Workgroup OKRs + Phenotype Phebruary Update #4

OHDSI Community Call
Feb. 27, 2024 • 11 am ET

Workgroups: FHIR + OMOP, Health Equity, the Africa Chapter, CDM Vocabulary, Electronic Animal Health Records, Phenotype Development & Evaluation, Dentistry, Medical Devices, Medical Imaging, GIS – Geographic Information System, and Clinical Trials

Week 4 Phenotype Focus: pulmonary arterial hypertension
# Upcoming Community Calls

<table>
<thead>
<tr>
<th>Date</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Feb. 27</td>
<td>Workgroup OKRs / Phenotype Phebruary Update 4</td>
</tr>
<tr>
<td>Mar. 5</td>
<td>New Vocabulary Release Update</td>
</tr>
<tr>
<td>Mar. 12</td>
<td>TBA</td>
</tr>
<tr>
<td>Mar. 19</td>
<td>NO MEETING</td>
</tr>
<tr>
<td>Mar. 26</td>
<td>Recent OHDSI Publications</td>
</tr>
<tr>
<td><em>coming in April</em></td>
<td></td>
</tr>
<tr>
<td><em>CDM Month</em></td>
<td></td>
</tr>
</tbody>
</table>
Three Stages of The Journey

Where Have We Been?
Where Are We Now?
Where Are We Going?
OHDSI Shoutouts!

Congratulations to the team of Christine Mary Hallinan, Roger Ward, Graeme K Hart, Clair Sullivan, Nicole Pratt, Ashley Ng, Daniel Capurro, Anton Van Der Vegt, Siaw-Teng Liaw, Oliver Daly, Blanca Gallego Luxan, David Bunker and Douglas Boyle on the publication of Seamless EMR data access: Integrated governance, digital health and the OMOP-CDM in BMJ Health & Care Informatics.

Seamless EMR data access: Integrated governance, digital health and the OMOP-CDM

Christine Mary Hallinan,1 Roger Ward,1 Graeme K Hart,2 Clair Sullivan,3 Nicole Pratt,4 Ashley P Ng,5 6 Daniel Capurro,7 Anton Van Der Vegt,8 Siaw-Teng Liaw,9 Oliver Daly,10 Blanca Gallego Luxan,10 David Bunker,4 Douglas Boyle1


Received 29 October 2023 Accepted 14 January 2024

ABSTRACT

Objectives In this overview, we describe the Observational Medical Outcomes Partnership Common Data Model (OMOP-CDM), the established governance processes employed in EMR data reporitng, and demonstrate how OMOP transformed data provides a lever for more efficient and secure access to electronic medical record (EMR) data by health service providers and researchers.

Methods Through pseudonymisation and common data quality assessments, the OMOP-CDM provides a robust framework for converting complex EMR data into a standardised format. This allows for the creation of shared end-to-end analysis packages without the need for direct data exchange, thereby enhancing data security and privacy. By securely sharing de-identified and aggregated data and conducting analyses across multiple OMOP-connected databases, patient-level data is securely shared within its respective local site.

Results By simplifying data management processes and governance, and through the promotion of interoperability, the OMOP-CDM supports a wide range of clinical, epidemiological, and translational research projects, as well as health service operational reporting.

Discussion Adoption of the OMOP-CDM improves and locally enables conversion of vast amounts of complex, and heterogeneous EMR data into a standardised structured data model, simplifies governance processes, and facilitates rapid repeatable cross-institution analysis through shared end-to-end analysis packages, without the sharing of data.

Conclusion The adoption of the OMOP-CDM has the potential to transform health data analytics by providing a common platform for analysing EMR data across diverse healthcare settings.

Electronic medical record (EMR) data into a standardised structured data model. The conversion of data has the potential to provide hospitals, health departments, auditors, regulators and universities valuable insights tailored to each institution’s needs, both for operational and research purposes. This is achievable as long as the secure utilisation of an institution’s EMR clinical and administrative data for purposes beyond its initial collection, known as ‘secondary use’, is effectively managed and employed.

Such data can be transformative, especially if used to monitor, evaluate and audit healthcare to improve clinical practice, reduce inefficiencies, contribute to the evidence base and develop a ‘learning healthcare system’ for improved patient care.1,3,4 However, this potential is often not realised due to the inherent complexity of EMR databases—what comprise thousands of data elements across thousands of proprietary tables—where vast amounts of data needs to be transformed, cleansed and restructured to make it fit for ‘secondary use’.5 For highly powered collaborative research, where large volumes of EMR data are combined, use is further constrained by the heterogeneity of each institution’s EMR schema; concerns over data sharing and privacy breaches and lack of clarity over governance and consent.6
Three Stages of The Journey

Where Have We Been?
Where Are We Now?
Where Are We Going?
## Upcoming Workgroup Calls

<table>
<thead>
<tr>
<th>Date</th>
<th>Time (ET)</th>
<th>Meeting</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wednesday</td>
<td>9 am</td>
<td>OMOP CDM Oncology Outreach/Research Subgroup</td>
</tr>
<tr>
<td>Wednesday</td>
<td>10 am</td>
<td>Surgery and Perioperative Medicine</td>
</tr>
<tr>
<td>Wednesday</td>
<td>12 pm</td>
<td>Latin America</td>
</tr>
<tr>
<td>Wednesday</td>
<td>3 pm</td>
<td>Vulcan/OHDSI Meeting (ZOOM)</td>
</tr>
<tr>
<td>Thursday</td>
<td>7 pm</td>
<td>Dentistry</td>
</tr>
<tr>
<td>Friday</td>
<td>10 am</td>
<td>GIS – Geographic Information System</td>
</tr>
<tr>
<td>Friday</td>
<td>11:30 am</td>
<td>Steering Group</td>
</tr>
<tr>
<td>Friday</td>
<td>11:30 am</td>
<td>Clinical Trials</td>
</tr>
<tr>
<td>Monday</td>
<td>10 am</td>
<td>Africa Chapter</td>
</tr>
<tr>
<td>Tuesday</td>
<td>10 am</td>
<td>Common Data Model</td>
</tr>
</tbody>
</table>
OHDSI Global Symposium

The 2024 OHDSI Global Symposium will be held Oct. 22-24 at the Hyatt Regency Hotel in New Brunswick, NJ.

