



Welcome OHDSI Newcomers!

OHDSI Community Call
July 28, 2026 • 11 am ET



Upcoming Community Calls

Date	Topic
July 28	OHDSI Newcomer Introductions
Aug. 4	Clinical Definitions
Aug. 11	Oncology Research in OHDSI
Aug. 18	Workgroup Spotlight: Eyecare and Vision Research, Rehabilitation
Aug. 25	Asia-Pacific Regional Updates
Sept. 1	Vocabulary Refresh, Summer 2026 Version
Sept. 8	Workgroup Spotlight: Medical Imaging, TBA
Sept. 15	Global Symposium Preview



Three Stages of The Journey

Where Have We Been?

Where Are We Now?

Where Are We Going?



OHDSI Shoutouts!



Congratulations to the team of **Josh Lemieux, Emily Pfaff, Richard Moffitt, Jay Nakashima, Raymond Duncan, Erica Galvez, John Giannini, Julie McMurry, Anita Walden, and Melissa Haendel** on the recent publication of **From clinics to discoveries: harnessing national EHR interoperability for research in JAMIA Open.**

JAMIA Open, 2026, 9(4), ooag137
<https://doi.org/10.1093/jamiaopen/ooag137>
Brief Communications

AMIA OXFORD
ADVANCING PROFESSIONAL, CLINICAL AND RESEARCH

From clinics to discoveries: harnessing national EHR interoperability for research

Josh Lemieux, BA¹, Emily R. Pfaff, PhD², Richard A. Moffitt, PhD³, Jay Nakashima, MBA⁴, Raymond Duncan, MD⁵, Erica Galvez, MS⁶, John P. Giannini, PhD⁷, Julie A. McMurry, MPH¹, Anita Walden, MS¹, Melissa Haendel, PhD^{*1}

¹Department of Genetics, University of North Carolina, Chapel Hill, NC 27599, United States

²Department of Medicine, University of North Carolina, Chapel Hill, NC 27599, United States

³Department of Hematology and Medical Oncology, Department of Biomedical Informatics, Emory University, Atlanta, GA 30032, United States

⁴eHealth Exchange, Denver, CO 80210, United States

⁵Enterprise Information Services, Cedars-Sinai Medical Center, Los Angeles, CA 90066, United States

⁶Manifest MedEx, Riverside, CA 92506, United States

⁷All of Us Research Program, National Institutes of Health, Bethesda, MD 20817, United States

*Corresponding author: Melissa Haendel, PhD, Department of Genetics, University of North Carolina, UNC-Chapel Hill, 120 Mason Farm Road, 5000 D, Genetic Medicine Building CB#7264, Chapel Hill, NC 27599-7264, United States (mhaendel@unc.edu)

Abstract

Objectives: To assess whether existing nationwide Health Information Networks (HINs) and Health Information Exchanges (HIEs) can be leveraged for electronic health record (EHR) data acquisition for research—specifically, for the NIH's *All of Us* Research Program.

Materials and Methods: The *All of Us* Center for Linkage and Acquisition of Data (CLAD) collaborated with eHealth Exchange, the nation's largest HIN, to transmit participant-authorized queries to an HIE and a hospital system. Returned FHIR and C-CDA records were mapped to the OMOP Common Data Model and compared with existing *All of Us* EHR data.

Results: Newly obtained records provided complementary information, enhancing completeness and research utility.

Discussion: This pilot demonstrates the feasibility of using HINs/HIEs for patient-authorized data exchange for research. Barriers remain—including inconsistent technical capacity to transact authorizations and data quality variations.

Conclusion: HINs and HIEs can support secure, patient-authorized EHR exchange for research, offering a pathway to more comprehensive, participant-driven biomedical discovery.

Key words: real-world data; health information exchange; EHR data; interoperability

Lay Summary

This pilot study explores how to best fill gaps in people's health records for use in research. Patients are often seen at many different clinics and hospitals, making it hard for researchers to see a complete picture of someone's health. This is a challenge for large programs like the National Institutes of Health's *All of Us* Research Program, which aims to collect comprehensive health data from over one million volunteer participants.

We evaluated whether nationwide health data-sharing networks used for clinical care data exchange could be adapted for research. We sent secure requests for medical records through one of the country's largest networks. The returned data were then compared with existing records supplied originally by the study's recruitment sites.

Our results show not only that the approach works, but also that we can acquire important missing information, including for participants who previously had no medical records in the system.

However, challenges remain, such as differences in data quality and technical limitations in sharing information. Overall, the study demonstrates that using existing health data networks, with patient consent, could significantly improve the quality and completeness of data used in medical research.



OHDSI Shoutouts!



Congratulations to the team of Irene López-Sánchez, Anna Palomar-Cros, Agustina Giuliodori, Laura Granés, Laura Pérez-Crespo, Berta Raventós, Edward Burn, Anton Barchuk, Annelies Verbiest, Caroline Eteve-Pitsaer, Danielle Newby, Elin Rowlands, Espen Enerly, Gargi Jadhav, Harlinde De Schutter, Jason Hsu, Jelle Evers, Josip Vrancic, Laia Peruchet-Noray, Lillian Sung, Loretta Kiss, Maaïke Van Swieten, Mario Šekerija, Mees Mosseveld, Patricia L Mabry, Raivo Kolde, Samuel Patnoe, Seng Chan You, Subin Kim, Sulev Reisberg, Thi Kim Hien Nguyen, Toni Lehtonen, Tor Åge Myklebust, Zsolt Bagyura, Asieh Golozar, Elena Roel, and Talita Duarte-Salles on the recent publication of **Temporal trends in epidemiology and patient characteristics of 36 cancers: a protocol for a multinational population-based cohort study using OMOP-standardised databases to investigate CANcer (OMOPCAN)** in *BMJ Open*.

Open access

Protocol

BMJ Open Temporal trends in epidemiology and patient characteristics of 36 cancers: a protocol for a multinational population-based cohort study using OMOP-standardised databases to investigate CANcer (OMOPCAN)

Irene López-Sánchez ^{1,2}, Anna Palomar-Cros,¹ Agustina Giuliodori,¹ Laura Granés,¹ Laura Pérez-Crespo,¹ Berta Raventós,³ Edward Burn ⁴, Anton Barchuk,³ Annelies Verbiest,^{5,6} Caroline Eteve-Pitsaer,⁷ Danielle Newby,⁴ Elin J Rowlands,⁴ Espen Enerly,⁸ Gargi Jadhav,⁹ Harlinde De Schutter,¹⁰ Jason C Hsu ^{11,12}, Jelle Evers,¹³ Josip Vrancic ¹⁴, Laia Peruchet-Noray,¹⁵ Lillian Sung ¹⁶, Loretta Kiss,¹⁷ Maaïke Van Swieten,¹³ Mario Šekerija,^{18,19} Mees Mosseveld,³ Patricia L Mabry,²⁰ Raivo Kolde,²¹ Samuel Patnoe ²⁰, Seng Chan You,^{22,23} Subin Kim,^{22,23} Sulev Reisberg,^{21,24} Thi Kim Hien Nguyen,²⁵ Toni Lehtonen,²⁶ Tor Åge Myklebust ²⁷, Zsolt Bagyura,¹⁷ Asieh Golozar,²⁸ Elena Roel,¹ Talita Duarte-Salles ^{1,3}

To cite: López-Sánchez I, Palomar-Cros A, Giuliodori A, et al. Temporal trends in epidemiology and patient characteristics of 36 cancers: a protocol for a multinational population-based cohort study using OMOP-standardised databases to investigate CANcer (OMOPCAN). *BMJ Open* 2026;16:e119069. doi:10.1136/bmjopen-2026-119069

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (<https://doi.org/10.1136/bmjopen-2026-119069>).

