Oncology Research Spotlight: Advancing Cancer Care in OHDSI

OHDSI Community Call
Aug. 11, 2026 • 11 am ET

OHDSI
OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS

Upcoming Community Calls

Date	Topic
Aug. 11	Oncology Research Spotlight: Advancing Cancer Care 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
Sept. 22	TBA
Sept. 29	OHDSI/OMOP Research Spotlight

Three Stages of The Journey

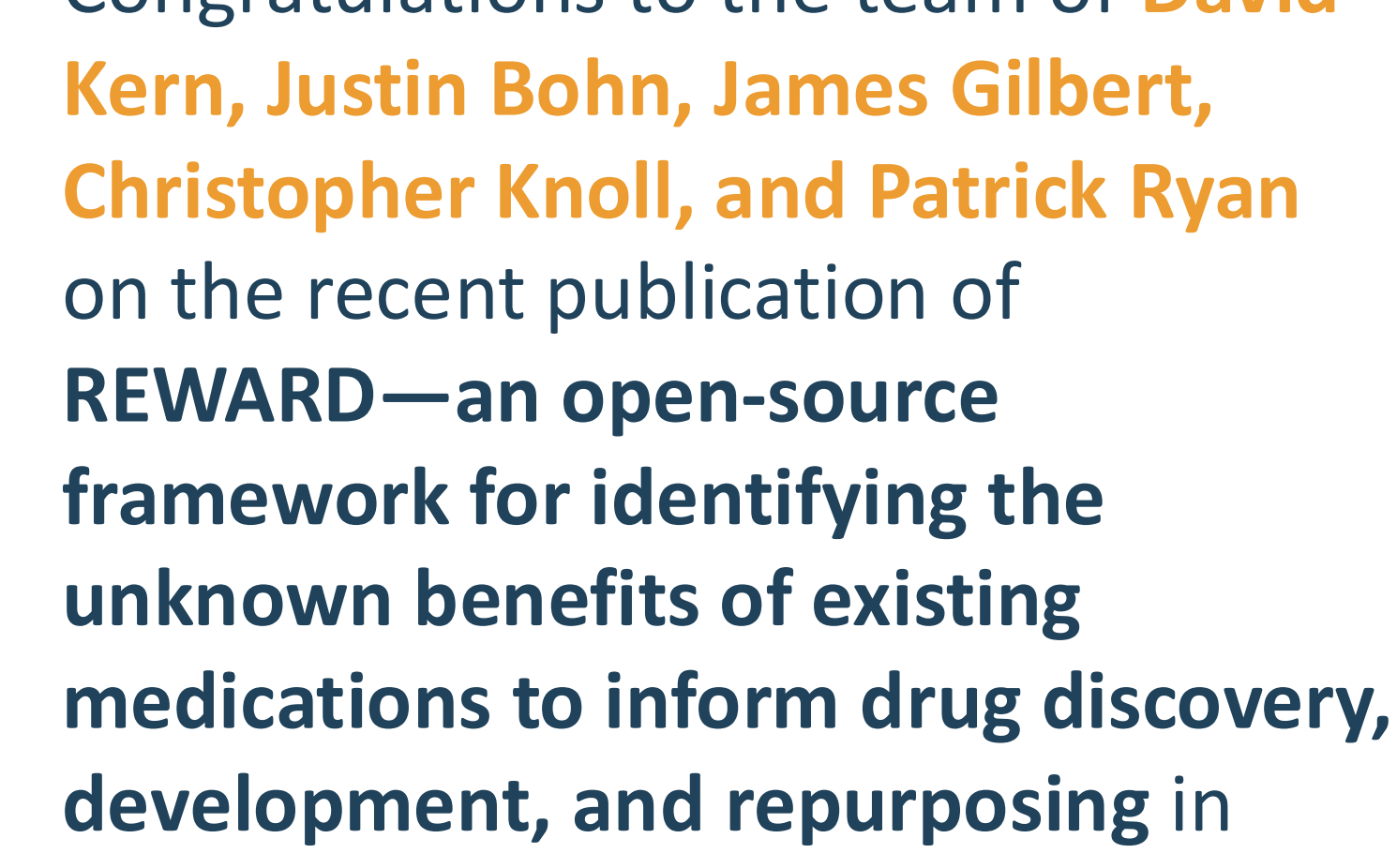
Where Have We Been?

