SCHOOL OF MEDICINE

North Carolina Translational and Clinical Sciences Institute

2023 PPMH Data Literacy Workshop: Electronic Health Data Basics

Peter Leese
The Data Science Lab in
NC TraCS Institute

Section I

health care data background

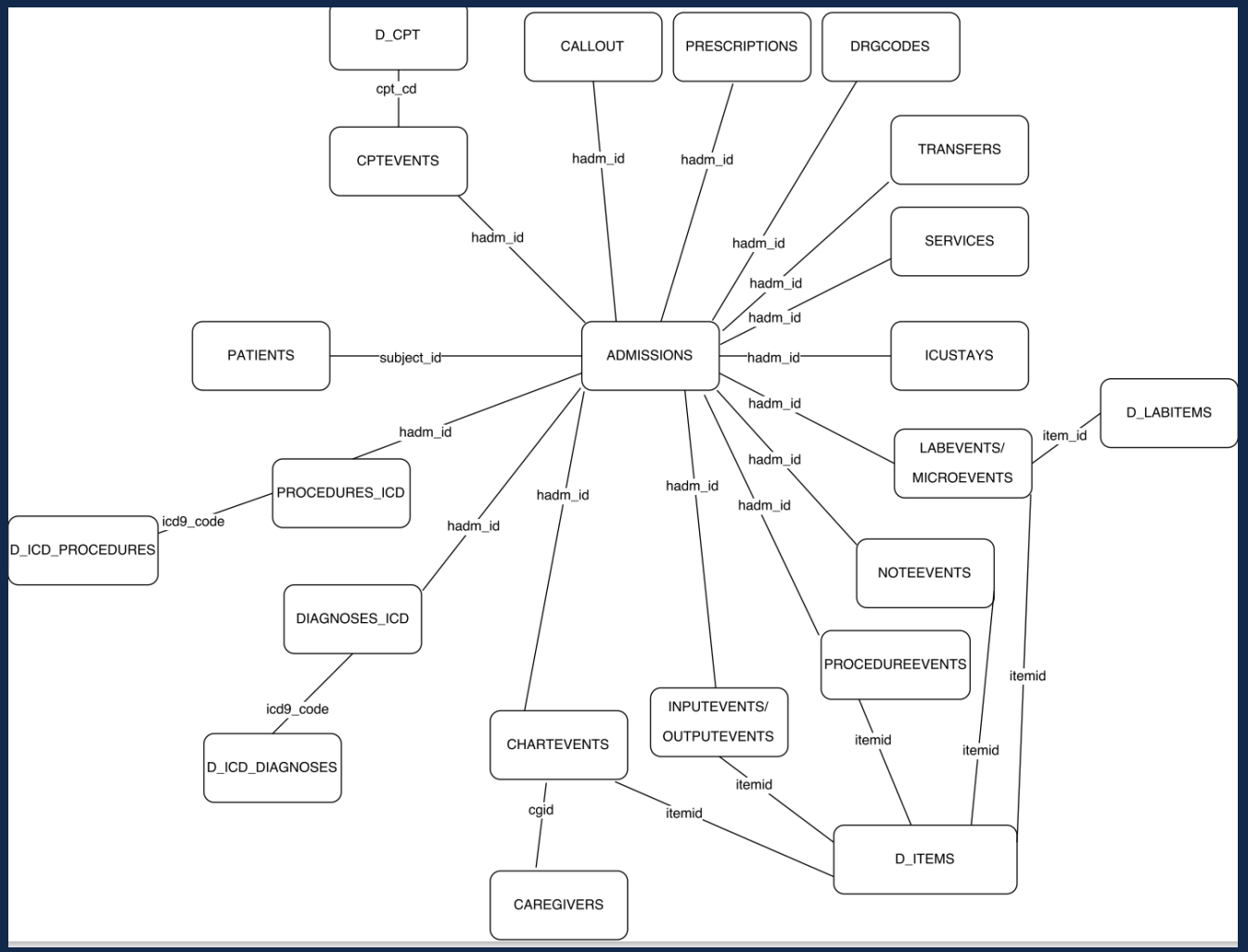
Database concepts

- Dataset
- Database (bigger, more complex)
- Data warehouse (bigger, more complex)
- Data model – design and structure of dataset

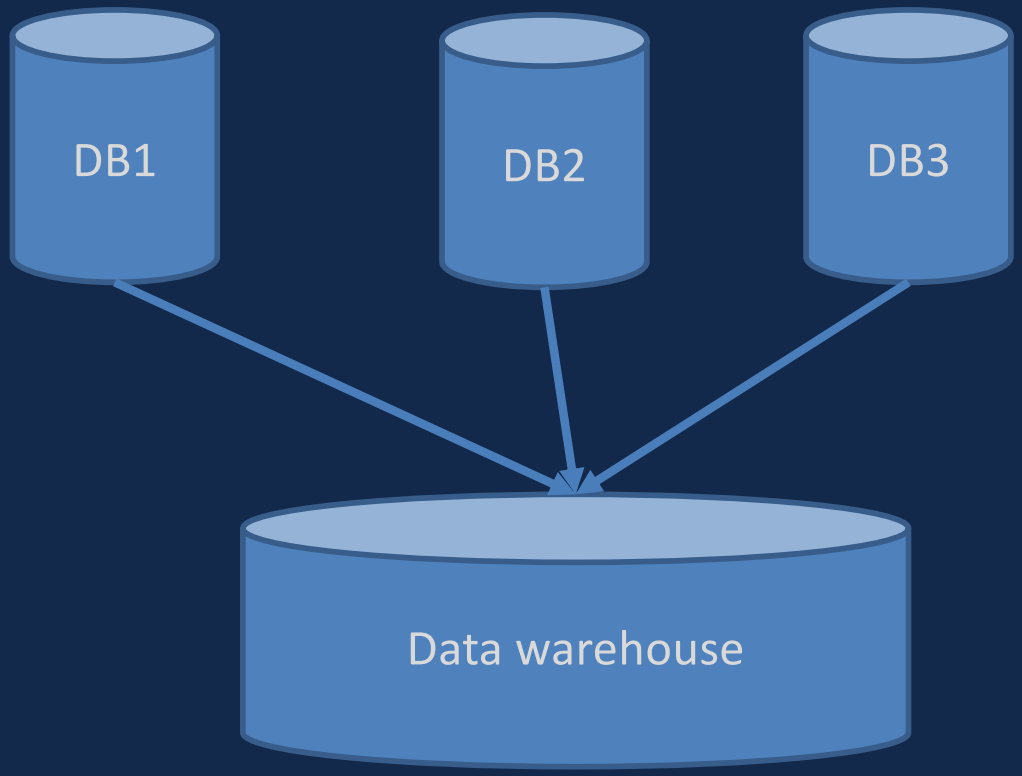
Dataset

39	State-gov	77516	Bachelors	13	Never-married	Adm-clerical	Not-in-family	White	Male	2174	0	40	United-States	<=50K
50	Self-emp-not-inc	83311	Bachelors	13	Married-civ-spouse	Exec-managerial	Husband	White	Male	0	0	13	United-States	<=50K
38	Private	215646	HS-grad	9	Divorced	Handlers-cleaners	Not-in-family	White	Male	0	0	40	United-States	<=50K
53	Private	234721	11th	7	Married-civ-spouse	Handlers-cleaners	Husband	Black	Male	0	0	40	United-States	<=50K
28	Private	338409	Bachelors	13	Married-civ-spouse	Prof-specialty	Wife	Black	Female	0	0	40	Cuba	<=50K
37	Private	284582	Masters	14	Married-civ-spouse	Exec-managerial	Wife	White	Female	0	0	40	United-States	<=50K
49	Private	160187	9th	5	Married-spouse-absent	Other-service	Not-in-family	Black	Female	0	0	16	Jamaica	<=50K
52	Self-emp-not-inc	209642	HS-grad	9	Married-civ-spouse	Exec-managerial	Husband	White	Male	0	0	45	United-States	>50K
31	Private	45781	Masters	14	Never-married	Prof-specialty	Not-in-family	White	Female	14084	0	50	United-States	>50K
42	Private	159449	Bachelors	13	Married-civ-spouse	Exec-managerial	Husband	White	Male	5178	0	40	United-States	>50K
37	Private	280464	Some-college	10	Married-civ-spouse	Exec-managerial	Husband	Black	Male	0	0	80	United-States	>50K
30	State-gov	141297	Bachelors	13	Married-civ-spouse	Prof-specialty	Husband	Asian-Pac-Islander	Male	0	0	40	India	>50K
23	Private	122272	Bachelors	13	Never-married	Adm-clerical	Own-child	White	Female	0	0	30	United-States	<=50K
32	Private	205019	Assoc-acdm	12	Never-married	Sales	Not-in-family	Black	Male	0	0	50	United-States	<=50K
40	Private	121772	Assoc-voc	11	Married-civ-spouse	Craft-repair	Husband	Asian-Pac-Islander	Male	0	0	40	?	>50K
34	Private	245487	7th-8th	4	Married-civ-spouse	Transport-moving	Husband	Amer-Indian-Eskimo	Male	0	0	45	Mexico	<=50K
25	Self-emp-not-inc	176756	HS-grad	9	Never-married	Farming-fishing	Own-child	White	Male	0	0	35	United-States	<=50K
32	Private	186824	HS-grad	9	Never-married	Machine-op-inspct	Unmarried	White	Male	0	0	40	United-States	<=50K
38	Private	28887	11th	7	Married-civ-spouse	Sales	Husband	White	Male	0	0	50	United-States	<=50K
43	Self-emp-not-inc	292175	Masters	14	Divorced	Exec-managerial	Unmarried	White	Female	0	0	45	United-States	>50K
40	Private	193524	Doctorate	16	Married-civ-spouse	Prof-specialty	Husband	White	Male	0	0	60	United-States	>50K
54	Private	302146	HS-grad	9	Separated	Other-service	Unmarried	Black	Female	0	0	20	United-States	<=50K
35	Federal-gov	76845	9th	5	Married-civ-spouse	Farming-fishing	Husband	Black	Male	0	0	40	United-States	<=50K
43	Private	117037	11th	7	Married-civ-spouse	Transport-moving	Husband	White	Male	0	2042	40	United-States	<=50K
59	Private	109015	HS-grad	9	Divorced	Tech-support	Unmarried	White	Female	0	0	40	United-States	<=50K

