

COU era®

cloudera

Precision Public Health

Shawn Dolley

Larkin Kay

cloudera®

cloudera

Precision Public Health

Shawn Dolley | Industry Leader, Health & Life Science
November, 2017

Best-in-class organizations use Cloudera

#1

**LARGEST PAYER
IN THE U.S.**

123 million lives and
\$950B to providers

130+

Health & Life Science
organizations use
Cloudera

#1

RANKED

Biotech company

#1

LARGEST

Hospital system in Europe

7/10

Medical device companies

#1

**HEALTH DATA
COMPANY**

500M+ anonymous patient
data records

#1

LARGEST

Health IT company in the
world

HIMSS
STAGE 7

HIGHEST DISTINCTION

This Children's Hospital
received the highest
possible distinction using
Cloudera to perform
machine learning, IoT,
streaming data, and
genomics*

9 of top 10

Global Pharma companies

2 of top 2

U.S. genomic institutes

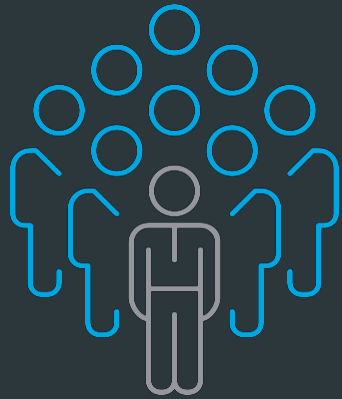
#1

Most utilized patient
centered medical home
program

We believe
data can make what
is impossible today,
possible tomorrow



Precision Health



- n of 1
- 10 MD's per 1 citizens
- 10:1 ratio

Public Health

- Large population
- 20 MD's per 1M citizens
- 1:50,000 ratio

Evolution of precision public health

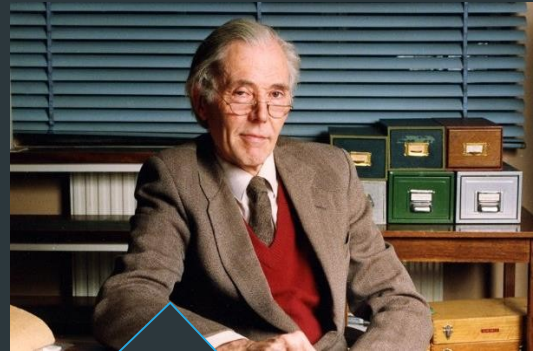
“Precision public health requires robust primary surveillance data, rapid application of sophisticated analytics to track the geographical distribution of disease, and the capacity to act on such information.” (2016)



“...programs are needed which will reduce risk factors among high-risk populations... the target is a particular part of the overall population, rather than the individual episode of sickness.” (1974)



“[Precision public health is] the application and combination of new and existing technologies, which more precisely describe and analyze individuals and their environment...to tailor preventive interventions for at-risk groups and improve the overall health of the population.” (2016)



“...a large number of people at a small risk may give rise to more cases of disease than the small number who are at a high risk.” (2001)



“By the mid-2030s, I believe we can send humans to orbit Mars...” (2010)

“That’s the promise of precision medicine -- delivering the right treatments...to the right person.” (2015)

“And one of the things that has worked so far is us putting Special Forces in...” (2016)