Tentative symposium format:

- **Oct. 22** – tutorials/workshops
- **Oct. 23** – main conference
- **Oct. 24** – workgroup activities
Registration is now OPEN for the 2024 OHDSI Europe Symposium, which will be held June 1-3 in Rotterdam, Netherlands.

**June 1** – tutorial/workshop
**June 2** – tutorial/workshop
**June 3** – main conference
Brazilian administrative data for real-world research: a deterministic linkage procedure and OMOP CDM harmonization

(Jessica Mayumi Maruyama, Julio Cesar Barbour Oliveira)

1. Background

- Brazilian population: nearly 214 millions of inhabitants
- 75% depends exclusively on Brazilian National System (SUS)
- DATASUS- national administrative database publicly available

Challenges and opportunities for DATASUS databases use in Real-World Evidence (RWE) studies:

- No integration between systems
- Lack of unique key identifier at individual level
- Lack of data standardization
- Missing data
- Technical and methodological difficulties related to linkage procedure

Aim: To describe methods and partial results of the minimal viable product (MVP) for parameter setting in creating a dataset from Brazilian claims data, assessing data quality against an OMOP CDM

2. Methods

- Ambulatory Information System
- Hospital Information System

Outpatient procedures, consultations, ICD-10 codes of a primary and secondary diagnosis, medicines, and personal data

Inpatient personal information, procedures, treatments and separation (hospital discharges, transfers, and deaths)

Record linkage and OMOP CDM harmonization

Exclusions: inconsistencies in basic information (date of birth or gender); Patients with different primary keys but with matching basic information

Key Information: zip code, date of birth, and gender

Exclusion: Patients from zip codes with more than 2500 distinct individuals linked to them, patients with more than 3 distinct zip codes, and patient's id before 2012

Data Quality Dashboard: overall pass rate of 94.5% (out of 960 tests)

3. Results

Standardized dataset encompassing the complete health history of 5.82 million patients

Table 1. Data Quality Dashboard of our MVP OMOP DATASUS

<table>
<thead>
<tr>
<th>% Pass</th>
<th>Plausibility</th>
<th>Conformance</th>
<th>Completeness</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>92.6</td>
<td>92.5</td>
<td>100.0</td>
</tr>
</tbody>
</table>

4. Conclusions

We showed a data treatment methodology for DATASUS that produces a high-quality dataset. Moving forward, our objectives include crafting a manuscript that delineates the methodology, seeking validation from the scientific community. Furthermore, we anticipate conducting and publishing Real-World Evidence studies utilizing this dataset in the upcoming months.
Estimating model performance on external data sources from their summary statistics: a real-world benchmark

(Tal El-Hay, Jenna M Reps, Chen Yanover)

TUESDAY

Title: Estimating model performance on external data sources from their summary statistics: a real-world benchmark

PRESENTER: Tal El-Hay

INTRO:

- External validation is often costly or even infeasible as access to patient-level data is typically limited.
- Here, we test a novel method that estimates model performance on external data sources from their limited statistical characteristics.
- We use data from five US datasets and prediction models for various outcomes in individuals with major depression.

METHODS

1. Datasets: CCAE, MedcoCare, Medicare, Optum-CDM, Optum-DH
2. Clinical prediction task: risk of developing fracture, sepsis, diabetes, ischemic, and gastrointestinal bleed in patients within 1 to 365 days after initial diagnosis of major depressive disorder.
3. Estimation method: fit first weights that induce internal weighted statistics that are similar to the external ones. Estimate AUROC and Binet (calibration) scores using the weighted sample and internal model predictions.

Summary statistics of external datasets may allow detection of poor model generalization before patient-level data is accessible

Results

Absolute difference between internal and external AUROC versus absolute difference between estimated and external ones. Every pair of bars correspond to analysis of five outcome models in a single internal-external dataset combination.

Detailed AUROC results

Brier score estimation are very accurate for both model types.

Additional notes:

- Additional use: Interrogate sources of data that affect performance.
- Ongoing effort: explore cases where AUROC estimation is fair and improve accuracy.

Tal El-Hay1, Jenna M Reps3, Chen Yanover1

1 Silverstein Institute, Kfar Habad, Israel
2 Jerusalem Research and Development, Kfar Saba, IL, USA
Harmonization of OMOP vaccine-related vocabularies through the Vaccine Ontology

(Yuanyi Pan, Warren Manuel, Rashmie Abeysinghe, Xubing Hao, Alexander Davydov, Qi Yang, Asiyah Yu Lin, Licong Cui, Yongqun Oliver He)

Harmonization of OMOP vaccine-related vocabularies through the Vaccine Ontology

Yuanyi Pan, Warren Manuel, Rashmie Abeysinghe, Xubing Hao, Alexander Davydov, Qi Yang, Asiyah Yu Lin, Licong Cui, Yongqun Oliver He

1. University of Michigan Medical School, Ann Arbor, MI, USA. 2. McEwan’s School of Biomedical Informatics, The University of Texas Health Science Center at Houston, Houston, TX, USA. 3. Department of Neurology, The University of Texas Health Science Center at Houston, Houston, TX, USA. 4. Otsuka Data Services, Inc., Cambridge, MA, USA. 5. OHDSI, LLC, King of Prussia, PA, USA. 6. National Institutes of Allergy and Infectious Diseases, Bethesda, MD, USA.