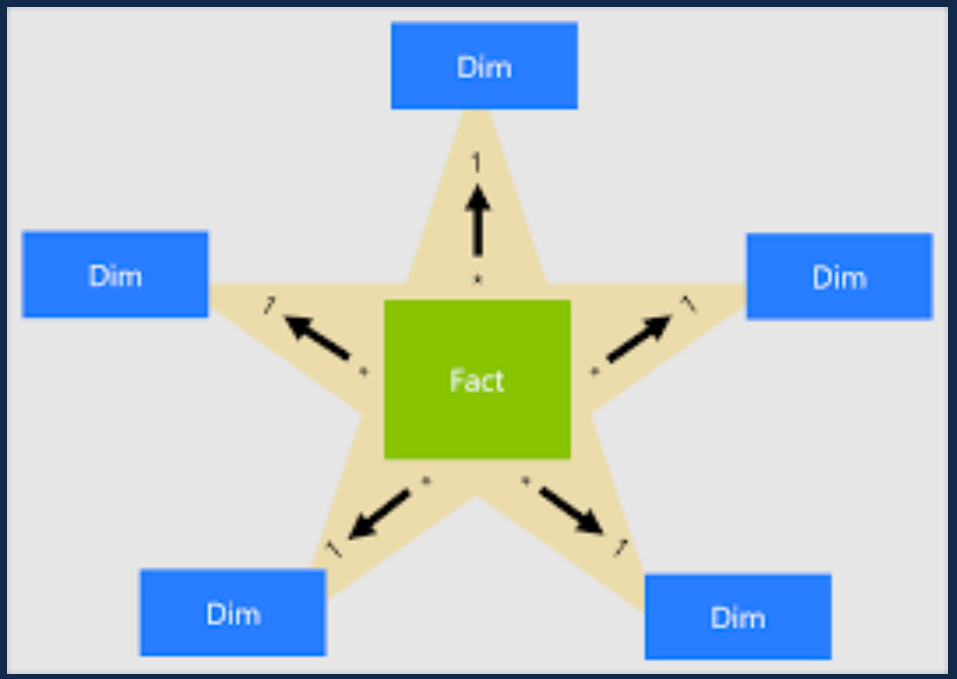
Database

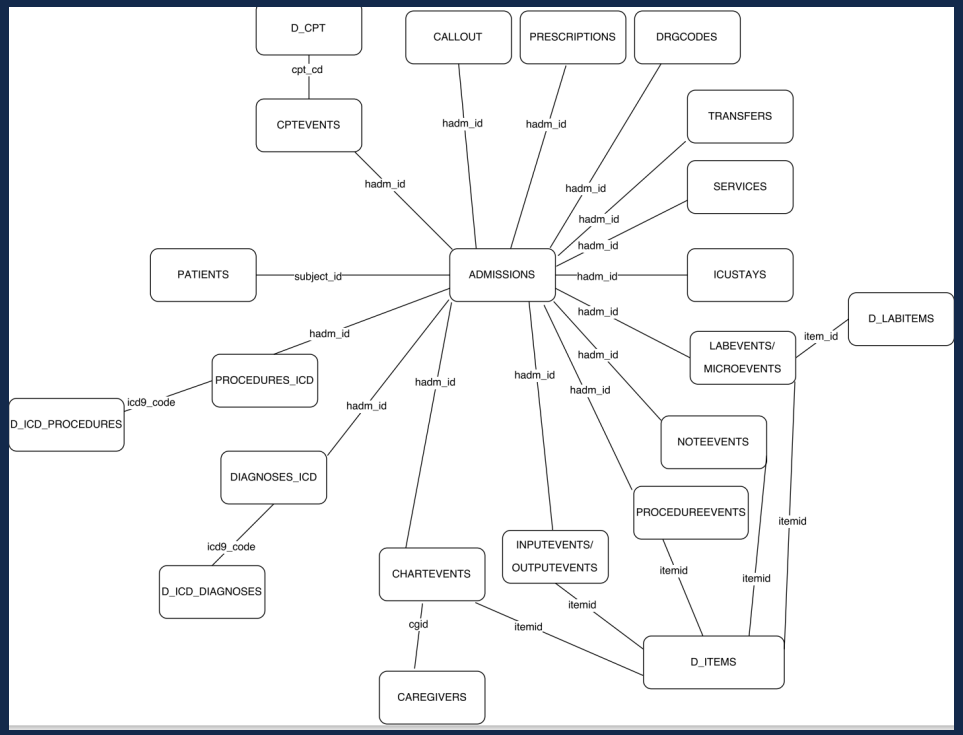

D_CPT
CALLOUT
PRESCRIPTIONS
DRGCODES
cpt_cd
CPTEVENTS
hadm_id
TRANSFERS
SERVICES
PATIENTS
subject_id
ADMISSIONS
ICUSTAYS
D_LABITEMS
item_id
LABEVENTS/
MICROEVENTS
PROCEDURES_ICD
icd9_code
D_ICD_PROCEDURES
NOTEEVENTS
DIAGNOSES_ICD
PROCEDUREEVENTS
itemid
INPUTEVENTS/
OUTPUTEVENTS
CHARTEVENTS
D_ICD_DIAGNOSES
cgid
D_ITEMS
CAREGIVERS

Data Warehouse



Data Model


Dim
Dim
Dim
Fact
Dim
Dim
1
*


D_CPT
CALLOUT
PRESCRIPTIONS
DRGCODES
cpt_cd
CPTEVENTS
TRANSFERS
SERVICES
hadm_id
PATIENTS
subject_id
ADMISSIONS
ICUSTAYS
D_LABITEMS
item_id
LABEVENTS/
MICROEVENTS
PROCEDURES_ICD
icd9_code
D_ICD_PROCEDURES
NOTEEVENTS
DIAGNOSES_ICD
PROCEDUREEVENTS
itemid
icd9_code
D_ICD_DIAGNOSES
CHARTEVENTS
INPUTEVENTS/
OUTPUTEVENTS
cgid
D_ITEMS
CAREGIVERS

Health Data Concepts

Types of healthcare data

Primary or ‘research’

- Clinical trials
- BRFSS, MEPS, NHIS

Real World Data

- EHR
- Claims, ‘discharges’, ‘administrative’
- From Patient(sensor, search engine)

Health Data Concepts

- Types of healthcare data

  Primary or ‘research’

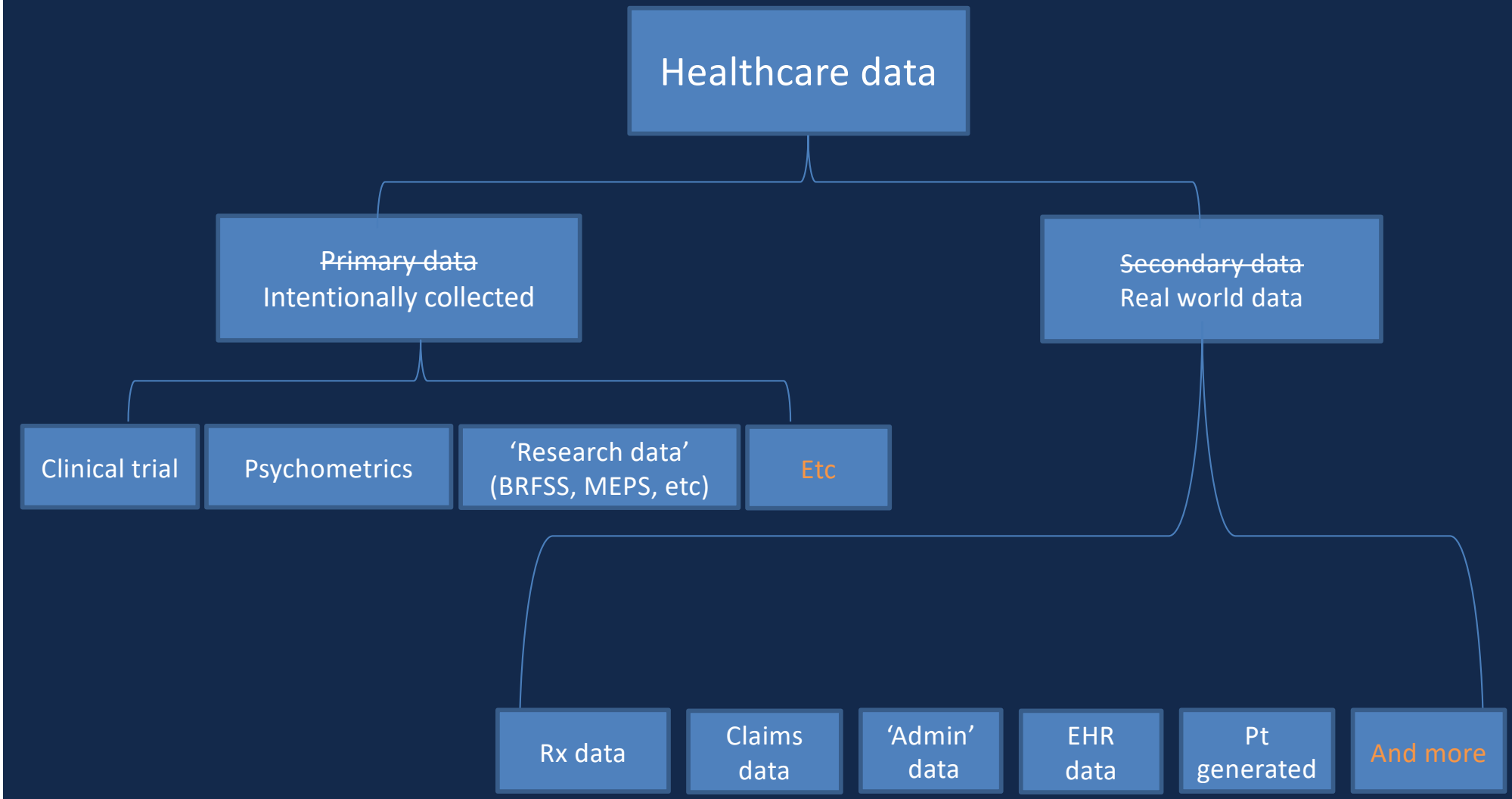
  - Clinical trials
  - BRFSS, MEPS, NHIS

