# Aug. 22 — OHDSI and Clinical Registries: Sanity for Health Systems

**Paul Nagy**
Program Director for Graduate Training in Biomedical Informatics and Data Science, Deputy Director of the Johns Hopkins Medicine Technology Innovation Center

**Lee Evans**
Founder, LTS Computing LLC

**DuWayne Willett**
Chief Medical Informatics Officer, University of Texas Southwestern Health System

**Jeff Weaver**
Director of Data Solutions for Emory University

@OHDSI     www.ohdsi.org     #JoinTheJourney     ohdsi

# What is a clinical registry?

A clinical registry is a database that collects and organizes information about a specific group of patients, such as those with a particular disease or condition, who have received a particular treatment.

Cohort Definition

Standard Data Specification

# How many clinical registries are there?

There are thousands of clinical registries in the world. There is not a central registry of clinical registries. WHO has a list of 200.

| Disease | Surveillance |
|---|---|
| Outcome | Pharmacovigilance |

# Clinical registries have a high human cost in chart abstraction

## Trauma registry methodology: A survey of trauma registry custodians to determine current approaches

Gerard M. O'Reilly [a,b,*], Belinda Gabbe [a], Peter A. Cameron [a,b,c]

[a] Department of Epidemiology and Preventive Medicine, School of Public Health and Preventive Medicine, Monash University, Commercial Rd, Melbourne, 3004, Australia
[b] Emergency and Trauma Centre, Alfred Health, Commercial Rd, Melbourne, Victoria 3004, Australia
[c] Emergency Services, Hamad Medical Corporation, Doha, Qatar

0.5 FTE for every 200-300 patients.

**Table 3**
Human resources–Single hospital registries (n = 40).

| Staffing type | Number of persons | | | |
|---|---|---|---|---|
| | 0 | 1 | 2 | >2 |
| All (total) staff | 0 | 2 | 4 | 34 |
| Director/Head | 14 | 25 | 1 | 0 |
| Manager | 18 | 21 | 1 | 0 |
| Data manager | 17 | 21 | 0 | 2 |
| Database programmer | 33 | 2 | 4 | 1 |
| Database analyst | 31 | 7 | 1 | 1 |
| Trauma nurse coordinator | 17 | 15 | 6 | 2 |
| Data collector | 12 | 2 | 4 | 22 |
| Data entry clerk | 23 | 9 | 2 | 6 |
| ICD coder | 26 | 4 | 0 | 10 |
| AIS coder | 21 | 4 | 2 | 13 |
| Data analyst | 28 | 9 | 2 | 1 |
| Office administrator | 29 | 11 | 0 | 0 |

# Standardizing registry data to the OMOP Common Data Model: experience from three pulmonary hypertension databases

Patricia Biedermann, Rose Ong, Alexander Davydov, Alexandra Orlova, Philip Solovyev, Hong Sun, Graham Wetherill, Monika Brand & Eva-Maria Didden ✉

Mapping registry data to the OMOP CDM facilitates more efficient collaborations between researchers and establishment of federated data networks

# Goal 1: Clinical Registry OHDSI feasibility checklist

| Question | Example | Answer |
|---|---|---|
| How many of the data elements correspond to core OMOP concepts? | Meds, Labs, Procedures, Conditions, Devices | Out of the Box |
| How many of data elements will need custom transformation scripts? | Vitals, Epic Flowsheets, Epic SmartForms | Minor Effort |
| How many of the data elements will need a custom concept? | Questions that are not in any standard lexicon. | Minor Effort |
| Will the OMOP data model needs to be extended to support the registry? | Imaging | Significant Effort |
| How many of the data elements will require NLP? | Symptoms at the time of admission. | Significant Effort |

