Welcome

Thank you for spending your time with us today.
<table>
<thead>
<tr>
<th>Introductions</th>
<th>9:00 - 9:15</th>
</tr>
</thead>
<tbody>
<tr>
<td>A Brief History</td>
<td>9:15 - 10:00</td>
</tr>
<tr>
<td>Architecture Overview</td>
<td>10:00 - 11:00</td>
</tr>
<tr>
<td>Break</td>
<td>11:00 - 11:15</td>
</tr>
<tr>
<td>BroadSea</td>
<td>11:15 - 11:45</td>
</tr>
<tr>
<td>Community Collaboration</td>
<td>11:45 - 12:30</td>
</tr>
<tr>
<td>Lunch</td>
<td>12:30 - 1:30</td>
</tr>
<tr>
<td>Demonstrations</td>
<td>1:30 - 3:30</td>
</tr>
<tr>
<td>Break</td>
<td>3:30 - 3:45</td>
</tr>
<tr>
<td>OHDSI Environments</td>
<td>3:45 - 4:00</td>
</tr>
<tr>
<td>Roadmap</td>
<td>4:00 - 4:30</td>
</tr>
<tr>
<td>Wrap Up</td>
<td>4:30 - 4:45</td>
</tr>
</tbody>
</table>

Agenda
Introductions

Frank DeFalco
Janssen Research & Development

Greg Klebanov
Odysseus Data Services

Lee Evans
LTS Computing

Sigfried Gold
University of Maryland, HCIL
A Brief Introduction

Premier: It starts with a single study...

A Patient Story

Health Informatics Center of Excellence

Adopting the OMOP CDM

A Benevolent Dictator is born...

Amazing science, individual contributors

The need for a formal architecture...

2007

2010

2013

2017

OMOP

OHDSI
Computational Epidemiology

coming soon
Early Feedback

“Everyone is at a different stage…” - Patrick Ryan, OHDSI 2017
How would you describe yourself?

- Software Developer
- Academic Researcher
- Data Manager
- Statistician
- Clinician
- Technologist
- Computational Epidemiologist

Start the presentation to activate live content

If you see this message in presentation mode, install the add-in or get help at PollEv.com/app
Tell us about your access to patient level data?

- No Data
- Data in proprietary format
- Data in CDM v4
- Data in CDM v5

Start the presentation to activate live content.

If you see this message in presentation mode, install the add-in or get help at PollEv.com/app.
<table>
<thead>
<tr>
<th>What is your data storage platform?</th>
</tr>
</thead>
<tbody>
<tr>
<td>PostgreSQL</td>
</tr>
<tr>
<td>Microsoft SQL Server</td>
</tr>
<tr>
<td>Impala / Hadoop</td>
</tr>
<tr>
<td>Oracle</td>
</tr>
<tr>
<td>Google Big Query</td>
</tr>
<tr>
<td>Amazon Redshift</td>
</tr>
</tbody>
</table>

Start the presentation to activate live content

If you see this message in presentation mode, install the add-in or get help at PollEv.com/app
Do you have ATLAS deployed?

- Yes
- No

Start the presentation to activate live content
If you see this message in presentation mode, install the add-in or get help at PollEv.com/app
Do you have WebAPI deployed?

Yes

No

Start the presentation to activate live content
If you see this message in presentation mode, install the add-in or get help at PollEv.com/app
Architecture Overview
The OMOP CDM

CDM Schema v5.0
Patient Level Data
Vocabulary Content
WebAPI

- Java
- REST web service
- Job Manager
- Leverages SQLRender
- APACHE Shiro
- Spring Boot
- Spring Batch
ATLAS

- Web Client
- HTML
- CSS
- Javascript
- Knockout
- Leverages WebAPI
- Demo...
Achilles

- Data Quality (Achilles Heel)
- Data Source level Characterization
- R Package
- Web Interface
- Originally JSON now WebAPI
Data Flow

- Source to CDM to Analytic Data Set to Analysis
- A single CDM or multi-CDM environment
- Standardized analytics across N data sources.
- Across a community the architecture provides a unifying approach to data management, methods and analysis
Data Architecture

- OHDSI Schema (configuration database)
- CDM Schema (patient level data, read only)
- Results Schema (analysis results, read / write)
15 Minute Break
Broadsea - Easily deploy anywhere

- Docker container - package up an app with all it’s dependencies
- Scaleable - one container on a single pc or multiple on a multi-server cluster
- Cross-platform - Windows, Linux, OS X, cloud(s)
- Scripted run-time environment - docker-compose.yml / kubernetes pod
- Broadsea Web Tools container = OHDSI Atlas + WebAPI + tomcat server
- Broadsea Methods container = OHDSI R Methods + RStudio server
- Quickstart: https://github.com/OHDSI/Broadsea
Broadsea API