  Real World Data

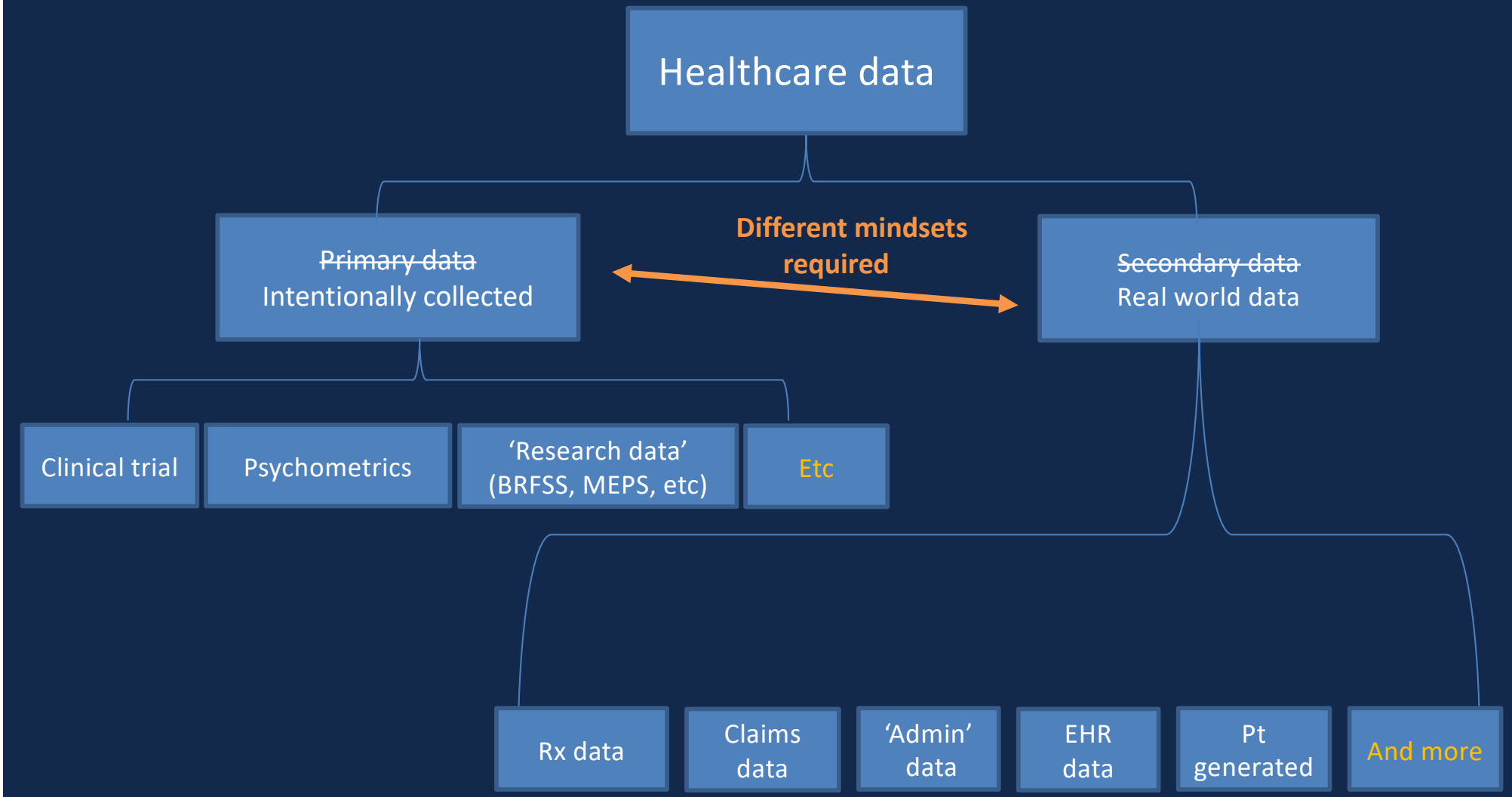
  - EHR
  - Claims, ‘discharges’, ‘administrative’
  - From Patient(sensor, search engine, social media)

Intention is key

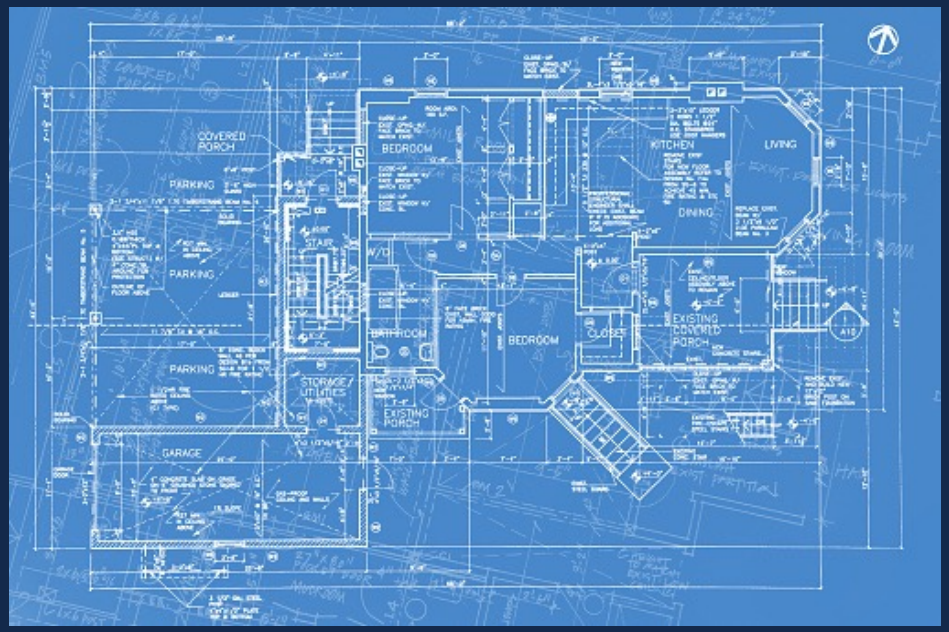
Healthcare Data



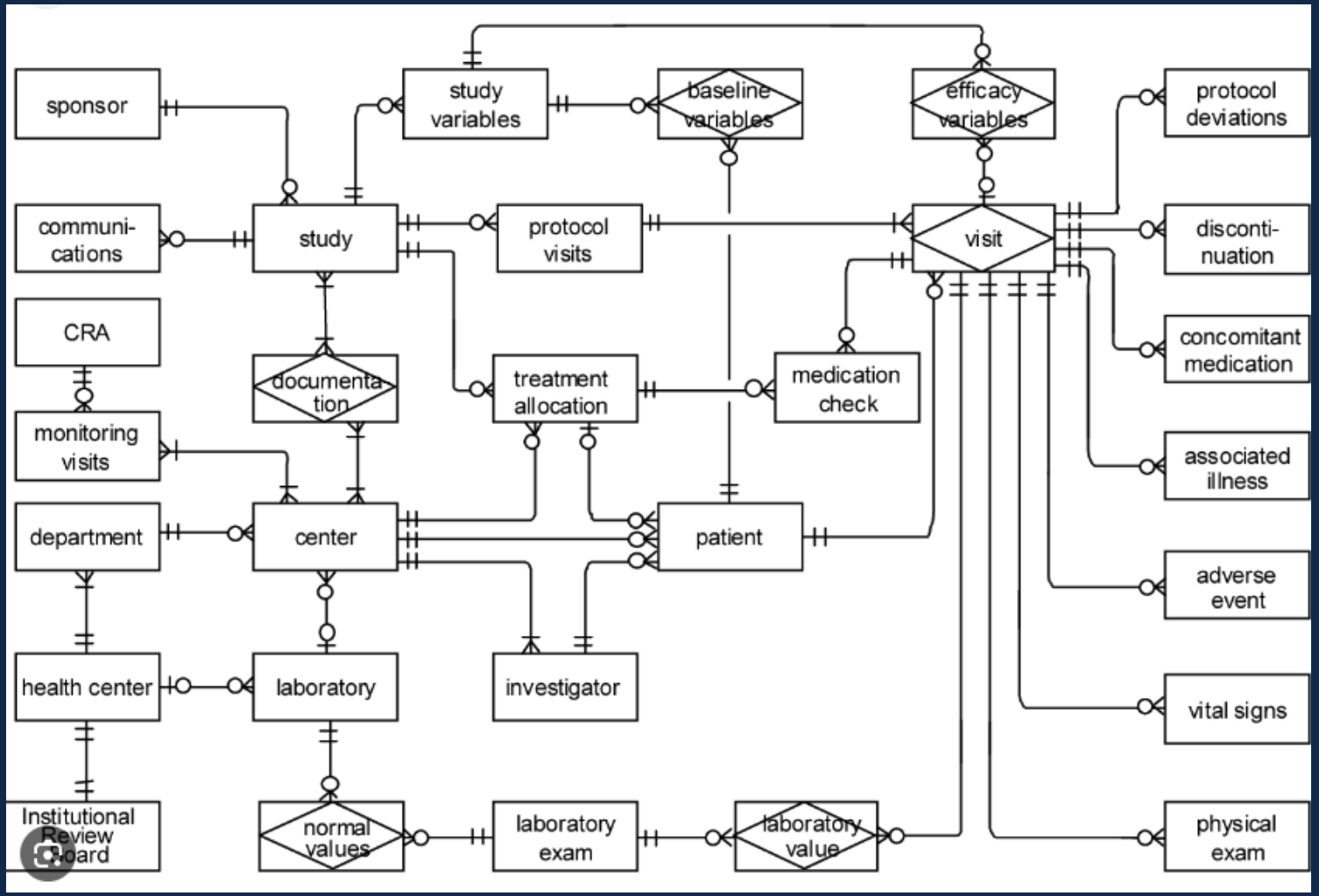
Healthcare Data



Primary / Research Data - Architecture



Applying the metaphor – Research Data



EHR Data – Archaeology (+/- Anthropology)

Applying the metaphor – EHR Data

Basic EHR data model for simple domain.

Full analysis of EHR might involve dozens of data domains from full database of >18,000 tables.

Just identifying, planning, and extracting the data can be an excavation project on its own.

Over time we will also need to factor in systematic and random heterogeneity of care & documentation.

RWD is not new

Claims data & analysis

- Used for decades
- Robust and deep literature of work & methods

- **Can still very difficult to learn about the data**

- **Same is true of EHR data**

RWD Sources Today

- **Claims** (Marketscan, CMS, etc)
- **Claims add-ons** (labs, etc)
- **Commercial products** (DRG/Clarivate, IQVIA, etc)
- **EHR-specific** (Datavant, TriNetX, Sensyne, Epic, etc)
- **Research data networks** (PCORnet, STAR CRN, N3C, etc)
- **Local** (University & hospital warehouses)

EHR vs Claims data

EHR vs Claims Data

- Billing activities
- Billing
  - Diagnoses
  - Procedures
  - Providers
  - Locations
  - Dates
- Clinical & operational
  - Patients
  - Care stays
  - Charting
  - Labs
  - Notes
  - Meds
  - Prescriptions
  - Diagnose
  - Procedures, and More!

Publicly available RWD

Claims

- SYNPuf synthetic Medicare claims

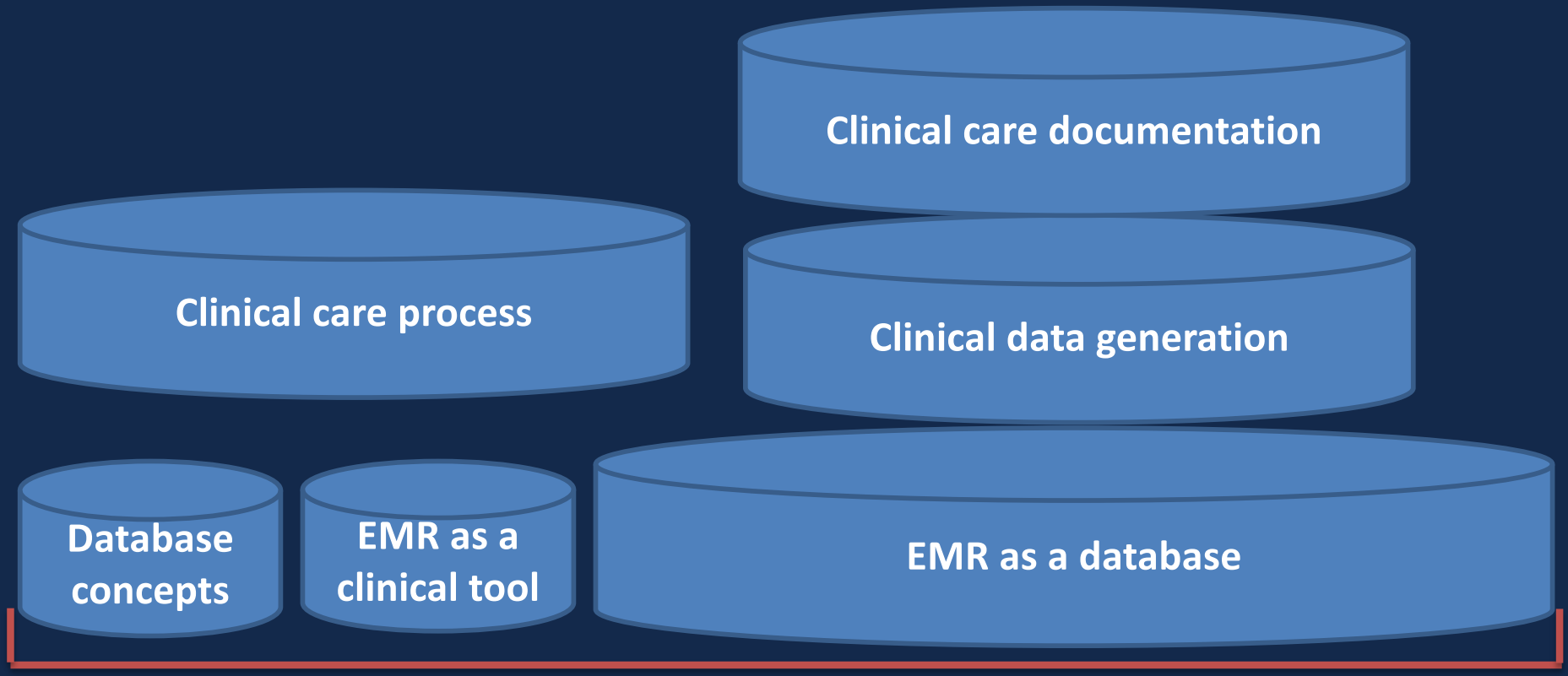
EHR

- Synthea synthetic EHR data
- MIMIC III, IV (requires access through MIT)