* These authors share first authorship; † Co–corresponding authors.

Vaccines have an important role in fighting against infectious diseases such as COVID-19. OMOP/OMOP CDI associated vocabularies (e.g., CDC Vaccine Administered CSV, ReKomm, and OMOP-CTS) include a variety of vaccine-related terms. However, these vaccine vocabularies have different coverages and use different design patterns and representation styles. As a result, the vaccine terms in these vocabularies could not be easily mapped and integrated.

To address the above challenge, we have formed an OMOP Vaccine Vocabulary Working Group (Vaccine Vocab WG) to map and integrate different vaccine vocabularies. Our basic strategy is to use the Vaccine Ontology (VO), a community-based biomedical ontology in the vaccine domain, as the platform to systematically represent the mapped results for vaccine terms in individual vocabularies. We started with mapping of CVX vaccine terms to the VO using manual and semi-automated strategies.

Our method is divided into manual mapping and updating, and semi-automated approach (Fig. 1). For semi-automated mapping approach, all these methods generated a similarly score for a CVX and VO term pair. A threshold was set by experimentation where all the CVX and VO term pairs above this threshold were considered as "matched." Furthermore, each method generated up to 30 candidate sets of matching VO terms for a CVX term which could be further reviewed by domain experts for confirmation.

Prior to similarity calculation across all the methods, normalization of textual information was performed via lowercasing and ASCII conversion, and expansion of common vaccine-related abbreviations and trade names sourced from the CDC.

Results

CVX-VO mapping and VO updating:

A total of 88 CVX-VO mapping pairs were identified. Additionally, we identified 69 CVX terms that have corresponding terms in VO but have no direct mapping annotations; these mappings were added using "RedirectHash" annotation property in VO (Fig. 2). Our study found 134 CVX terms not initially present in the VO, which were then added to VO accordingly.

Semi-automated mapping approach: The 4,302 vaccine terms under the VO concept ‘vaccine’ (VO:00000001) and all CVX terms were considered here. The results of the semi-automated method were compared with the manually annotated mappings. With the manual annotation gold standard, we evaluated the performance of the approaches in terms of precision, recall, and F1-score. The results are given in Table 1. Overall, in terms of F1 score, the hybrid method was found to be the best out of the three methods. Table 2 shows 5 examples for valid mappings obtained with the hybrid method.

Table 1: Performance of each model considering manual evaluation as a gold standard.

<table>
<thead>
<tr>
<th>Method</th>
<th>Precision</th>
<th>Recall</th>
<th>F1 score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manual manual mapping</td>
<td>0.17503</td>
<td>0.586</td>
<td>0.31515</td>
</tr>
<tr>
<td>ReKomm</td>
<td>0.45104</td>
<td>0.586</td>
<td>0.53193</td>
</tr>
<tr>
<td>Hybrid</td>
<td>0.88882</td>
<td>0.586</td>
<td>0.73448</td>
</tr>
</tbody>
</table>

Table 2: Five valid CVX to VO mappings identified by our hybrid method.

<table>
<thead>
<tr>
<th>CVX Term</th>
<th>VO Term</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>CVX:00000001 - Vaccine</td>
<td>VO:00000001 - Vaccine</td>
<td>Vaccine</td>
</tr>
<tr>
<td>CVX:00000002 - Vaccine Administration</td>
<td>VO:00000002 - Vaccine Administration</td>
<td>Vaccine Administration</td>
</tr>
<tr>
<td>CVX:00000003 - Vaccine Administration</td>
<td>VO:00000003 - Vaccine Administration</td>
<td>Vaccine Administration</td>
</tr>
<tr>
<td>CVX:00000004 - Vaccine Preparation</td>
<td>VO:00000004 - Vaccine Preparation</td>
<td>Vaccine Preparation</td>
</tr>
<tr>
<td>CVX:00000005 - Vaccine Distribution</td>
<td>VO:00000005 - Vaccine Distribution</td>
<td>Vaccine Distribution</td>
</tr>
</tbody>
</table>

Conclusions

Overall, we applied both manual and semi-automated methods to map CVX and VO vaccine terms and updated VO correspondingly. The hybrid method used in this study was shown to outperform the other two methods. The semi-automated methods can be promising as they require significantly less human effort than purely manual approaches. With expanded coverage and interoperability, the updated VO will further be used for systematic and integrative analysis of vaccine-related clinical data available in the OMOP/OMOP-compliant systems.

Acknowledgments

This study is supported by NIH through grants U44AI137008 and R03AI126287.
OHDSI on Databricks: A Complete Guide to Implementing OHDSI on Databricks

(John Gresh, Brad Rechkemmer)

THURSDAY

OHDSI on Databricks: A Complete Guide to Implementing OHDSI on Databricks

John Gresh1, Brad Rechkemmer1
1National Association of Community Health Centers (NACHC), 2Oregon

Background

Methods

OHDSI on Databricks: A Complete Guide to Implementing OHDSI on Databricks

John Gresh1, Brad Rechkemmer1
1National Association of Community Health Centers (NACHC), 2Oregon

Background

Methods

Results

Conclusions

References

The complete OHDSI on Databricks guide provides an in-depth tutorial on implementing OHDSI on Databricks. The guide is based on an extensive review of the available literature and practical experience gained from implementing OHDSI on Databricks at the National Association of Community Health Centers (NACHC). The guide also includes a comprehensive list of references for further reading.

