

Introduction to AIR·MS

Health Data Literacy

Ashwin Sawant
Andrew Deonarine

Credits: Eugenia Alleva

Sept 15th, 2026



Hasso Plattner Institute for Digital Health at Mount Sinai

Overview

1. What is AIR•MS?
2. Mount Sinai Health System & Epic EHR
3. AIR•MS & OMOP
4. Coding Systems & Concept Mapping
5. Content Standardization via Concept Mapping
6. Data Contents of Mount Sinai's OMOP Research Data Repository
7. Putting It Together: Queries & Phenotypes (Data2Evidence, AI Agents)



<https://labs.ica hn.mssm.edu/airms/>

What is AIR·MS?



Artificial Intelligence-Ready Mount Sinai (AIR·MS) is a research platform that is composed of:

- 1) A (very fast) integrated database including Mount Sinai Data Warehouse (MSDW), Pathology and Radiology metadata; and the included data is growing
- 2) A Research Environment that allows interactions with the AIR·MS database from Python or R
- 3) An Application Tier to host a growing number of applications – including cohort building tools and annotation apps

Data Modalities in AIR·MS



Currently Available Modalities:

- Mount Sinai Data Warehouse (MSDW), both containing protected health information (PHI) and DeID (deidentified) Observational Medical Outcomes Partnership (OMOP)-mapped electronic health record (EHR)
- Pathology Metadata
- Radiology Metadata
- BioMe/Sinai Million
- electrocardiogram (EKG)
- Echocardiography
- GI Research Database

Work in progress: Electroencephalogram (EEG), Endoscopy & Colonoscopy Reports

All modalities are stored in separate database schemas, and access is granted to each schema individually based on Institutional Review Board (IRB)

Request Access to AIR·MS in Sailpoint



AIR·MS

- 1) Obtain Institutional Review Board (IRB) approval for your project if you want to access PHI data, including this indication “*we will use the AIR·MS platform (IRB # 20-01288) to access and store our data*”
- 2) Request Minerva/High Performance Computing (HPC) account
- 3) Request access to specific modalities (i.e. schemas) on [SailPoint](#)
- 4) After access approval you can get started with the data using our Getting Started Guides:

```
git clone https://github.mountsinai.org/AIRMS/airms-researcher-tutorials-minerva.git
```

Working with AIR·MS in Python



Connection: The `airms_connect` library enables easy connection to AIR·MS and is pre-installed on Minerva; install it separately when using conda/venv.

On Minerva vs Local: From a Minerva compute node, first establish an SSH tunnel via `.on_minerva()`.

Interacting with HANA: The `airms_connection()` class exposes a `conn` attribute, a `hana_ml.ConnectionContext` object, enabling direct use of hana-ml.

- `conn.sql()` executes raw SQL queries.
- `conn.table()` accesses specific tables.

Both return a hana-ml DataFrame, which mimics pandas syntax and supports in-database operations. To pull data locally, use `.collect()` to return a pandas DataFrame.

```
# import
from airms_connect.connection import airms_connection

# initialize connection
airms = airms_connection()

# establish tunnel on Minerva
login_host_name='li04e04'
airms.on_minerva(login_host_name=login_host_name)

# establish connection
airms.connect()

# query airms
print(airms.conn.table('PERSON', 'CDMDEID').head(10).collect())
# or
query = "SELECT TOP 10 * FROM CDMDEID.PERSON"
print(airms.conn.sql(query).collect())
```

Working with AIR·MS in R



Install Packages

To use AIR·MS with R, you need to install the following packages in your environment:

- `install.packages("odbc")`
- `install.packages("RJDBC")`

Download Java driver

You then need to point to the [Java driver](#) in your R script.

- On Minerva the driver can be found at `/sc/arion/projects/airms/lib/ngdbc.jar`

```
library(odbc)
library(RJDBC)
library(getPass)

driver_path <- path/to/driver
# Initialize the driver
jdbcDriver <- JDBC(driverClass="com.sap.db.jdbc.Driver", classPath=driver_path)
# Define connection context
hana_url <- "db.airms.mssm.edu:30041"
database_name <- "AIRMS"
username <- "user-name"
connection_string <- sprintf( "jdbc:sap://%s/?databaseName=%s&user=%s
                               &encrypt=TRUE&validateCertificate=FALSE&sslHostnameInCertificate=%s
                               &connectTimeout=0&sslTrustStore=None",
                               hana_url, database_name, username, hana_url)
# Connect
conn <- dbConnect(jdbcDriver, connection_string, username,
                  password = getPass("Enter your password: "))
result <- dbGetQuery(conn, "SELECT TOP 10 * FROM CDMDEID.PERSON")
```

Working with AIR·MS in R vs R Studio on Minerva



Working with R

Launch an interactive R session:

```
$ bsub -q interactive -P your_account -n 1 -W 1:00 -R rusage[mem=8000] -XF  
-Is /bin/bash
```

Load R module and start R

```
$ ml R/4.2.0
```

```
$ R
```

Download with R Studio

Launch R studio via `$ minerva-rstudio-web-airms.sh`

Rstudio is started on compute node lh06c28, port 8788

Access the RStudio Web using your web browser: `http://10.95.46.94:53543 ...`

Change the hana_url from `hana_url <- "db.airms.mssm.edu:30041"` to

```
hana_url <- paste("localhost", Sys.getenv("db_port"), sep=":")
```

Querying AIR·MS - SQL

We will give an introduction to SQL in Session II!

Most basic components:

- **Schema**: Is like a folder – each stores a collection of tables that can be connected to each other via keys (e.g. patient mrns)
- **Clause**: control the structure of a query

```
SELECT * FROM CDMDEID.PERSON  
WHERE YEAR_OF_BIRTH > 1988
```

- **Functions**: perform calculations (AVG, COUNT, ROUND, MIN, MAX, ...)

```
SELECT AVG(2025-YEAR_OF_BIRTH) FROM CDMDEID.PERSON  
WHERE YEAR_OF_BIRTH > 1988  
  
GROUP BY GENDER
```

-

Mount Sinai Health System & Epic Electronic Health Record (EHR)

The Health Data Ecosystem



- The Mount Sinai Hospital was founded in 1852, one of the oldest and largest teaching hospitals in the United States
- The Mount Sinai Health System is one of the largest health systems in North America, and there are over 12 million patients in the MSHS information system