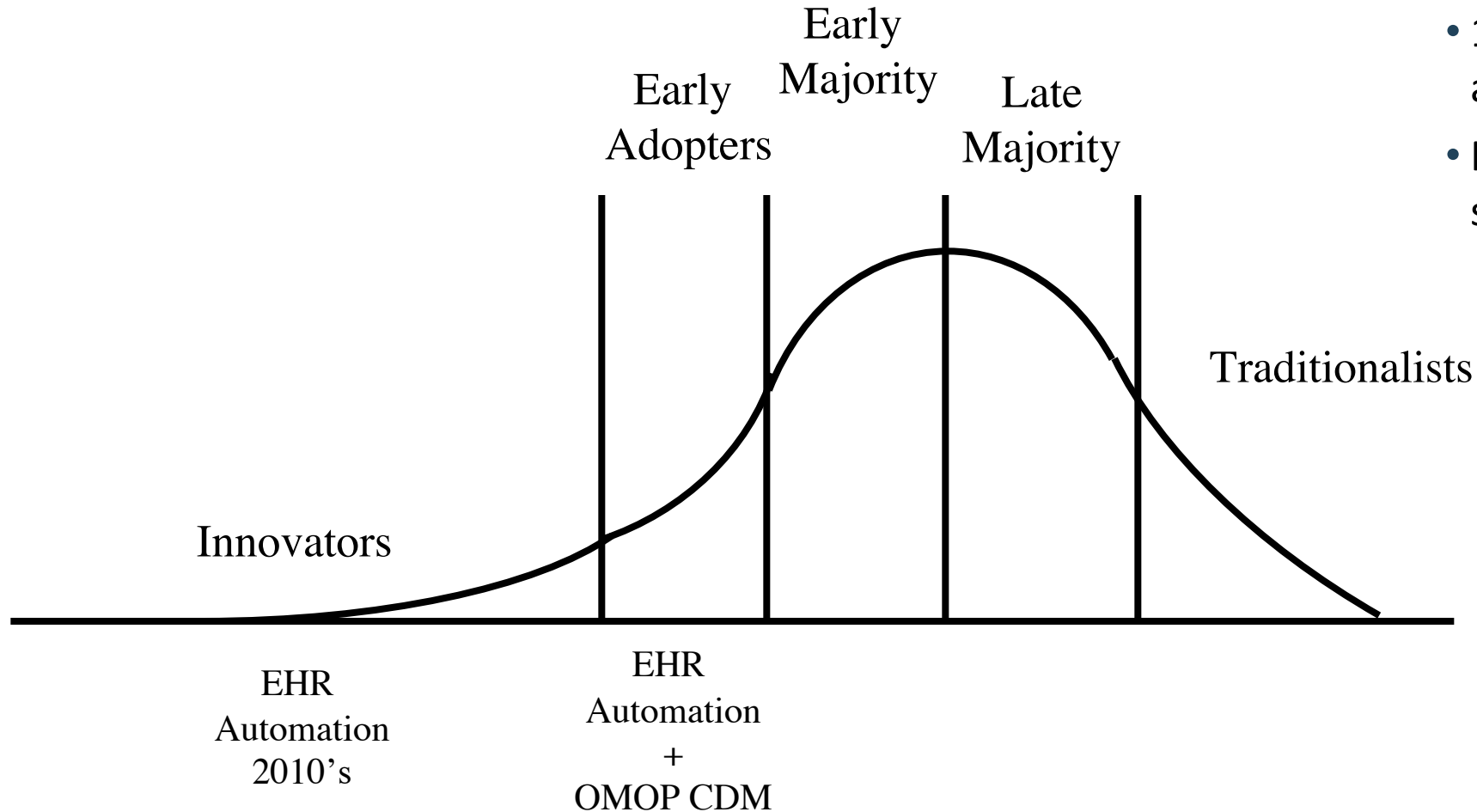# OHDSI Based Clinical Registries

## Join the OHDSI Clinical Registries Working Group to learn how to transform your registry! Ohdsi.org

1. NIH All of US
2. UK Biobank
3. National Covid Collaborative Consortium (N3C)
4. Cure Infectious Disease (Cure ID)
5. Registre National du Cancer du Luxembourg
6. CancerDataNet
7. American Society of Hematology Research Collaborative
8. AOA National Joint Replacement Registry
9. Netherlands Cancer Registry
10. Finnish Hematology Registry/ HUS

11. Basilicata Cancer Registry
12. Prostate Cancer Registry of South West Finland
13. Norwegian Cancer Registry
14. European Rare Kidney Disease Registry
15. Geneva Cancer Registry
16. Advocate Aurora Health & University of Madison Health Non-Muscle Invasive Bladder Cancer
17. Sloan Kettering Cancer Center Surveillance, Epidemiology, and End Results Program (SEER): B-Cell
18. IQVIA- MMI Specialty EMR

# EHR automation for Clinical Registries
# Diffusion of innovation model



- 1943 model of Iowa farmers adoption of hybrid seeds
- Five personality and socioeconomic categories

Early Majority

Early Adopters

Late Majority

Traditionalists

Innovators

EHR Automation 2010's

EHR Automation + OMOP CDM

Factors
- Reference implementations
- Adoption of standards
- De-risking the initial and ongoing costs
- Reducing technical barriers
- Availability of skilled people
- Clear financial incentives

# CURE ID and Virus COVID-19 Registry

- **CURE ID**
  - Joint initiative between FDA, NIH/NCATS, Critical Path Institute
  - Aims to identify repurposed drug candidates to treat infectious diseases
  - Started as an online/app-based registry for clinicians to enter case reports
  - COVID-19 expands mandate: automated extraction from electronic health records
- **Viral Infection & Respiratory Illness Universal Study (VIRUS) COVID-19 Registry**
  - Launched by the Society of Critical Care Medicine within weeks of pandemic onset
  - Rapidly described COVID-19 clinical course
  - Global reach to 306 sites in 28 countries
  - Started with manual data entry – hundreds of variables, ~4 hours per patient

# Cure ID
## Next Generation Clinical Registry

- Build in synergy with OMOP and the OHDSI Community
  - Teams group under Health System Interest Group
  - Github repo under OHDSI
- Build capacity at health systems to lower the cost of the ETL and management of OMOP data.
- Encourage the use of the OHDSI software analytics tools.
  - Atlas, DQD, Perseus, HADES
- Invest in the open-source tool development to lower the cost of ETL for sites.
  - Broadsea, Perseus

# Finding respiratory support devices in flowsheets

- Find flowsheets entries (context and entry names) by partial string match & inspection of name & value

- Generate frequency table of source values – trim tail

| Value | Count | Concept ID |
|---|---|---|
| Nasal cannula | 95231 | 4224038 |
| Ventilator | 23183 | 45768197 |
| High flow | 5912 | 4139525 |
| …. | 23 | Trim tail |

- Map to OMOP Concepts

| Flowsheet name (context) | Entry name | Display name | Value |
|---|---|---|---|
| Inpat vitals | LH Resp Dev | Oxygen therapy | Nasal cannula |
| Inpat vitals | RN HR | Pulse | 78 |
| RT Doc | O2 Delivery H32 | Oxygen therapy | BiPAP |
| Inpat vitals | O2 flowrate | Flow L/min | 2 |
| Inpat vitals | FS Gluc | Fingerstick | 425 |
| OR Anesthes | O2 Delivery H32 | Oxygen therapy | Ventilator |
| OR Anesthes | LH Resp Dev | Oxygen therapy | Ventilator |
| RT Doc | RT Secr | Secretions | Thick |
| OR Anesthes | Cath 916 | Cath position | R ventricle |
| OR Anesthes | FS Gluc | Fingerstick | 193 |
| Inpat vitals | O2 flowrate | Oxygen therapy | 4 |