The OHDSI on Databricks guide is intended for users who are interested in implementing OHDSI on Databricks for their healthcare organizations. The guide is suitable for users with varying levels of experience with Databricks and OHDSI. The guide is available for free download on the OHDSI website.

The OHDSI on Databricks guide is the first of its kind and is designed to help users implement OHDSI on Databricks successfully. The guide includes step-by-step instructions, best practices, and troubleshooting tips to ensure a smooth implementation.

For more information, please visit www.ohdsi.org or contact the authors at john.gresh@nachc.org or brad.rechkemmer@nachc.org.
Antihypertensive medication use in pregnancy: A pilot OHDSI network analysis in electronic health record data

(Stephanie A. Leonard, Louisa H. Smith, Sara Siadat, Karthik Natarajan, Brian T. Bateman, Thomas Falconer, John DiPalazzo, Alison Callahan)
Opening: Three Positions at Gilead

Sr. Director, Head of Data Office

**Job Description:**
As a Senior Director in our Data Office, you will play a pivotal role in shaping and executing our data strategy. In this leadership position, you will oversee and drive activities related to data sharing, governance, and access across the organization. Working closely with cross-functional teams, you will define and implement data acquisition policies and practices, ensuring the efficient and effective use of data to support our scientific and business objectives.

Director, Data Acquisition - Clinical Data Science

**Responsibilities:**
This role reports to the Head of Gilead data office, RWE Generation, Clinical Data Science and is based at different Gilead sites. This individual has responsibility for acquiring all data across clinical, development, medical affairs function and Gilead affiliates. This individual will work in close collaboration with the Development organization, Commercial, Procurement, Medical Affairs, IT, and other functions at Gilead in implementing data acquisition processes and is expected to operate with a "one Gilead" mindset & play a key role in the global Gilead Data Office set up.

Director, RWE - Data Science - OHDSI

**Responsibilities:**
Collaborate with researchers and data scientists to understand project requirements and translate them into OHDSI-compatible solutions. Work with databases, ensuring data integrity and optimization for OHDSI-related queries and analyses. Perform data analyses in OHDSI-related tools like ATLAS. Customize and extend OHDSI tools and applications to meet specific project needs. Collaborate with cross-functional teams to troubleshoot and resolve technical issues related to OHDSI implementations. Stay informed about OHDSI community updates, best practices, and emerging trends in observational health data research. Contribute to the development and documentation of data standards and conventions within the OHDSI community.
The Zhang Lab at Washington University School of Medicine in St. Louis has one postdoc/senior data analyst position to work on causal machine learning and responsible AI for reliable real-world evidence generation.

- More details at https://linyingzhang.com
  - Postdoc: https://linyingzhang.com/files/Postdoc.pdf
  - Data analyst: https://linyingzhang.com/files/Analyst.pdf

- If interested, please send CV and cover letter to linyingz@wustl.edu
Opening: Epidemiology UX/Web Design Intern at J&J

Career Programs

Epidemiology UX/Web Design Intern

<table>
<thead>
<tr>
<th>JOB TITLE</th>
<th>Epidemiology UX/Web Design Intern</th>
</tr>
</thead>
<tbody>
<tr>
<td>FUNCTION</td>
<td>Career Programs</td>
</tr>
<tr>
<td>SUB FUNCTION</td>
<td>Non-LDP Intern/Co-Op</td>
</tr>
<tr>
<td>LOCATION</td>
<td>Raritan, New Jersey, United States</td>
</tr>
<tr>
<td>DATE POSTED</td>
<td>Jan 19 2024</td>
</tr>
<tr>
<td>REQUISITION NUMBER</td>
<td>2406163977W</td>
</tr>
</tbody>
</table>

DESCRIPTION

Janssen Research & Development, LLC., a division of Johnson & Johnson's Family of Companies is recruiting for Epidemiology UX/Web Design Intern. This position is a member of the Observational Health Data Analytics (OHDA) team. OHDA's mission is to improve the lives of individuals and quality of healthcare by efficiently generating real-world evidence from the world's observational health data, transparently disseminating evidence-based insights to real-world decision-makers, and objectively advancing the science and technology behind reliability.
## Opening: Research Information Specialist at UNC

**Research Informatics Specialist**

Please see Special Instructions for more details.

Posting hours are Monday-Friday, 8:00 am – 6:00 pm EST with flexibility available within that window.

### Position Information

<table>
<thead>
<tr>
<th>Department</th>
<th>TraCS Institute-429801</th>
</tr>
</thead>
<tbody>
<tr>
<td>Career Area</td>
<td>Information Technology</td>
</tr>
<tr>
<td>Posting Open Date</td>
<td>12/13/2023</td>
</tr>
<tr>
<td>Application Deadline</td>
<td>01/30/2024</td>
</tr>
<tr>
<td>Open Until Filled</td>
<td>No</td>
</tr>
<tr>
<td>Position Type</td>
<td>Permanent Staff (EHRA NF)</td>
</tr>
<tr>
<td>Working Title</td>
<td>Research Informatics Specialist</td>
</tr>
<tr>
<td>Appointment Type</td>
<td>EHRA Non-Faculty</td>
</tr>
<tr>
<td>Position Number</td>
<td>20660002</td>
</tr>
<tr>
<td>Vacancy ID</td>
<td>NH0007640</td>
</tr>
<tr>
<td>Full Time/Part Time</td>
<td>Full-Time Permanent</td>
</tr>
<tr>
<td>FTE</td>
<td>1</td>
</tr>
</tbody>
</table>