48,000
Employees

>12 Million
Patients

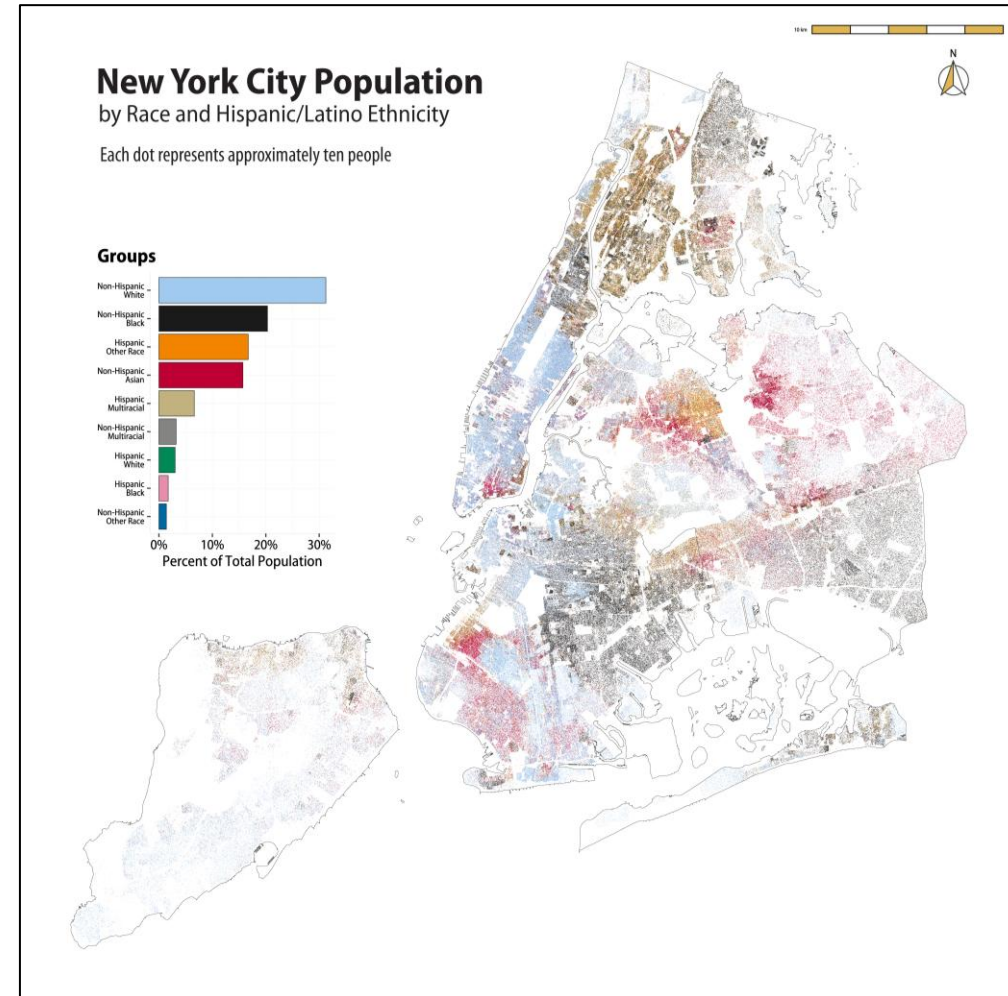
3,221 Beds

48 Institutes

7 Hospitals

New York City

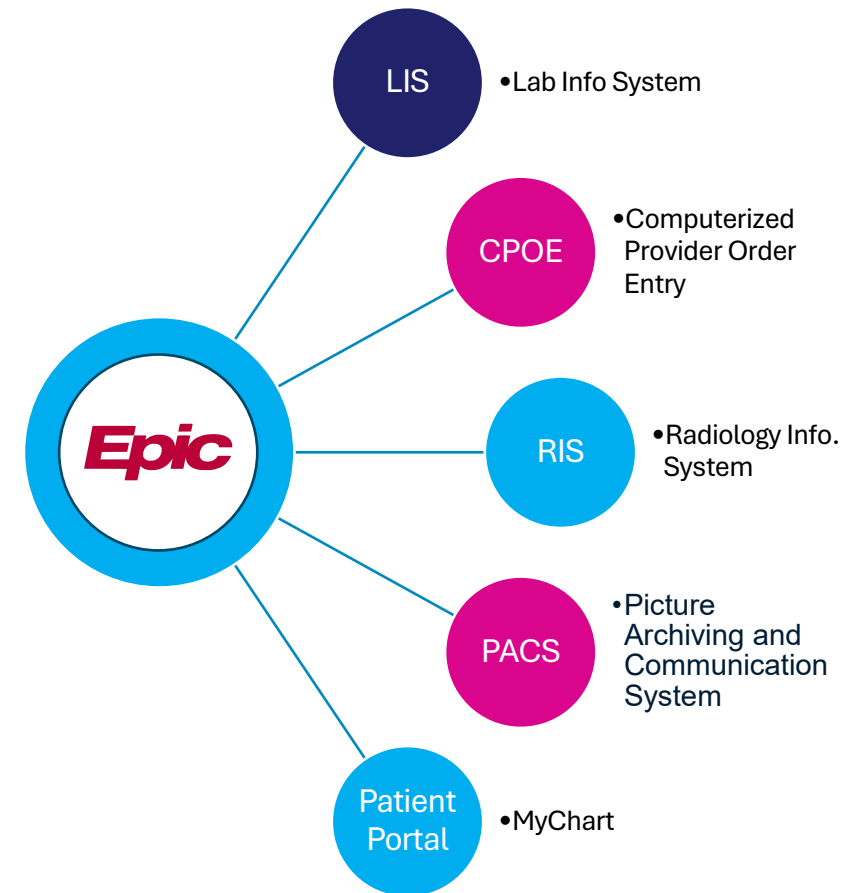
- New York is one of the largest, most diverse cities in the world
- This is reflected in Mount Sinai's patient population as well
- Mount Sinai services the New York City area
- Models and research developed here can be applied to many other locations



The Health Data Ecosystem: Epic

In MSHS, there are software systems we use:

- Electronic health record system (EHR) or electronic medical record system(EMR); biggest systems are Epic and PowerChart (Oracle Cerner)
- Picture Archiving and Communication System (PACS) – or storing imaging data like CTs and MRIs
- Patient Portals (like MyChart)
- At the state level – there are HIEs (health information exchanges)
- Epic is used in the MSHS



