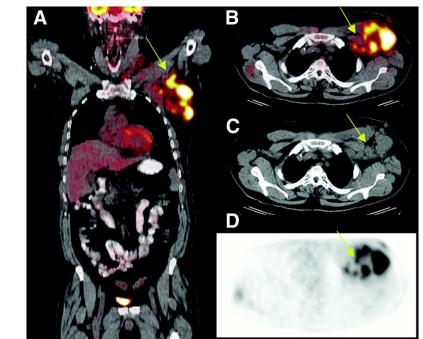
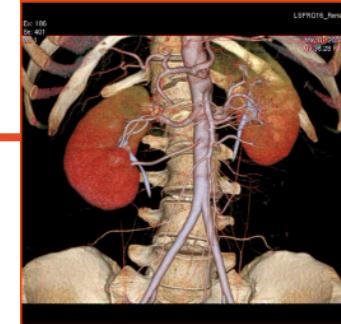
Noisy: sensor placement, interference

Large volume: vast amount raw data need to be processed in real-time

Imaging data



- Many different imaging data are available
 - X-ray
 - Computed tomography (CT)
 - magnetic resonance imaging (MRI)
 - positron emission tomography (PET-CT)
- Storage size per image
 - Full body PET 9GB
 - CT Cardiac 36GB
 - fMRI 300GB
- Size estimates
 - Imaging data in US 2014: 100 PB

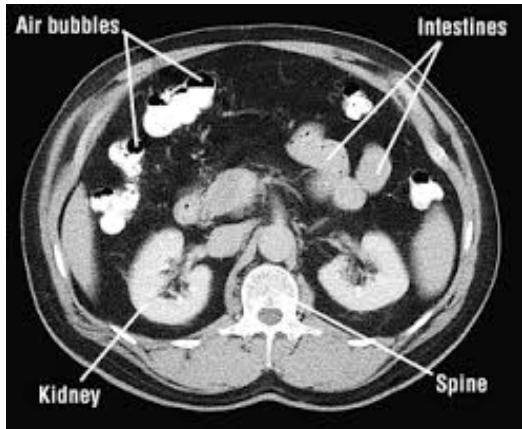


Quiz:



- Which types of imaging are they?

- a) X-ray
- b) Computed tomography (CT)
- c) magnetic resonance imaging
- d) positron emission tomography (PET-CT)



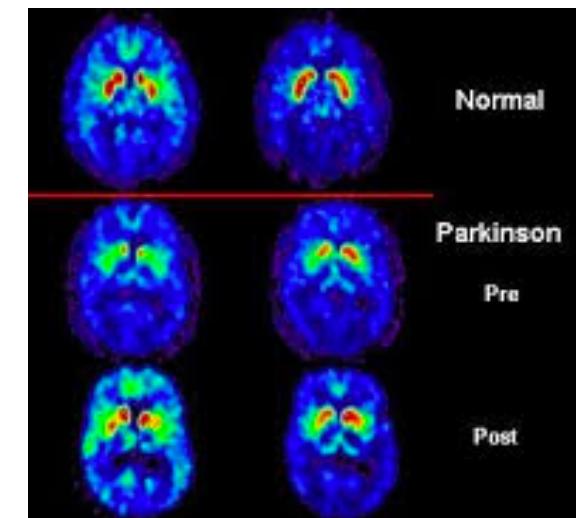
(b)



(c)



(a)



(d)

Properties of Imaging data



Pros

- **Objective measures:** no human opinion
- **Data standard:** ensure generalizability of trained models
- **Detailed:** high resolution



Cons

- **Label insufficiency:** difficult to acquire high quality labels from experts
- **High dimensional:** high resolution imaging data are difficult to analyze

Medical literature



- Medical knowledge are documented in medical literature, such as publications and clinical guidelines
 - Text data



- Medical literature search engine
- Medline/PubMed data – medical publication abstract, author, journal, time



- Clinical guideline database

Properties of Medical literature



Pros



Cons

High quality: well written articles or documents

Comprehensive: covers diverse conditions

- **Difficult to parse:** necessary expertise to locate and understand the relevant content
- **Not machine friendly:** designed for human consumption which limited structured data and can be hard for machine to consume



Medical ontology

- Medical ontology are knowledge graphs of different medical terminologies such as diseases, symptoms and treatments
- Popular medical ontologies include
 - CPT for procedure codes
 - RxNORM for drug
 - SNOMED CT for general clinical terms
 - MESH for medical subject headings for literatures
 - ATC for drug

Properties of medical ontology



Pros



Cons

Machine readable: directed acyclic graphs

Easy to integrate: it is easy to integrate ontology with other data sources such as EHR or claims

Limited coverage: The labor intensive effort to create

Noise: The error is difficult to control as the usage of ontology is often limited

Easily out of date: Due to labor intensive effort to maintain

Clinical trial data

- Clinical trial is the process to test the efficacy and safety of a new drug
- Data sources
 - Clinical trial protocols (text document)
 - Clinicaltrials.gov
 - including trial eligibility criteria
 - Recruitment networks between trials, investigators, patients
 - Safety reporting
 - FDA Adverse event database
 - Clinical trial management system (CTMS)

Properties of trial data



Pros

- **Important:** valuable data for getting drug approved
- **Heterogeneous:** both structured and unstructured, longitudinal data

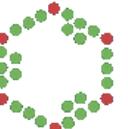


Cons

- **Difficult to match:** unstructured protocols need to map to structured patient records
- **Data integration:** integration of heterogeneous data with variable quality

Drug data



- Existing drugs
 - Drugbank data
- Chemical database for drug discovery
 - ChEMBL  large-scale bioactivity database for drug discovery
 - **ZINC** commercially available chemical compounds for virtual screening
 - QM9 dataset: quantum chemistry benchmark data for property prediction

Properties of drug data



Pros

Standard: data formats are standard

Free: many freely available (but not all)



Cons

- **Lack of 3D structure:** 3D molecule data are limited
- **Latest novelty chemical data are lacking**