IL-S and AP-C are joint first authors.

Received 03 March 2026

ABSTRACT

Introduction Cancer registries remain the gold standard for global cancer monitoring, yet complementing them with electronic health records and claims can significantly enhance the understanding of the cancer burden by providing a more complete picture of the patient journey. The main aim of this project is to serve as a proof of concept for using real-world data mapped to the Observational Medical Outcomes Partnership (OMOP) common data model (CDM) to monitor cancer epidemiology over time and characterise patients' clinical history and outcomes.

Methods and analysis This study will be conducted as an observational cohort study using a multinational network of large real-world data sources mapped to the OMOP CDM. Electronic health records (EHR) from primary and secondary care, health insurance claims and cancer registry data will be included. To date, 20 databases from 16 countries, mainly from Europe but also North America

STRENGTHS AND LIMITATIONS OF THIS STUDY

- ⇒ Large sample from broadly representative and heterogeneous settings across multiple geographical regions.
- ⇒ Longitudinal design including data extending to the most recent years available and covering 36 solid and haematologic cancers.
- ⇒ Use of data mapped to Observational Medical Outcomes Partnership common data model facilitates the harmonisation of diverse data sources, leveraging their complementary strengths and supporting federated and standardised analysis without sharing individual-level data, ensuring patient privacy.
- ⇒ As electronic health record and medical insurance claims data are collected primarily for clinical or administrative purposes rather than research, there will be a potential risk of outcome misclassification



OHDSI Shoutouts!



Congratulations to the team of **Kevin Haynes, Jenna Reys, Jill Hardin, Eva-Maria Didden, James Gilbert, Joel Swerdel, Justin Bohn, Mitchell Conover, Amir Sarayani, Anthony Sena, Emily Yost, Valerie van Baalen, Kourtney Davis, Patrick Ryan, and Martijn Schuemie** on the recent publication of **Active Safety Surveillance Using Real-World Evidence (ASSURE) Implementation: Transparent and Reproducible Real-World Evidence Standardized Framework to Support Safety Signal Evaluation in Pharmacoepidemiology & Drug Safety.**

Pharmacoepidemiology and Drug Safety

WILEY

ORIGINAL ARTICLE OPEN ACCESS

Active Safety Surveillance Using Real-World Evidence (ASSURE) Implementation: Transparent and Reproducible Real-World Evidence Standardized Framework to Support Safety Signal Evaluation

Kevin Haynes¹ | Jenna Reys² | Jill Hardin² | Eva-Maria Didden¹ | James Gilbert² | Joel Swerdel² | Justin Bohn¹ | Mitchell M. Conover² | Amir Sarayani¹ | Anthony G. Sena² | Emily Yost¹ | Valerie van Baalen¹ | Kourtney Davis¹ | Patrick Ryan² | Martijn Schuemie²

¹Johnson & Johnson, Horsham, Pennsylvania, USA | ²Johnson & Johnson, Raritan, New Jersey, USA | ³Johnson & Johnson, Allschwil, Switzerland

Correspondence: Kevin Haynes (khaynes2@its.jnj.com)

Received: 30 October 2025 | Revised: 11 June 2026 | Accepted: 7 July 2026

Keywords: pharmacovigilance | real world evidence | safety surveillance

ABSTRACT

Purpose: To summarize experience with Johnson & Johnson's novel Active Safety Surveillance Using Real-world Evidence (ASSURE) program for producing efficient, transparent, applicable, and impactful RWE to support routine pharmacovigilance processes.

Methods: ASSURE follows a stepwise framework that encompasses database diagnostics to assess fit-for-purpose real-world data (RWD) sources in the US, France, Germany, Australia, and Japan, phenotype development, query specification, and analytic implementation, objective study validity diagnostics, and standardized reporting. Licensed RWD sources are transformed into the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM), from which we generate evidence for safety signal evaluation questions using a suite of open-source analytic tools made publicly available through the OHDSI community. For each signal evaluation request involving a medication exposure and outcome(s), we conducted population characterization and population-level causal effects estimation. Target medication, comparator medication, and indication cohorts utilize predefined phenotypes, when available, for rapid analytics. We provide an evaluation of the initial implementation of this framework, including results from database diagnostics, phenotype development, and objective study validity diagnostics applied prior to unblinding evidence to prevent exposing biased estimates. We defined efficiency as the time from query to results and integration of results into safety decisions.

Results: The initial 19 months of the ASSURE program supported 110 safety signal evaluations across 24 unique products that arose from 47 requests. 74 evaluations (67%) passed study diagnostics and yielded effect estimation results with 62 (54%) providing propensity-score adjusted comparative cohort method results and 40 (36%) providing self-controlled case series results. Many analyses failed one or more diagnostics, preventing the generation and unblinding of potentially biased results: database (7 evaluations), phenotype specification (3 evaluations), or the requisite objective study validity diagnostics (26 evaluations). These 26 evaluations (24%) all provided population characterizations of exposure, indication, and outcome, including incidence rates.

Conclusions: The ASSURE framework and implementation process, in collaboration with safety management teams, enables rapid response to medical product safety signals under evaluation and produces actionable RWE that supports pharmacovigilance decision making within regulatory timeframes.



OHDSI Shoutouts!



Congratulations to the team of **Obinwa Ozonze, Ogechukwu Okonor, Uzundu Dike, and Taiwo Adedeji** on the recent publication of **Enabling interoperability across disparate health data sources using Large Language Models** in *PLOS Digital Health*.



RESEARCH ARTICLE

Enabling interoperability across disparate health data sources using Large Language Models

Obinwa Ozonze¹, Ogechukwu Okonor², Uzundu Dike³, Taiwo Adedeji^{1*}

¹ School of Computing, University of Portsmouth, Portsmouth, Hampshire, United Kingdom, ² Department of Computing & Emerging Technologies, Ravensbourne University London, London, United Kingdom, ³ Insight and Analytics, Emcor Group, Eastleigh, Hampshire, United Kingdom

* taiwo.adedeji@port.ac.uk



OPEN ACCESS

Citation: Ozonze O, Okonor O, Dike U, Adedeji T (2026) Enabling interoperability across disparate health data sources using Large Language Models. *PLOS Digit Health* 5(7): e0001511. <https://doi.org/10.1371/journal.pdig.0001511>

Editor: Hadi Amiri, UMass Lowell: University of Massachusetts Lowell, UNITED STATES OF AMERICA

Received: July 1, 2025

Accepted: June 5, 2026

Published: July 23, 2026

Copyright: © 2026 Ozonze et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data availability statement: All research materials are publicly available in a GitHub repository (https://github.com/taiwoaddejil/Automate_Health_Data_Integration) and are accompanied by a Zenodo DOI (<https://doi.org/10.5281/zenodo.18407899>) upon

