Welcome

Thank you for spending your time with us today.

Agenda

Introductions	9:00 - 9:15
A Brief History	9:15 - 10:00
Architecture Overview	10:00 - 11:00
Break	11:00 - 11:15
BroadSea	11:15 - 11:45
Community Collaboration	11:45 - 12:30
Lunch	12:30 - 1:30
Demonstrations	1:30 - 3:30
Break	3:30 - 3:45
OHDSI Environments	3:45 - 4:00
Roadmap	4:00 - 4:30
Wrap Up	4:30 - 4:45

Introductions
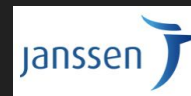
Frank DeFalco
Janssen Research & Development

Greg Klebanov
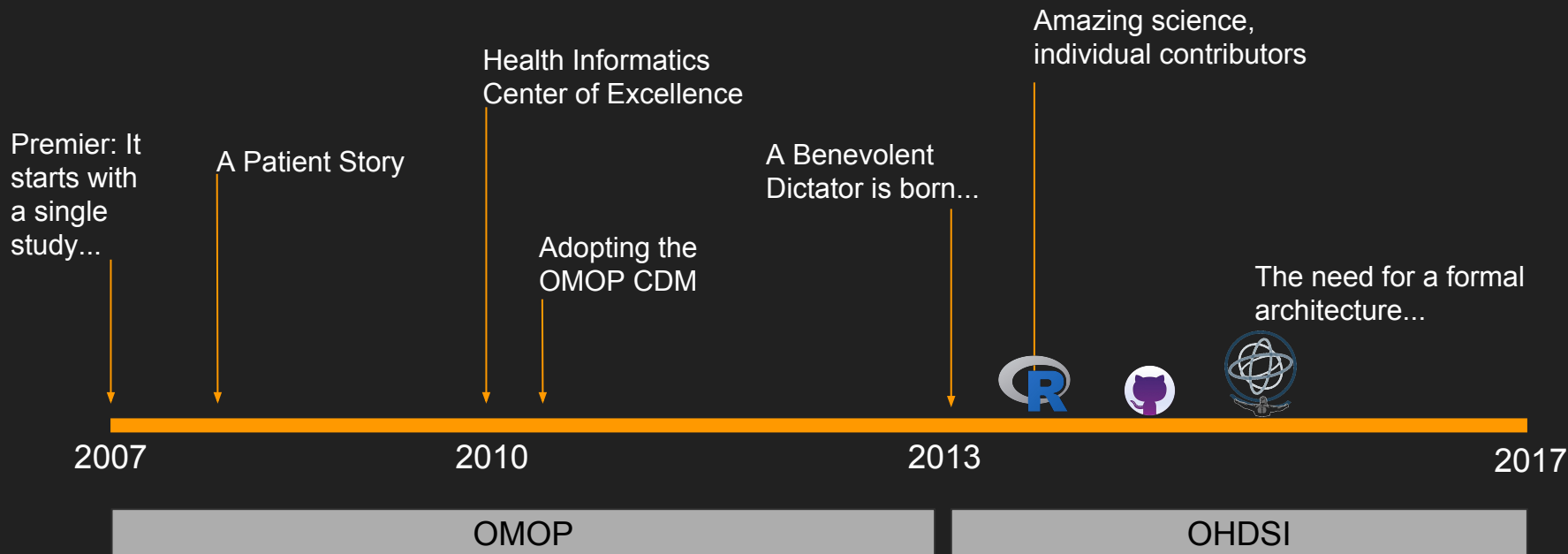Odysseus Data Services

Lee Evans
LTS Computing

Sigfried Gold
University of Maryland, HCIL

OHDSI

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

Other

Tell us about your access to patient level data?

No Data

Data in proprietary format

Data in CDM v4

Data in CDM v5

What is your data storage platform?

PostgreSQL

Microsoft SQL Server

Impala / Hadoop
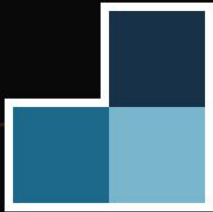
Oracle

Google Big Query

Amazon Redshift

Do you have ATLAS deployed?

