

Introduction to AIR·MS

Health Data Literacy

Ashwin Sawant
Andrew Deonarine

Credits: Eugenia Allewa

April 7th, 2026



Hasso Plattner Institute for Digital Health at Mount Sinai

Welcome and Introduction



Patricia Kovatch
Professor and Dean for Scientific
Computing and Data



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

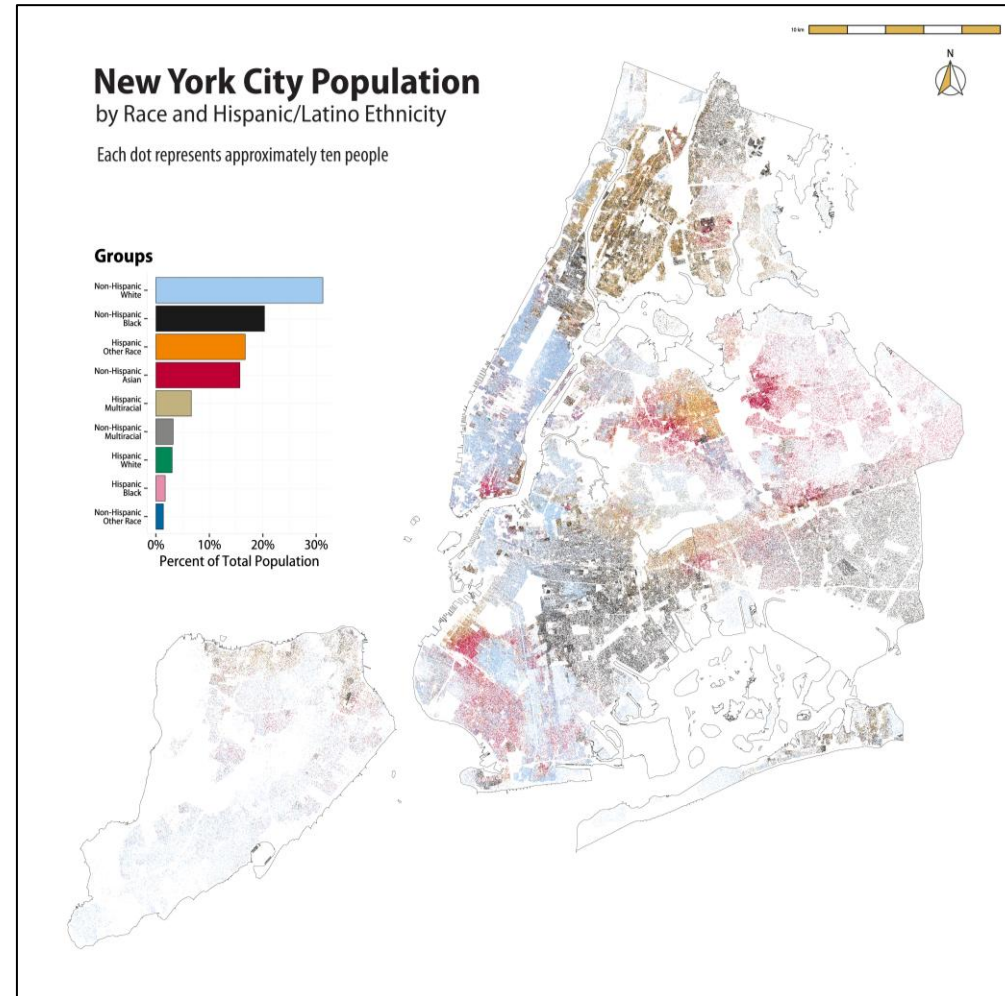


- The Mount Sinai Hospital was founded in 1852, one of the oldest and largest teaching hospitals in the United States
- The Mount Sinai Health System is one of the largest health systems in North America, and there are over 12 million patients in the MSHS information system