Developer API Portal
http://broadseaapi.org
Community Collaboration

- Github
- OHDSI Forums
- OHDSI Community Call
- OHDSI workgroups (our favorite is the Architecture call)
- OHDSI Slack
GitHub Exercises

- Everyone will have a GitHub account when this is over
- Everyone will watch / star the ATLAS and WebAPI repositories

OHDSI Collaboration

- Forums
  - Register on the forums, possibly with your GitHub account.
  - Start a thread
- Post questions throughout the day on the ‘OHDSI Architecture Tutorial Thread’
- create “What’s missing in the architecture?” “What haven’t we covered that you want to hear about? “ We’ll read these during breaks and exercises.
- Wiki - register / edit / contribute
Get Involved
Long-term importance of OHDSI as F/OSS for health records analysis

Advantages of CDMs and digitally defined study protocols:

- Faster research cycles
- Validity by replication
- Validity by coordination/aggregation across research networks

Advantages of free/open-source software platform:

- Transparency. Legitimacy in setting standards for data structures, methodologies, etc.
- Even experimental work can integrate with functioning system and real users with real use cases
Wastefulness of health records research without CDMs

Completely wild guesses:

- Clinical researchers:
  - Comparative effectiveness
  - Outcomes research
  - Precision medicine
  - Epidemiological studies
  - New drug or treatment development

- Government agencies and regulators (U.S.):
  - FDA, AHxRQ, NIH, CDC, CMS, HRSA, ONC, BEA

- Hospitals and health care systems
  - Quality of care
  - Guideline compliance
  - JCAHO reporting
  - Patient safety
  - Cost effectiveness

- Payers
  - Insurers
  - Pharmacy benefit providers
  - CMS

10,000 researchers
20,000 stats/tech support
200 institutions
$100 billion spent
$90 billion wasted

20,000 administrators
40,000 analysts
6,000 institutions
$300 billion spent
$100 billion wasted

Opportunity cost of wasting all these people’s time: $2.5 gazillion
Vision for the future of health data research

Any organization (or citizen researcher) with analytic tasks to be performed on patient data should have access to free, open-source tools that solve the most common of these tasks, and that provide a platform for solving more unique problems. No one doing health records research should be reinventing these wheels:

- ETL
- Data quality
- Population characterization
- Phenotyping
- Basic effect estimation
- Patient-level prediction
- Digital protocol replication, aggregation, and publishing
- Visualization to make all these functions intuitively accessible to researchers
Is OHDSI the only game in town?

It shouldn’t be.

We, along with PCORNet, Sentinel, i2b2, ?, are transforming healthcare research and health data analysis.

Too important to be limited to (or controlled by) one organization (including OHDSI), or limited to one API, one DBMS, one programming language or Javascript framework.

Other CDMs and their software platforms are open-source, but they don’t have active, diverse developer communities. We need to be courting them, not competing with them.

As a developer community we are, and must be, both open and—sortof, kindof—unified. It’s hard to build technologies that allow extension in unexpected directions. It’s hard to coordinate or integrate tools written in different languages and frameworks. But that’s what we’re working towards.
What can we as developers do?

- Innovate
  - APIs
  - UI/UX
  - Machine learning
  - Methods
  - Scale
  - Application-level standards

- but — NOT ALONE
  - Even—especially— if you use tools not currently integratable into OHDSI stack
    - Use or help develop app-level standards
    - Come to architecture call
    - Make it open source
    - Make it reusable
    - Reuse as much as you can
Standard V5.2 Open-Source Lunch
Demonstrations / Discussions

- ARACHNE
- Dataprint
- Helios
Enabled reproducible, scalable and collaborative research studies within multiple teams in a single organizations or across multiple healthcare organizations
ARACHNE key features

• Study lifecycle and workflow management
• Discover federated data sets in RWE data catalog
• Build study team
• Federated analysis across organizational boundaries
• Secure, compliant and trusted data access
• Exchange and store analysis results
• Support for R, SQL and complex packages
• Integration with OHDSI Platforms (ATLAS, Achilles)
• Support for OHDSI OMOP CDM
ARACHNE Technical Architecture

ARACHNE Central
- Study Manager
- RWD Data Catalog
- Expert Finder
- Insights Library
- Athena - Vocabulary Explorer and Download

ARACHNE Data Node
- Data Node Manager
- ARACHNE Execution Engine
- OHDSI ATLAS/WebAPI and Achilles
- Sample CDM Instance (SynPuf 2.3mil)

Back-end:
- Java 8
- Spring + Hibernate
- PostgreS
- Solr
- Docker

Front-end:
- JS (ES6) in Community / Typescript in Enterprise
- React + Redux + Reselect
- SASS + BEM methodology
- D3.js
- Webpack
Developing Dataprint