Yes

No

Do you have WebAPI deployed?

Yes

No

Architecture Overview

The OMOP CDM

CDM Schema v5.0
Patient Level Data
Vocabulary Content

DRUG_STRENGTH
- DRUG_CONCEPT_ID
- INGREDIENT_CONCEPT_ID
- AMOUNT_VALUE
- AMOUNT_UNIT_CONCEPT_ID
- NUMERATOR_VALUE
- NUMERATOR_UNIT_CONCEPT_ID
- DENOMINATOR_UNIT_CONCEPT_ID
- VALID_START_DATE
- VALID_END_DATE
- INVALID_REASON

CONCEPT_RELATIONSHIP
- CONCEPT_ID_1
- CONCEPT_ID_2
- RELATIONSHIP_ID
- VALID_START_DATE
- VALID_END_DATE
- INVALID_REASON

DOMAIN
- DOMAIN_ID
- DOMAIN_NAME
- DOMAIN_CONCEPT_ID

VOCABULARY
- VOCABULARY_ID
- VOCABULARY_NAME
- VOCABULARY_REFERENCE
- VOCABULARY_VERSION
- VOCABULARY_CONCEPT_ID

CONCEPT_ANCESTOR
- ANCESTOR_CONCEPT_ID
- DESCENDANT_CONCEPT_ID
- MIN_LEVELS_OF_SEPARATION
- MAX_LEVELS_OF_SEPARATION

CONCEPT
- CONCEPT_ID
- CONCEPT_NAME
- DOMAIN_ID
- VOCABULARY_ID
- CONCEPT_CLASS_ID
- STANDARD_CONCEPT
- CONCEPT_CODE
- VALID_START_DATE
- VALID_END_DATE
- INVALID_REASON

CONCEPT_CLASS
- CONCEPT_CLASS_ID
- CONCEPT_CLASS_NAME
- CONCEPT_CLASS_CONCEPT_ID

CONCEPT_SYNONYM
- CONCEPT_ID
- CONCEPT_SYNONYM_NAME
- LANGUAGE_CONCEPT_ID

RELATIONSHIP
- RELATIONSHIP_ID
- RELATIONSHIP_NAME
- IS_HIERARCHICAL
- DEFINES_ANCESTRY
- REVERSE_RELATIONSHIP_ID
- RELATIONSHIP_CONCEPT_ID

SOURCE_TO_CONCEPT_MAP
- SOURCE_CODE
- SOURCE_VOCABULARY_ID
- SOURCE_CODE_DESCRIPTION
- TARGET_CONCEPT_ID
- TARGET_VOCABULARY_ID
- VALID_START_DATE
- VALID_END_DATE
- INVALID_REASON