Abstract

The heterogeneity of health data systems remains a major barrier to interoperability and data integration, and a challenge for clinical decision support and AI-driven research. The coexistence of multiple Common Data Models (CDMs) complicates integration, requiring complex manual or rule-based schema mapping, which is labor-intensive, error-prone, and difficult to scale. We evaluated three leading Large Language Models (LLMs): ChatGPT, Gemini, and DeepSeek, for their ability to match data elements from four widely used CDMs (PCORnet, OMOP, Sentinel, and i2b2) to FHIR R4 via web-based chat interfaces. Reference mappings from the HL7 Common Data Models Harmonization Implementation Guide were used as the benchmark. Standardized prompts were submitted across 36 total runs (3 LLMs × 4 schemas × 3 repetitions), and accuracy was measured using F1-scores with 95% confidence intervals, alongside strict (AND) and relaxed (OR) match criteria. DeepSeek obtained the highest numerical mean F1-scores (0.78-0.99) with strong consistency (AND Match=0.87). Gemini showed good coverage (OR Match=0.96) but greater variability (AND Match=0.64), while ChatGPT achieved moderate accuracy with lower consistency (AND Match=0.51). A two-way ANOVA indicated no statistically significant differences by LLM type ($p=0.138$), CDM schema ($p=0.194$), or their interaction ($p=0.451$). Performance varied by domain, with Demographics and Condition mapped most accurately, while Observation showed the highest error rates (44% of total errors). Out of 2,583 total matchings, 314 errors were observed and categorized into four failure patterns: semantic confusion (42%), structural misalignment (28%), ambiguous terminology (18%), and terminology variance (12%). This study demonstrates the feasibility of using LLMs to automate health data integration as a complementary approach to traditional ontology-based methods, particularly for well-structured, context-rich CDMs. However, safe deployment requires human-in-the-loop validation, governance, prompt optimization, metadata integration, and validation to ensure clinical and research reliability.



OHDSI Shoutouts!



Congratulations to the team of **Dong Han Yu, Sun Mi Lim, Hyun Chul Shin, Moosung Kim, Seng Chan You, Rae Woong Park, and Chungsoo Kim** on the recent publication of **Data Resource Profile: Health Insurance Review and Assessment Service Korean nationwide claims OMOP CDM (2015–24) database, HIRA K-OMOP** in the *International Journal of Epidemiology*.

International Journal of Epidemiology, 2026, 55(4), dyag131
<https://doi.org/10.1093/ije/dyag131>
Data Resource Profile



Data Resource Profile: Health Insurance Review and Assessment Service Korean nationwide claims OMOP CDM (2015–24) database, HIRA K-OMOP

Dong Han Yu^{1,✉}, Sun Mi Lim¹, Hyun Chul Shin¹, Moosung Kim¹, Seng Chan You², Rae Woong Park^{3,✉}, Chungsoo Kim^{3,4,★}

¹Big Data Department, Health Insurance Review and Assessment Service, Wonju, 26465, Republic of Korea

²Department of Biomedical Systems Informatics, Yonsei University College of Medicine, Seoul, 03722, Republic of Korea

³Department of Biomedical Informatics, Ajou University School of Medicine, Suwon, 16499, Republic of Korea

⁴Center for Outcomes Research and Evaluation, Yale New Haven Hospital, New Haven, Connecticut, 06510, United States

★Corresponding author: Department of Biomedical Informatics, Ajou University School of Medicine, 164 World cup-ro, Hongjae Bld, 5th fl, Suwon 16499, Republic of Korea. E-mail: ted9219@ajou.ac.kr

Keywords claims, OMOP, common data model, data standardization, observational study

Key Features

- HIRA K-OMOP is a nationwide Observational Medical Outcome Partnership Common Data Model resource derived from South Korea's National Health Insurance claims.
- It covers the entire insured population from 2015 to 2024 (56 416 773 persons), enabling population-scale health-services and pharmacoepidemiology research.
- The database captures diagnoses, procedures, costs, and medication data, with prescribing and pharmacy dispensing recorded as separate events to improve drug-exposure validity.
- Access is provided via a privacy-preserving secure research environment under Health Insurance Review and Assessment Service (HIRA) governance and requires approval from the Institutional Review Board and authorization from HIRA.



OHDSI Shoutouts!



Congratulations to the team of **Haeun Lee, Benjamin Martin, Andreea Creanga, Evan Minty, Asieh Golozar, Cindy Cai, Cynthia Sung, Sean O'Reilly, Khyzer Aziz, and Paul Nagy** on the recent publication of **Assessment of real-world evidence research competencies in federated research networks: a maternal health fellowship program evaluation** in *JAMIA Open*.

JAMIA Open, 2026, 9(4), ooag128
<https://doi.org/10.1093/jamiaopen/ooag128>
Research and Applications

AMIA OXFORD
ADVANCING MEDICAL RESEARCH LEADING THE WAY

Assessment of real-world evidence research competencies in federated research networks: a maternal health fellowship program evaluation

Haeun Lee, MS^{*1}, Benjamin Martin, PhD¹, Andreea A. Creanga, MD, PhD², Evan Minty, MD, MSc³, Asieh Golozar, MD, PhD^{1,4}, Cindy X. Cai, MD^{1,5}, Cynthia Sung, PhD⁶, Sean O'Riley, BS¹, Khyzer Aziz, MD^{1,7}, Paul Nagy, PhD¹

¹Biomedical Informatics and Data Science, Johns Hopkins School of Medicine, Johns Hopkins University, Baltimore, MD, 21205, United States

²Department of Gynecology and Obstetrics, The Johns Hopkins University School of Medicine, Baltimore, MD, 21287, United States

³Faculty of Medicine, O'Brien Institute for Public Health, University of Calgary, Calgary, Alberta, 2500, Canada

⁴Nemesis Health, New York, NY, 10009, United States

⁵Wilmer Eye Institute, Johns Hopkins School of Medicine, Johns Hopkins University, Baltimore, MD, United States

⁶Center of Regulatory Excellence, Duke-NUS Medical School, Singapore

⁷Department of Neonatology, The Johns Hopkins University School of Medicine, Baltimore, MD, United States

*Corresponding author: Haeun Lee, MS, Biomedical Informatics and Data Science, Johns Hopkins School of Medicine, Johns Hopkins University, 2024 E. Monument Street, Baltimore, MD, 21205, United States (hlee292@jh.edu)

Abstract

Background and Significance: Generating real-world evidence (RWE) approaches across federated evidence networks based on real-world data requires learning new methodological skills. However, standardized frameworks for defining and assessing competencies for large-scale RWE research remain limited. This study aims to develop an assessment framework for RWE studies and evaluate the competencies of RWE research professionals.

Methods: We adapted the Joint Task Force (JTF) Core Competency Framework for RWE research using the Observational Medical Outcomes Partnership (OMOP) Common Data Model. Through iterative expert review, 4 new domains were created, including Ethics and Governance, Protocol Development, Study Operations, and Data Analysis and Informatics, and 4 original JTF domains were adapted. Participants completed pre- and post-training self-assessments on a 10-point scale. Paired t-tests with effect sizes were used to evaluate changes in competency scores.

Results: Seventeen participants from academic, data, and clinical roles completed the assessment. Mean self-reported competency scores increased by 2.5 points across the 8 competency domains, with statistically significant changes observed in all domains. The largest gain was observed in Leadership and Professionalism (mean increase 3.66; 95% CI, 2.69-4.62).

Discussion: The RWE research competency assessment tool captured changes in self-reported skill proficiency, perceived readiness to conduct RWE studies, and methodological knowledge gaps. These findings suggest the potential to identify competency gaps, tailor educational interventions, and inform quality benchmarks for RWE research training programs.



OHDSI Shoutouts!