Where Are We Now?

Where Are We Going?

OHDSI Shoutouts!

Congratulations to the team of **David Kern, Justin Bohn, James Gilbert, Christopher Knoll, and Patrick Ryan** on the recent publication of **REWARD—an open-source framework for identifying the unknown benefits of existing medications to inform drug discovery, development, and repurposing** in *JAMIA*.



Journal of the American Medical Informatics Association, 2026, 1–12
https://doi.org/10.1093/jamia/ocag137
Research and Applications

AMIA OXFORD

REWARD—an open-source framework for identifying the unknown benefits of existing medications to inform drug discovery, development, and repurposing

David M. Kern, PhD, MS*,[1], Justin Bohn, ScD, ScM[2], James P. Gilbert, PhD[2], Christopher Knoll, BS[3], Patrick B. Ryan, PhD[1]

[1]Global Epidemiology, Johnson & Johnson, Horsham, PA 19044, United States
[2]Global Epidemiology, Johnson & Johnson, Titusville, NJ 08560, United States
[3]Global Epidemiology, Johnson & Johnson, Raritan, NJ 08869, United States
*Corresponding author: David M. Kern, PhD, MS, Global Epidemiology Organization, Johnson & Johnson Innovative Medicine, 850 Ridgeview Dr, Horsham, PA 19044, United States (dkern2@its.jnj.com)

Abstract

Objectives: **The potential of "big data" in health research remains largely untapped, particularly concerning real-world data sources such as administrative health data and electronic medical records. While healthcare insurance claims data have been essential for assessing medication safety and effectiveness within indicated patient populations, exploring broader drug-outcome associations could uncover significant insights.**

Materials and Methods: **The REWARD (REal-World Evidence and Research of Drug performance) framework was established to perform large-scale analytics on medication benefits beyond their original indications. Employing an "all-by-all" approach, it investigates all medication-outcome pairs using standardized vocabularies from the OMOP common data model and implements causal inference methods, including self-controlled cohort and active comparator new-user designs. Negative control calibration and large-scale propensity scores are used to control for systematic bias in the study designs.**

Results: **Our framework enables the identification of new benefits associated with thousands of medications across millions of patients. In particular, REWARD facilitates insights into disease mechanisms to guide the development of novel interventions, identifies opportunities for drug repurposing, and informs potential additional indications for drugs in clinical development. REWARD has led to various publications with the goal of informing new drug development.**

Discussion: **REWARD is distinguished by its open-source implementation, use of standardized OMOP CDM vocabularies, and integration of best-practice pharmacoepidemiologic methods. Results are best interpreted as hypothesis-generating signals, with the active comparator new-user design providing higher causal rigor for prioritized drug-outcome pairs.**

Conclusion: **The REWARD framework demonstrates how real-world evidence can be harnessed to address unmet medical needs, particularly for diseases lacking effective approved treatments. By making the REWARD analytic package open-source and accessible, we promote an open scientific approach, while maintaining best practices in pharmacoepidemiology.**

Key words: epidemiologic methods; drug repositioning; drug discovery; electronic health records; big data.

OHDSI Shoutouts!

Congratulations to the team of **Elizabeth Sherwin, Benjamin Martin, Xiao Xu, Anna Booman, Alyssa Howren, Kimberlee McKay, Sadaf Kazi, Sean O'Reilly, Katie Kolla, Khyzer Aziz, Katherine Bianco, and Stephanie Leonard** on the recent publication of **Congenital heart disease in pregnancy and severe maternal morbidity: A distributed data network study** in *Pregnancy*.