OHDSI
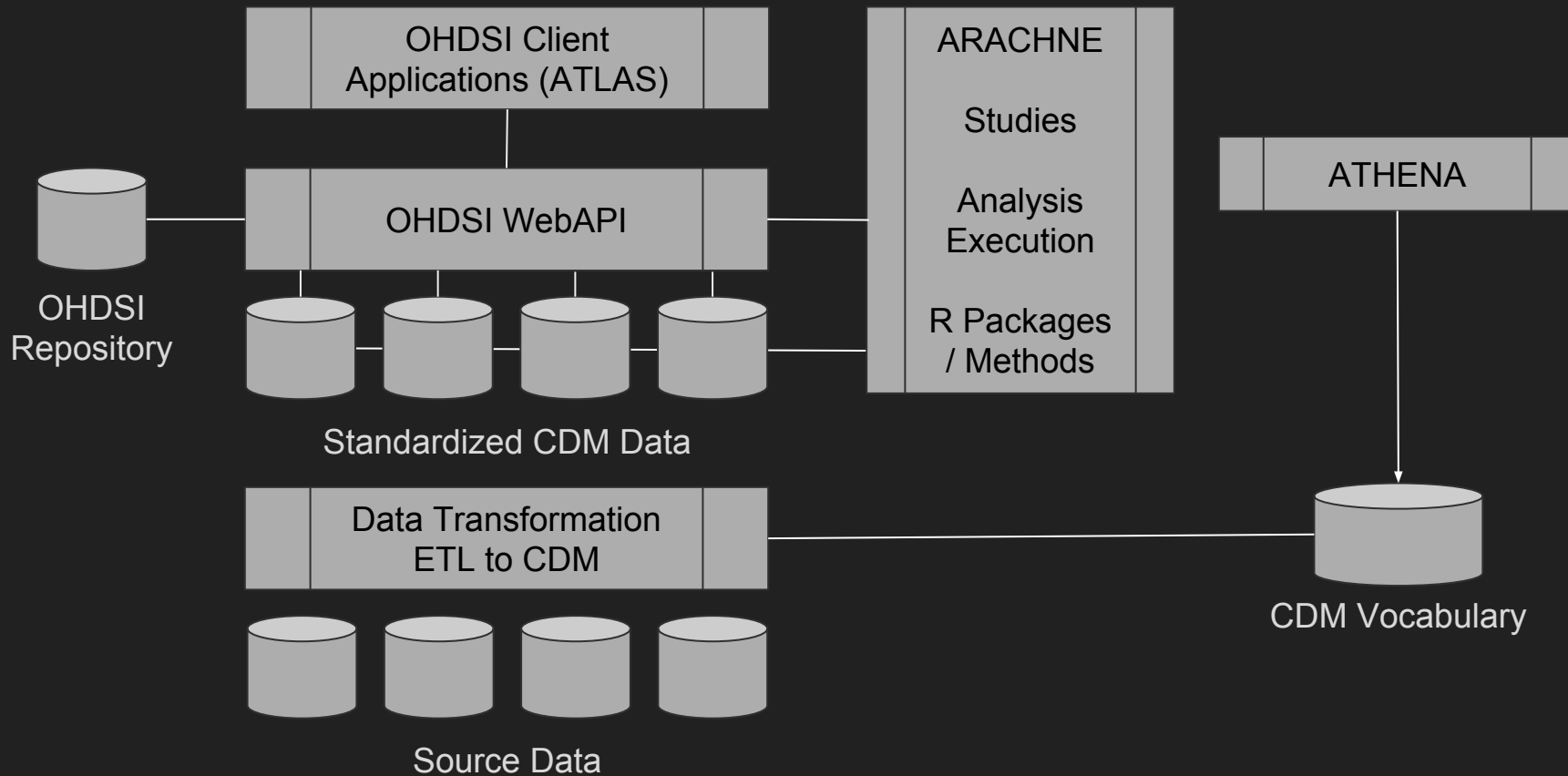Observational Health Data Sciences and Informatics