MRCT CENTER HOMEPAGE | MRCT CENTER CLINICAL TRIAL COMPETENCY PROJECT | CONTACT

JOINT TASK FORCE FOR CLINICAL TRIAL COMPETENCY

Competency Domains for the Clinical Research Professional

- 1 Scientific Concepts and Research Design
- 2 Ethical and Participant Safety Considerations
- 3 Investigational Products Development and Regulation
- 4 Clinical Study Operations (Good Clinical Practice)
- 5 Study and Site Management
- 6 Data Management and Informatics
- 7 Leadership and Professionalism
- 8 Communications and Teamwork

MRCT CENTER CLINICAL TRIALS
THE MRCT CENTER OF BRIGHAM AND WOMEN'S HOSPITAL AND HARVARD

Assessment of real-world evidence research competencies in federated research networks: a maternal health fellowship program evaluation

Harun Lee, MS, Benjamin Martin, PhD, Andros A. Coarasa, MD, PhD, Evan Minty, MD, MS, Ashish Goelkar, MD, PhD, Cindy X Cai, MD, Cynthia Sung, PhD, Sean O'Sullivan, BS, Khayr Aziz, MD, Paul Hagg, PhD

JAMA Open, Volume 9, Issue 4, August 2026, oae128, <https://doi.org/10.1093/jamaopen/oae128>

Published: 14 July 2026 Article history

RWE Research Competency Domains

- 1 Scientific Concepts & Research Design
- 2 RWE Research Ethics & Governance
- 3 Research Protocol Development
- 4 RWE Study Operations
- 5 Study & Site Management
- 6 Data Analysis & Informatics
- 7 Leadership & Professionalism
- 8 Communication & Teamwork

<https://acrpnet.org/competency-domains-clinical-research-professionals>

OHDSI Education WG

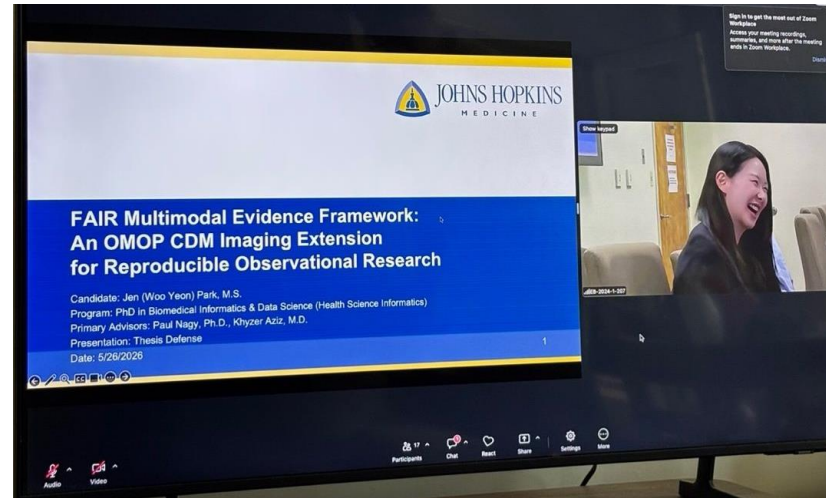


OHDSI Shoutouts!



Congratulations to **Jen Wooyean Park** on completing her Ph.D. recently at Johns Hopkins University.

Her dissertation was entitled *FAIR Multimodal Evidence Framework: An OMOP Common Data Model Imaging Extension for Reproducible Observational Research*.





Three Stages of The Journey

Where Have We Been?

Where Are We Now?

Where Are We Going?



Upcoming Workgroup Calls



Date	Time (ET)	Meeting
Tuesday	12 pm	CDM Vocabulary
Wednesday	10 am	Women of OHDSI
Thursday	6:30 am	India Community Call
Thursday	9 am	Phenotype Development and Evaluation
Thursday	10 am	GIS-Geographic Information System
Thursday	10 am	Africa Chapter
Thursday	11 am	Perinatal & Reproductive Health
Friday	11:30 am	Steering
Monday	10 am	Health Systems Interest Group
Tuesday	10 am	CDM Survey



First Latin America Symposium – July 30-31

Registration is open
for the first OHDSI
Latin America
Symposium, taking
place July 30-31 in
Salvador, Brazil.

Day 1

**Strategic panels with government,
academia and industry**

Thursday, July 30, 2026



Opening and keynote

Common Data Model for Health Equity: the Role of Latin America.



Panel 1— Health data interoperability and standards

Panelists from the Ministry of Health, Bahia State Health Department, PAHO and Latin American Governments.



Panel 2 — The power of administrative data for health research

Panelists from the Ministry of Health, CONASS, Fiocruz, Latin American Governments, Industry and OHDSI Global.



Panel 3 — The future of interoperability in healthcare in Latin America

A public-private debate.

Panelists from the Ministry of Health, CONASS, Fiocruz, private hospitals and Latin American Governments.

Day 2

**Hands-on workshops and scientific
collaboration**

Friday, July 31, 2026



Introductory OMOP CDM workshops

- Introduction to OMOP
- Building cohorts with OHDSI tools



Parallel tracks of specialized workshops

- ETL to OMOP
- Scientific collaboration



Closing

Future perspectives and next steps for the OHDSI Latin America community.

ohdsilatam.org



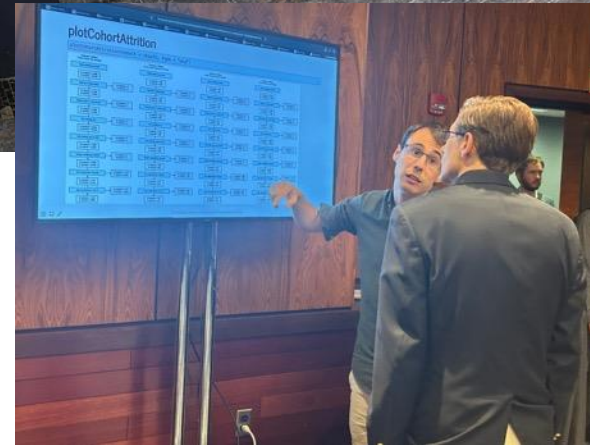
2026 OHDSI Global Symposium

Registration is OPEN for the **2026 OHDSI Global Symposium**, which will be held Oct. 20-22 in New Brunswick, N.J., USA.