The Health Data Ecosystem: Epic

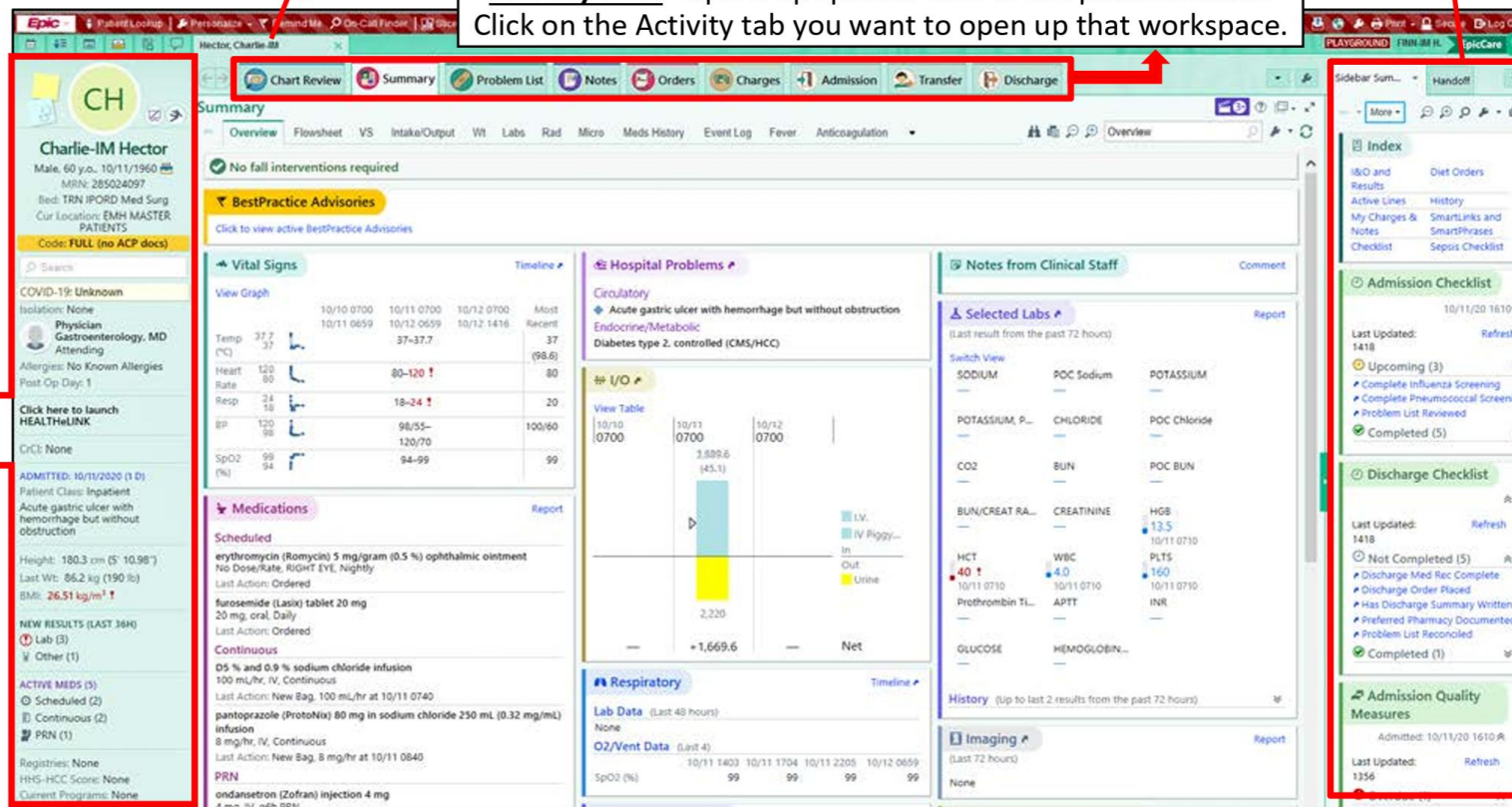
Story Board

This displays useful patient information and hyperlinks that open up another part of the chart when you click on it. Hover over something and it opens up a window to discover more information.

Patient Name

Activity Tabs - opens up specific view of the patients chart. Click on the Activity tab you want to open up that workspace.

Sidebar



Summary

Overview | Flowsheet | VS | Intake/Output | Wt | Labs | Rad | Micro | Meds History | Event Log | Fever | Anticoagulation

✓ No fall interventions required

BestPractice Advisories

Click to view active BestPractice Advisories

Vital Signs

	10/10 0700	10/11 0700	10/12 0700	Most Recent
Temp (°C)	37.7	37	37-37.7	37 (98.6)
Heart Rate	120	80	80-120	80
Resp	24	18	18-24	20
BP	120/98	98/55	120/70	100/60
SpO2 (%)	99	94	94-99	99

Hospital Problems

Circulatory
Acute gastric ulcer with hemorrhage but without obstruction

Endocrine/Metabolic
Diabetes type 2, controlled (CMS/HCC)

Notes from Clinical Staff

Selected Labs

	POC Sodium	POTASSIUM
SODIUM	135	4.0
POTASSIUM, P...	CHLORIDE	POC Chloride
CO2	BUN	POC BUN
BUN/CREAT RA...	CREATININE	HGB
HCT	WBC	PLTS
Prothrombin TL...	APTT	INR
GLUCOSE	HEMOGLOBIN...	

Respiratory

Lab Data (Last 48 hours)

O2/Vent Data (Last 4)

SpO2 (%)