Section I
Questions

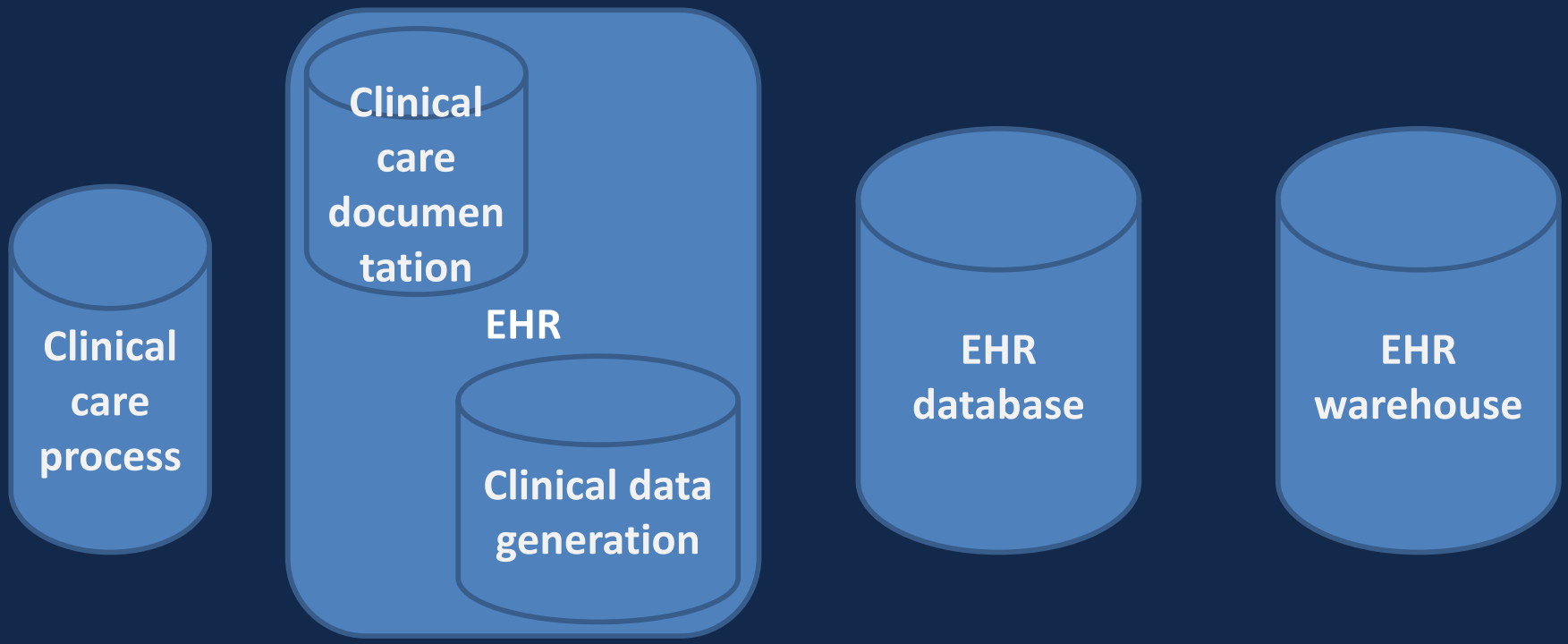
Section II:
EHR as data system

Multiple perspectives for EHR success



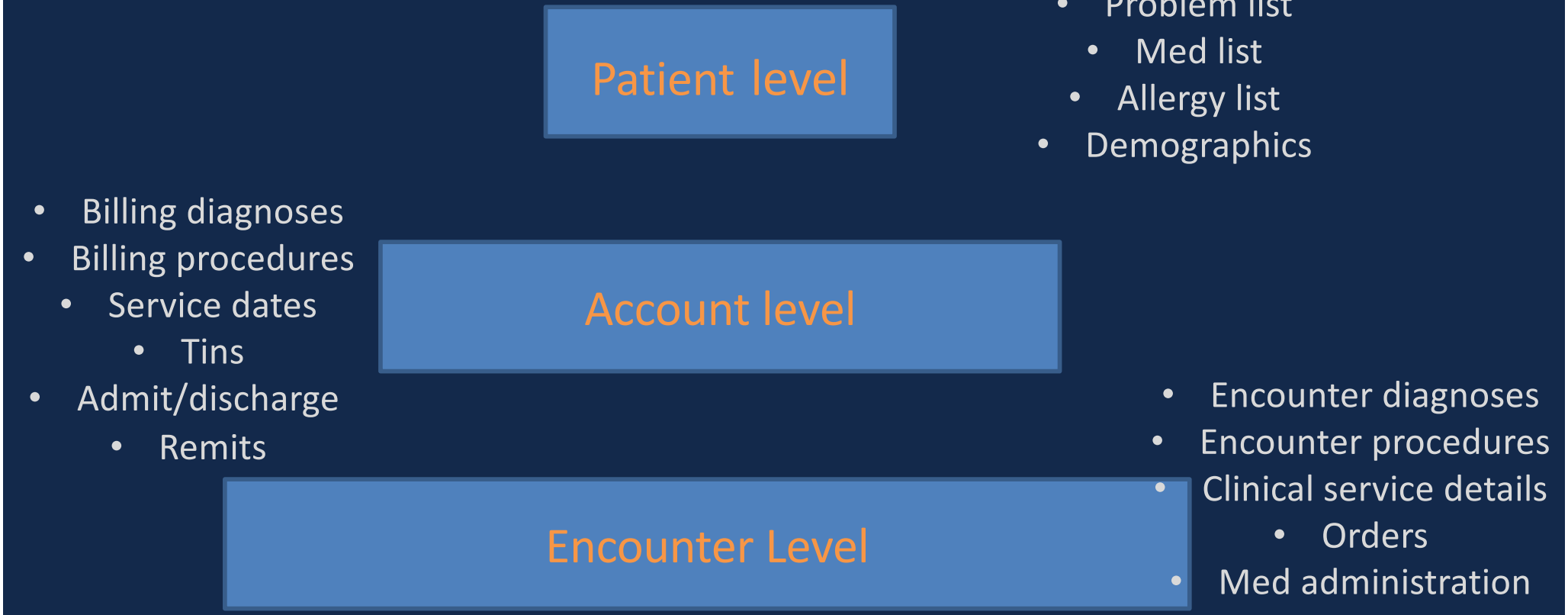
Breadth can be more helpful than depth.

Multiple perspectives for EHR success



Expertise to formulate question ⟷ Expertise to define question ⟷ Expertise to Obtain data

Information hierarchy in EHR



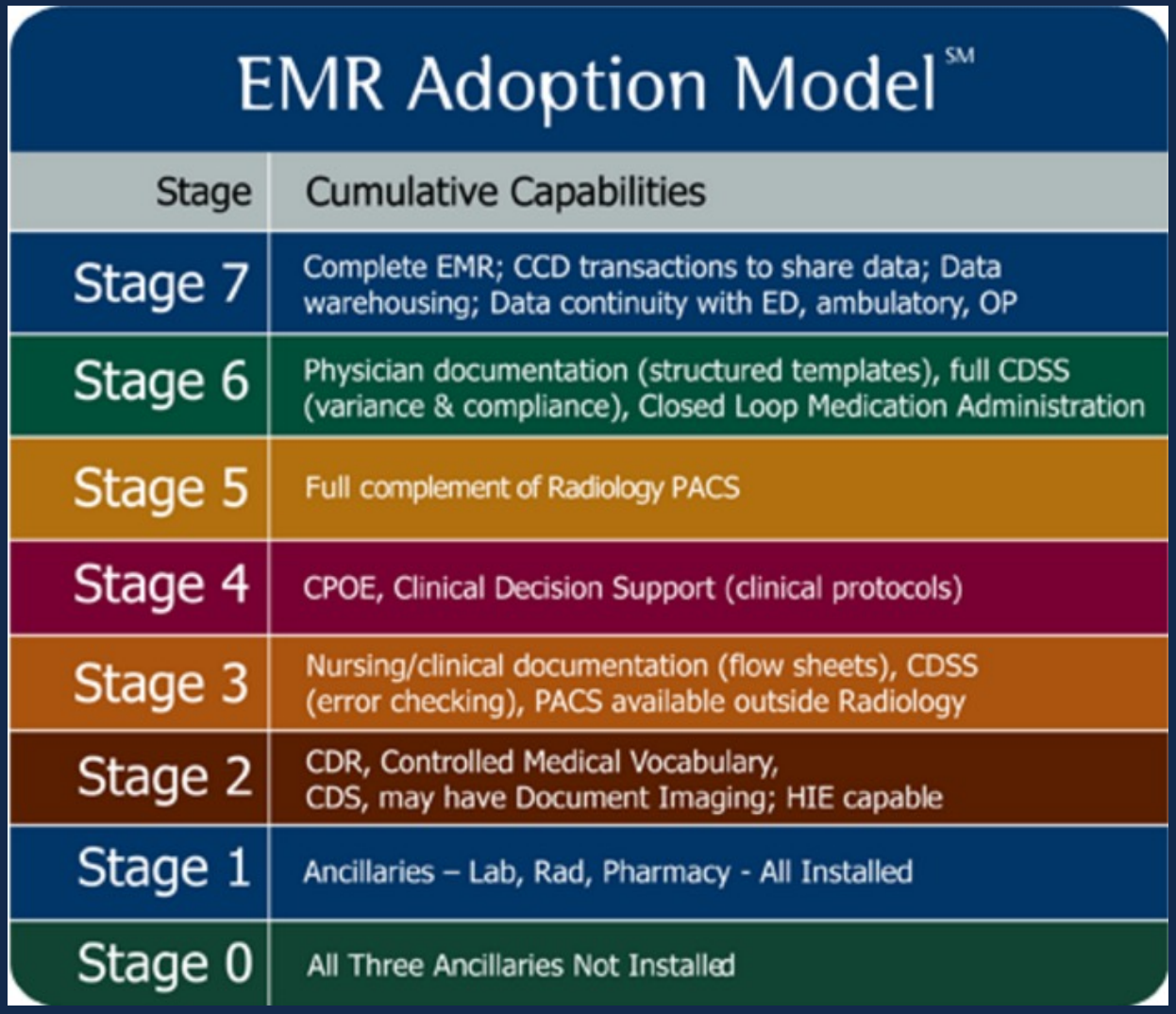
Care process & documentation

- LIP workflow: notes, CPOE, lists, decision support
- Nursing workflow: flowsheets, assessments, vitals
- Labs: results
- Pharmacy: supply chain, compounding/dispensing, administration
- Prescribing/medications: medication list, e-scripts
- Population health: registries, decision support, health maintenance