# JHU OHDSI Data Core

Johns Hopkins University

Paul Nagy, PhD

Khyzer Aziz, MD

Matt Robinson, MD

Danielle Boyce, DPA

Will Garneau, MD

Michael Cook

Tanner Zhang, MD

Ben Martin, PhD

Steve Miller, MD

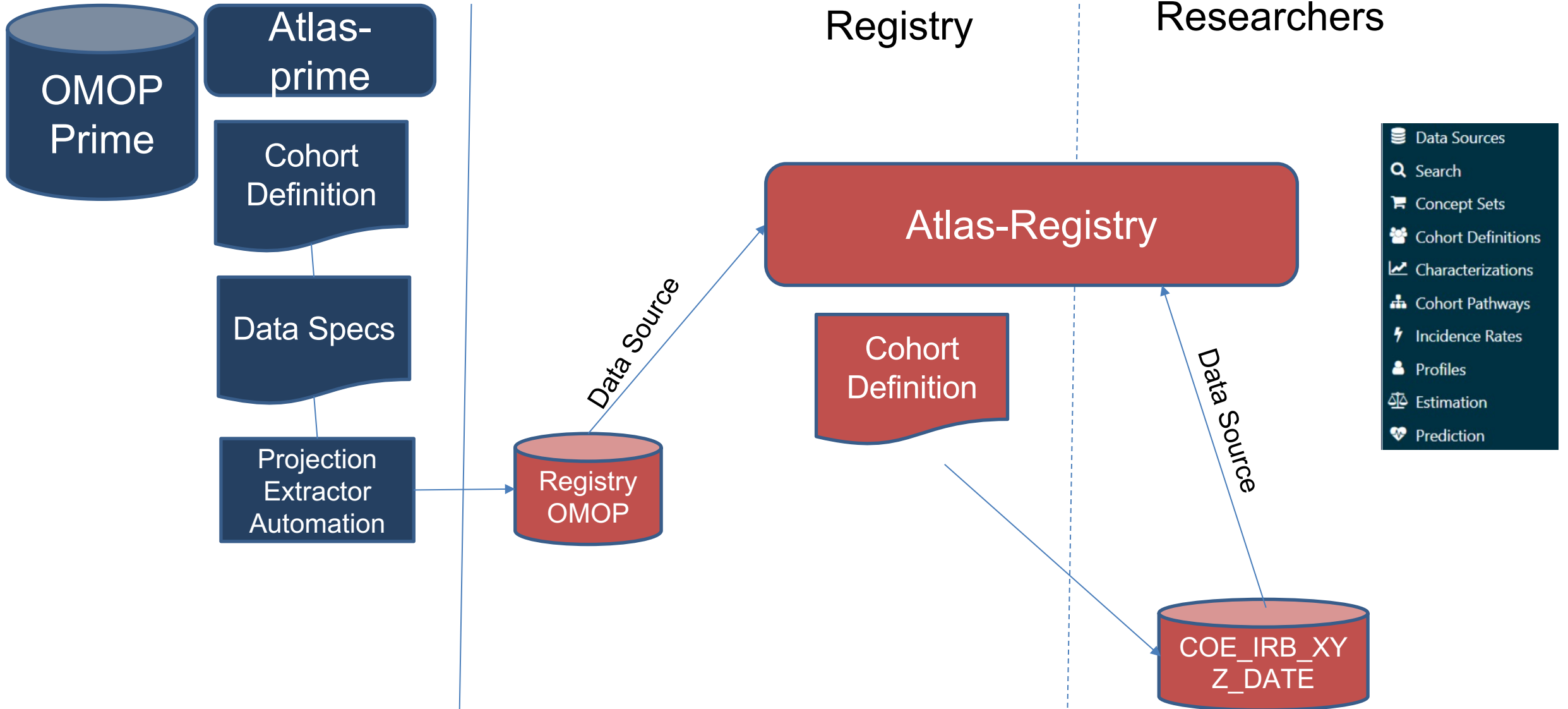OHDSI Community Experts

Lee Evans (Broadsea)

Nate Buesgens (Perseus)

Katy Sadowski - Trialspark (DQD)

Roger Carlson - Spectrum (Epic ETL)

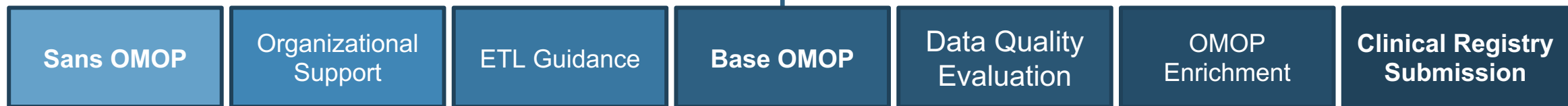Janos Hajagos, Phd - Stony Brook Medicine (Cerner)

# Registry Subsetting

# Data Partner Journey

OHDSI — OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS
Data Partner Journey

| Sans OMOP | Organizational Support | ETL Guidance | Base OMOP | Data Quality Evaluation | OMOP Enrichment | Clinical Registry Submission |
|---|---|---|---|---|---|---|
| • Intro to OMOP<br>• Connection with OHDSI community | • Connecting clinical and data teams<br>• IRB guidance<br>• Training and education from OHDSI members | • Understanding their data<br>• Guidance with OHDSI tools (Broadsea, Usagi) | • OMOP tables built<br>• Cohort creation & deidentification processes run | • Increasing capability to audit and refine data | • Flowsheets | Ready to participate in network studies |

# Top Issues from sites

1. The clinical principal investigator (Pulmonology and Critical Care) had no idea how to find their enterprise OMOP team.

2. The "OMOP Team" was not prepared to provide OMOP datasets as a service. Their OMOP instance had minimal clinical domain coverage, not refreshed, not supported, not validated, deemed a research project.  Gap between team and enterprise org.

3. The analytics/architecture teams didn't have a known pathway to bring in software and were constrained to traditional SQL environments. Gap between analytics and IT groups.

4. Mapping source concepts due to disconnect between analytics team and clinicians.  Gap between analytics and clinicians.

**DuWayne Willett**

Chief Medical Informatics Officer, University of Texas Southwestern Health System

Drawing Reproducible Conclusions from Observational Clinical Data with OHDSI
George Hripcsak, et al. Yearbook of Medical Informatics 2021; 30(01): 283-289
https://www.thieme-connect.de/products/ejournals/html/10.1055/s-0041-1726481

Filter components...

NavigatorSectionDim
NumbersDim
OmopCareSiteDimX
OmopLocationDimX
OmopPersonDimX
    OmopDeathDimX
    OmopObservationPeriodFactX
OmopProviderDimX
OmopVisitOccurrenceFactX
    OmopConditionOccurrenceFactX
    OmopDeviceExposureFactX
    OmopDrugExposureFactX
    OmopMeasurementFactX
    OmopNoteFactX
    OmopObservationFactX
    OmopProcedureOccurrenceFactX
    OmopVisitDetailFactX
OmopVocabularyRawDataX
    OmopConceptAncestorRawDataX
    OmopConceptClassRawDataX
    OmopConceptRawDataX
    OmopConceptRelationshipRaw...
    OmopConceptSynonymRawDat...
    OmopDomainRawDataX
    OmopDrugStrengthRawDataX
    OmopRelationshipRawDataX
    OmopSourceToConceptMapRa...
OpioidOrderDataMart
OrdersetDim

# OmopConditionOccurrenceFactX

Summary | **ER Diagram**

**DateDim**

| DateKey | PK | bigint |
| Other Columns ▷ | | |

**PatientDim**

| PatientKey | PK | bigint |
| DurableKey | DK | bigint |
| Other Columns ▷ | | |

**OmopConditionOccurrenceFactX**

| OmopConditionOccurrenceKey | PK | bigint |

Foreign Keys (a)

| ConditionEndDateKey | bigint |
| ConditionStartDateKey | bigint |
| EncounterKey | bigint |
| PatientKey | bigint |
| PatientDurableKey | bigint |
| ProviderKey | bigint |
| ProviderDurableKey | bigint |

Other Columns ▽ (b)

| ConditionConceptId | int |
| ConditionStartDate | date |
| ConditionStartDatetime | datetime |
| ConditionEndDate | date |
| ConditionEndDatetime | datetime |
| ConditionTypeConceptId | int |
| ConditionStatusConceptId | int |
| StopReason | nvarchar(50) |
| VisitDetailId | int |
| ConditionSourceValue | nvarchar(50) |
| ConditionSourceConceptId | int |
| ConditionStatusSourceValue | nvarchar(50) |
| RowSourceTable | nvarchar(50) |
| RowSourceKeyX | bigint |
| Count | tinyint |
| _CreationInstant | datetime |
| _LastUpdatedInstant | datetime |
| _IsInferred | tinyint |
| _IsDeleted | tinyint |
| _PrimaryPackageToImpactRecord | nvarchar(50) |
| _MostRecentPackageToImpactRecord | nvarchar(50) |
| _NumberOfSources | smallint |
| _HasSourceCaboodle | tinyint |
| _DeletedFromSourceCaboodle | tinyint |