	10/11 1403	10/11 1704	10/11 2205	10/12 0659
SpO2 (%)	99	99	99	99

Sidebar

Index

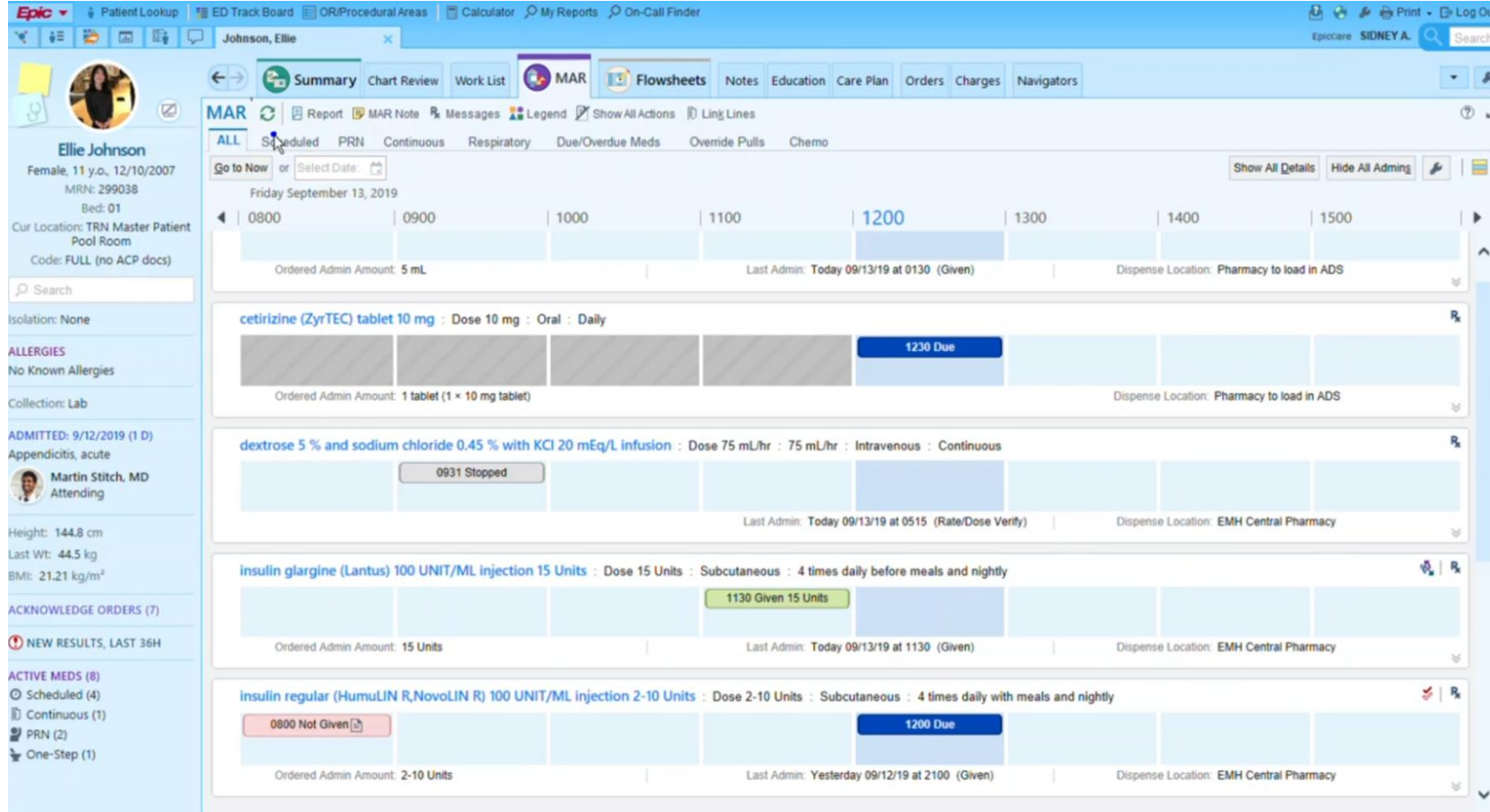
Admission Checklist

Discharge Checklist

Admission Quality Measures

A typical physician view in Epic, showing summary patient information

The Health Data Ecosystem: Epic



Ellie Johnson
Female, 11 y.o., 12/10/2007
MRN: 299038
Bed: 01
Cur Location: TRN Master Patient Pool Room
Code: FULL (no ACP docs)

ADMITTED: 9/12/2019 (1 D)
Appendicitis, acute
Martin Stich, MD
Attending

Height: 144.8 cm
Last Wt: 44.5 kg
BMI: 21.21 kg/m²

ACKNOWLEDGE ORDERS (7)
NEW RESULTS, LAST 36H

ACTIVE MEDS (8)
○ Scheduled (4)
○ Continuous (1)
○ PRN (2)
○ One-Step (1)

MAR (Medication Administration Record)
Report | MAR Note | Messages | Legend | Show All Actions | Link Lines

ALL | Scheduled | PRN | Continuous | Respiratory | Due/Overdue Meds | Override Pulls | Chemo

Go to Now or Select Date: Friday September 13, 2019

0800 | 0900 | 1000 | 1100 | 1200 | 1300 | 1400 | 1500

Ordered Admin Amount: 5 mL | Last Admin: Today 09/13/19 at 0130 (Given) | Dispense Location: Pharmacy to load in ADS

cetirizine (Zyrtec) tablet 10 mg : Dose 10 mg : Oral : Daily

Ordered Admin Amount: 1 tablet (1 x 10 mg tablet) | Dispense Location: Pharmacy to load in ADS

dextrose 5 % and sodium chloride 0.45 % with KCl 20 mEq/L infusion : Dose 75 mL/hr : 75 mL/hr : Intravenous : Continuous

0931 Stopped | Last Admin: Today 09/13/19 at 0515 (Rate/Dose Verify) | Dispense Location: EMH Central Pharmacy

insulin glargine (Lantus) 100 UNIT/ML injection 15 Units : Dose 15 Units : Subcutaneous : 4 times daily before meals and nightly

1130 Given 15 Units | Last Admin: Today 09/13/19 at 1130 (Given) | Dispense Location: EMH Central Pharmacy

insulin regular (Humulin R, NovoLIN R) 100 UNIT/ML injection 2-10 Units : Dose 2-10 Units : Subcutaneous : 4 times daily with meals and nightly

0800 Not Given | 1200 Due | Last Admin: Yesterday 09/12/19 at 2100 (Given) | Dispense Location: EMH Central Pharmacy

Nurses see different views than physicians – MAR (medication administration record) <https://www.youtube.com/watch?v=XTstvjyrcE>

Health Information Privacy Regulations

State

NYHIPA

SHIELD

Public Health 27F

...

National

HIPAA

GINA

NIST 800-171

...

International

GDPR (EU/UK)

PIPEDA (Canada)

HIPAA (USA)

...

Health Regulations: HIPAA

- Epic contains several modules and tens of thousands of data fields which capture data from doctors, nurses, and other health professionals
- Health information in the United States is governed by HIPAA regulations
- PHI = private health information
- Health information can be "de-identified" by review by an expert ("expert determination") or using "safe harbor" (removing 18 specific identifiers)