Received: 13 April 2026 | Revised: 27 June 2026 | Accepted: 13 July 2026

DOI: 10.1002/pmf2.70370

PREGNANCY

BRIEF REPORT

Congenital heart disease in pregnancy and severe maternal morbidity: A distributed data network study

Elizabeth B. Sherwin[1] | **Benjamin Martin**[2] | **Xiao Xu**[3] | **Anna Booman**[4] | **Alyssa Howren**[5] | **Kimberlee McKay**[6,7] | **Sadaf Kazi**[8,9] | **Sean O'Reilly**[2] | **Katie S. Kolla**[10] | **Khyzer Aziz**[2,11] | **Katherine Bianco**[1] | **Stephanie A. Leonard**[1]

[1]Division of Maternal-Fetal Medicine and Obstetrics, Department of Obstetrics and Gynecology, Stanford University School of Medicine, Stanford, California, USA

[2]Biomedical Informatics and Data Science, The Johns Hopkins University School of Medicine, Baltimore, Maryland, USA

[3]Department of Obstetrics and Gynecology, Columbia University Irving Medical Center, New York, New York, USA

[4]Department of Anesthesiology, Perioperative, and Pain Medicine, Stanford University School of Medicine, Stanford, California, USA

[5]Department of Epidemiology and Population Health, Stanford University School of Medicine, Stanford, California, USA

[6]Department of Obsetrics and Gynecology, Avera McKennan Hospital, Sioux Falls, South Dakota, USA

[7]Department of Obstetrics and Gynecology, Sanford School of Medicine, University of South Dakota, Sioux Falls, South Dakota, USA

[8]National Center for Human Factors in Healthcare, MedStar Health Research Institute, Columbia, Maryland, USA

[9]Department of Emergency Medicine, Georgetown University Medical Center, Washington, District of Columbia, USA

[10]Department of Cardiothoracic Surgery, Stanford Children's Health, Stanford, California, USA

[11]Department of Pediatrics, The Johns Hopkins University School of Medicine, Baltimore, Maryland, USA

Correspondence
Elizabeth B. Sherwin, Division of Maternal-Fetal Medicine and Obstetrics, Department of Obstetrics and Gynecology, Stanford University School of Medicine, Stanford, CA, USA.
Email: esherwin@stanford.edu

Funding information
Eunice Kennedy Shriver National

Abstract

Introduction: Pregnant people with congenital heart disease (CHD) are a growing patient population in obstetrics, yet evidence on the risk for severe maternal morbidity (SMM) has largely been limited to studies that lack specificity for CHD. We conducted this study to demonstrate the utility of distributed data networks for obstetric research and to characterize pregnant patients with CHD and their risk of SMM.

OHDSI Shoutouts!

Congratulations to the team of **Jing Wang-Silvanto, Mansee Jajoo, He Guo, Rishi Ohri, Judith Nelissen, and Carolina Casañas i Comabella** on the recent publication of **Evaluating AI Performance in Systematic Literature Reviews for HEOR: A Case Study** in the *Journal of Health Economics and Outcomes Research*.

Wang-Silvanto J, Jajoo M, Ohri R, Guo H, Nelissen J, Casañas i Comabella C. Evaluating AI performance in systematic literature reviews for HEOR: a case study. *JHEOR.* **2026;13(2):46-54.** doi:10.36469/jheor.2026.165173

JHEOR Journal of Health Economics and Outcomes Research

Methodology and Healthcare Policy

Check for updates

Evaluating AI Performance in Systematic Literature Reviews for HEOR: A Case Study

Jing Wang-Silvanto[1*], Mansee Jajoo[2], Rishi Ohri[2], He Guo[2], Judith Nelissen[3], Carolina Casañas i Comabella[4]

[1]Astellas Pharma A/S, filial I Finland, Espoo; Observational Health Data Sciences and Informatics (OHDSI), Finland
[2]Astellas Pharma, Astellas Pharma Global Development Inc. Northbrook, USA
[3]Astellas Pharma Europe BV, Leiden, The Netherlands
[4]PPD Evidera Health Economics & Market Access, Thermo Fisher Scientific, London, UK

ARTICLE INFORMATION

Accepted July 15, 2026

Keywords: artificial intelligence; machine learning; systematic literature reviews; large language models; automation tools; evidence synthesis

*Corresponding author:
Email address: jingwang@ohdsi.org

Supplementary Material

ABSTRACT

Background: Systematic literature reviews (SLRs) are central to evidence generation in health economics and outcomes research, but traditional SLR workflows are labor-intensive. Artificial intelligence (AI) is increasingly being explored to support evidence synthesis tasks.