Oct. 20: Tutorials

Oct. 21: Plenaries, Showcase

Oct. 22: Workgroup Activities



ohdsi.org/OHDSI2026



2026 Symposium Tutorials – Session 1

- **An Introduction to the Journey from Data to Evidence Using OHDSI**
- **An Introduction to ATLAS**
- **Bringing FAIR to Imaging Research with the Medical Imaging OMOP Extension**
- **Complex Phenotyping at Scale with and without LLMs Using PhenotypeR**
- **OHDSI Leadership Storytelling Workshop**
- **Mastering OMOP: Transforming EHR Data with Practical Strategies, Best Practices, and OHDSI Integration**



2026 Symposium Tutorials – Session 2

- **Building and Using the OHDSI Evidence Network: From Data Partner to Federated Study Execution**
- **From Multi-Modal Data to Real-World Evidence: Hands-on with the Data2Evidence Platform for OMOP Data Curation and Analytics**
- **Integrating Geospatial Data Into OMOP CDM**
- **Introduction to OHDSI Phenotype Development & Evaluation**
- **OHDSI Standardized Vocabularies on FHIR: A Deep Dive Using the Echidna Terminology Server**
- **Using OMOP Model in Registry Context & Clinical Trials Standardization Context: Conventions, Past Use Cases, SDTM & Regulatory Consideration, Challenges**



2026 Global Symposium Agenda

Start	End	Topic	Presenter/Lead
8:00 am	8:30 am	State of the Community	George Hripcsak
8:30 am	9:15 am	OHDSI Year In Review	Early-Stage Researcher WG
9:15 am	10:00 am	Collaborator Showcase: Posters and Demos (Session 1)	
10:00 am	11:00 am	Plenary 1: Federated Learning Meets Negative Control Calibration: Toward Reliable Multi-Site Evidence Generation	Yong Chen
11:00 am	12:00 pm	Plenary 2: Beyond the Defaults: How the OHDSI Community is Adapting, Extending, and Reimagining Its Tools	Scott Duvall
12:00 pm	1:00 pm	Network & Lunch	
1:00 pm	2:00 pm	Plenary 3: The role of national initiatives in supporting sustainability, collaboration, and growth of OHDSI	Ed Burn
2:00 pm	2:45 pm	Collaborator Showcase: Lightning Talks (Session 1)	5 presenters
2:45 pm	3:30 pm	Collaborator Showcase: Posters and Demos (Session 2)	
3:30 pm	4:15 pm	Collaborator Showcase: Posters and Demos (Session 3)	
4:15 pm	5:00 pm	Collaborator Showcase: Lightning Talks (Session 2)	5 presenters
5:00 pm	6:00 pm	Titan Awards, Closing	Patrick Ryan
6:00 pm	8:30 pm	Dinner on your own	
8:30 pm	11:30 pm	OHDSI Jam Session	Martijn Schuemie



2026 Symposium Workgroup Activities

Session 1 (8 am – 10 am): Eyecare and Vision Research, GIS – Geographic Information Systems, Early-Stage Researchers, Industry, Tidy R Programming with OMOP, HADES Hackathon, Generative AI and Foundation Models, Phenotype Development and Evaluation, Oncology, Health Equity, Vocabularies, APAC

Session 2 (10:30 am – 12:30 pm): Perinatal and Reproductive Health, Waveform, Medical Imaging, Industry, Tidy R Programming with OMOP, HADES Hackathon, Generative AI and Foundation Models, Phenotype Development and Evaluation, Oncology, Health Equity, Vocabularies, Rare Disease

Session 3 (1:30 pm – 3:30 pm): Dentistry, GIS & Waveform Cross-Pollination Meeting, Evidence Network, Women of OHDSI, Psychiatry, HADES Hackathon, Natural Language Processing, Health Economics & Value Assessment, ATLAS/WebAPI, CDM Survey, Surgery & Perioperative Medicine

Session 4 (3:30 pm – 5 pm): Workgroup Summary Session



2026 OHDSI Symposium Jam Session

Whether you are an amateur hobbyist, a seasoned pro, an instrumentalist, or a vocalist, our very first OHDSI Jam Session is open to everyone. We'll do some free-form improvisation, but we will also vote on a few songs ahead of time so everyone has a chance to practice.

The Gear: The organizers will provide a backline with amplifiers, drums, and microphones.

Your Instruments: We encourage you to bring your own gear! If you're open to letting fellow community members plug in or play your instrument, please let us know.

The Vibe: Low pressure, high energy, and entirely collaborative.
Want to join?

Let us know if you're thinking about participating so we can coordinate the song list and gear!





APAC Symposium: Nov. 13-15

The 2026 APAC Symposium will be held Nov. 13-15 at Yonsei University in Seoul, South Korea.

Nov. 13: Main Conference

Nov. 14: Tutorials, Datathon

Nov. 15: Datathon Team Activities

Registration will open soon, and the showcase submission deadline is August 17 at 8 pm ET.

ohdsi.org/APAC2026





Atlas 3 Demo Site

- Presented by Peter Hoffmann at 7/7 OHDSI Community Call
- We are asking for community feedback
- Published to <https://atlas-preview.ohdsi.org/atlas/>
 - Db Authentication enabled: use ohdsi/ohdsi
 - Will enable anonymous access in a future update
- Discussions at <https://github.com/OHDSI/Atlas3/discussions>



Wanted: HADES Developers

Help wanted - developers for HADES shiny viewer

■ Developers



jreps

5d

We have a set of R packages for developing modular shiny apps to explore the standardized results returned via HADES packages.

- OhdsiShinyAppBuilder - this builds the shiny app and defines what modules to include (e.g., cohort explorer, characterization, estimation, ...)
- OhdsiShinyModules - the main shiny code for the key HADES packages are here
- OhdsiReportGenerator - this is a set of SQL that pulls out friendly data.frames with the results

However, there is a lot of potential for improvements in these packages including (but not limited to):

- adding bookmarking to the URL for easy sharing ([Chapter 11 Bookmarking | Mastering Shiny](#))
- improving display time (improving data extraction speed)
- improving the UI

If you are a developer with knowledge or interest in R or shiny or web development/design and UI and want to help out with HADES development please post below or message me.

I'm planning to set up regular meetings for this work group.



#OHDSISocialShowcase This Week

Monday

The REALM Data Catalogue and Federation Model for Cohort-Driven Evaluation of Medical Device Software

(Frederic Jung, Stef Rommes, Gökhan Ertaylan)

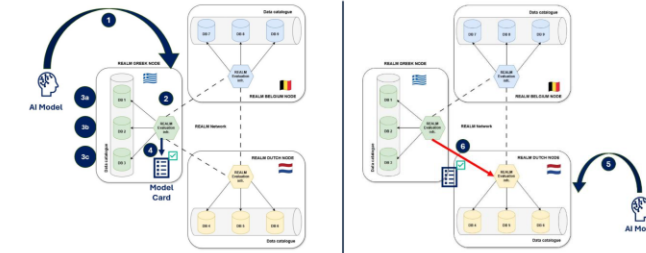
From Federated National Data Nodes to trustworthy medical AI evaluation

Title: *The REALM Data Catalogue and Federation Model for Cohort-Driven Evaluation of Medical Device Software*



Background: The evaluation of Medical Device Software (MDSW) and in particular those using AI increasingly require access to diverse, Real-World Data (RWD) across multiple data providers. The OHDSI community has established mature federation networks such as DARWIN EU, EHDS, and FHIN based on distributed analytics and local execution with proven track record of efficiency and robustness for observational research and real-world evidence generation. However, the evaluation of AI-based MDSW introduces additional requirements, including repeated execution of computationally intensive models, scenario-based testing, and detailed performance characterization, which motivate complementary evaluation infrastructures tailored to these workflows and regulatory needs. The REALM project addresses this gap by designing an evaluation-oriented data infrastructure tailored to evaluate AI from MDSW that enables controlled access to harmonized clinical data for the purpose of verifying manufacturer claims, benchmarking model performance, and supporting regulatory decision-making in Europe.

Figure 1: REALM national node & evaluation workflow



Belgium Node	Greek Node	Dutch Node
 Gökhan Ertaylan Frederic Jung	 Yiannis Subitidis Emmanuel D. Savellidis	 Michiel Overduin Gökhan Ertaylan
	 	

Data protection and governance are central to the REALM federation model. All patient-level data remain secure within the national nodes. Patient-level data are never transferred outside the evaluation environment, and AI models submitted for evaluation are always pre-trained and executed locally without any retraining, ensuring that data does not leave the node at any stage. Only high-level evaluation results such as performance metrics, provenance information and high-level data distributions are shared externally through structured model cards. This approach aligns conceptually with OHDSI's federated principles while enabling workflows that necessitate continuous access to multiple data sources.