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
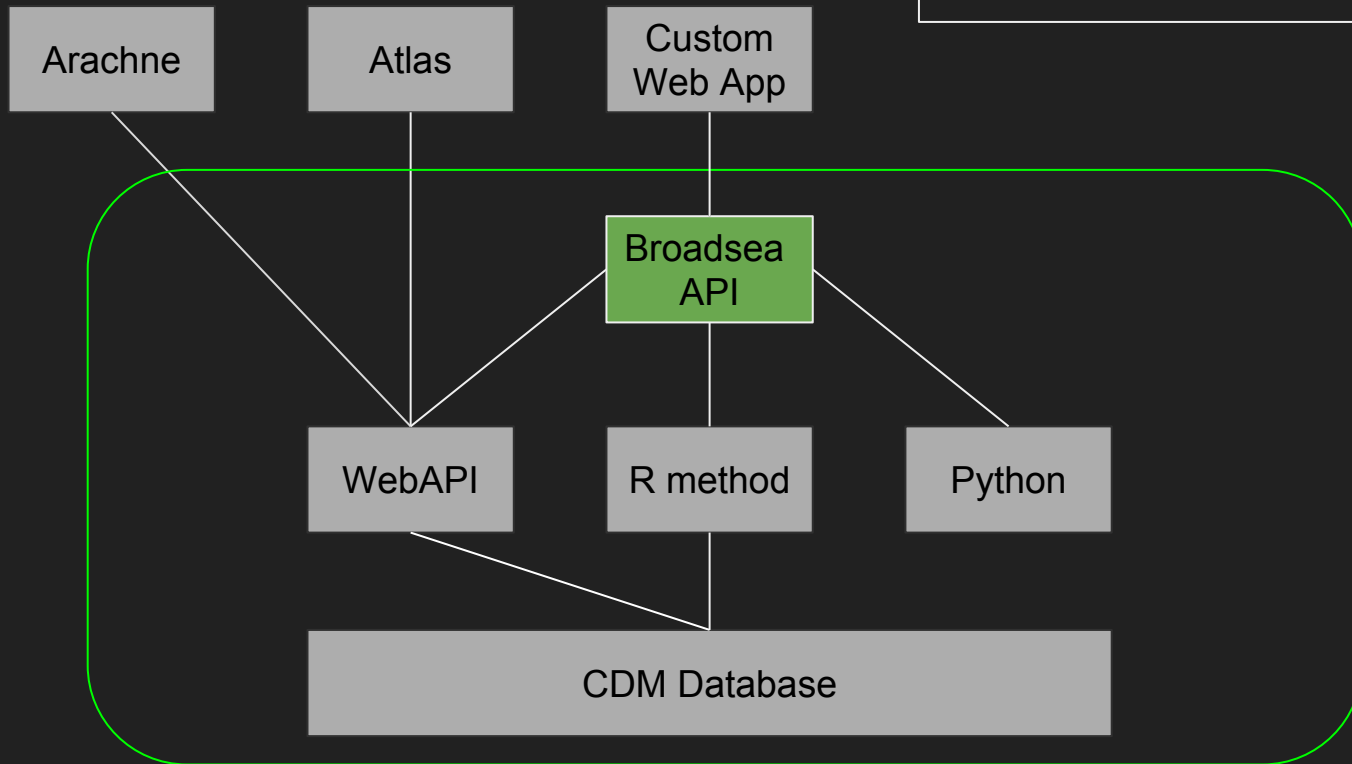
Community Collaboration

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

https://www.github.com/OHDSI

GitHub Exercises

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

http://www.ohdsi.org/web/wiki/doku.php?id=development:ohdsi_github_projects_v2

OHDSI

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
- Validity by simultaneously running multiple treatment / outcome combinations and performing p-value calibration (https://www.ohdsi.org/wp-content/uploads/2016/09/OHDSI-Sympsium-2016-Schuemie-estimation-23sept2016.pdf)

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

OHDSI

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
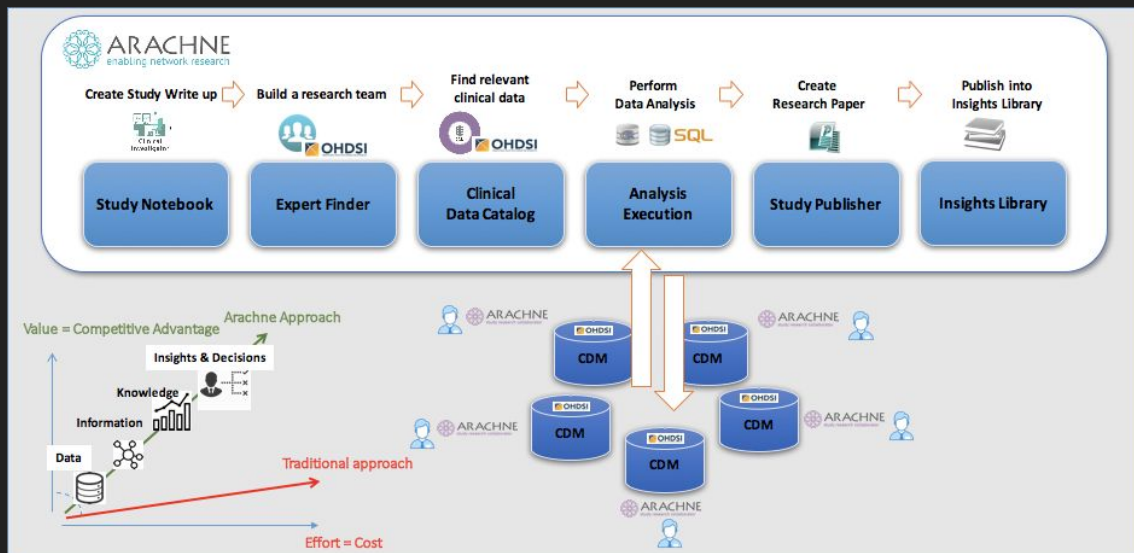   - Reuse as much as you can
    -