<table>
<thead>
<tr>
<th>Conditions</th>
<th>Deaths</th>
<th>Device Exposures</th>
<th>Drug Exposures</th>
<th>Measurements</th>
<th>Notes</th>
<th>Observation Periods</th>
<th>Observations</th>
<th>Procedures</th>
<th>Visit Presence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Troxene MSEC</td>
<td>0.5%</td>
<td>0.7%</td>
<td>1.2%</td>
<td>0.9%</td>
<td>1%</td>
<td>10.0%</td>
<td>0.0%</td>
<td>2%</td>
<td>0.0%</td>
</tr>
<tr>
<td>Troxene SSEC</td>
<td>2.5%</td>
<td>2.7%</td>
<td>3.2%</td>
<td>2.8%</td>
<td>3%</td>
<td>11.0%</td>
<td>1.0%</td>
<td>3%</td>
<td>0.0%</td>
</tr>
<tr>
<td>Troxene GECM              1:1.9%</td>
<td>2.5%</td>
<td>2.7%</td>
<td>3.2%</td>
<td>2.8%</td>
<td>3%</td>
<td>11.0%</td>
<td>1.0%</td>
<td>3%</td>
<td>0.0%</td>
</tr>
<tr>
<td>Option SSD</td>
<td>77.9%</td>
<td>77.9%</td>
<td>77.9%</td>
<td>77.9%</td>
<td>77.9%</td>
<td>77.9%</td>
<td>77.9%</td>
<td>77.9%</td>
<td>77.9%</td>
</tr>
<tr>
<td>Option XSD</td>
<td>15.7%</td>
<td>15.7%</td>
<td>15.7%</td>
<td>15.7%</td>
<td>15.7%</td>
<td>15.7%</td>
<td>15.7%</td>
<td>15.7%</td>
<td>15.7%</td>
</tr>
<tr>
<td>Other options</td>
<td>5.6%</td>
<td>5.6%</td>
<td>5.6%</td>
<td>5.6%</td>
<td>5.6%</td>
<td>5.6%</td>
<td>5.6%</td>
<td>5.6%</td>
<td>5.6%</td>
</tr>
<tr>
<td>Other options</td>
<td>15.7%</td>
<td>15.7%</td>
<td>15.7%</td>
<td>15.7%</td>
<td>15.7%</td>
<td>15.7%</td>
<td>15.7%</td>
<td>15.7%</td>
<td>15.7%</td>
</tr>
<tr>
<td>Other options</td>
<td>5.6%</td>
<td>5.6%</td>
<td>5.6%</td>
<td>5.6%</td>
<td>5.6%</td>
<td>5.6%</td>
<td>5.6%</td>
<td>5.6%</td>
<td>5.6%</td>
</tr>
</tbody>
</table>
Helios Solr
15 Minute Break
Maintaining your Local Environment

Addressing concerns session

- CDM data updates
- Vocabulary updates
- CDM schema conversions & updates
- R package management
- Security - (Shiro enabled?)
With a growing number of network studies, there is a need to have a controlled, production grade research environment that will ensure quality, reproducible and transparent research. The environment must contain key OHDSI platforms and components:

1. ARACHNE
2. ATLAS
3. ATHENA

All OHDSI Network Research platforms (deployed at ohdsi.org) will utilize a single consistent Authentication and Authorization (A&A) mechanism, where key user record is stored within ARACHNE's user database.
OHDSI Network Research Environment (cont.)

Production

- ARACHNE (Network Studies) - arachne.ohdsi.org
- ATLAS/WebAPI (Analysis Design) - atlas.ohdsi.org
- ATHENA (OMOP Standard Vocabularies) - athena.ohdsi.org

Demo (Test)

- ARACHNE (Network Studies) - ohdsi.org/arachne
- ATLAS/WebAPI (Analysis Design) - ohdsi.org/atlas
- ATHENA (OMOP Standard Vocabularies) - ohdsi.org/athena
Looking back and to the Road Ahead
OHDSI Architecture

It is successfully deployed at large pharmaceutical organizations, payers, academic medical centers…

… but it isn’t easy to do, nor as widely adopted as we would like.
Criticism: Deployment

- Install relational database engine (your choice), pull WebAPI from GitHub, configure settings, build with maven, establish WAR container (tomcat), deploy WAR, opaque auto-configuration of OHDSI repository, pull ATLAS from GitHub, configure, deploy to web server, manually insert rows into OHDSI repository to identify your CDM data sources and their daemons, pull ACHILLES from GitHub, install R, run R package for each CDM manually
Solution
Criticism: Specifications

Opaque, unpublished specifications.

Requirement for burdensome relational mapping and service layer integration.
Solutions
Gauging Interest
Wrapping Up...
"If I have seen further, it is by standing on the shoulders of giants."

- Sir Isaac Newton
Thanks
Backup slides...
Early Model Specification Discussion w/Greg
Architectural Overview