(c)

**EncounterFact**

| EncounterKey | PK | bigint |
| Other Columns ▷ | | |

**ProviderDim**

| ProviderKey | PK | bigint |
| DurableKey | DK | bigint |
| Other Columns ▷ | | |

# Nightly Execution

## Nightly

### Execution Details

⚠️ **Nightly**

**Properties**

| | | | |
|---|---|---|---|
| Start time | 07/16/2023 01:05:21 | Duration | 3h 34m |
| End time | 07/16/2023 04:39:45 | Processing time | 32h 58m |
| Status | Completed | Rows extracted | 1,102,155,498 |
| Priority | Nightly | Records created | 8,469,646 |
| SQL Threads | 9 | Records modified | 54,621,612 |
| SSIS Threads | 16 | Records deleted | 63,111 |
| Extract mode | Default | Record pairs matched | 0 |
| Created from | Nightly | | |

| | | |
|---|---|---|
| ▷ ✓ NavigatorSectionDim | 07/16/2023 01:44:42 | < 1m |
| ▷ ✓ NppesInfo | 07/16/2023 01:10:25 | < 1m |
| ▷ ✓ OmopCareSiteDimX | 07/16/2023 01:44:11 | < 1m |
| ▷ ✓ OmopConceptAncestorRawDataX | 07/16/2023 01:29:30 | 11m |
| ▷ ✓ OmopConceptClassRawDataX | 07/16/2023 01:32:44 | < 1m |
| ▷ ✓ OmopConceptRawDataX | 07/16/2023 01:22:58 | 8m |
| ▷ ✓ OmopConceptRelationshipRawDataX | 07/16/2023 01:28:01 | 8m |
| ▷ ✓ OmopConceptSynonymRawDataX | 07/16/2023 01:29:30 | 1m |
| ▷ ✓ OmopConditionOccurrenceFactX | 07/16/2023 04:23:34 | 1h 47m |
| ▷ ✓ OmopDeathDimX | 07/16/2023 01:49:55 | 4m |
| ▷ ✓ OmopDeviceExposureFactX | 07/16/2023 02:35:00 | 3m |
| ▷ ✓ OmopDomainRawDataX | 07/16/2023 01:32:57 | < 1m |
| ▷ ✓ OmopDrugExposureFactX | 07/16/2023 04:31:45 | 1h 52m |
| ▷ ✓ OmopDrugStrengthRawDataX | 07/16/2023 01:29:53 | 1m |
| ▷ ✓ OmopLocationDimX | 07/16/2023 01:43:23 | 12m |
| ▷ ✓ OmopMeasurementFactX | 07/16/2023 04:39:07 | 3h 36m |
| ▷ ✓ OmopNoteFactX | 07/16/2023 03:31:01 | 1h 05m |
| ▷ ✓ OmopObservationFactX | 07/16/2023 03:12:15 | 31m |
| ▷ ✓ OmopObservationPeriodFactX | 07/16/2023 02:37:56 | 10m |
| ▷ ✓ OmopPersonDimX | 07/16/2023 01:56:16 | 7m |
| ▷ ✓ OmopProcedureOccurrenceFactX | 07/16/2023 02:37:53 | 7m |
| ▷ ✓ OmopProviderDimX | 07/16/2023 01:48:17 | 2m |
| ▷ ✓ OmopRelationshipRawDataX | 07/16/2023 01:33:01 | < 1m |
| ▷ ✓ OmopSourceToConceptMapRawDataX | 07/16/2023 01:32:45 | < 1m |
| ▷ ✓ OmopVisitDetailFactX | 07/16/2023 02:29:12 | < 1m |
| ▷ ✓ OmopVisitOccurrenceFactX | 07/16/2023 02:20:06 | 23m |
| ▷ ✓ OmopVocabularyRawDataX | 07/16/2023 01:33:00 | < 1m |
| ▷ ✓ OpioidOrderDataMart | 07/16/2023 02:22:43 | 2m |

- Above are stats for a full UTSW Caboodle nightly ETL execution.
  - This weekend execution completed before 5 am.
  - Our weekday executions typically complete before 6 am.

- OMOP tables process *following* their source Epic-released Caboodle tables. Thus, delivery of regular clinical/operational Caboodle data to users is not delayed by addition of OMOP DMCs

# OMOP Database with CDM Schema Views
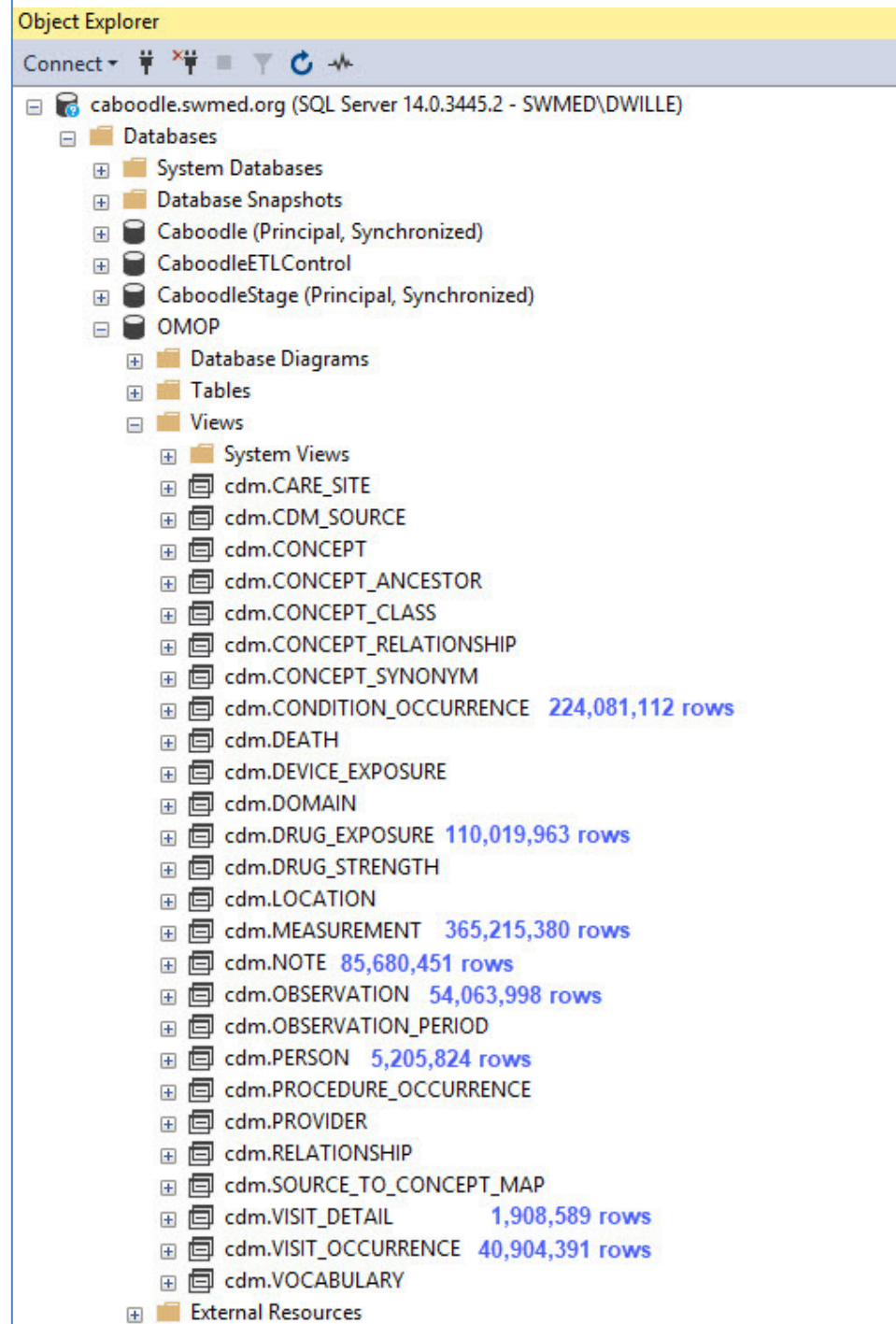
# Encounter diagnoses from clinic visits 1 day prior to query