Administrative process & documentation

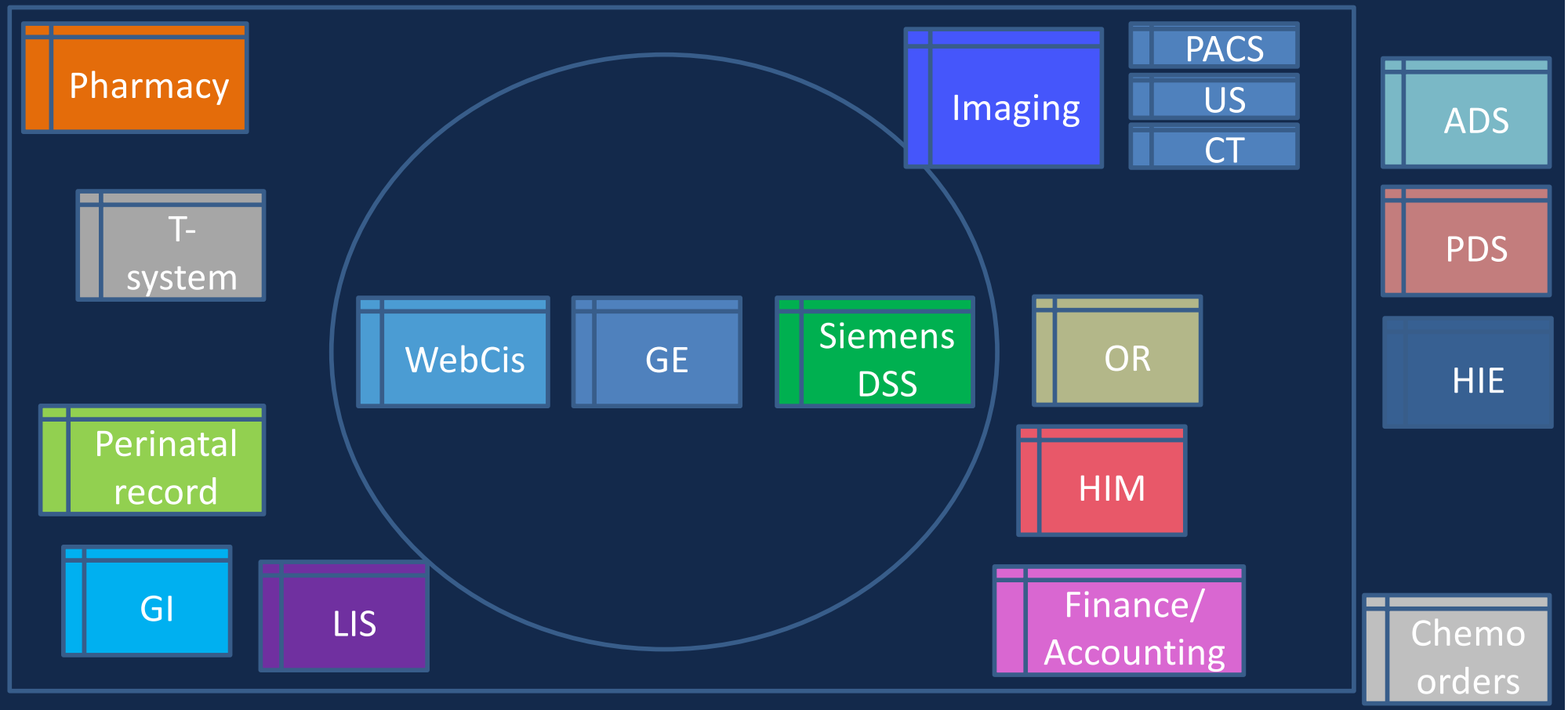
- Registration & scheduling
- Hospital operations: ADT, bed tracking, transfer center
- HIM: patient management, compliance, coding
- Billing: inpatient & outpatient, professional and facility
- Managed care

EHR "Stuff" Score

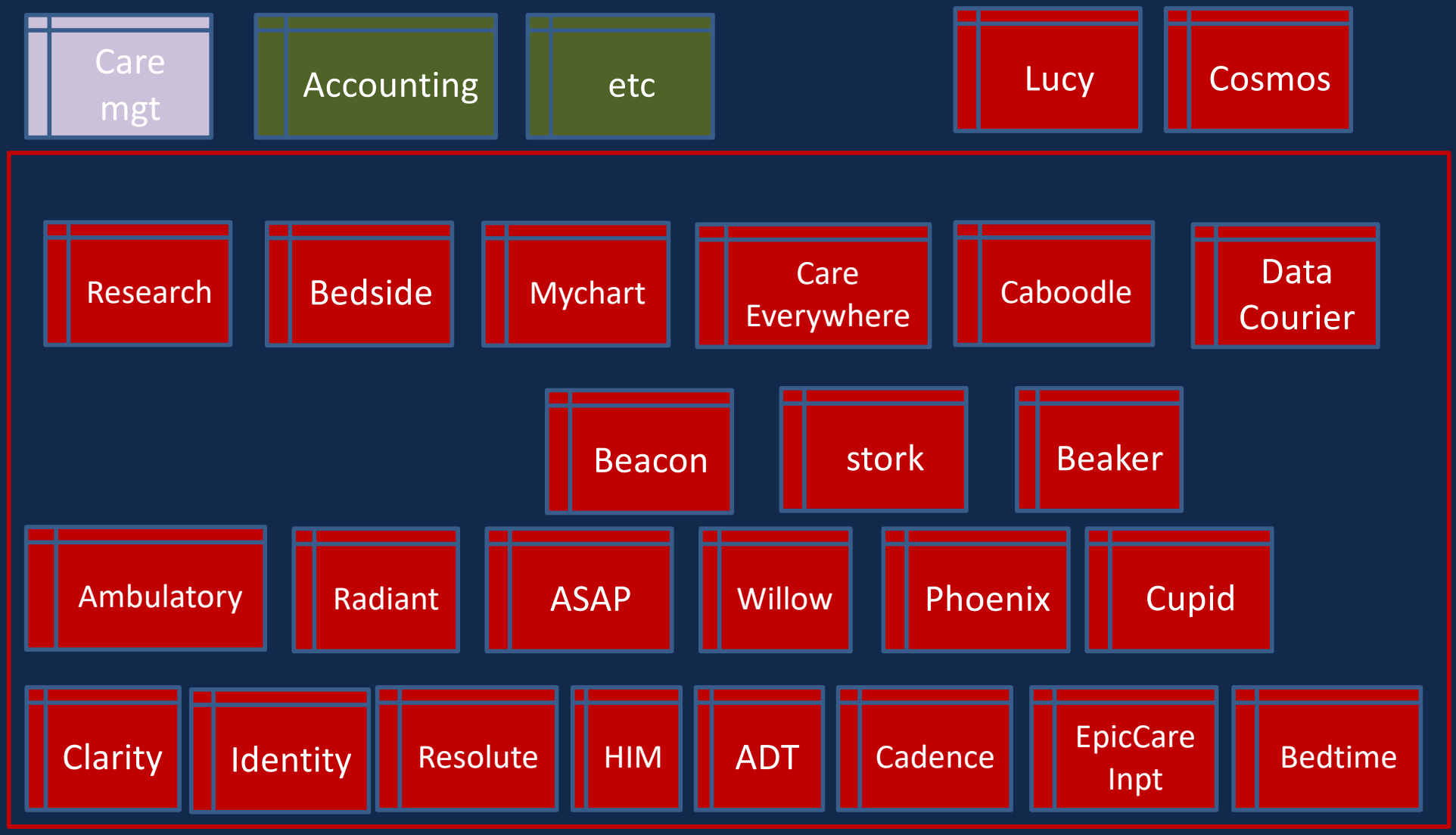


EMR Adoption Model℠	
Stage	Cumulative Capabilities
Stage 7	Complete EMR; CCD transactions to share data; Data warehousing; Data continuity with ED, ambulatory, OP
Stage 6	Physician documentation (structured templates), full CDSS (variance & compliance), Closed Loop Medication Administration
Stage 5	Full complement of Radiology PACS
Stage 4	CPOE, Clinical Decision Support (clinical protocols)
Stage 3	Nursing/clinical documentation (flow sheets), CDSS (error checking), PACS available outside Radiology
Stage 2	CDR, Controlled Medical Vocabulary, CDS, may have Document Imaging; HIE capable
Stage 1	Ancillaries – Lab, Rad, Pharmacy - All Installed
Stage 0	All Three Ancillaries Not Installed