New York City

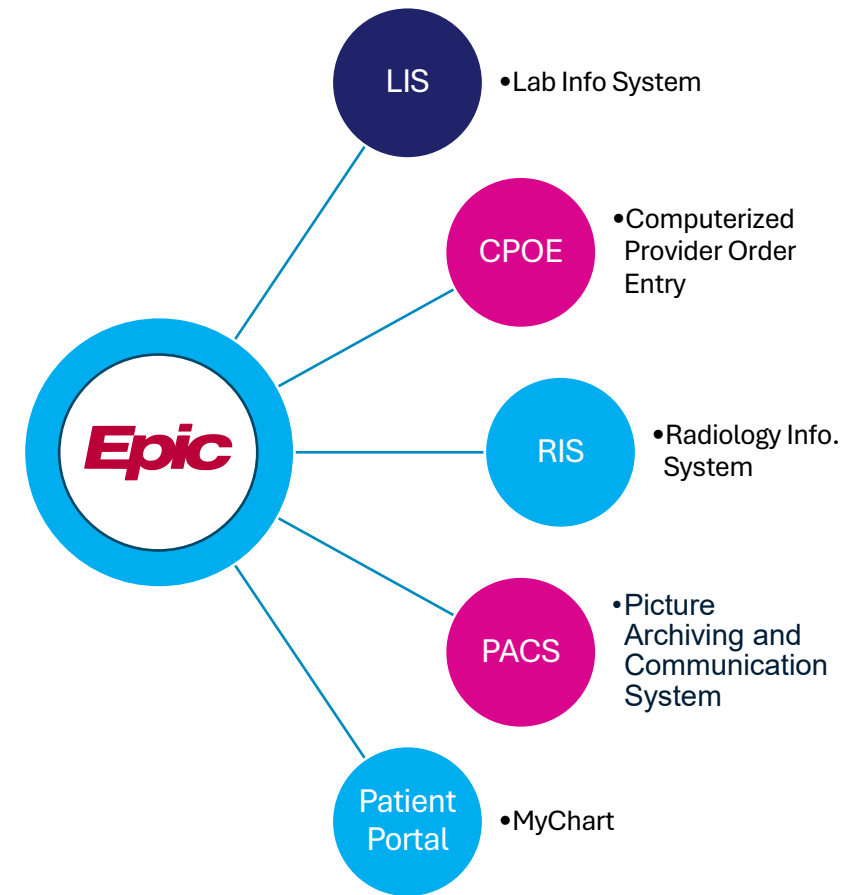
- New York is one of the largest, most diverse cities in the world
- This is reflected in Mount Sinai's patient population as well
- Mount Sinai services the New York City area
- Models and research developed here can be applied to many other locations



The Health Data Ecosystem: Epic

In MSHS, there are software systems we use:

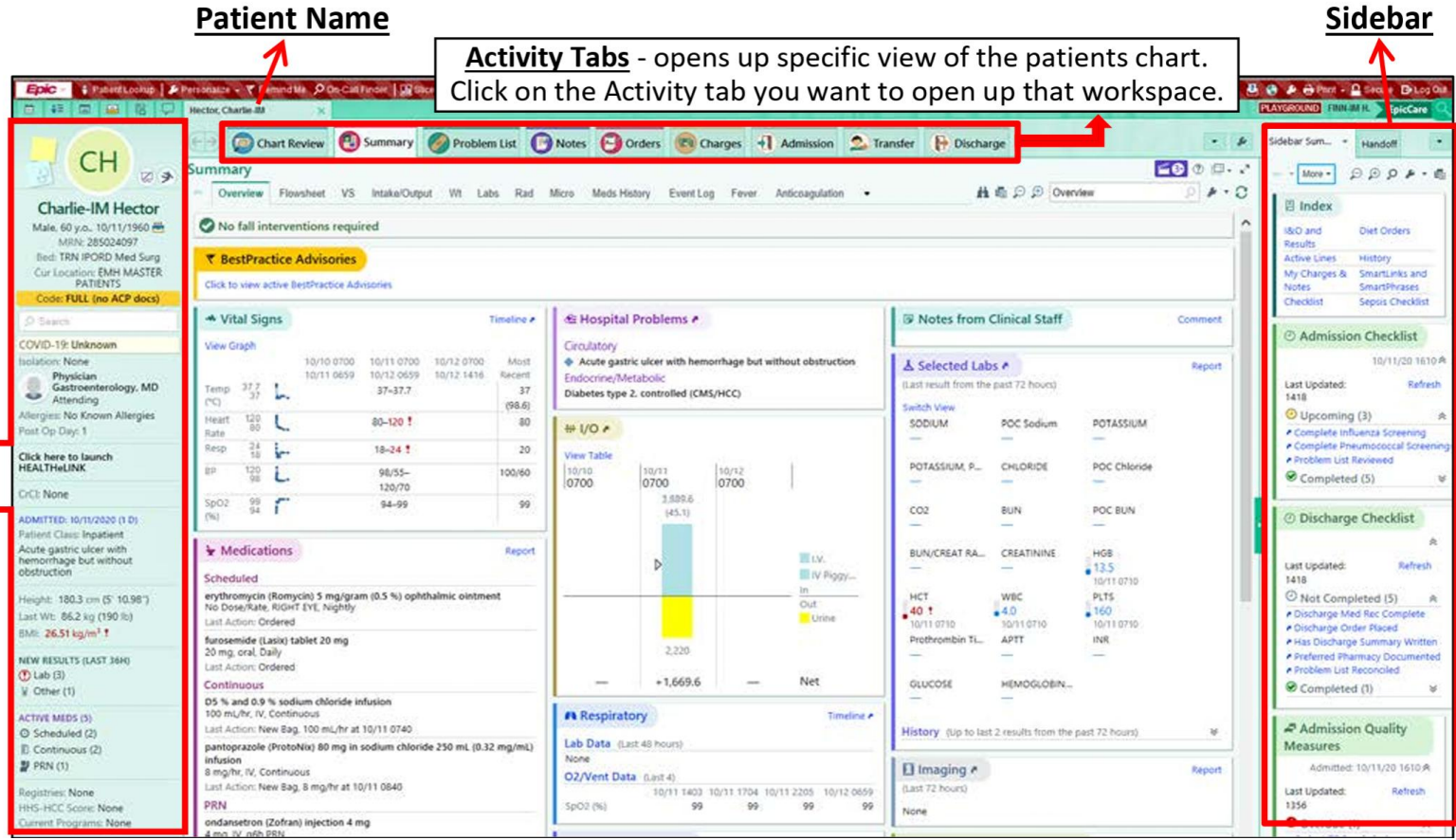
- Electronic health record system (EHR) or electronic medical record system(EMR); biggest systems are Epic and PowerChart (Oracle Cerner)
- Picture Archiving and Communication System (PACS) – or storing imaging data like CTs and MRIs
- Patient Portals (like MyChart)
- At the state level – there are HIEs (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 (points to Charlie-IM Hector)