# State of OMOP at Emory (Early 2022)

- Mapping/repositories created for specific projects by different groups
  - All of Us
  - Winship (CARS- Cancer Analytics and Reporting System)
- New registries began requiring data in OMOP format
  - CURE ID
  - ASH Network
  - N3C
- I2B2 implemented but limited
  - Not widely used for feasibility
  - Decision made to not map Epic data
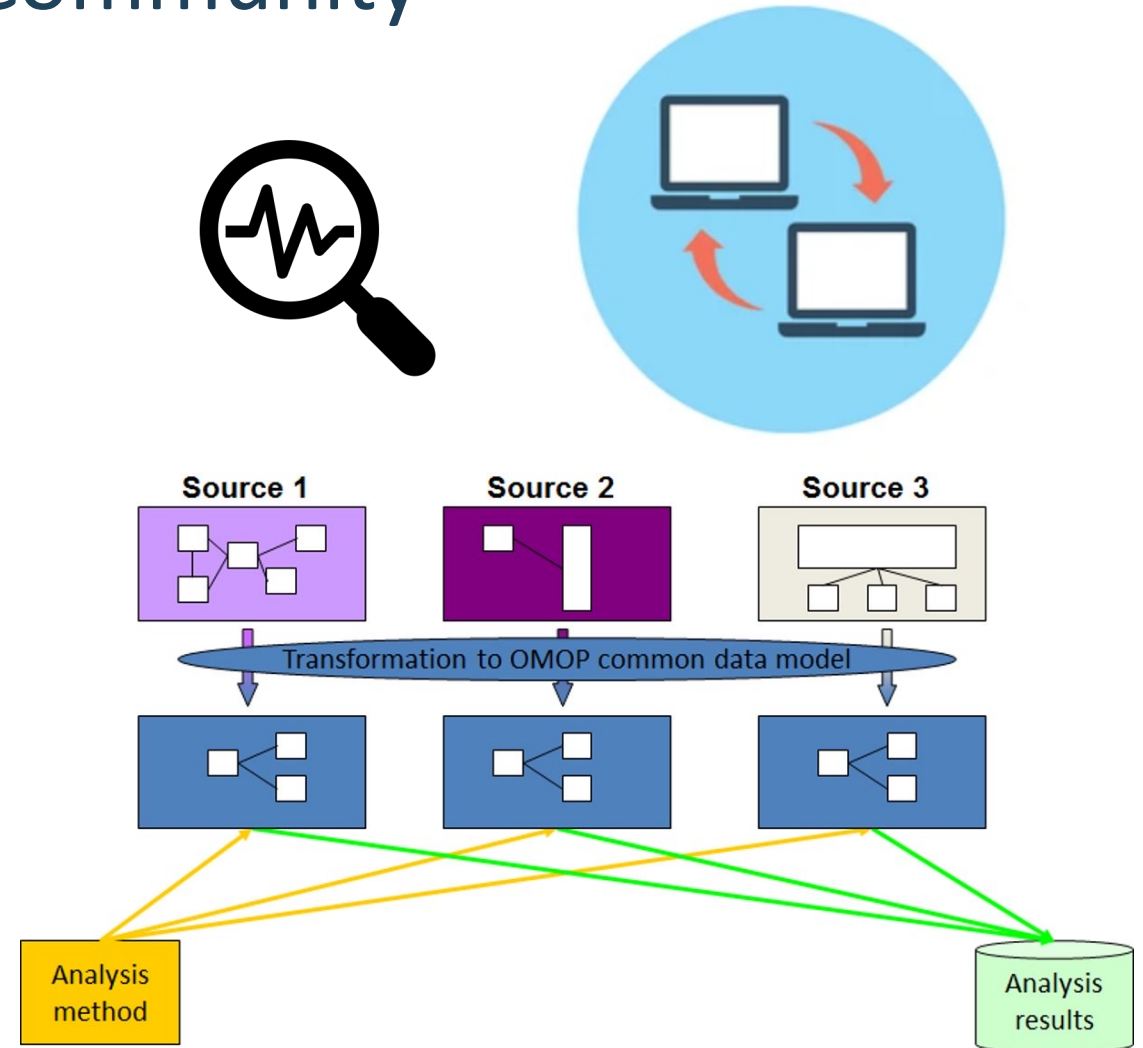- Most Cerner data (Pre-October 2022) not migrated to Epic Caboodle database