### Position Summary

**Responsibilities include:**

* Perform SQL-based programming against UNC’s clinical data warehouse to identify patient cohorts and develop patient datasets.
* Consult with and collaborate with researchers to ensure programming work aligns with project needs.
* Develop ETL (extract, transform, and load) and data integration processes to support common data models (OMOP, PCORnet) using appropriate technologies (SQL, Python, or R).
* Carefully following UNC’s regulatory and governance policy to ensure data integrity and security.
* In collaboration with IDSci team, identify potential enhancements in current workflows and data architecture.
* Implement quality assurance strategies, such as data validation and peer code review.
* Write and maintain up-to-date supporting documentation. Ensure code is well-commented and use GitLab/GitHub to manage code changes and track data lineage.
* Provide technical leadership and direction for assigned projects and/or data requests.

### Minimum Education and Experience Requirements

Master’s and 1-2 years’ experience; or Bachelors and 2-4 years’ experience; or will accept a combination of related education and experience in substitution.

### Required Qualifications, Competencies, and Experience

* Expert-level knowledge of SQL programming, data modeling, and relational database systems such as Oracle, Microsoft SQL Server, MySQL, etc.
* Demonstrable past experience in developing technical projects in terms of length of time, competencies and cost. Individual will be expected to manage multiple projects at once while delivering high-quality work on time.
* Excellent written and oral business communication skills. Public speaking at meetings and conferences may be required. The ability to clearly convey technical concepts to non-technical clients is a must.
Opening: Data Steward at EBMD

Description

Are you looking for a job where you can make a difference and work in a non-profit? Would you like to be a part of an ambitious and international organisation on the cutting edge of science? Then this position might be right up your alley.

The EBMT is a non-profit medical and scientific organisation which hosts a unique patient registry providing a pool of data to perform studies and assess new trends.

OUR MISSION
Save and improve the lives of patients with blood-related disorders.

The Registry
Holding the data of over half a million patients, the EBMT registry is the starting point for all studies carried out through the EBMT working parties. The department focuses on data collection processes, data quality monitoring, and maintenance of the database.

YOUR MISSION
Responsible for collecting, collating, and evaluating issues and problems with data and enforcing data usage policies.

RESPONSIBILITIES AND TASKS

Data Stewardship:
- Design, implementation and testing of new data collection processes including data collection forms (DCFs) development.
- Take care of the mapping of new items from DCFs to the OMOP CDM
- Providing input on data quality reports
- Check and clean data on request and ad hoc.
- Data retrieval including designing data reports and data report running.
- Carry out computerized system validation activities.
- Supporting consolidation/harmonization of data
- Creating standard data definitions, and maintain a consistent use of data assets across the organization
- Documenting data policies and data standards
Where Are We Going?

Any other announcements of upcoming work, events, deadlines, etc?
Three Stages of The Journey

Where Have We Been?
Where Are We Now?
Where Are We Going?
Learn more about all of the OHDSI workgroups

ohdsi.org/workgroups
OMOP + FHIR Working Group
2024 OKRs

Davera Gabriel
Evidentli, LLC
OMOP + FHIR WG Co-Lead
FHIR+ OMOP WG 2024 Purpose

To facilitate the collaboration between OHDSI and HL7 agreed by both parties in 2021. The work group will develop and validate standard transformation specifications and canonical maps between data conformant to FHIR to OMOP CDM, and from OMOP CDM to FHIR.
FHIR + OMOP 2024 Objectives

- Complete a draft specification transforming FHIR to OMOP v5.4 for core EMR data elements as a joint effort with the Vulcan accelerator FHIR to OMOP project.

- Convene or participate in at least one Hack- / Transform-athon meeting(s) to validate and improve generated specifications.

- Co-develop SDoH section of FHIR to OMOP Implementation Guide with Health Equity WG focused on
  - Gender Harmony
  - Occupational Data for Health: Industry & Occupation
Join Us!

OHDSI OMOP + FHIR WG / Vulcan FHIR-to-OMOP Project Joint Meeting

Weekly on Wednesdays 3p (EASTERN US)

Zoom Link

https://us02web.zoom.us/j/87666493433?pwd=TzMyWncyRVYxSVNsT29WMXVUZEs3QT09

Meeting ID: 876 6649 3433
Passcode: 8675309

Meeting Minutes on Confluence
OHDSI Health Equity Working Group OKRs

2024
Objectives

1. Generate and disseminate real-world evidence about the substantial public health issue of health inequities

2. Operationalize individual-level Social Determinants of health, Risk factors, and Needs (SDRN), and other data elements relevant to health equity work in OHDSI network studies

3. Operationalize place-based public data sources in OHDSI network studies

4. Extend OHDSI tools to make a health equity perspective the default and/or an option

5. Engage the broader community on issues related to health equity

6. Support the work of the group
2024 OHDSI Health Equity Working Group Goals

- **Research**: Advance Health Equity Research
- **Engagement**: Community Engagement and Awareness
- **Integration**: OHDSI Community Development and Integration
2024 OHDSI Health Equity Working Group Goals

**Research**

**Milestones**
- Conduct at least 2 health equity-oriented studies
- Invite OHDSI and non-OHDSI speakers present their work at part of on-going Journal clubs
- \((STRETCH)\) Collaborate on at least 1 paper/abstract/conference submission based on health equity research by the end of the year.

**Engagement**

**Milestones**
- Host at least 2 webinars or workshops on health equity topics throughout the year, inviting experts and community members.
- \((STRETCH)\) Establish a feedback mechanism to actively involve OHDSI community members in shaping health equity initiatives.

