Knowledge sharing to advance health: perspectives from OHDSI Data Collaborative

Dr. Juan M. Banda
12/04/2020
What is OHDSI?

- *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.

- Vision: OHDSI collaborators access a network of 1 billion patients to generate evidence about all aspects of healthcare. Patients and clinicians and other decision-makers around the world use OHDSI tools and evidence every day.
OHDSI Collaborators:
- 2,770 users
- 25 workgroups
- 18,700 posts on 3,250 topics

OHDSI Network:
- 152 databases
- 18 countries
- approx. 600M patient records
Current Approach: “One Study – One Script”

"What's the adherence to my drug in the data assets I own?"

Analytical method: Adherence to Drug

Application to data

One SAS or R script for each study

- Not scalable
- Not transparent
- Expensive
- Slow
- Prohibitive to non-expert routine use
Solution: Data Standardization Enables Systematic Research

- Adherence
- Mortality
- Source of Business
- Safety Signals

OHDSI Tools

OMOP CDM

- North America
- Southeast Asia
- China
- Europe
- UK
- Japan
- India
- So Africa
- Switzerland
- Italy
- Israel

Standardized data
Analytics can be remote
Analytics can be behind firewall
Network Studies
Networks of networks

Coordinating Center

Another Network

Network

ISDN
University Medical Center
Inpatient Hospital
Outpatient Hospital
OHDSI strategy

- Standardize the data
- Standardize the analytics

... simple right?
So, what is the cost of admission?
OHDSI provides all the tools... Open Source!

- Standard Common Data Model is the cornerstone
What OHDSI is not

- A centralized data network
- A federated query mechanism
- An NLP pipeline / data conversion tool
- A big dataset of publicly available data
OHDSI provides all the tools... Open Source!

• Standardized Common Vocabulary – Athena
  • https://athena.ohdsi.org/search-terms/start

• Data ETL Design – White Rabbit
  • https://www.ohdsi.org/analytic-tools/whiterabbit-for-etl-design/

• Data Quality - With Achilles and Data Quality Dashboard
  • https://data.ohdsi.org/DataQualityDashboard/
Standardized Phenotyping in OHDSI

• You can use the traditional rule-based way of building phenotypes (like PheKB) and build them in ATLAS
  • https://www.ohdsi.org/atlas-a-unified-interface-for-the-ohdsi-tools/

• Brand new phenotype library contains a hefty list of several dozen phenotypes freely available
  • https://data.ohdsi.org/PhenotypeLibrary/

• Looking for machine learning, probabilistic phenotypes?
  • APHRODITE is the tool - https://github.com/OHDSI/Aphrodite
And you can SHARE all these objects!

- ATLAS cohorts are JSON objects that can be shared
- Phenotype Library comes as an R package
- APHRODITE phenotypes can be shared and are FAIR
And plenty more tools/software available

- R Packages – CohortMethod, PatientLevelPrediction,
Characterizing treatment pathways at scale using the OHDSI network

George Hripcsak\(^{a,b,c,}\), Patrick B. Ryan\(^{c,}\), Jon D. Duke\(^{c,}\), Nigam H. Shah\(^{c,}\), Rae Woong Park\(^{c,}\), Vojtech Huser\(^{c,}\), Marc A. Suchard\(^{c,}\), Martijn J. Schuemie\(^{c,}\), Frank J. DeFalco\(^{c,}\), Adler Perotte\(^{a,}\), Juan M. Banda\(^{c,}\), Christian G. Reich\(^{c,}\), Lisa M. Schilling\(^{c,}\), Michael E. Matheny\(^{c,}\), Daniella Meeker\(^{c,}\), Nicole Pratt\(^{c,}\), and David Madigan\(^{c,}\)

Medical Informatics Services, NewYork-Presbyterian 0032; \(^{a}\)Epidemiology Analytics, Janssen Research and Development, IN 46205; \(^{b}\)Center for Biomedical Informatics Research, University of Medicine, Suwon, South Korea, 443-380; \(^{c}\)Laboratories of Health, Bethesda, MD 20894; \(^{d}\)Department of Informatics, University of California, Los Angeles, CA 90095; \(^{e}\)Department of Health Informatics, Vanderbilt University Medical Center, Nashville, TN 37212; \(^{f}\)Department of Biostatistics, University of Southern California, Los Angeles, CA 90033; \(^{g}\)Department of Statistics, Columbia University, New York, NY 10027.
COVID-19 efforts – Study-A-thon

In only 88 hours, we have:

- Convened 351 participants from 30 countries
- Held 12 Global Huddles, >100 calls, >13,000 chat messages
- Engaged 15 concurrent channels
- Reviewed >10,000 publications
- Drafted 9 protocols
- Released 13 study packages
- Designed 355 cohort definitions
- Assembled a distributed data network with 37 partners signed on to execute studies

(source: Panacea Lab)
APHRODITE phenotypes in multiple sites and countries

Development and validation of phenotype classifiers across multiple sites in the observational health data sciences and informatics network

Mehr Kashyap, Martin Seneviratne, Juan M Banda, Thomas Falconer, Borim Ryu, Sooyoung Yoo, George Hripcsak, Nigam H Shah


Published: 06 May 2020  Article history

https://doi.org/10.1093/jamia/ocaa032
More resources

• The OHDSI Bible:
  • [https://ohdsi.github.io/TheBookOfOhdsi/](https://ohdsi.github.io/TheBookOfOhdsi/)

• Questions? Need some help joining the OHDSI journey?
  • [jbanda@gsu.edu](mailto:jbanda@gsu.edu)