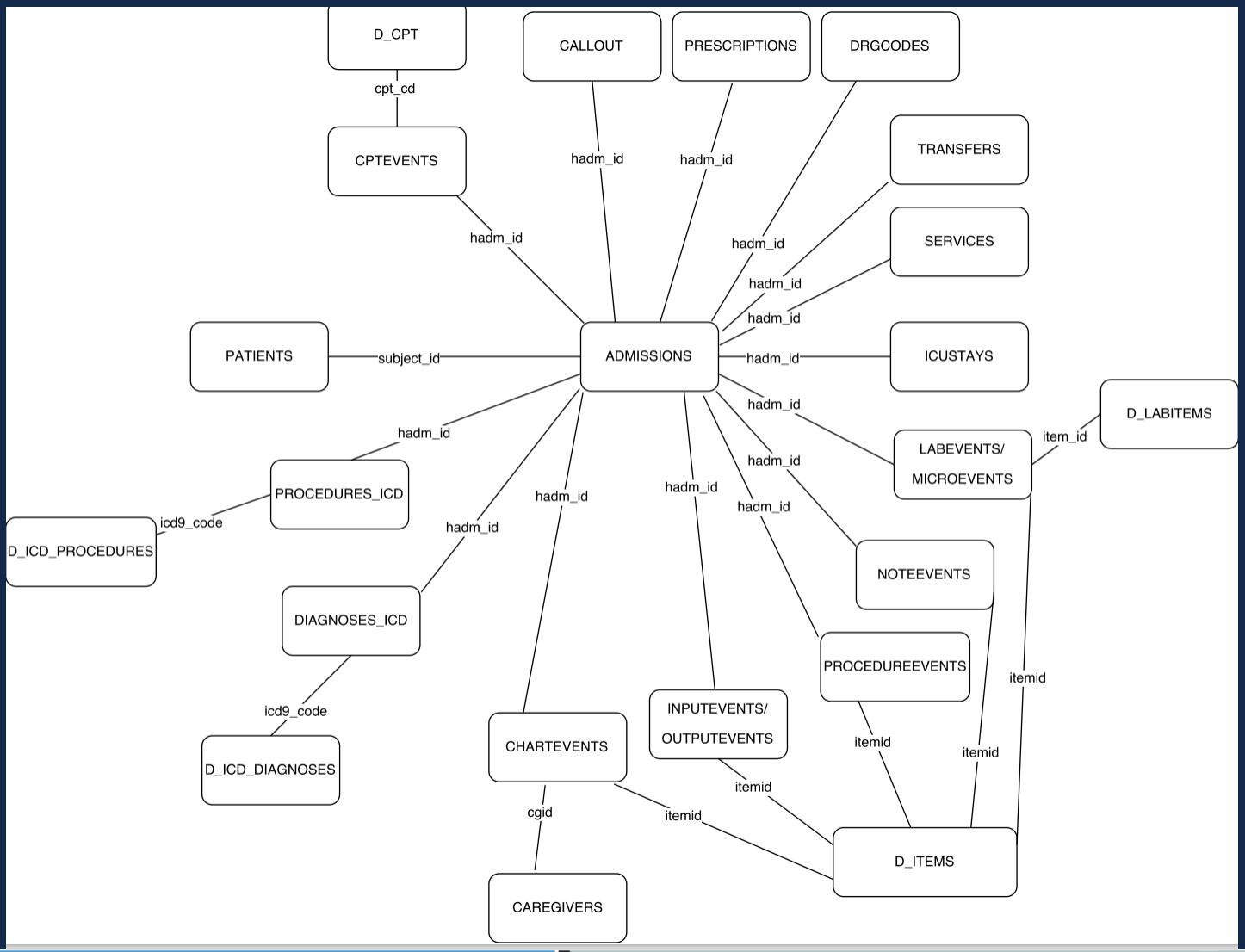
‘best of breed’ strategy



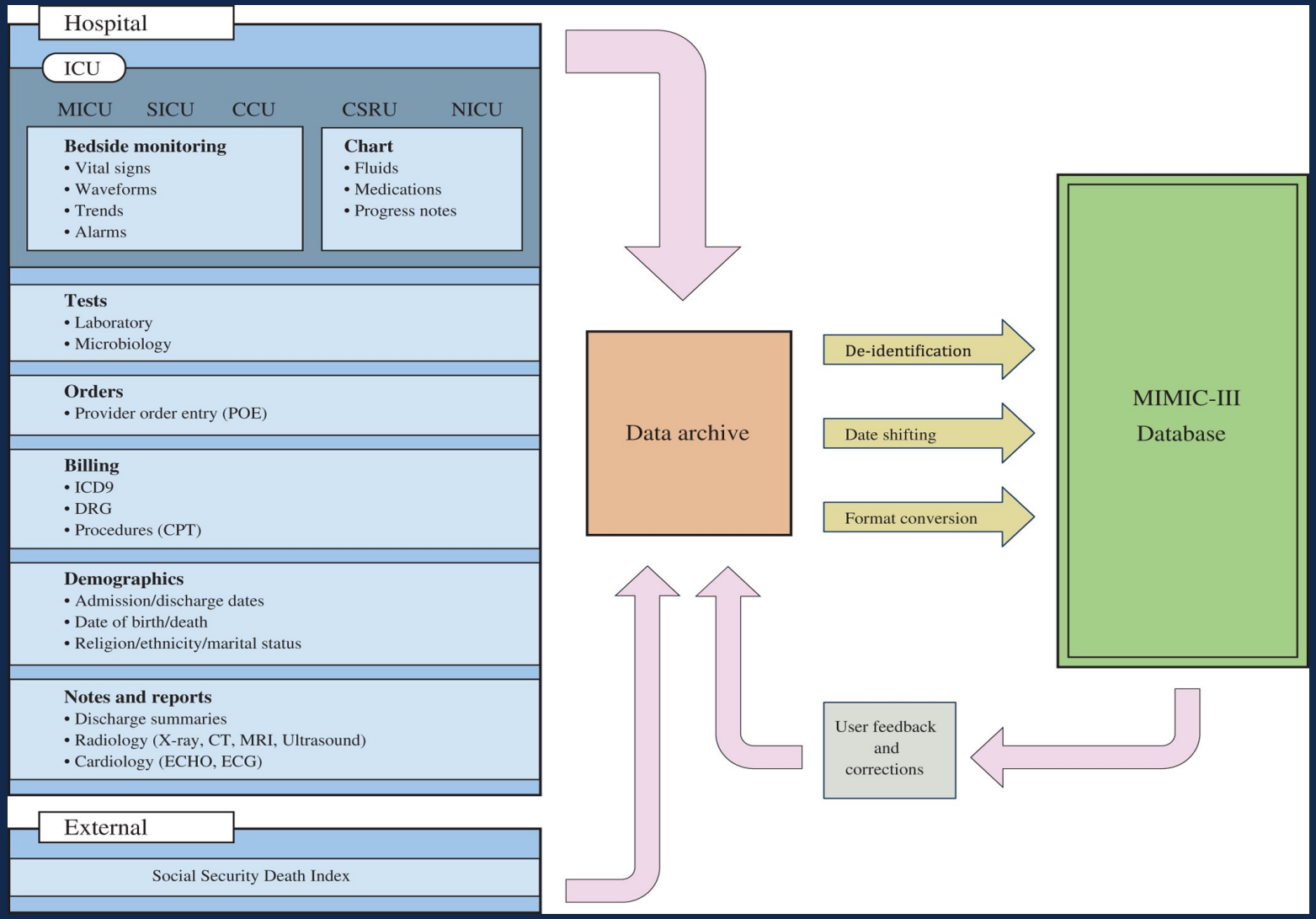
Single Vendor EHR Platform



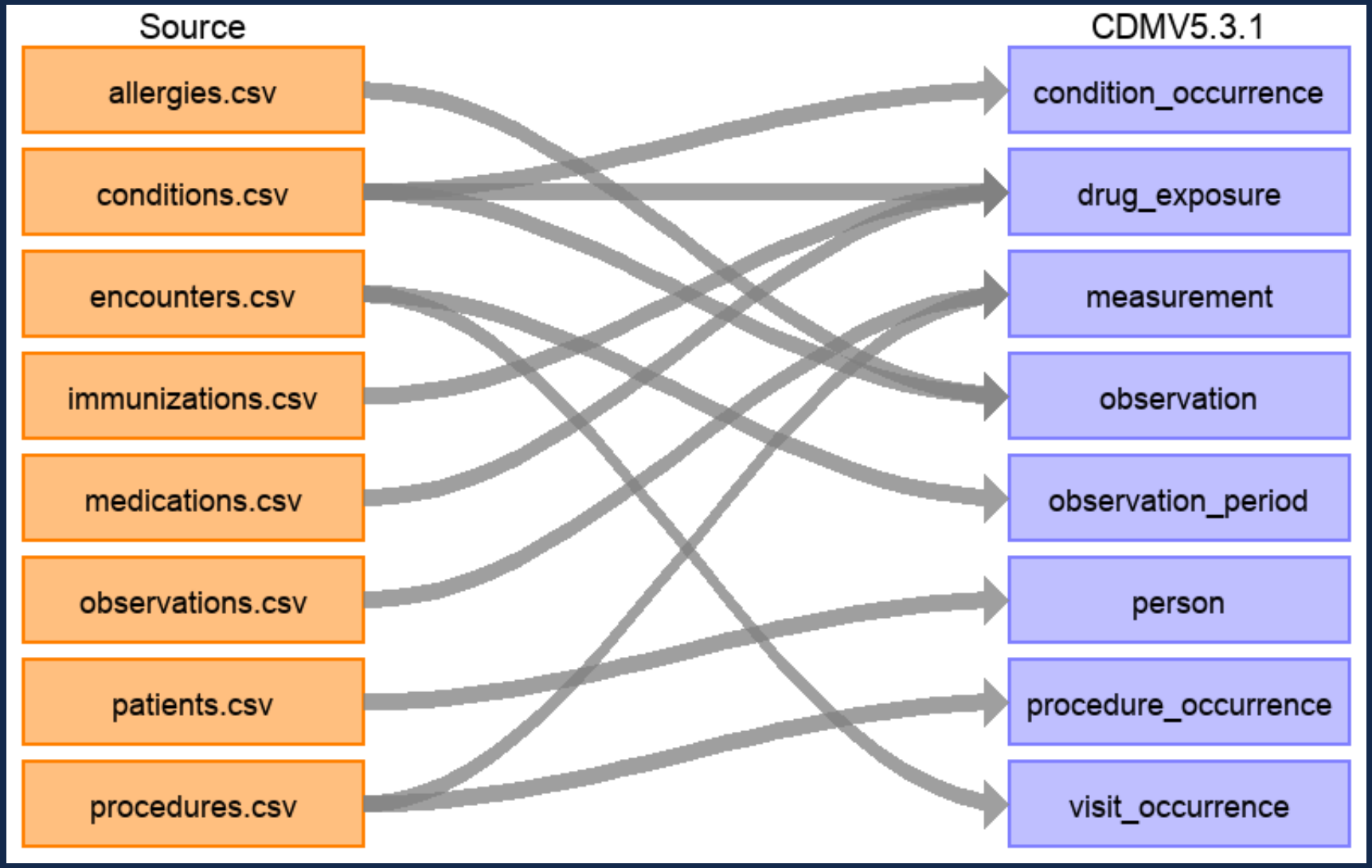
EHR data becomes domains



Domains to dataset



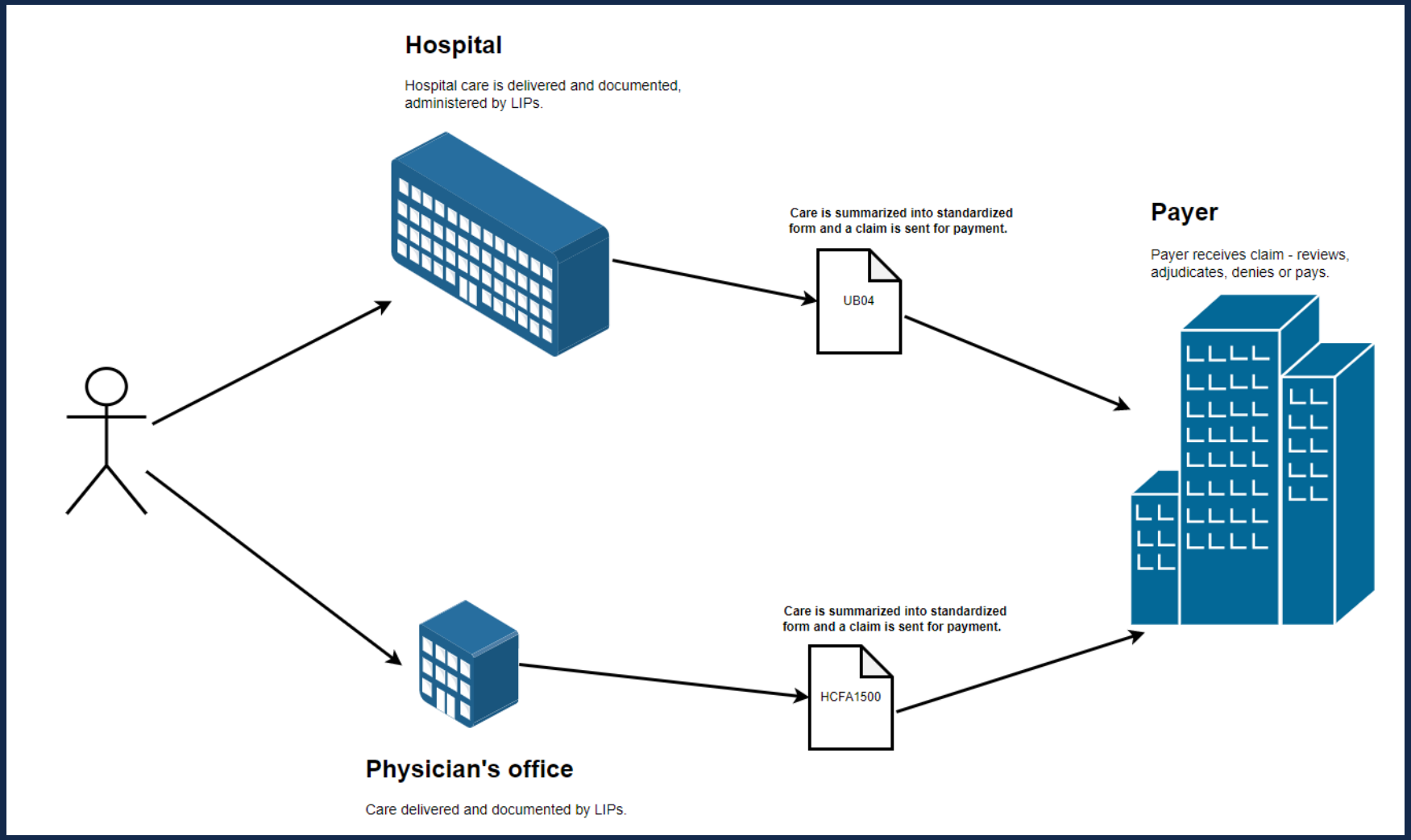
CDMs briefly



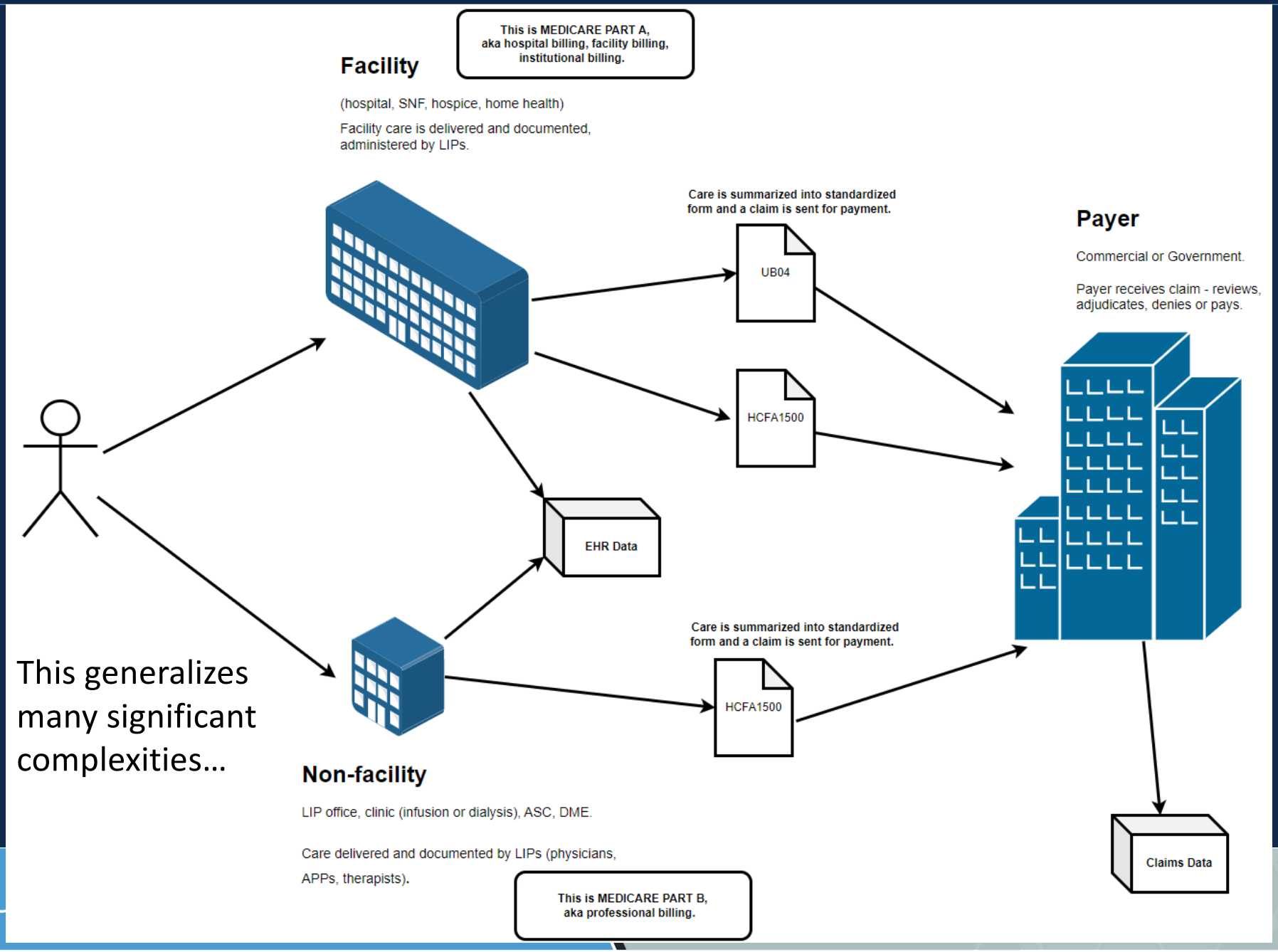
Section II
Questions

Section III:
Documentation to data

US healthcare system in one slide

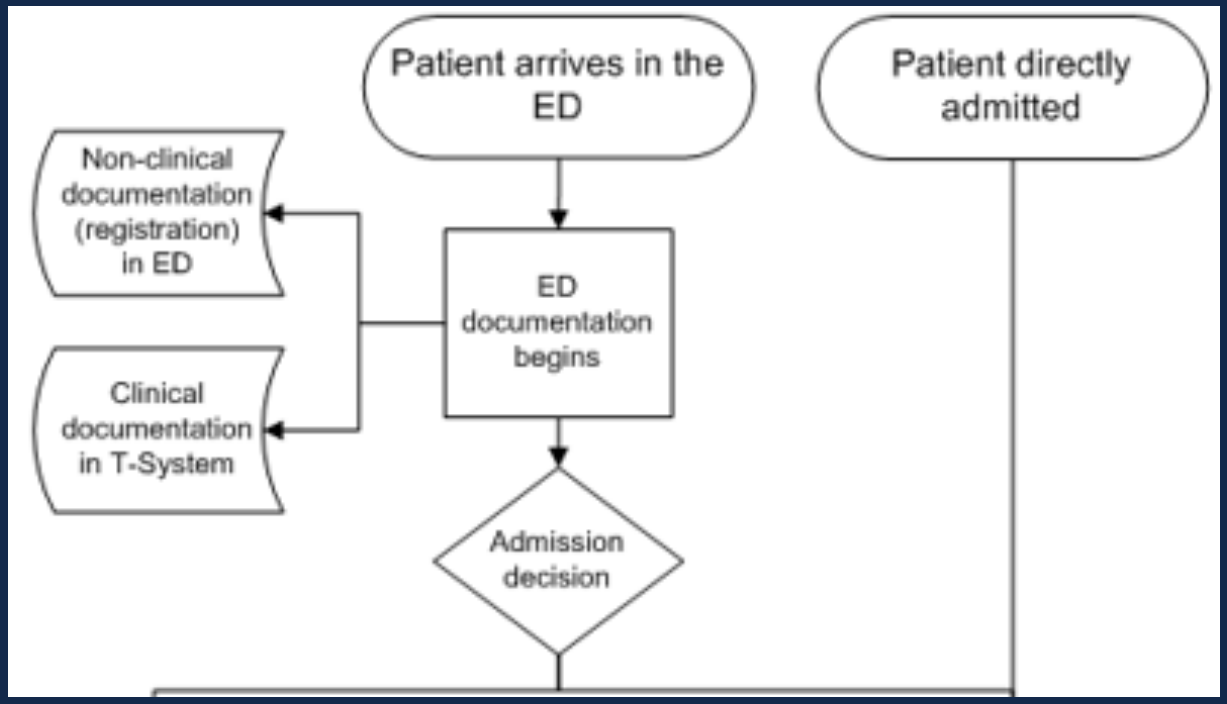


Two Systems: Facility VS Professional

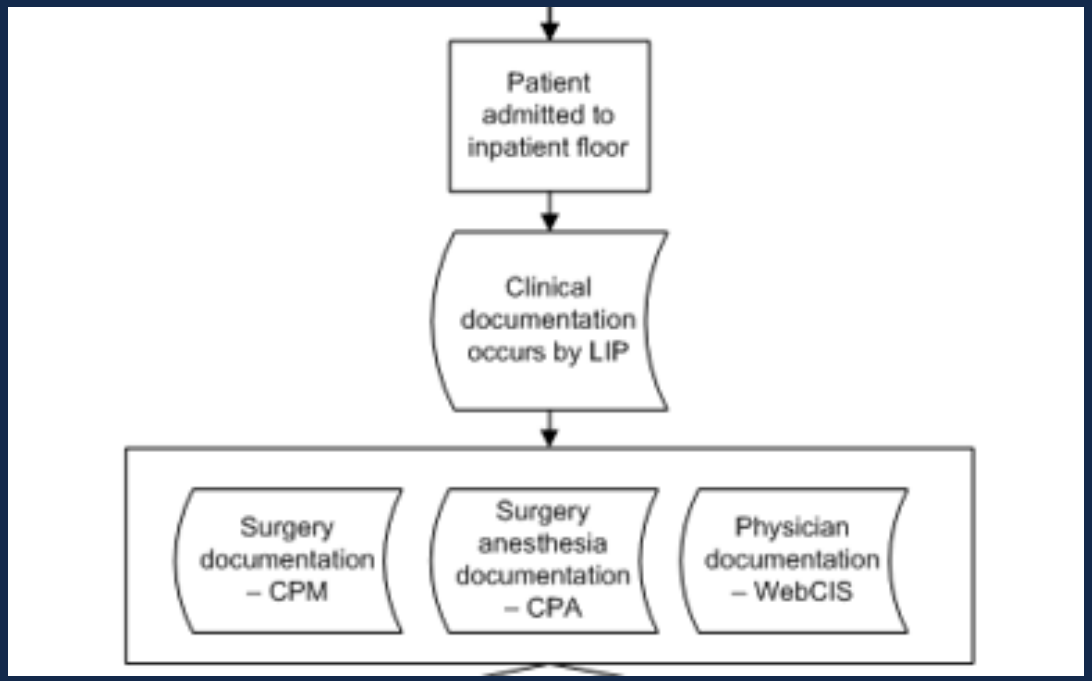


This generalizes many significant complexities…

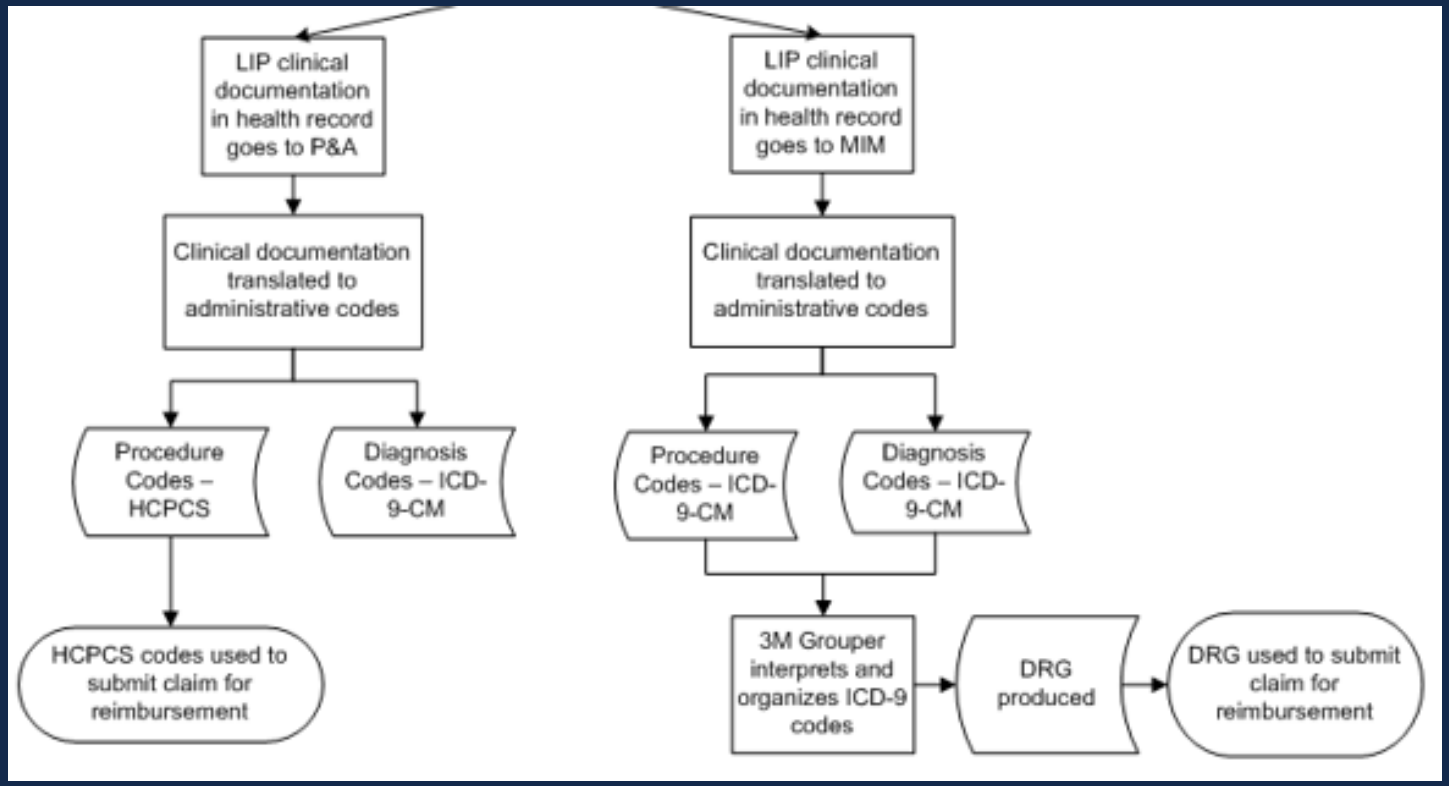