**Integration**

**Milestones**
- Collaborate with \textbf{GIS WG} to quantify the impact of social and environmental determinants of health (SEDoH) on healthcare resource allocation
- Co-develop SDoH section of FHIR to OMOP Implementation Guide with \textbf{OMOP+FHIR WG} focused on Gender Harmony and Occupational Data for Health: Industry & Occupation
Africa Chapter
Chapter Leads: Cynthia Sung, Agnes Kiragga

Purpose
• To strengthen awareness and capacity for data harmonization and analysis using OHDSI tools to meet the data-driven evidence needs of African researchers, health providers, and governments

Biweekly Meeting – Monday 10 AM ET
next ones: Mar 4, Mar 18, Apr 1 . . .
# Africa Chapter 2024 Objectives and Key Results

<table>
<thead>
<tr>
<th>Objective 1</th>
<th>Key Result</th>
<th>OKR Lead(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Create a guideline for African Data Access</td>
<td>Guideline with recommended steps &amp; examples of approved access requests*</td>
<td>Cynthia Sung, Marc Twagirumukiza (Rwanda), Katherine Johnston (S. Africa)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Objective 2</th>
<th>Key Results</th>
<th>OKR Lead(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ETL new databases in Africa</td>
<td>• ETL of DBs where Chapter member support exists: target 2 DBs*</td>
<td>Mack Kigada (Kenya), Henry Ogoe (Ghana), Kofi Agyare (Ghana), Aize Cao (US, TN)</td>
</tr>
<tr>
<td></td>
<td>• Use INSPIRE Training dataset to demonstrate ETL steps, apply to identified DBs</td>
<td>Agnes Kiragga (Uganda), APHRC staff (Kenya)</td>
</tr>
<tr>
<td></td>
<td>• Identify Vocabulary needs for prioritized use cases</td>
<td>Andy Kanter, David Amadi (Kenya), Jared Houghtaling</td>
</tr>
</tbody>
</table>

*Leverage 2023 Value Proposition documents*
## Africa Chapter 2024 Objectives and Key Results

<table>
<thead>
<tr>
<th>Objective 3</th>
<th>Key Result</th>
<th>OKR Lead(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Collaboration on grant proposals</td>
<td>• Grant opportunities identified &amp; prioritized</td>
<td>Marty Alvarez, Andrew Williams (US)</td>
</tr>
<tr>
<td></td>
<td>• Multi-institutional, multi-national collaborative proposal submitted*</td>
<td>Cynthia Sung, Aize Cao</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Objective 4</th>
<th>Key Results</th>
<th>OKR Lead(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Solidify collaboration with the Africa CDC</td>
<td>Action plan developed by Chapter members to strengthen ties among Africa CDC, US CDC and USAID*</td>
<td>Agnes Kiragga (Uganda), Nega Gebreyesus (Ethiopia), Chidi Asuzu (Nigeria)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Objective 5</th>
<th>Key Results</th>
<th>OKR Lead(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participation of African member in OHDSI network study</td>
<td>Participation can be reviewing literature, contributing to cohort or phenotype definitions, providing statistical support, or running algorithms on African data</td>
<td>Individual choice</td>
</tr>
</tbody>
</table>

*Leverage 2023 Value Proposition documents*
Africa Chapter
Chapter Leads: Cynthia Sung, Agnes Kiragga

Purpose

• To strengthen awareness and capacity for data harmonization and analysis using OHDSI tools to meet the data-driven evidence needs of African researchers, health providers, and governments

Biweekly Meeting – Monday 10 AM ET
next ones: Mar 4, Mar 18, Apr 1 . . .
Vocabulary WG OKRs 2024
Statement of purpose

Support OHDSI community by:

- *answering questions* related to Vocabularies content, development process, community contribution and use
- engaging community in *discussions around Vocabularies* changes and improvements with such discussions being the required step to introduce any changes
- helping community to address their Vocabularies-related *needs*
What we achieved in 2023

Had sessions on: SNOMED overhaul, MedDRA overhaul, race and ethnicity, impact of vocabulary changes on concept sets, LOINC-SNOMED hierarchy

Published and talked about (7+ sessions) community contribution, currently 20+ contributions

SSSOM (meta-data) implemented in community contribution and pending in Usagi

Beyond WG:

- Performed landscape assessment
- Established committee, roadmap, vocabulary cadence [https://github.com/OHDSI/Vocabulary-v5.0/wiki](https://github.com/OHDSI/Vocabulary-v5.0/wiki)
- Published The Vocabulary Paper [https://doi.org/10.1093/jamia/ocad247](https://doi.org/10.1093/jamia/ocad247)
- Regular maintenance and improvement by the Vocab Team + 100+ hours on forums/GitHub/individual calls
Objective 1: **Support** vocabulary-related activities of workgroups and contributors

- Hold and document landscape assessment of **needs** across the workgroups
- Hold regular **office hours** to address any questions of vocabulary users and contributors
Objective 2: Increase Vocabularies content and process transparency

- Run sessions for all *proposed changes* and improvements, both for community contributions and roadmap
- Increase *visibility* of Vocabularies’ documentation (release, development process and content of individual vocabularies)
- Have at least two sessions with *external speakers* highlighting their experience with Vocabularies
Objective 3: Enable vocabulary-related collaboration in the community

- Hold a joint community activity:

1. OHDSI Europe Symposium [planning in progress]
2. OHDSI Global Symposium Advanced Tutorial

Vocabularies joint community activity

aostroplets Anna Ostroplets

In the past couple of Vocabulary WG meetings we discussed a joint vocabulary community activity we would like to carry this year.