# Value of OMOP to Emory Community

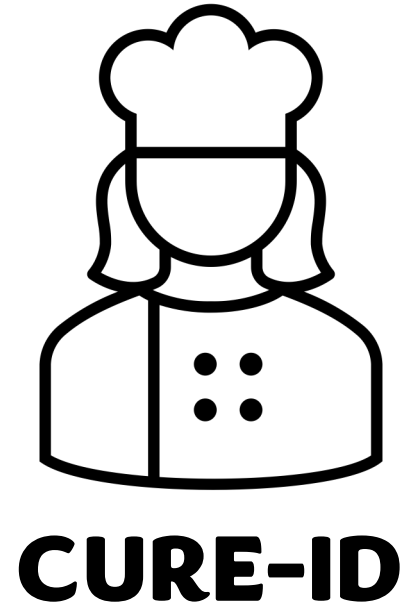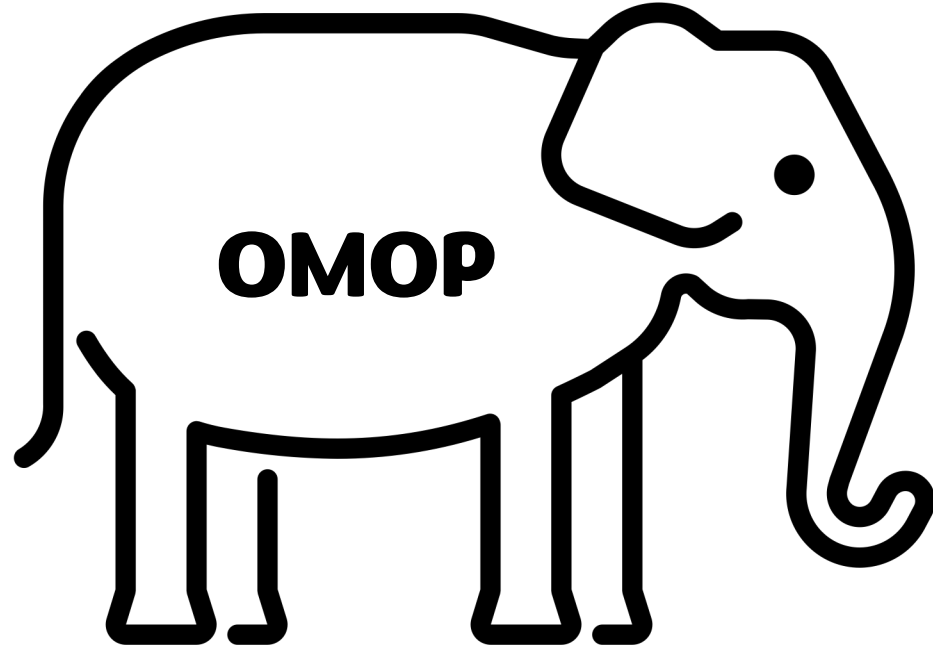Powerful interface (ATLAS) for cohort discovery and analysis

Submitting data to funded initiatives and registries: *FDA CURE-ID covid data submission*

Sharing analysis with other institutions: *"Are my results at Emory reproducible at other institutions"*

Combining disparate medical record data into a regional and longitudinal repository: *Cerner + Epic and Emory + Childrens*
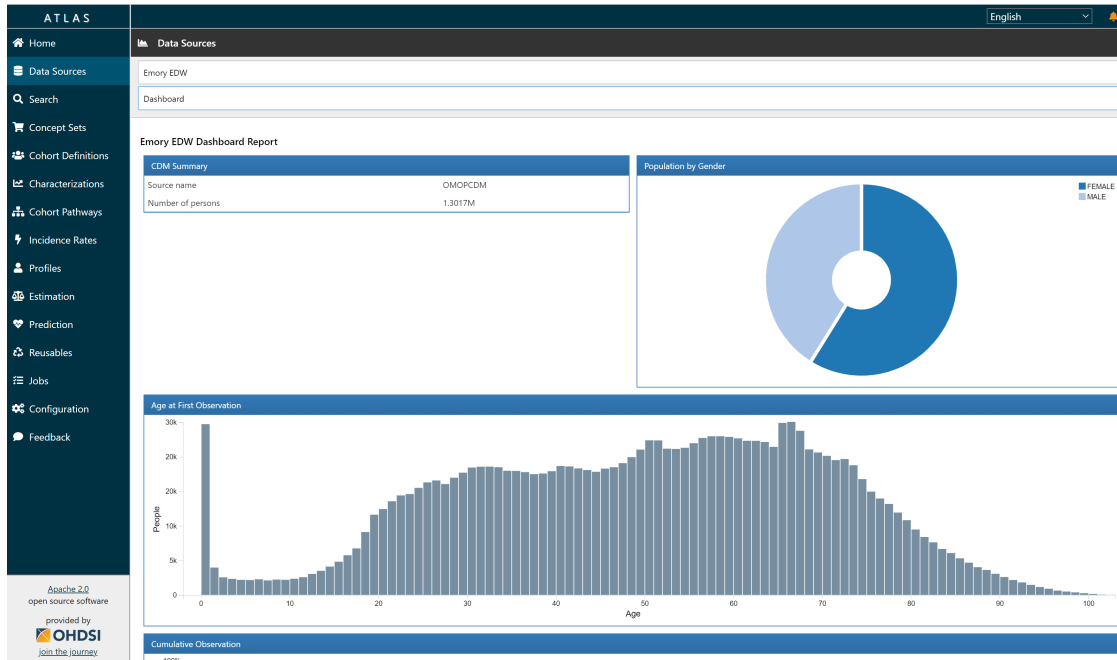


Source 1    Source 2    Source 3

Transformation to OMOP common data model

Analysis method

Analysis results

# How did we eat this elephant?

# Approach and lessons learned

- Identify a specific project
- Only bring in the timeframe of data you need
  - Keeps queries fast
  - Refresh of ETL is easy
  - Allows you to iterate
- Focus on depth instead of breadth
  - What labs or meds do you need?
  - What flowsheet data is important?
  - Mapped terms aren't useful without mapped results

# Enterprise OMOP Initiative



## Purpose

- Create an Enterprise instance of the OMOP that incorporates key data sources including historical (Cerner) clinical data, prospective (Epic) clinical data, and others

## Team

- Technical Lead
- Data Engineer
- Data Analyst
- Clinical SMEs
- OMOP SME (Consultant)

## Focus (Phase 1)

- Data submission for research registries: CURE-ID, ASH registry, N3C
- Clinical domains (no billing or cost data)
- Cerner data (No Epic in phase 1)