- 1 Model developers import AI medical software in a REALM node
- 2 Model developers submit their claims
- 3 A cohort definition is created and submitted against the data sources available (a, b, c)
- 4 Model and evaluation results are described in a Model card
- 5 Model is submitted to another node, steps 1,2,3,4 are repeated
- 6 New results are written in the model card



Frédéric Jung¹, Stef Rommes¹, Gökhan Ertaylan¹
1. VITO, Vlaamse Instelling voor Technologisch Onderzoek, Mol, Belgium





#OHDSISocialShowcase This Week

Tuesday

Comparing the representation of medicinal products in RxNorm Extension and SNOMED CT

(**Maryia Khitrun**, Anna Ostropolets)

RxNorm Extension offers broader global drug coverage than SNOMED CT, with **98% overlap** at the clinical drug level

Comparing the representation of medicinal products in RxNorm Extension and SNOMED CT

Background

Real-world pharmacovigilance studies benefit from large, diverse, multinational datasets. Network studies require harmonizing drug exposure data across heterogeneous sources and markets. RxNorm Extension (RxÉ) was developed to extend the RxNorm framework to global markets. SNOMED CT has also developed a parallel International Medicinal Product Model. This study systematically compares their structural philosophies, coverage, and overlap.

Methods

RxNorm Extension (RxÉ)

Open-world assumption

Ingredient-based hierarchy → Clinical Drug Form → Clinical Drug → Branded Drug → Marketed Product

12 source vocabularies across 10 countries:

AMIS (Germany), AMT (Australia), BOPM (France), dm+d, Genscript (UK), DPD (Canada), EDI, KDC (Korea), GGR (Belgium), GRR (Brazil), JMDC (Japan), Z-index (Netherlands)

Attributes: dose form, drug strength, brand name, box size, supplier/manufacture

SNOMED CT Drug Model

Closed-world assumption

Drugs with only ingredient of interest occupy separate hierarchy branches from those with other ingredients

Data from 5 national drug extensions: UK, US, Australia, Netherlands, New Zealand

Attributes: ingredient, dose form, drug strength (shared with RxNorm)

Results

79,646

Clinical Drugs in RxÉ
(+ 35,756 in RxNorm)

8,348

Clinical Drugs in SNOMED CT
from 3,914 active ingredients

98%

SNOMED monodrug forms
present in RxNorm or RxÉ

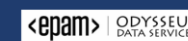
Feature	RxNorm Extension	SNOMED CT Intl.
World assumption	Open-world	Closed-world
Clinical drugs	79,646 (RxÉ) + 35,756 (RxNorm)	8,348
Ingredients	6,098 (1,871 non-US)	3,914 active
Market coverage	10 countries, 12 vocabularies	5 national drug extensions
Marketed product	Yes (supplier, box size)	In local extensions

Conclusion

Both RxÉ and SNOMED CT offer comprehensive, internationally-oriented drug vocabularies — but differ in structural philosophy and scale. With 98% overlap at the clinical drug level, the models show strong complementarity. RxÉ provides broader coverage for global observational research, while SNOMED CT's closed-world framework offers distinct classificational precision. Future work will explore incorporating SNOMED drugs into RxÉ and comparing SNOMED classes to ATC for therapeutic grouping.

Maryia Khitrun¹ • Anna Ostropolets^{2,3}

¹Odyssey, an EPAM Company, ²J&J Observational Health Data Analytics, ³Columbia University Medical Center





#OHDSISocialShowcase This Week

Wednesday

Explainable, Governance-Ready Clinical NLP Pipeline for OMOP-CDM Using Locally Deployed LLMs

(**Alisson Licon**, Muhammad Waseem, Alex Loleit)

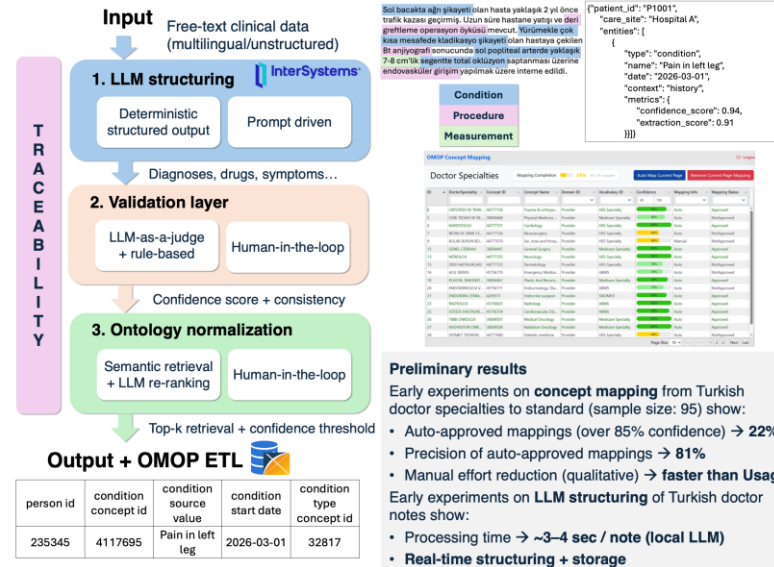
Explainable, Governance-Ready Clinical NLP Pipeline for OMOP-CDM Using Locally Deployed LLMs

Alisson Licon¹, Muhammad Waseem¹, Alex Loleit¹
¹Traverse Health

Clinical free-text notes contain **rich patient information** but remain difficult to **reuse for large-scale research, limiting integration** into standardized frameworks such as OMOP-CDM. While NLP methods have improved clinical information extraction, traditional rule-based systems lack flexibility, and modern machine learning approaches often face **challenges in interpretability, scalability, and cross-lingual adaptation**. Mapping local clinical concepts —especially from non-English data— to standardized vocabularies remains a critical bottleneck.

We propose an **explainable, end-to-end clinical NLP pipeline** that transforms free-text into standardized, research-ready data aligned with OMOP-CDM.

A preliminary implementation has been developed, focusing on the ontology normalization stage. In parallel, an LLM-based structuring module has been implemented to convert free-text into structured representations, currently undergoing validation.



Conclusion

We present an **explainable, end-to-end framework** for structuring clinical free text into OMOP-aligned data. The pipeline demonstrates the potential to enhance standardization, **transparency**, and scalable research. Local LLMs ensure **data governance and accessible deployment**.

Future work will focus on **full pipeline integration, large-scale validation, and benchmarking** against existing OMOP mapping approaches.