Five HIPAA Rules

HIPAA Privacy Rule
PHI Disclosure Rules



HIPAA Security Rule

Standards to safeguard ePHI



Omnibus Rule

Merges HITECH rules into HIPAA



Breach Notification Rule

60 Days to notify HHS



Enforcement Rule

How investigations are conducted

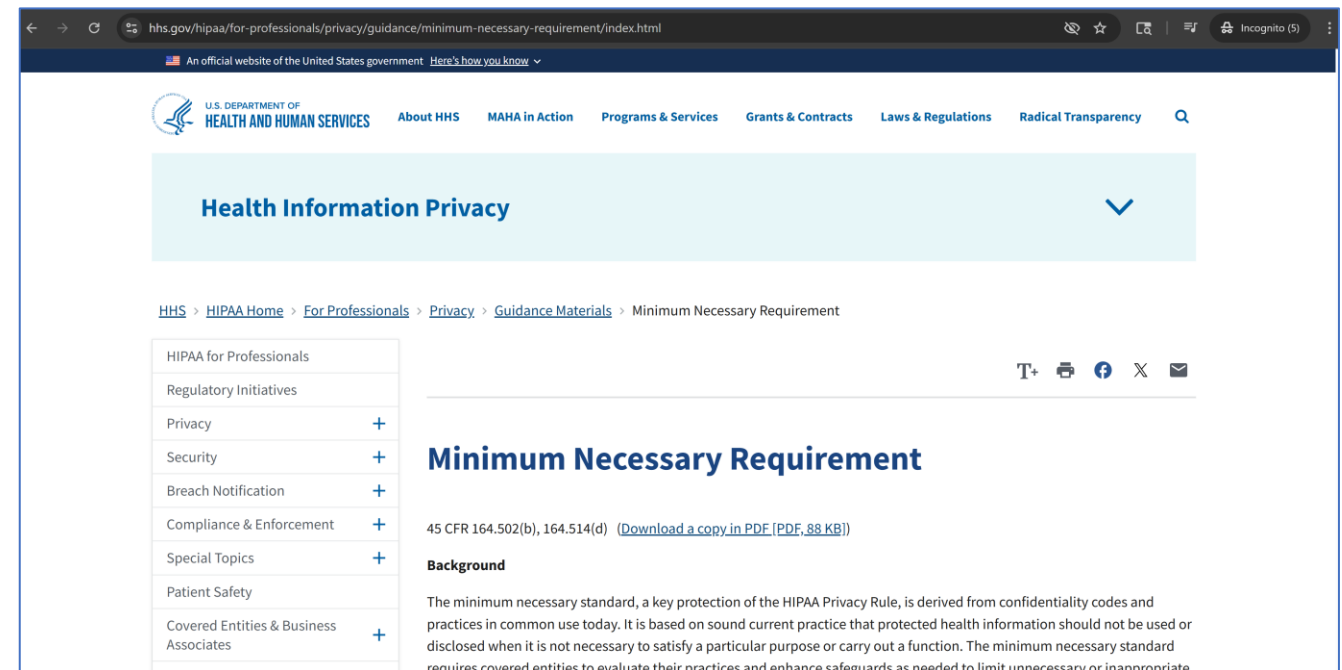


Copyright © 2018 The HIPAA Guide

<https://www.hipaaguide.net/hipaa-for-dummies/>

Health Regulations: HIPAA

- The "Minimum Necessary Requirement" states that researchers should only use the minimal amount of data necessary to address a research question or use case
- This means making sure data is only provided for well-defined inclusion / exclusion criteria for a specific question
- Exploring broad datasets usually happens later in a study after initial limited dataset delivery
- If there's an incident or data is leaked – reduces the number of people impacted



<https://www.hhs.gov/hipaa/for-professionals/privacy/guidance/minimum-necessary-requirement/index.html>

The Patient Journey & Epic

- Let's take a look at a typical patient walkthrough with Epic
- Key workflows: admission, discharge, and transfer (ADT)



The Patient Journey & Epic



Epic

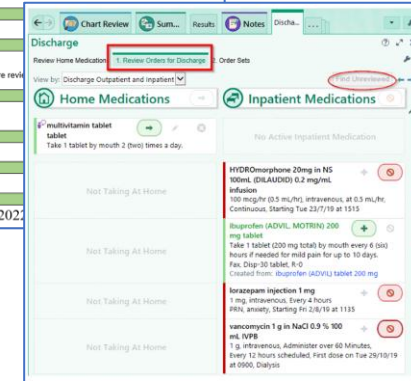
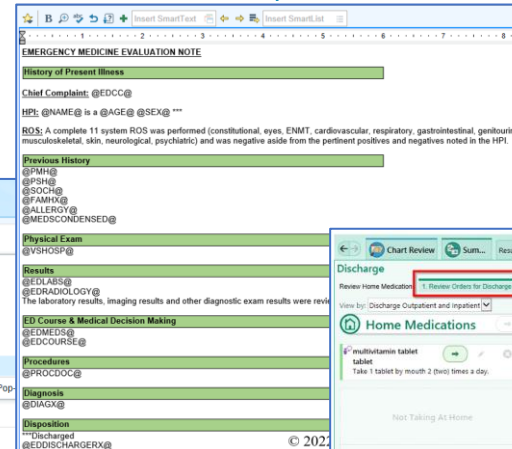
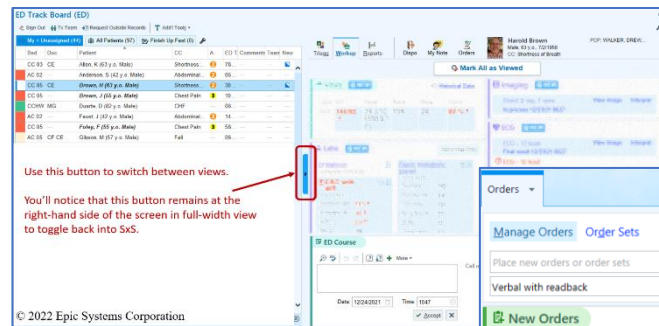
Admission

Orders

Document

Discharge

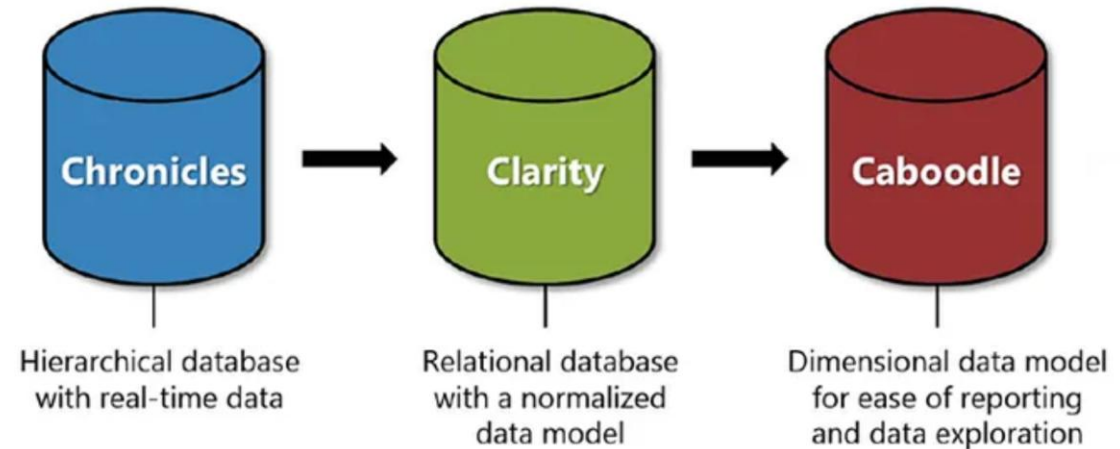
Pop. Health



- Margaret Smith is a 75 year old female who has arrived at the emergency room today with dyspnea (shortness of breath) and angina (chest pain)
- Patient has a previous history of pneumonia, anxiety

The Health Data Ecosystem

- Once we have all this data, it then goes into a series of databases in Epic
- Epic has multiple databases that store information
- SQL = structured query language – Clarity, Caboodle
- NOSQL = other types of databases (ex. hierarchical, graph)- Chronicles



<https://it.uclahealth.org/about/ohia/services/reports-dashboards/webi-dashboards>