Use cases in Precision Public Health with big data

air quality

Antibiotic resistance

blood lead levels

child abuse

child asthma

child obesity

diabetes

drowning

drug safety

heat wave

Hepatitis C

HIV

Influenza A, H1N1

injectable drug use

Lyme disease

opioid abuse

pre-term birth

smoking

tuberculosis

vaccines

Zika

CLEARPATH: Cleveland Area Platform for Advancing Translational Healthcare

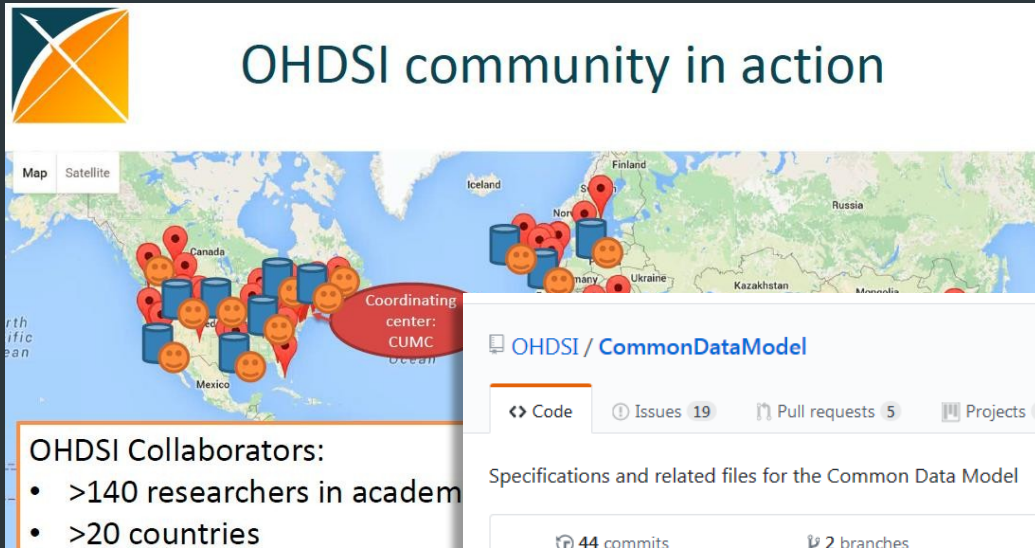
WHAT: We are planning to map data from the EMR systems of hospitals in the Cleveland area as well as data from consented research studies, biospecimen data, disease registry data, exomes, imaging data, other genomic data, and non-medical data such as environmental, for the purpose of creating a Cleveland Area Platform for Advancing Translational Healthcare (CLEARPATH). Data from the hospitals will be aggregated, as a limited data set (LDS) under HIPAA, into a University owned data warehouse and de-duplicated. Linkage of the clinical data sets to the genomic data will be key to the success of this project. Parallel research may include NLP, indexing and other

WHY: In order to do longitudinal studies of patients in the Cleveland metropolitan area and continue to contribute to the knowledge base which is required to achieve evidence based medicine

WHERE: This would be stood up on-premise both behind hospital firewalls and in a university controlled data center. An aggregated secure research environment owned by Case Western Reserve University

WHO: 3 million patient records, initially up to 1000 exomes with RNA seq (focused on whole exome or targeted sequencing panels such as HIV and Cancer)

Open source data model for healthcare

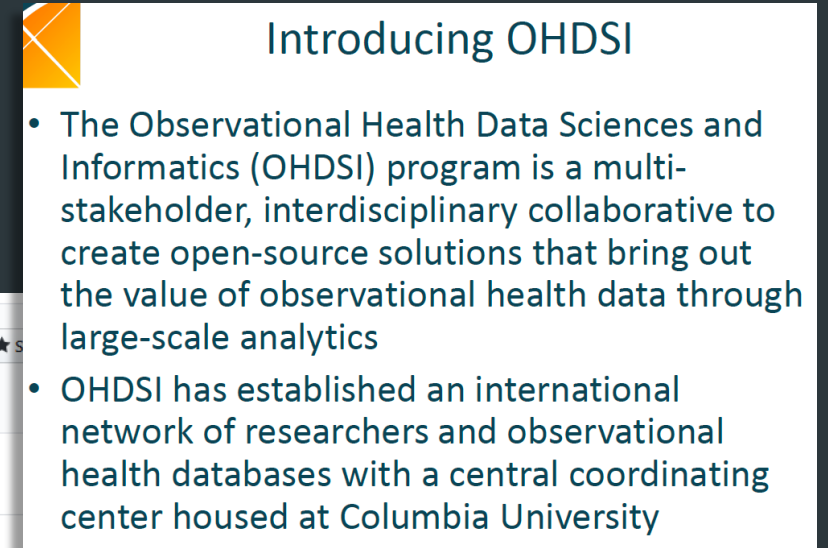


OHDSI community in action

Map showing OHDSI collaborators across various countries, including Canada, Mexico, Iceland, Norway, Sweden, Finland, Russia, Ukraine, and Kazakhstan. A red callout bubble indicates the Coordinating center: CUMC.