#OHDSISocialShowcase This Week

Thursday

OMOP SQL Generator Website Based on Generative AI

(Angela Leis, Miguel-Angel Mayer, Juan Manuel Ramírez-Anguita, Carla Arxé)

- ✱ This tool represents a step toward **democratising access to structured clinical data** by removing technical barriers using a Generative AI-based SQL tool
- ✱ Adding this tool to national and collaborative projects would greatly **help clinical researchers and data scientists work with the OMOP CDM** by improving their ability to query these databases

OMOP SQL Generator Website Based on Generative AI

Background:

Querying OMOP databases remains a practical bottleneck: many clinically meaningful questions require non-trivial SQL (multiple joins, temporal logic, inclusion/exclusion criteria, and concept-set logic), plus familiarity with OMOP conventions. In parallel, platforms such as OpenAI GPTs enable the creation of custom conversational agents that combine instructions, structured guidance, and domain knowledge, which is particularly relevant for specialised schemas like OMOP CDM. However, in clinical research, using free-form text-to-SQL can lead to mistakes. To address this gap, we developed a publicly accessible custom GPT that supports users in generating PostgreSQL queries over OMOP CDM using natural language plus a guided, menu-based interaction pattern. This approach aims to democratise access to OMOP data for feasibility analyses and cohort extraction while ensuring that outputs remain aligned with OMOP structure and conventions in federated and multi-institutional research contexts, such as the European Health Data Space platforms.

Results

Figure 1 shows the custom ChatGPT created. A dedicated web interface for the "SQL generator OMOP-CDM" custom GPT was deployed, providing one-click example prompts and a menu-driven workflow to help users generate OMOP-CDM PostgreSQL queries. Figure 2 displays the first menu of the SQL generator.



Figure 1. SQL Generator web interface



Figure 2. SQL generator OMOP CDM menu

Methods

- Assistant Development:** A conversational agent was developed based on OpenAI's custom GPT 5.4 technology and made publicly accessible. The system helps users define groups and criteria using structured menus and automatically converts a plain-language question into an SQL statement that matches the OMOP-CDM structure. It is applied to a complete, standardised OMOP CDM schema in PostgreSQL:
 - The "OMOP-CDM Explorer" assistant helps users explore the OMOP CDM database by using a menu to perform tasks such as checking demographics, analysing conditions and drugs, defining groups, looking at trends over time, and searching for vocabulary.
 - To improve the accuracy and portability of natural-language-to-SQL generation, we embedded an explicit representation of the OMOP schema in the assistant's knowledge base:
 - omop_schema.sql: the source of truth (tables, columns, datatypes, PK/FK constraints)
 - omop_schema_index.md: human-readable map of the schema
 - omop_schema_index.js: machine-readable index
- Evaluation:** We performed several tests using multiple queries to gauge the assistant's feasibility and usefulness. We designed a set of natural language cohort requests reflecting use cases a clinical researcher might have and queried our database. Below an example of the table generated based on the query created by the SQL generator (using the generated query of 135 lines):

Table	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207	208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239	240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255	256	257	258	259	260	261	262	263	264	265	266	267	268	269	270	271	272	273	274	275	276	277	278	279	280	281	282	283	284	285	286	287	288	289	290	291	292	293	294	295	296	297	298	299	300	301	302	303	304	305	306	307	308	309	310	311	312	313	314	315	316	317	318	319	320	321	322	323	324	325	326	327	328	329	330	331	332	333	334	335	336	337	338	339	340	341	342	343	344	345	346	347	348	349	350	351	352	353	354	355	356	357	358	359	360	361	362	363	364	365	366	367	368	369	370	371	372	373	374	375	376	377	378	379	380	381	382	383	384	385	386	387	388	389	390	391	392	393	394	395	396	397	398	399	400	401	402	403	404	405	406	407	408	409	410	411	412	413	414	415	416	417	418	419	420	421	422	423	424	425	426	427	428	429	430	431	432	433	434	435	436	437	438	439	440	441	442	443	444	445	446	447	448	449	450	451	452	453	454	455	456	457	458	459	460	461	462	463	464	465	466	467	468	469	470	471	472	473	474	475	476	477	478	479	480	481	482	483	484	485	486	487	488	489	490	491	492	493	494	495	496	497	498	499	500	501	502	503	504	505	506	507	508	509	510	511	512	513	514	515	516	517	518	519	520	521	522	523	524	525	526	527	528	529	530	531	532	533	534	535	536	537	538	539	540	541	542	543	544	545	546	547	548	549	550	551	552	553	554	555	556	557	558	559	560	561	562	563	564	565	566	567	568	569	570	571	572	573	574	575	576	577	578	579	580	581	582	583	584	585	586	587	588	589	590	591	592	593	594	595	596	597	598	599	600	601	602	603	604	605	606	607	608	609	610	611	612	613	614	615	616	617	618	619	620	621	622	623	624	625	626	627	628	629	630	631	632	633	634	635	636	637	638	639	640	641	642	643	644	645	646	647	648	649	650	651	652	653	654	655	656	657	658	659	660	661	662	663	664	665	666	667	668	669	670	671	672	673	674	675	676	677	678	679	680	681	682	683	684	685	686	687	688	689	690	691	692	693	694	695	696	697	698	699	700	701	702	703	704	705	706	707	708	709	710	711	712	713	714	715	716	717	718	719	720	721	722	723	724	725	726	727	728	729	730	731	732	733	734	735	736	737	738	739	740	741	742	743	744	745	746	747	748	749	750	751	752	753	754	755	756	757	758	759	760	761	762	763	764	765	766	767	768	769	770	771	772	773	774	775	776	777	778	779	780	781	782	783	784	785	786	787	788	789	790	791	792	793	794	795	796	797	798	799	800	801	802	803	804	805	806	807	808	809	810	811	812	813	814	815	816	817	818	819	820	821	822	823	824	825	826	827	828	829	830	831	832	833	834	835	836	837	838	839	840	841	842	843	844	845	846	847	848	849	850	851	852	853	854	855	856	857	858	859	860	861	862	863	864	865	866	867	868	869	870	871	872	873	874	875	876	877	878	879	880	881	882	883	884	885	886	887	888	889	890	891	892	893	894	895	896	897	898	899	900	901	902	903	904	905	906	907	908	909	910	911	912	913	914	915	916	917	918	919	920	921	922	923	924	925	926	927	928	929	930	931	932	933	934	935	936	937	938	939	940	941	942	943	944	945	946	947	948	949	950	951	952	953	954	955	956	957	958	959	960	961	962	963	964	965	966	967	968	969	970	971	972	973	974	975	976	977	978	979	980	981	982	983	984	985	986	987	988	989	990	991	992	993	994	995	996	997	998	999	1000	1001	1002	1003	1004	1005	1006	1007	1008	1009	1010	1011	1012	1013	1014	1015	1016	1017	1018	1019	1020	1021	1022	1023	1024	1025	1026	1027	1028	1029	1030	1031	1032	1033	1034	1035	1036	1037	1038	1039	1040	1041	1042	1043	1044	1045	1046	1047	1048	1049	1050	1051	1052	1053	1054	1055	1056	1057	1058	1059	1060	1061	1062	1063	1064	1065	1066	1067	1068	1069	1070	1071	1072	1073	1074	1075	1076	1077	1078	1079	1080	1081	1082	1083	1084	1085	1086	1087	1088	1089	1090	1091	1092	1093	1094	1095	1096	1097	1098	1099	1100	1101	1102	1103	1104	1105	1106	1107	1108	1109	1110	1111	1112	1113	1114	1115	1116	1117	1118	1119	1120	1121	1122	1123	1124	1125	1126	1127	1128	1129	1130	1131	1132	1133	1134	1135	1136	1137	1138	1139	1140	1141	1142	1143	1144	1145	1146	1147	1148	1149	1150	1151	1152	1153	1154	1155	1156	1157	1158	1159	1160	1161	1162	1163	1164	1165	1166	1167	1168	1169	1170	1171	1172	1173	1174	1175	1176	1177	1178	1179	1180	1181	1182	1183	1184	1185	1186	1187	1188	1189	1190	1191	1192	1193	1194	1195	1196	1197	1198	1199	1200	1201	1202	1203	1204	1205	1206	1207	1208	1209	1210	1211	1212	1213	1214	1215	1216	1217	1218	1219	1220	1221	1222	1223	1224	1225	1226	1227	1228	1229	1230	1231	1232	1233	1234	1235	1236	1237	1238	1239	1240	1241	1242	1243	1244	1245	1246	1247	1248	1249	1250	1251	1252	1253	1254	1255	1256	1257	1258	1259	1260	1261	1262	1263	1264	1265	1266	1267	1268	1269	1270	1271	1272	1273	1274	1275	1276	1277	1278	1279	1280	1281	1282	1283	1284	1285	1286	1287	1288	1289	1290	1291	1292	1293	1294	1295	1296	1297	1298	1299	1300	1301	1302	1303	1304	1305	1306	1307	1308	1309	1310	1311	1312	1313	1314	1315	1316	1317	1318	1319	1320	1321	1322	1323	1324	1325	1326	1327	1328	1329	1330	1331	1332	1333	1334	1335	1336	1337	1338	1339	1340	1341	1342	1343	1344	1345	1346	1347	1348	1349	1350	1351	1352	1353	1354	1355	1356	1357	1358	1359	1360	1361	1362	1363	1364	1365	1366	1367	1368	1369	1370	1371	1372	1373	1374	1375	1376	1377	1378	1379	1380	1381	1382	1383	1384	1385	1386	1387	1388	1389	1390	1391	1392	1393	1394	1395	1396	1397	1398	1399	1400	1401	1402	1403	1404	1405	1406	1407	1408	1409	1410	1411	1412	1413	1414	1415	1416	1417	1418	1419	1420	1421	1422	1423	1424	1425	1426	1427	1428	1429	1430	1431	1432	1433	1434	1435	1436	1437	1438	1439	1440	1441	1442	1443	1444	1445	1446	1447	1448	1449	1450	1451	1452	1453	1454	1455</
-------	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	--------