AIR•MS & OMOP

Using the Data: Introduction to OMOP

- Data is stored in “OMOP” format
- OMOP = Observational Medical Outcomes Partnership
- After the initial work completed in 2013, the OMOP group then became the Observational Health Data Sciences and Informatics (OHDSI) group (pronounced “Odessey”)
- The common data model (CDM) for OHDSI is OMOP
- A significant project is the OMOP vocabulary
- In the US – OMOP leadership is located at Columbia U.
- Visit <https://ohdsi.org>
- Mount Sinai's OMOP data is maintained by Dr. Timothy Quinn (Chief Data Architect)



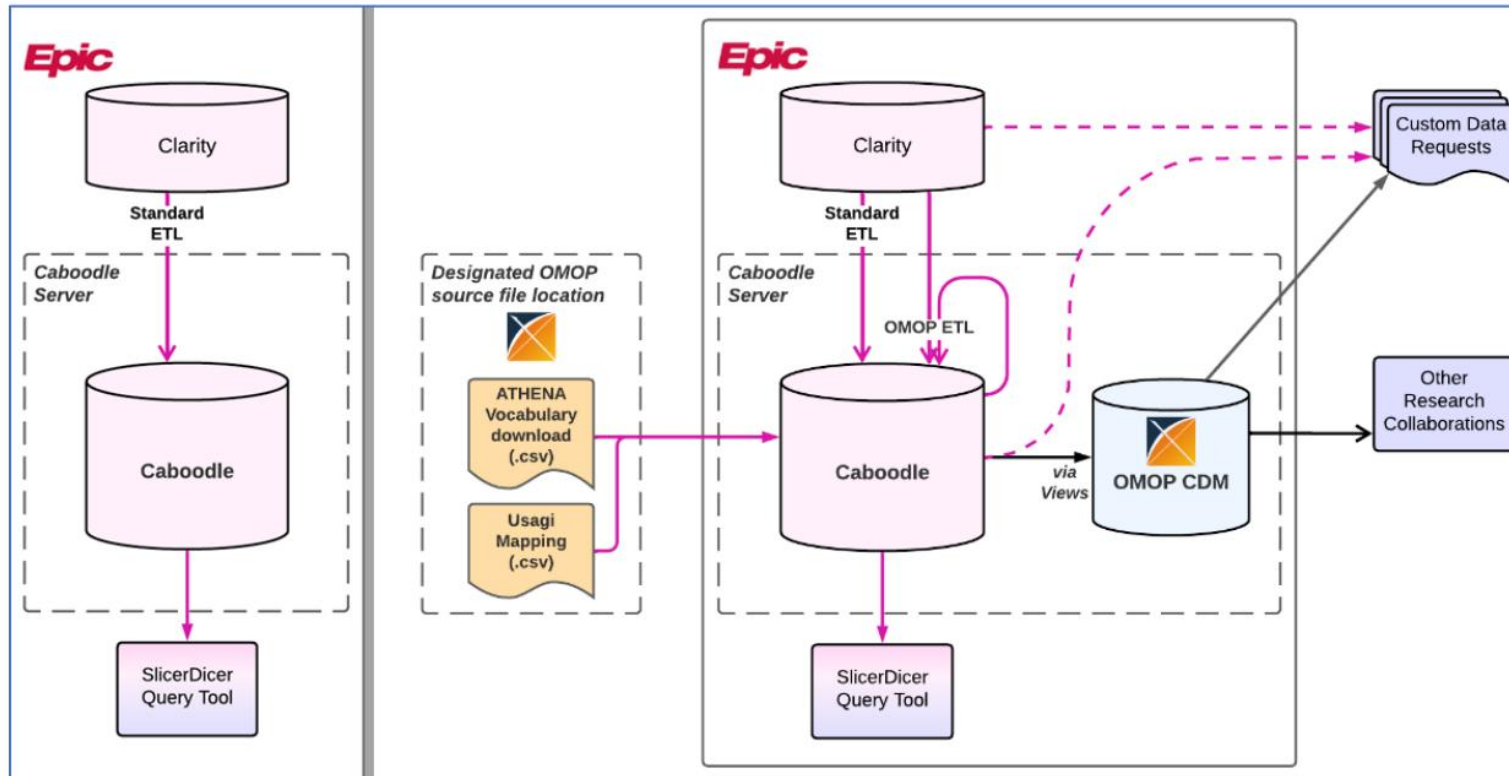
Using the Data: Who Uses OMOP?

- Several large collaborations use OMOP:



- Federal observational data research usually involves OMOP formatted data

Using the Data: How Does OMOP Work?