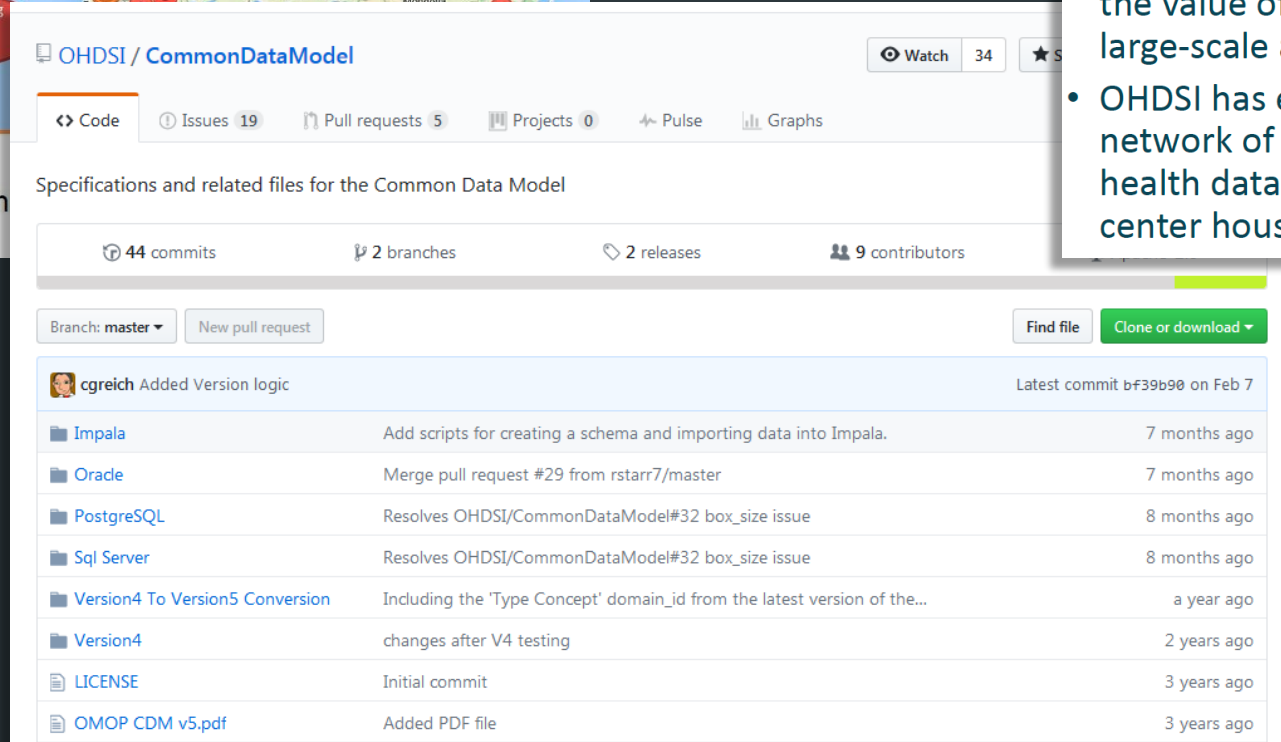
OHDSI Collaborators:

- >140 researchers in academia
- >20 countries



Introducing OHDSI

- The Observational Health Data Sciences and Informatics (OHDSI) program is a multi-stakeholder, interdisciplinary collaborative to create open-source solutions that bring out the value of observational health data through large-scale analytics
- OHDSI has established an international network of researchers and observational health databases with a central coordinating center housed at Columbia University



OHDSI / CommonDataModel

Specifications and related files for the Common Data Model

44 commits | 2 branches | 2 releases | 9 contributors

Branch: master | New pull request | Find file | Clone or download

Commit	Description	Time
cgreich	Added Version logic	Latest commit bF39b90 on Feb 7
Impala	Add scripts for creating a schema and importing data into Impala.	7 months ago
Oracle	Merge pull request #29 from rstarr7/master	7 months ago
PostgreSQL	Resolves OHDSI/CommonDataModel#32 box_size issue	8 months ago
Sql Server	Resolves OHDSI/CommonDataModel#32 box_size issue	8 months ago
Version4 To Version5 Conversion	Including the 'Type Concept' domain_id from the latest version of the...	a year ago
Version4	changes after V4 testing	2 years ago
LICENSE	Initial commit	3 years ago
OMOP CDM v5.pdf	Added PDF file	3 years ago

Embrace precision public health,
Embrace open science,
Pick a use case...



Being data-driven is a journey

Think big. Start small. Iterate often.

5 keys to success

- 1) Build a **data-driven culture**
- 2) Develop the **right team and skills**
- 3) Be **agile/lean** in development
- 4) Leverage **DevOps** for production
- 5) Right-size **data governance**



Start the conversation about data

Catalog available data across agencies

Brainstorm use cases

Align to the strategic initiative

Prioritize work

Leverage agile principles to prioritize data ingest and use cases/insights

Thank you

Shawn Dolley
Larkin Kay