Mapping Our Future Health data to OMOP

Mapping Our Future Health Data to OMOP

Magda Meier, Kevin Garwood
Our Future Health, Manchester, United Kingdom.

Background

Our Future Health is a non-profit organisation which aims to recruit 5 million adults to a prospective cohort that will catalyse artificial and translational research for both common and rare diseases. It is already the world's largest connected cohort dataset due to its scale, depth and breadth and the cohort data is a crucial resource for population health studies. We currently make health care records, genetics, clinical measurements, and questionnaire data available for approved research studies. The health care records will span a variety of health services across the UK's nations. Our Future Health seeks to map its data holdings to OMOP to better support data discoverability and a wider range of studies. It also seeks to leverage tools and data sets that have already been developed by the OMOP community. We are aware that some of our researchers will want to reproduce analyses conducted elsewhere. Supporting these needs requires data interoperability.

Our Future Health is not just interested in OMOP mappings. We are interested in the stability and scalability of infrastructure that can support these mappings in a production setting as well. Rather than try to build it all ourselves, we are seeking to collaborate with different OMOP groups and in pool collective experience. As part of our early efforts, we have been examining the use of various OMOP tools, libraries and reference data sets that would support whole cohort processing in a cloud-native environment. Currently we are evaluating how tools such as Camt, White Rabbit, and Rabbit in a Hat would support our data processing characteristics. We are also investigating how OMOP could better support AI-related activities.

Methods

We aim to use the following approach:

1. **Identify example research studies.** We will engage researchers who work with our future OMOP data sets who will have the greatest impact on healthcare. This will help us understand the priorities and how we can have the most impact.
2. **Limit the scope of data to map.** We envision that OMOP suits some data modalities better than others. We will initially focus on mapping the participant routinely collected health data gathered from England, Scotland and Wales based on community input from other subjects who have had to do similar mappings. We will assess whether OMOP can help reduce the work of harmonising data from the UK's nations that would need to be done across a broad range of heart disease. We will then prioritise other OMOP mappings based on their complexity and stability.
3. **Partition data processing tasks and evaluate suitability of OMOP tools for each.** We will separate concerns related to those of domain experts defining the mappings and those of data engineers implementing them at scale in production. We are specifically interested in the interplay between these two areas of interest.

4. **Prototype OMOP workflows.** Based on our scoping priorities for data, we will develop exemplar data flows that will map participant data. We are specifically interested in how community informed iterations of mappings can be supported.
5. **Assess the performance of the OMOP workflows.** Because our cohort is so large, we will need to assess the compute performance of code used to process the mappings in the UK data release activities.
6. **Assess maintainability of OMOP workflows.** We are less interested in learning to do one-off OMOP mappings and more in how those would work in regular production activities.

Results

Much of our efforts spent so far have been gathering a community of OMOP partners and scoping the data sets that are likely to generate the most return value for our researchers. We have recently examined how our OMOP community partners could work together and how mappings could be managed (see Figure 1).

Figure 1. Supporting a community-evolved OMOP mapping effort.

Conclusions

From our work so far, we have concluded:

- ✓ Most OMOP mapping projects have been observed so far appear to be one-off efforts that support a specific study rather than provide a long-term, general purpose service. Most of those projects emphasize the characteristics of particular mappings rather than the software maintenance of those mappings. Maintainability remains an important aspect of a sustainable OMOP effort.
- ✓ Groups we have consulted who have significantly invested in OMOP advocate rationalising its coverage by business need rather than by attempting to map entire data assets as a goal in itself.
- ✓ Bearing in mind the cost of maintaining mappings in production, we will emphasize the oldest and most stable use cases for which OMOP was originally developed.
- ✓ We believe in investing in using a basket of common data models which best suits specific data modalities. OMOP appears well-suited for mapping routinely collected health data and to a lesser extent questionnaires. Other data modalities such as genetic, imaging, and geospatial data may warrant being aligned with more domain-specific common data models.
- ✓ Groups we have consulted who use OMOP have an appreciation for how much work is involved when they try to do mappings on their own. This collective maintenance has provided the driver for us to engage in our OMOP efforts through a cooperative effort featuring shared experience and likely shared mappings among our potential partners.

References

1. Cook WB, Adams N, Adjirey A, Ananthim R, Balashanovic M, Blackwood B, Booth A, Cairns BJ, Cornall A, Ellis S, Edwards B. Cohort Profile: Our Future Health. *International Journal of Epidemiology*. 2023 Dec;54(8):dyt17.



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?



July 28: Newcomer Introductions

Ryan Chou

Louise Sun

András Novoszáth

Zongzhi Zach Liu

Marianna Macera

Pulkit Gairola

Asha Mahesh

Asif Syed

Michael Gaile

Anne Marie Chan Cheung Shing

Carla Laladula

Jonathan Lowe

Sandy Estremera-Zink

Beatrice Okrah

Welcome to the annual **Welcome OHDSI Newcomers** community call! If you are relatively new to the community or hope to have a bigger role, this is a great opportunity to introduce yourself.

Please tell us who you are and where you work, what are your research interests, how you hope to help the community, and how OHDSI can help your own research journey.



**The weekly OHDSI community call is held
every Tuesday at 11 am ET.**

Everybody is invited!

Links are sent out weekly and available at:
[ohdsi.org/community-calls-2026](https://www.ohdsi.org/community-calls-2026)



Find your workgroup.

Fuel our mission.

ohdsi.org/workgroups