This post is an opportunity to get involved for those who did not attend the meetings.

Right now, we have three options:

a) Vocabulary Tutorial (how to use the Vocabularies for ETL and studies)
b) Community Contribution Workshop (how to get the content you need into the Vocabularies)
c) Quality Assurance Study-a-thon (look into quality problems together).

If you are interested in participating in organizing and contributing to the content of such an activity, please review the proposals and sign up here. We are hoping that our experts who were the part of such activities in the past (@MPhilofsky, @DTek, @amvanandt, @clairblacketer and others) will be interested, but anybody is truly welcome.
Electronic Animal Health Records Workgroup

2024 Objectives and Key Results
Purpose

The Electronic Animal Record Workgroup exists to adopt the OMOP Common Data Model for electronic animal records to facilitate:

- Improved decision-making and care for animals and people through evidence generation
- Support comparative medicine, zoonotic and environmental disease analysis
- Analysis of human-animal bond on health and quality of life
2024 Annual Objectives

Objective #1: Facilitate Communication

• Hold monthly meetings to facilitate the exchange of ideas to use the OMOP CDM for animal records
Objective #2: Propose solutions to support use of animal health data for research in veterinary and human medicine

- Analyze OHDSI components (CDM/Vocabulary/software tools) as they relate to animal health data
- Document modifications necessary for successful ETL of animal data including methods, guidelines, experiences, and other artifacts that support successful ETL of animal data
2024 Key Results – First Quarter

• Identify current veterinary users of OMOP/CDM
• Support update of the veterinary extension to SNOMED CT in Athena
• Begin the review of current usage of the OMOP CDM at veterinary teaching hospitals especially looking at the data dictionaries at each institution
2024 Key Results – Second Quarter

• Continue the review of existing work on the data dictionary and current usage of the CDM at veterinary teaching hospitals with the goal of identifying difficulties encountered and ongoing needs

• Begin consensus development of a prototype for modifications/extension to OMOP CDM for animal records with guidance from the OHDSI CDM Workgroup
• Develop formal documentation of the recommended modifications/extension of the OMOP CDM needed to incorporate animal and veterinary data for presentation to OHDSI CDM Workgroup for review and guidance.
2024 Key Results – Fourth Quarter

• Finalize documentation for presentation to the OHDSI CDM Workgroup
• Initiate the development of a list of software potentially useful for ETL, cohort identification, and data quality assessment of animal records
• Identify any modifications needed to use software
Phenotype Development and Evaluation Workgroup

2024 OKR
• **Objective: Enhance the Science of Phenotyping and Best Practices.**
  o Publish a paper/communication on OHDSI Phenotype Library by Q2 2024
  o Conduct a network study to support methods on incorporating measurement error into background incidence rate estimates by Q4 2024

• **Objective: Community Engagement and Educational Outreach.**
  o Complete Phenotype Phebruary 2024 thru a collaborative set of activities on the theme of heterogeneity in phenotype algorithm in published literature Q1 2024.
  o Conduct 3 Atlas demo training/demonstration, open to the community Q1 2024.
  o Role in the new submission flow to the library to elicit contributions by Q3 2024.

• **Maintenance and development of tools**
  o Complete 3 refreshments of the Phenotype library with major releases by Q4 2024
  o Create a submission tool for community contributions to OHDSI Phenotype Library Q2 2024.
  o Develop objective failure criteria for cohort definitions by Q4 2024.
OHDSI Dentistry Workgroup
Objectives and Key Results (OKR)

Leads: Robert Koski and Danielle Boyce
Work Group Mission

To understand how dentistry can leverage observational research to improve oral health outcomes and further investigate the links between oral health and systemic disease.
Objectives and Key Results

1. Increase the workgroup’s involvement with the broader OHDSI community
   a. Publicize Dentistry WG efforts in the OHDSI community and dental profession
   b. Host guest speakers both from within OHDSI and from external organizations

2. Increase our understanding of observational research in dentistry
   a. Complete scoping review on observational research in dentistry and submit for publication (nearing completion)
   b. Further explore the challenges and opportunities for oral health observational research

3. Further develop the capabilities of the dental community to conduct observational research and identify opportunities and challenges for observational research in dentistry
   a. Phenotype Phebruary submission
   b. Save our Sisyphus Challenge submission
   c. Obtain a dental dataset and map to the OMOP-CDM
   d. Test use case on the acquired dataset
OHDSI Dentistry Workgroup

WG Meetings
Thursdays at 7PM ET on MS Teams
WG Name: OHDSI Medical Device WG
WG Lead: Asiyah Lin & subgroup leaders

Objective 1: Expand the leadership team and establish collaborations across OHDSI and beyond

Key results:

1. **1Q2023**: Establish **subgroups** (device generated data, device data and device adverse events) and leadership teams.
2. **1Q2023**: Respond to FDA medical device active surveillance **RFI** by Mar. 30, 2023.
3. **2-3Q2023**: Develop activities to establish **collaborations** with other related WG or efforts: Surgery WG and EHDEN
4. **3Q2023**: Plan Think-a-thon or Hackathon at the OHDSI annual **symposium**
5. **2024 ORK**: continue building partnership, including NESTcc, and cross function OHDSI WGs, hold one WG F2F meeting at the 2024 OHDSI annual symposium.