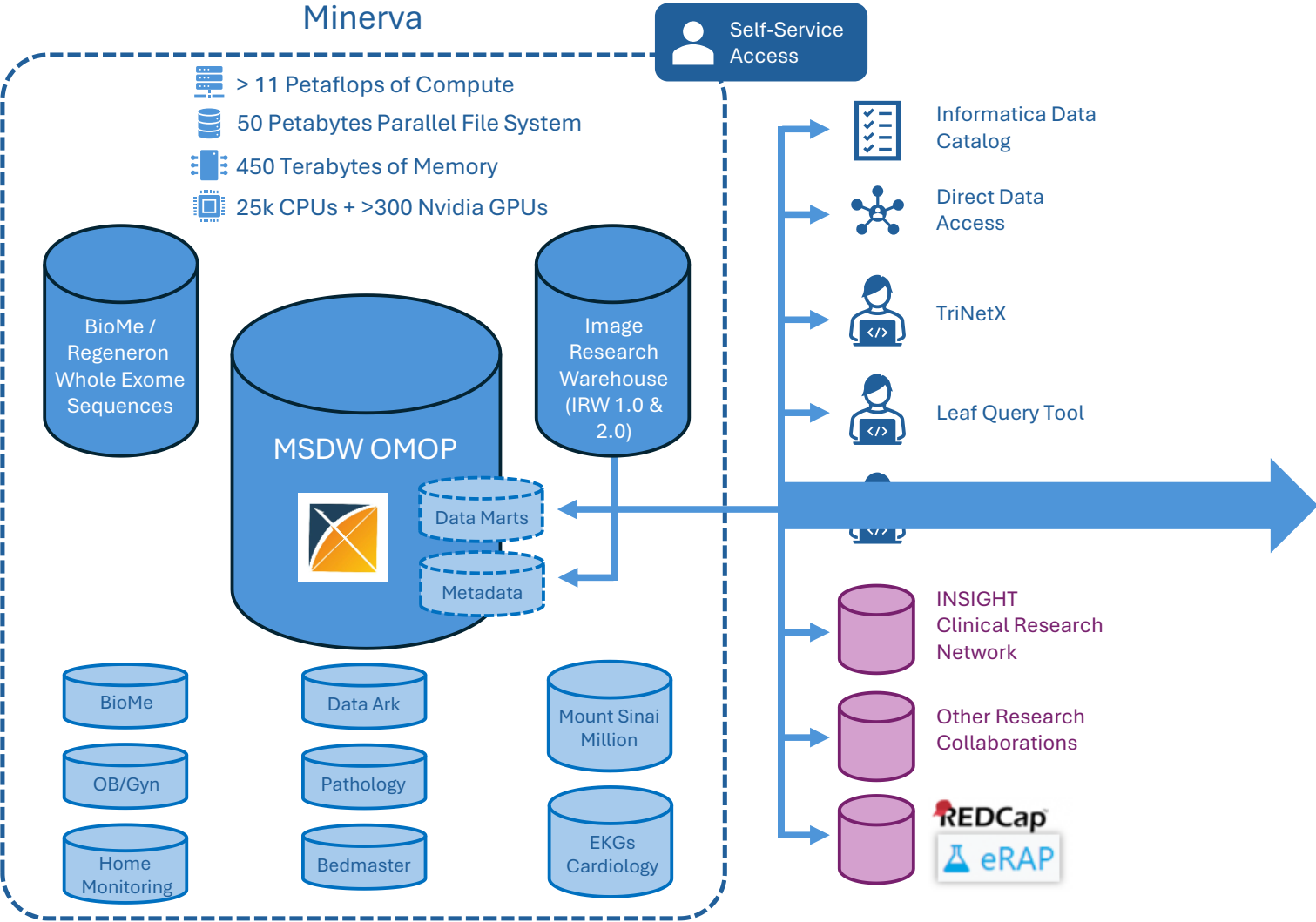
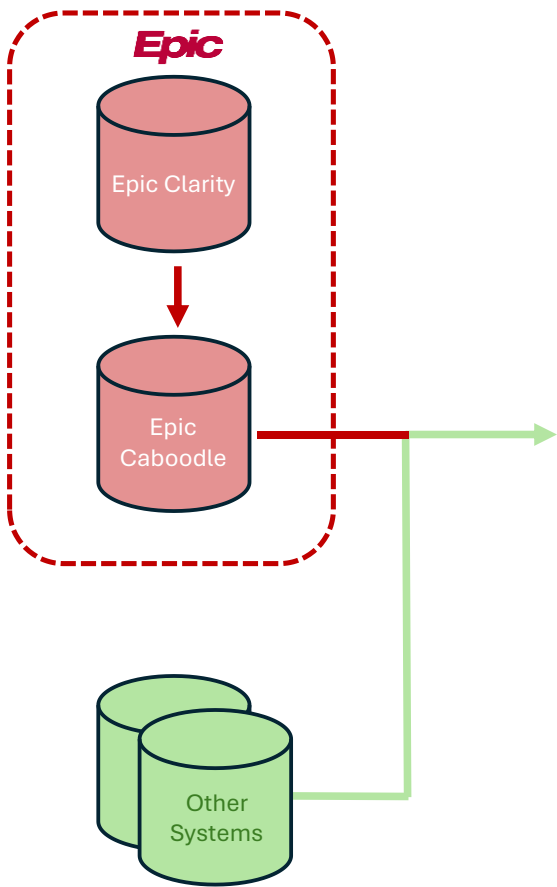


<https://www.ohdsi.org/wp-content/uploads/2023/10/10-Willett-BriefReport.pdf>

- Data is extracted from Caboodle into the OMOP Common Data Model (CDM)

Using the Data: How Does OMOP Work? (Cont'd)

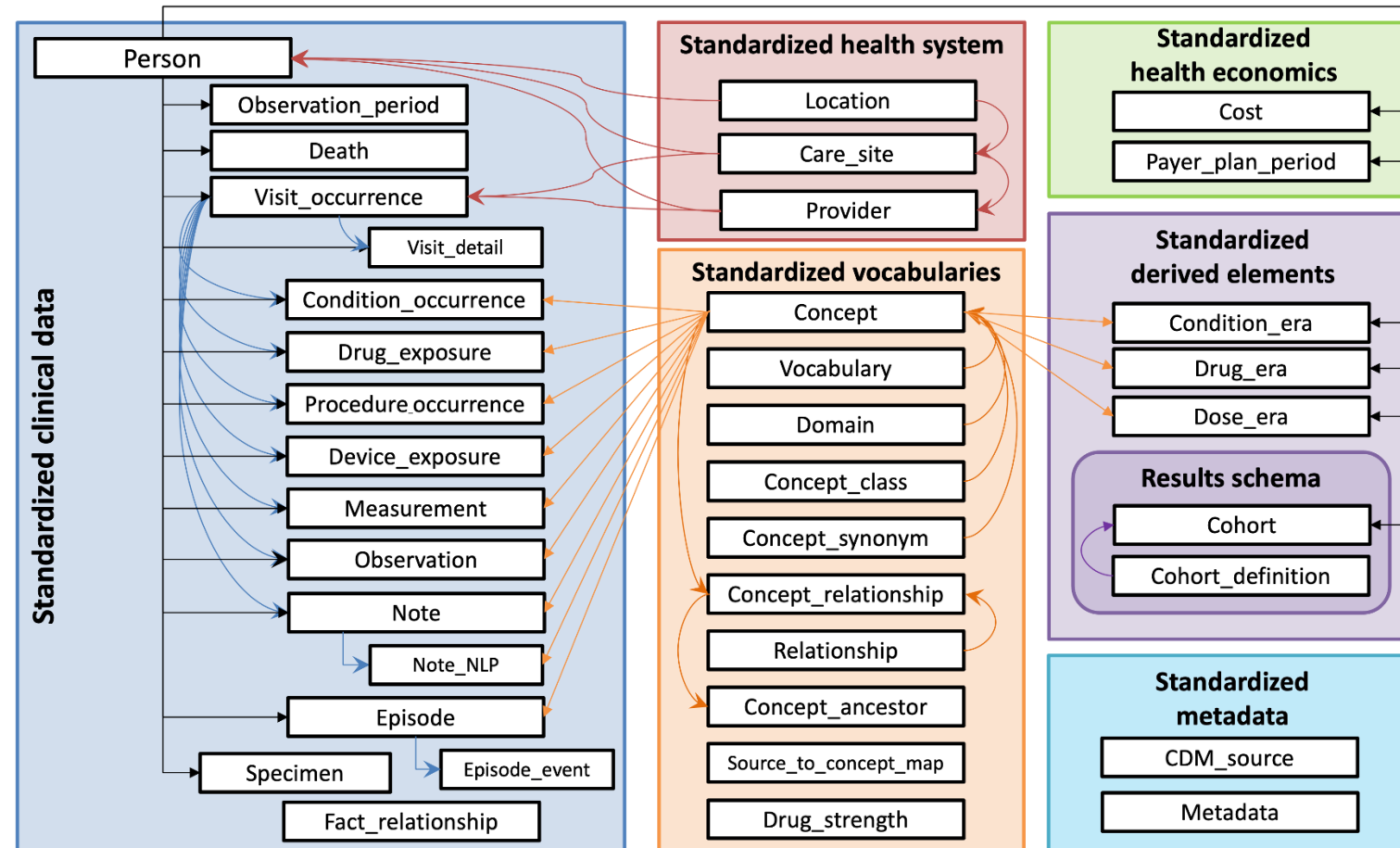
Electronic Health Record (EHR)