Standard V5.2 Open-Source Lunch

Demonstrations / Discussions

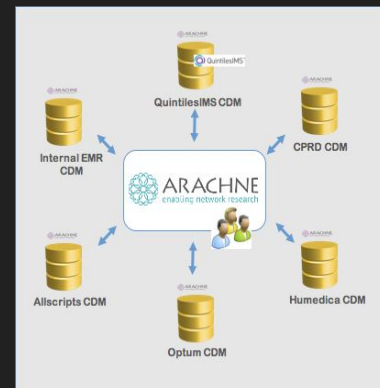- ARACHNE
- Dataprint
- Helios

ARACHNE

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
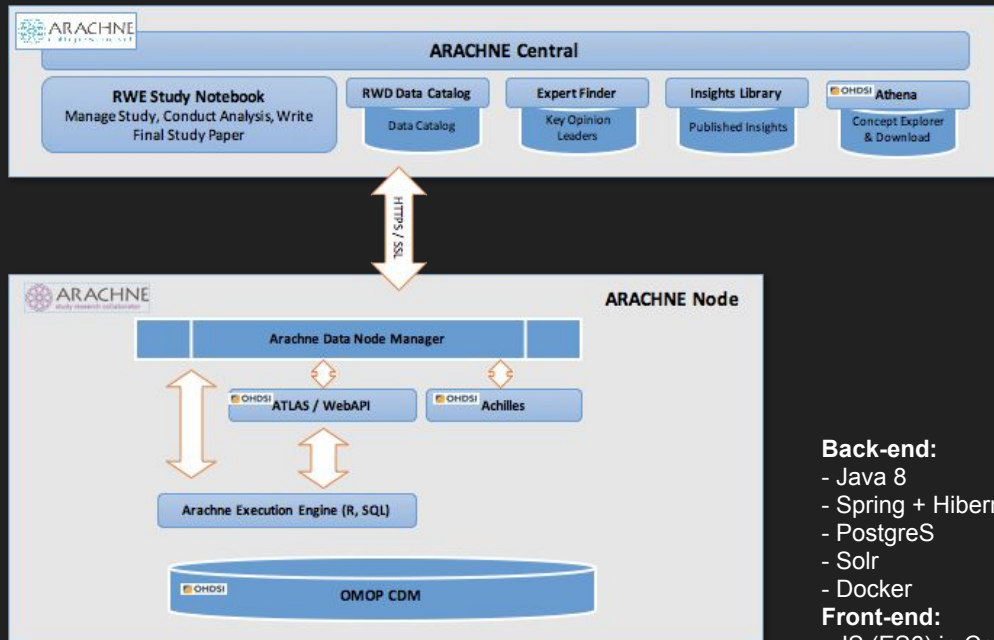
ARACHNE Central