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

Sidebar (points to the right sidebar)

Summary (points to the top navigation bar)

BestPractice Advisories (points to the BestPractice Advisories section)

Vital Signs (points to the Vital Signs table)

Hospital Problems (points to the Hospital Problems section)

Notes from Clinical Staff (points to the Notes from Clinical Staff section)

Selected Labs (points to the Selected Labs section)

Admission Checklist (points to the Admission Checklist section)

Discharge Checklist (points to the Discharge Checklist section)

Admission Quality Measures (points to the Admission Quality Measures section)

A typical physician view in Epic, showing summary patient information

The Health Data Ecosystem: Epic



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



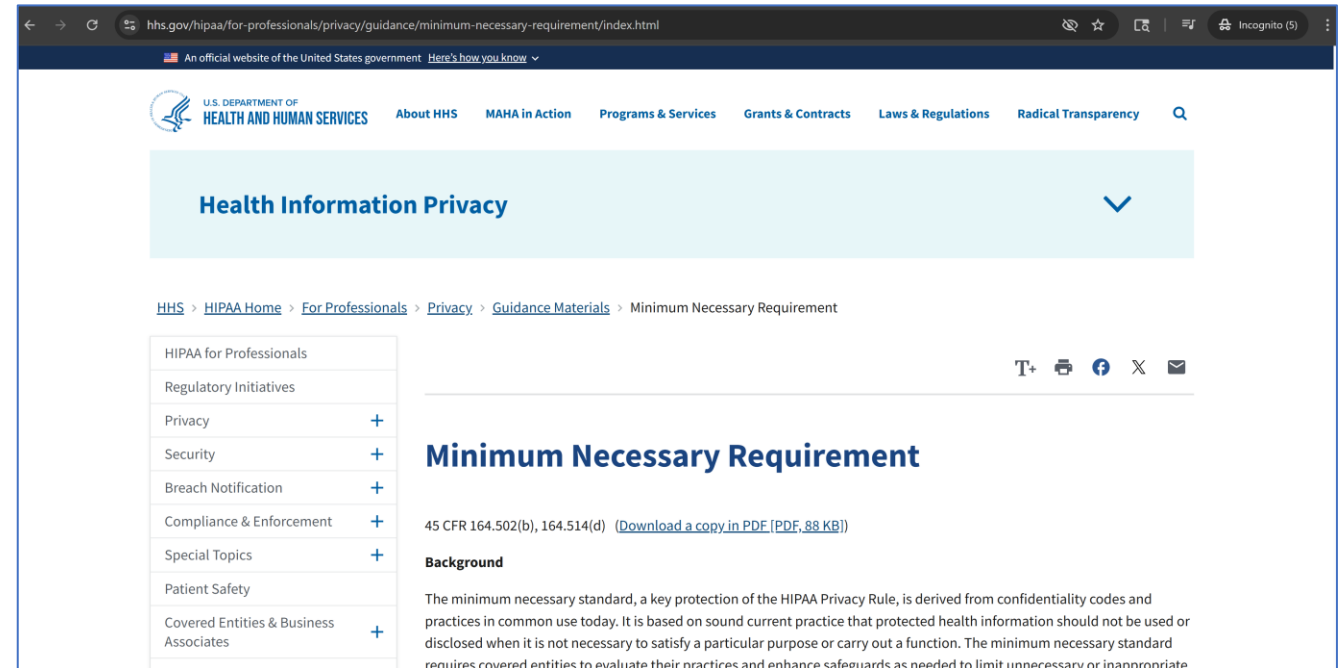
Copyright © 2018 The HIPAA Guide

<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)