- Informatica Data Catalog
- Direct Data Access
- TriNetX
- Leaf Query Tool
- INSIGHT Clinical Research Network
- Other Research Collaborations
- REDCap eRAP

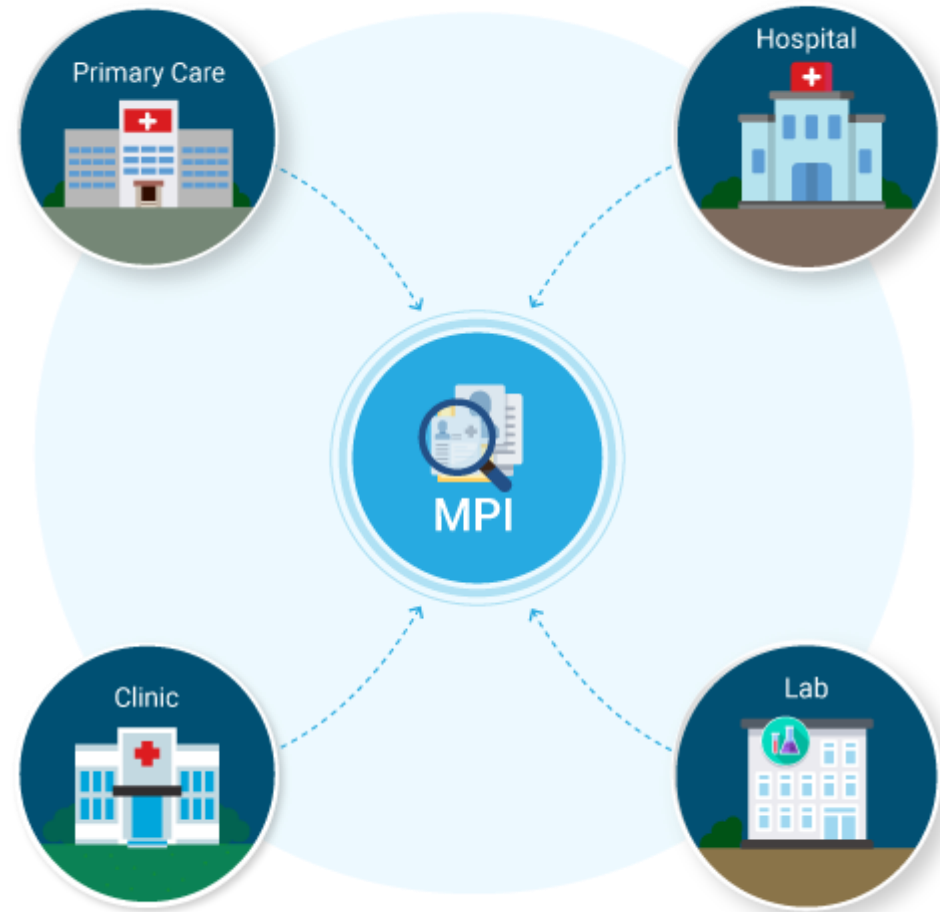
Using the Data: How Does OMOP Work? (Cont'd)

- The OMOP CDM consists of about 35 tables
- Very simplified compared to the ~20,000 tables in Clarity



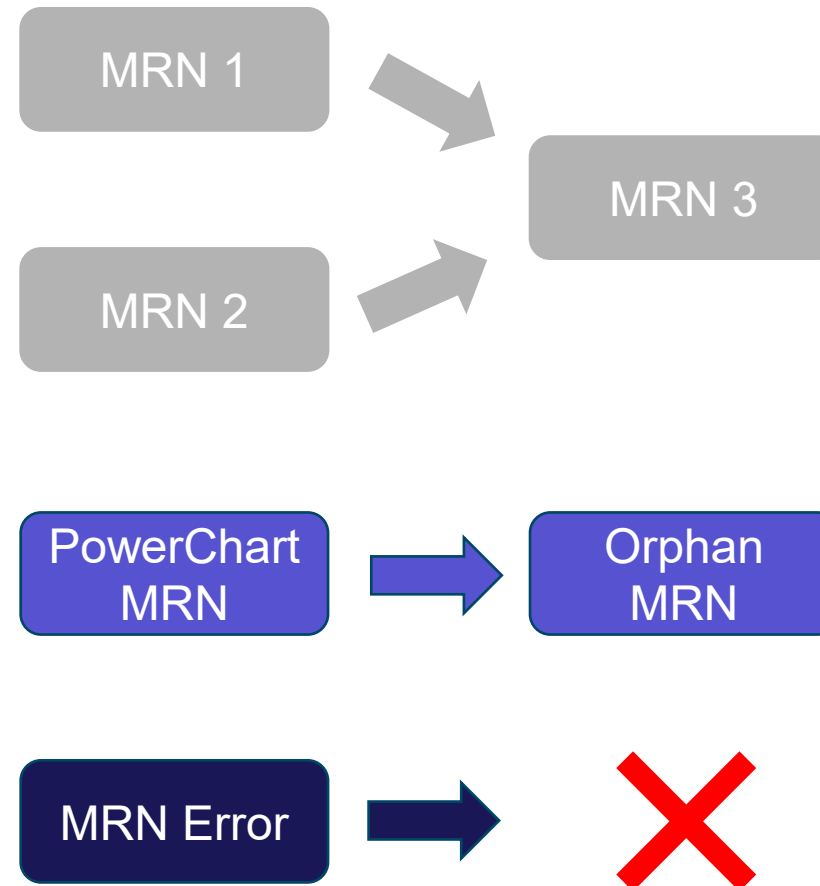
Medical Record Numbers (MRNs)

- Medical record numbers (MRNs) are the primary identifiers of patients
- They link information from across a health system
- Help with standardizing patient identifiers
- Identifiers can (1) merge, (2) be deleted, (3) be substituted/change
- These are tracked in the Master Patient Index (MPI)



Master Patient Index (MPI)

- In MSHS, several hospitals joined the system at different types, using different medical record systems (ex. PowerChart)
- Even though we have standardized over several years, there still may be old PowerChart MRNs in the system
- Additionally, some patient records have been merged across hospitals, so old identifiers may disappear



*Two MRNs
for same
patient from
separate
hospitals*

*Old EHR
System*

*Patient
Accidentally
Admitted*