- 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
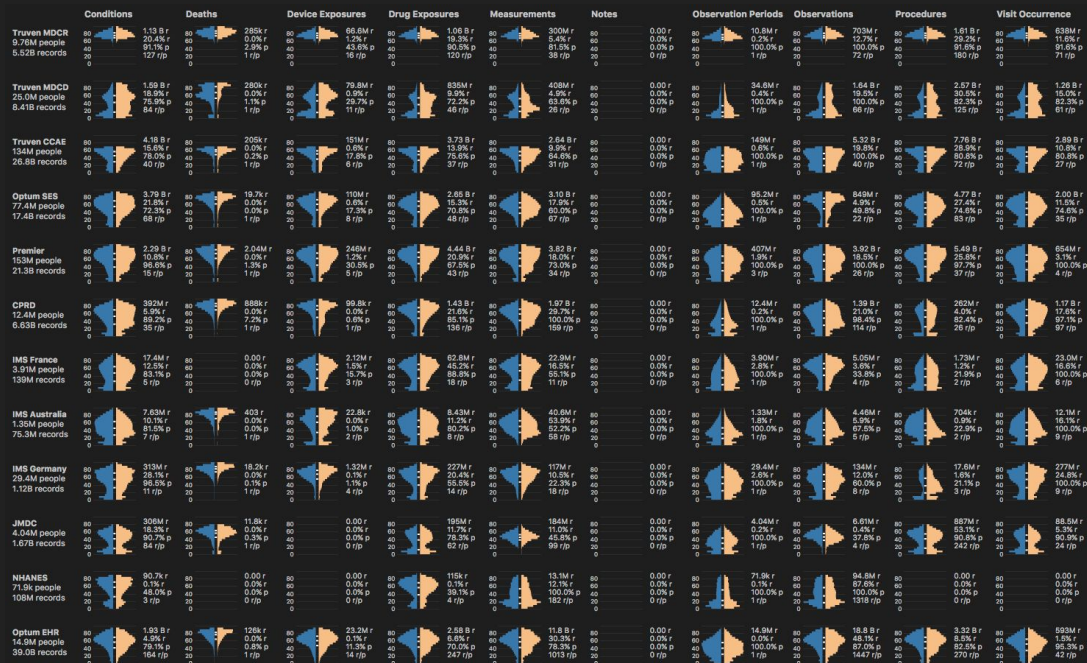- Sample CDM Instance (SynPuf 2.3mil)

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

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?)

OHDSI Network Research Environment

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 daimons, pull ACHILLES from GitHub, install R, run R package for each CDM manually
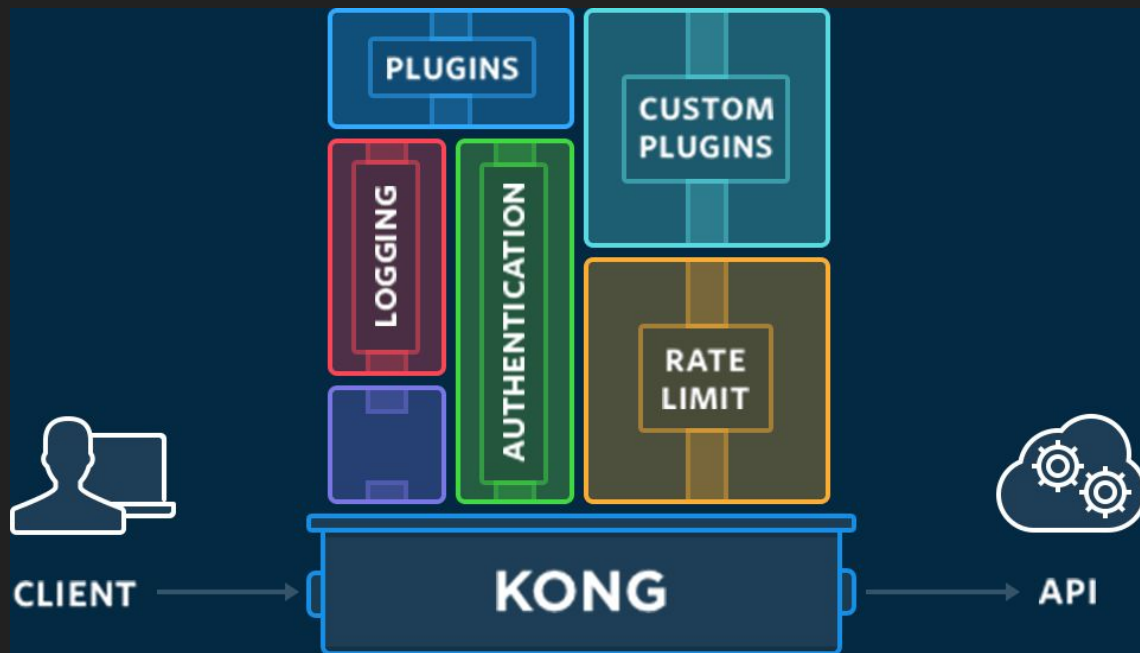
Solution
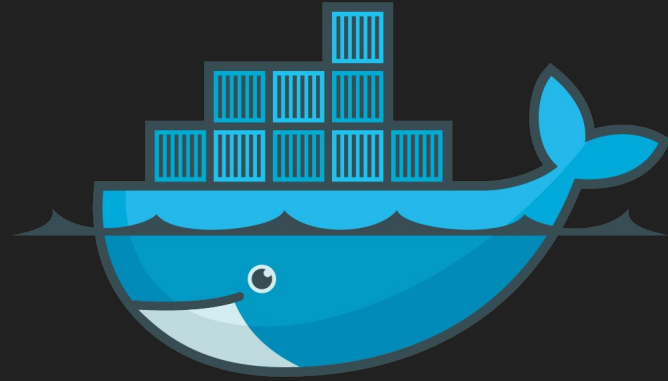
Criticism: Specifications