# Infrastructure (AWS)
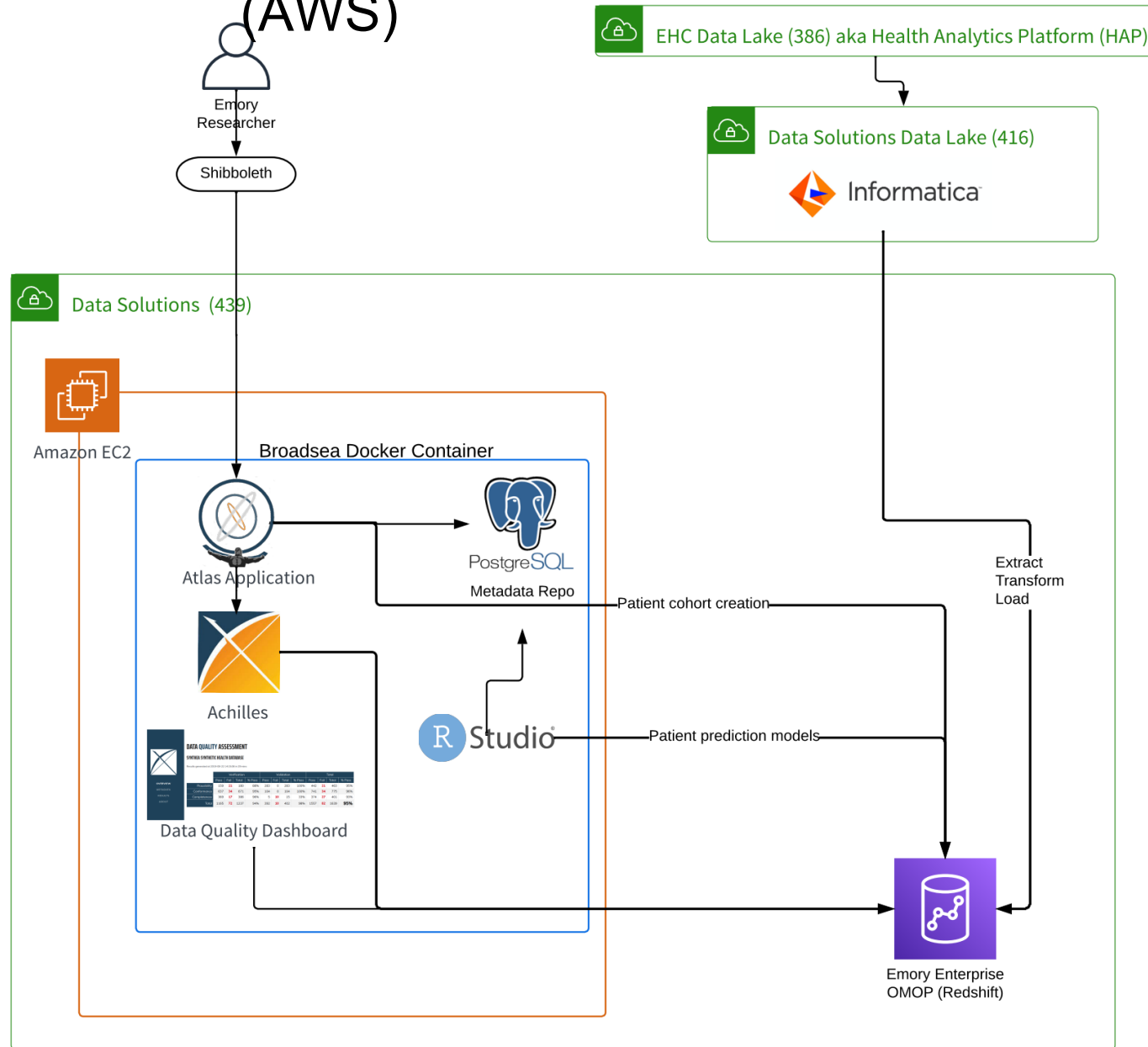
# Value realized

➢Reduce redundancies of mapping for multiple projects (CARS, AOU, CURE-ID, ASH, N3C)

➢Grant funding through CURE-ID and ASH network

➢Being used as clinical source for the Goizueta Institute Data Repository (GInDR)

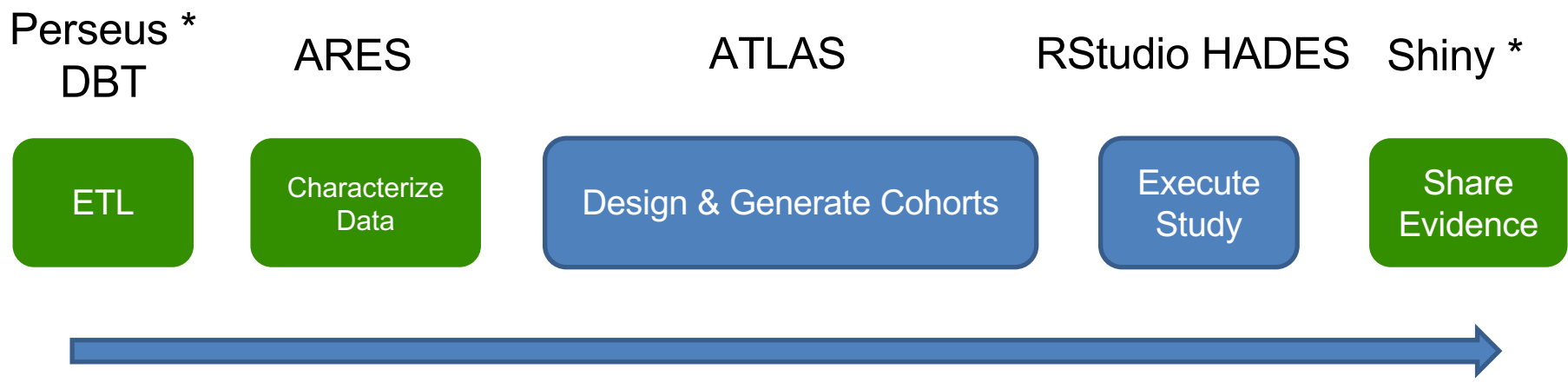➢Positioned to participate in additional registries when opportunity arises

# OHDSI Broadsea 3.0 – what's new
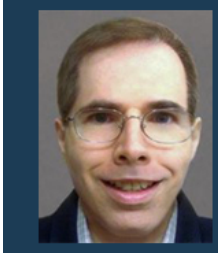
Lee Evans
Founder, LTS Computing LLC

Lead for Broadsea 3.0
development: Ajit Londhe

**Docker container-based deployment of core OHDSI tools**

Perseus *
DBT

ARES

ATLAS

RStudio HADES    Shiny *

| ETL | Characterize Data | Design & Generate Cohorts | Execute Study | Share Evidence |