Example 1: ED


Patient arrives in the ED
Patient directly admitted
Non-clinical documentation (registration) in ED
Clinical documentation in T-System
ED documentation begins
Admission decision

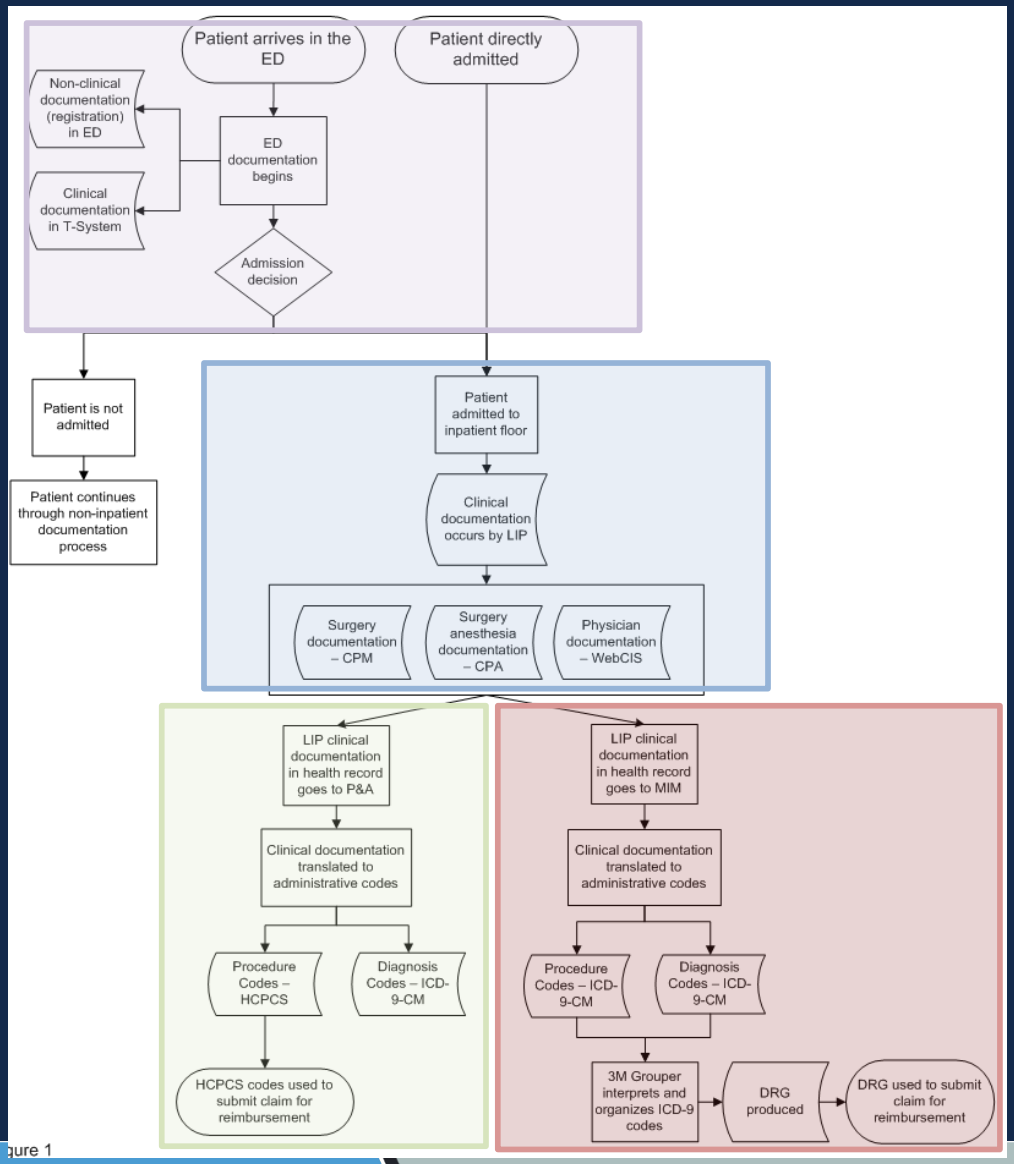
Example 2: Inpatient



Documentation to Data



Facility: care to documentation to data

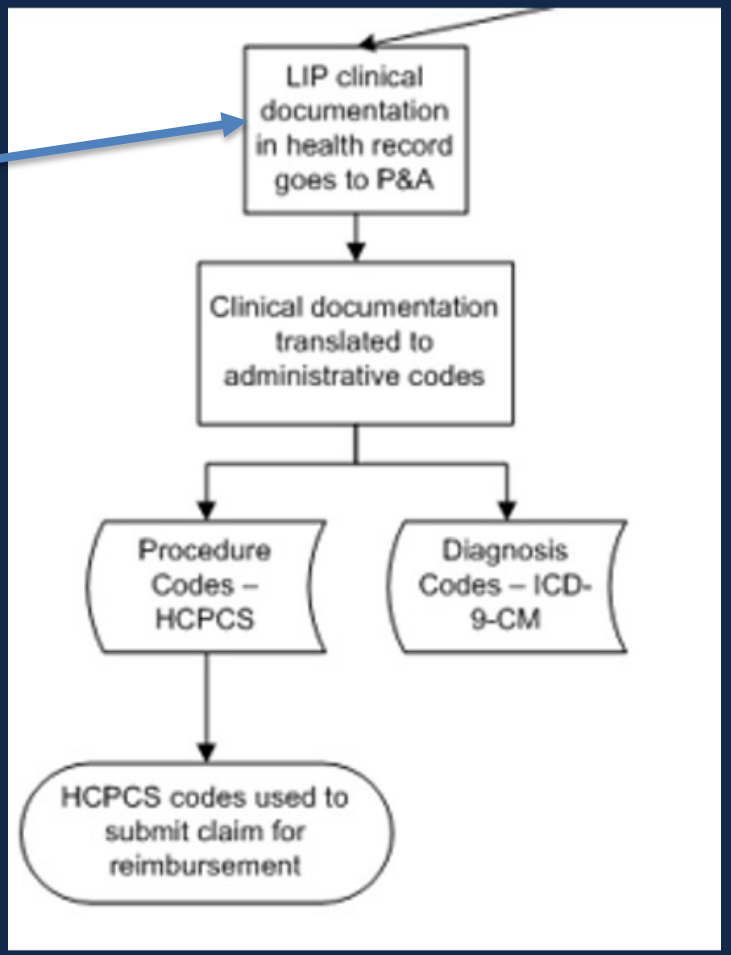


gure 1

Non-facility example

Care delivered

LIP documents services & justifies



In healthcare exceptions are the rule

- Infusions, dialysis -> recurring visits
- Clinics are outpatient until owned by facility

- There is variation to much of this by specialty, practice type, and geography

CLINICAL DATA LITERACY SERIES: ELECTRONIC HEALTH DATA BASICS

Date	Topic	Instructor(s)
Wed May 10, 2:30-4:00pm	How health care system generates data and how this data is stored in the EHR	Peter Leese
Wed May 17, 2:30-4:00pm	code sets used to record health care data	Emily Pfaff
Wed May 24, 2:30-4:00pm	fundamental units of how health care data is organized in the EHR	Peter Leese & Emily Pfaff
Wed May 31, 2:30-4:00pm	how to design a research question for clinical data	Michael Adams & Anna Jojic

Helpful Resources Handout

Download at bottom of series webpage

https://go.unc.edu/clinical-data-literacy