What We Achieved in 2023:

1. Michael Matheny leads device adverse events subgroup.
2. Done and more:
3. Responded two NESTcc RFP
4. Had presentations by Surgery and Perioperative Medicine WG
5. **2024 ORK**: continue building partnership, including NESTcc, and cross function OHDSI WGs, hold one WG F2F meeting at the 2024 OHDSI annual symposium.
Objective 2: Enable the device standardization efforts to be interoperable with OMOP to support large scale device data analysis

Key results:
1. 1.1-2Q 2023: Explore current OHDSI datasets for device data coverage.
2. 1.2-2Q 2023: Explore and evaluate by extending OMOP by adding a device table.
3. 2.3-3Q2023: Explore tools and method to include device data in OMOP vocabulary.

What We Achieved:
1. OHDSI device vocab and procedure vocab explore (Asiyah)
2. Carrie shared the Symmetric device attributes.
3. Seng Chan You provided UDI-EDI-SNOMED Mapping

Submit a poster to OHDSI symposium
OHDSI Medical Device WG
- Device ID Data subgroup

Subgroup Leads: Asiyah Lin & Carrie Bosela

• Year 2024 OKR: Deep dive into OHDSI vocab – device & procedure
  • Establish the medical device branch in OHDSI device vocab, establish UDI as standard concept, FDA ProCode as Classification. Add Seng Chan You’s EDI as non-standard. (decide a particular device to focus)
  • Compare the OHDSI vocab relationships with Asiyah’s medical device semantic model.
  • Map to FHIR

Collaboration with Vocab WG, and Tufts
WG Name: OHDSI Medical Device WG
Device-Generated Data subgroup

Subgroup Leads: Andrew Williams, Manlik Kwong

Objective 3: Develop standard strategy for managing and representing features waveform and other device-generated data:

1. Clarify OMOP Standard concept coverage gaps for features from 12-lead ECG Data and ICU monitor data
2. Develop strategy for addressing concept gaps
3. Test previously developed strategy for mapping covered concepts using MIMIC-4 Waveform Database waveform and "numerics" data

What We Achieved:

1. Reviewed prior efforts to cover concepts waveform acquisition and features
2. Developed strategy for addressing concept gaps
3. Began assessing gaps and developing candidate concepts to fill them
4. Began testing MIMIC IV waveform and numerics integration EHR strategy
OHDSI Medical Device WG
Device-Generated Data subgroup
Subgroup Leads: Andrew Williams & Manlik Kwong

Year 2024 OKR:

• Contribute extensions to the OMOP vocabulary to cover standard concepts for waveform and numerics acquisition and features

• Demonstrate the strategy for supporting OHDSI standardized analytics across integrated EHR and waveform/"numerics" data

Collaboration with CHoRUS and PhysioNet
WG Name: OHDSI Medical Device WG  
Device Adverse Event Event subgroup  
Subgroup Lead: Michael Matheny

Objective 4: Establish the subgroup, identify leaders, and develop OKR

Key results:
1. 1Q2023: Identify leader for this group.
2. 2Q2023: develop OKR

What We Achieved:
1. Michael Matheny joined as lead

Year 2024 OKR:
1. OHDSI members' capacity for medical device RWE research.
2. OHDSI network study led by VA/Vanderbilt
Imaging OHDSI WG

From pixels to Phenotypes

WG co-leads Paul Nagy and Seng Chan You
Imaging WG Goals

1. The ability to bring features derived from medical images into the OMOP data model while maintaining provenance.
2. Ability to perform cohort definitions in OHDSI for medical imaging research studies
3. Supporting deep learning research on medical images as part of the prediction modeling in Atlas.
4. Develop infrastructure for reproducible research on medical images.
5. Evaluate federated learning as part of a network study.
2024 OKRs

- Objectives: Identification of heterogeneity of DICOM metadata across institutions
- Key Results
  - Harvesting DICOM standards
    - Vocabulary
    - Data Quality tests
  - Evaluating real-world DICOM compared with the DICOM standards
    - Alzheimers
    - Covid
    - Lung cancer
OHDSI GIS

2024 OKRs

https://ohdsi.github.io/GIS/
Purpose

The goal of the OHDSI GIS WG is to enable studies of place-related data in conjunction with longitudinal patient-level data.

https://ohdsi.github.io/GIS/
Accomplishments

• Developed **OMOP Vocabularies for GIS, SDoH, and Environmental Toxins** concepts

• Created an **end-to-end example** use case that demonstrated integration of place-related data

• Developed and tested a **CDM extension table** for integrating environmental toxin and Social Determinants of Health (SDoH) data in OMOP

• Expanded metadata **catalog functionality** (local datasets) and corpus (ADI, Census BGs)

• Conducted a survey of offline (secure) **geocoding tools** fit for patient data

https://ohdsi.github.io/GIS/
Objective 1:

Create a cohesive and comprehensive body of documentation

Key Results:

• Complete vignette-style documentation to orient and on-ramp new users
• Improve transparency with development roadmaps and readiness-for-adoption metrics

https://ohdsi.github.io/GIS/
Objective 2:

Complete the workgroup’s transition to new organization and use case leadership structure

Key Results:

• Acquire three “guiding” use cases and work with collaborators to meet their project’s goals while advancing the OHDSI GIS mission
• Foster collaboration with other OHDSI workgroups

https://ohdsi.github.io/GIS/
Objective 3:

Continue to develop and mature Gaia functionality and extensions

Key Results:

• Secure a sustainable platform for the Gaia data catalog
• Propose and integrate a CDM extension table into the OHDSI ecosystem
• Expand metadata related functionality to enable automated retrieval from APIs
• Develop a HADES package that does appropriately adjusted analytics with spatiotemporal data

https://ohdsi.github.io/GIS/
The weekly OHDSI community call is held every Tuesday at 11 am ET.

Everybody is invited!

Links are sent out weekly and available at: ohdssi.org/community-calls