Observational Research Process

**Broadsea 3.0 adds tools covering more steps in the process**

**ATLAS Security Providers**
Basic, LDAP, AD, OpenID, SAML

**Build / Launch Options**
Build from GitHub
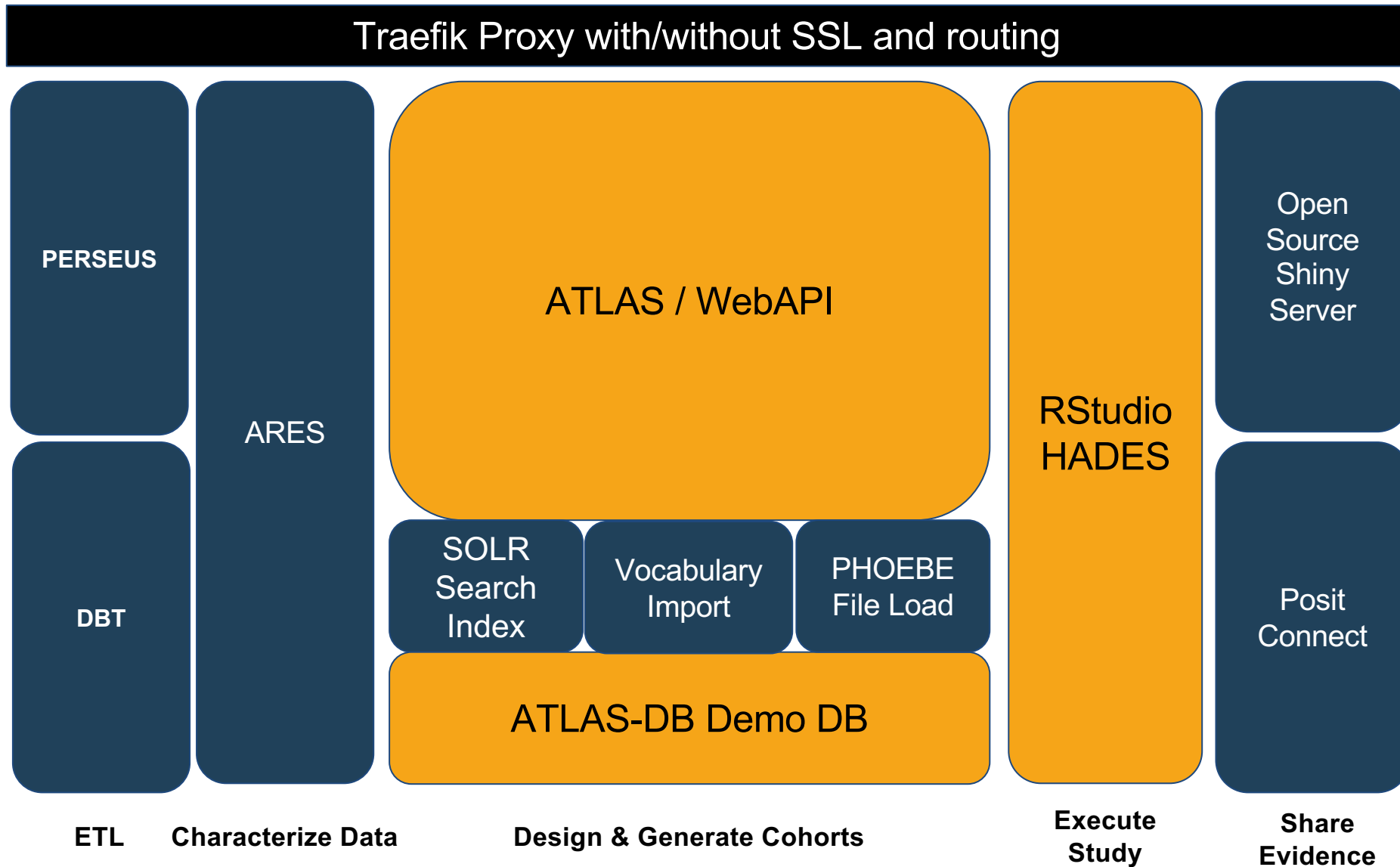Docker profiles

**Databases**
All OHDSI supported DBs

**Configuration**

One templated
.env file

* Actively under development

# Broadsea 3.0 config .env file sections

**Section 1:** Broadsea Host

**Section 2:** Atlas GUI configuration

**Section 3:** WebAPI Database configuration

**Section 4:** Atlas security provider configuration

**Section 5:** WebAPI security configuration

**Section 6:** Building Atlas or WebAPI from Git

**Section 7:** SOLR Vocab (optional)

**Section 8:** HADES credentials to use in RStudio

**Section 9:** Postgres and UMLS credentials for loading OMOP Vocab files into Postgres schema

**Section 10:** Postgres credentials for loading Phoebe file for Atlas 2.12+ Concept Recommendations into Postgres hosted OMOP Vocabulary schema

**Section 11:** Ares Data Folder config

**Section 12:** Broadsea Content Page config

# Benefits of Broadsea 3.0 for Emory University

| Quick Installation | Security Enabled | Simple Upgrades |
| --- | --- | --- |

- Hands-on Atlas demo
- Use Atlas demo to verify infrastructure configuration
- Use RStudio HADES to run Achilles for ETL

- HTTPS: Certificate
- Single Sign On

- Deploy latest OHDSI Broadsea Docker containers when new versions released

**Shibboleth.** Consortium

**NIH** National Institutes of Health
Turning Discovery Into Health

Sea

| Health Information | Grants & Funding | News & Events | Research & Training |

# NEWS RELEASES

Thursday, August 17, 2023

# NIH establishes Maternal Health Research Centers of Excellence

*Initiative to support research to reduce pregnancy-related complications and deaths and promote maternal health equity.*

The National Institutes of Health has awarded $24 million in first-year funding to establish Maternal Health Research Centers of Excellence. Part of NIH's Implementing a Maternal Health and Pregnancy Outcomes Vision for Everyone (IMPROVE) initiative, the centers will develop and evaluate innovative approaches to reduce pregnancy-related complications and deaths and promote maternal health equity. The grants are expected to last seven years and total an estimated $168 million, pending the availability of funds.

<u>Data Core Co-PIs</u>
Andreea Creanga, MD, PhD
Center for Women's Health, Sex, and Gender Research
Johns Hopkins School of Public Health

Paul Nagy, PhD
Johns Hopkins OHDSI Research Group
Section of Biomedical Informatics and Data Science
Johns Hopkins School of Medicine

# Clinical registry strategic alignment with OHDSI

**Aim 1:** Enhance coordination for better communication among Research Centers, IS Hub, and IMPROVE grantees.

**Aim 2:** Create tools and workflows for consistent data collection, analysis, and sharing, ensuring high-quality results.

**Aim 3**: Boost data science and innovation skills of Research Centers and IMPROVE grantees.

- Expand the OHDSI community by creating OHDSI teams and Github Repos for this registry.
- Enhance health systems' ETL capabilities and train scientists in OHDSI methods.
- Implement OHDSI analytics tools in organizations.
- Fund OHDSI open-source tool development.
- Support EHDEN educational initiatives.
- Execute Network Studies with the OHDSI Data Network.

**Implementing a Maternal health and PRegnancy Outcomes Vision for Everyone (IMPROVE) Initiative**

JOHNS HOPKINS
SCHOOL of MEDICINE