Opaque, unpublished specifications.

Requirement for burdensome relational mapping and service layer integration.
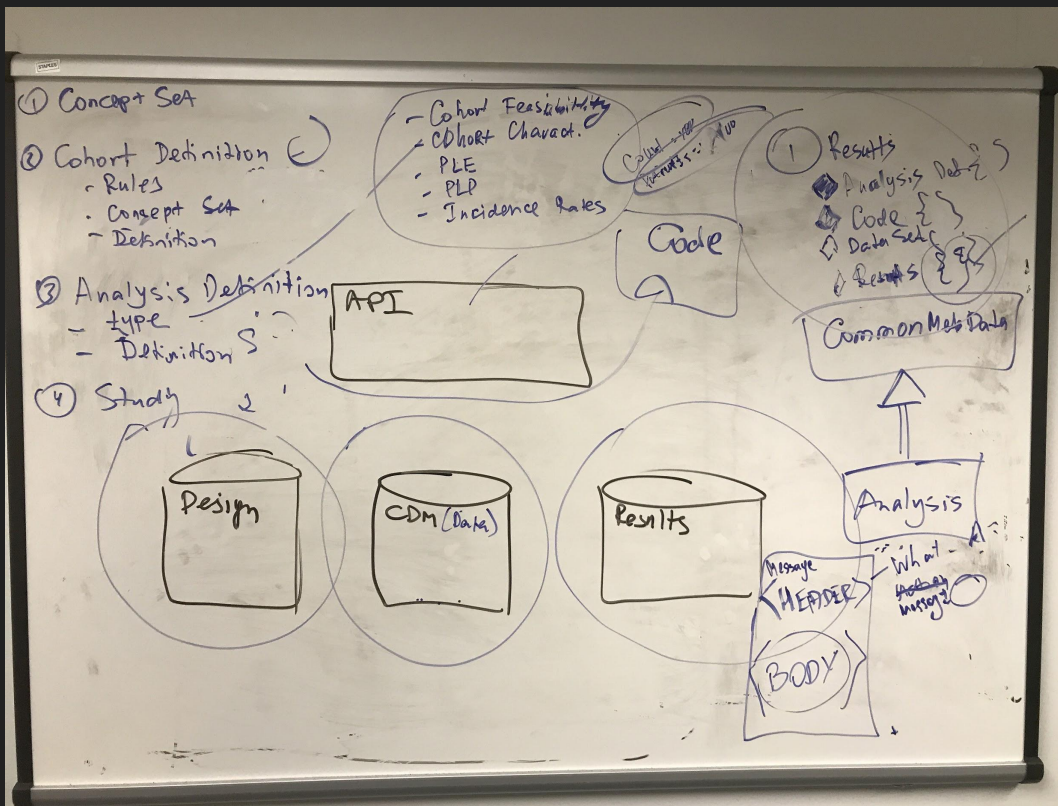
Solutions

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