Objectives: To describe methods used to assess AI performance in screening and data extraction in SLRs, and to provide recommendations on how to ensure appropriate use of AI in literature reviews.

Methods: We conducted an AI-assisted SLR that mirrored a traditional SLR performed by humans only and assessed the performance of the AI tools employed. AI was used to screen titles/abstracts, screen full texts, and conduct data extraction. Using the traditional SLR as the benchmark, AI performance for title/abstract screening was evaluated in terms of accuracy, recall, and precision at predefined screening milestones, followed by the accuracy assessment of full-text screening by AI. Data items extracted by AI were verified by humans and categorized as correct, incomplete, missing, incorrect, or requiring human checks. The average accuracy rate calculated on the basis of correct data items extracted was used as an indicator of AI performance in data extraction.

Results: During title/abstract screening, accuracy and recall remained high but precision was low, increasing false positives. Full-text screening identified a minority of studies included in the traditional SLR. Mean extraction accuracy was 72.93% (range, 57.69%-88.46%). No data were hallucinated by the AI model used.

Conclusions: AI can support some tasks in evidence synthesis workflows, but uneven results across stages reinforce the need for human oversight and clear validation methods. Transparent, metric-based evaluation and structured human verification are essential for AI use in evidence synthesis to ensure quality standards are met.

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	Generative AI and Analytics
Wednesday	9 am	Patient-Level Prediction (PLP)
Wednesday	2 pm	Canada Chapter
Wednesday	7 pm	Eyecare and Vision Research
Thursday	10 am	Rare Diseases
Thursday	10 am	Africa Chapter
Thursday	10 am	ATLAS/WebAPI
Thursday	10 am	GIS-Geographic Information System
Thursday	12 pm	Medical Devices
Friday	11 am	Clinical Trials
Friday	11:30 am	Steering
Friday	11 pm	China Chapter
Monday	9 am	Vaccine Vocabulary
Monday	2 pm	Electronic Animal Health Records
Tuesday	10 am	CDM Survey Subgroup

OHDSI LATAM Symposium

August Newsletter is Available

OHDSI

The Journey Newsletter (August 2026)

Check out the enhancements coming in ATLAS 3.0, learn more about a record-breaking Global Symposium showcase, check out an impressive collection of updates and publications from July, as well as spotlights on Maxim Moinat and Misha Khitrun, and plenty more in the August 2026 newsletter.
#JoinTheJourney

Podcast: ATLAS 3.0, Showcase Themes

In the August 2026 On The Journey podcast, Patrick Ryan and Craig Sachson discuss ATLAS as a community tool, enhancements being made for ATLAS 3.0 and why community support is critical now. They also discussed the 2026 collaborator showcase, including how AI was used in the review process, and some themes that came out of a record-breaking submission total. *(If video does not appear, please click 'View this email in your browser')*

Community Updates

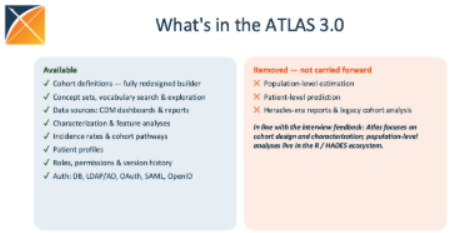
Where Have We Been?

- **Record Submission Total:** Acceptances went out to more than 160 submissions for the 2026 Global Symposium, including more than 130 posters and 24 demos. There will be 10 lightning talks this year, and those are listed later in this newsletter. Thank you to the 41 volunteers in our Scientific Review Committee for their time and efforts in the peer-review process.
- **Global Growth:** The first LATAM Symposium was held July 30-31 in Salvador, Brazil, with the goal of bringing together the public sector, industry and academia to advance health data standardization and strengthen an ethical, open and inclusive regional community. The symposium worked to accelerate OMOP Common Data Model adoption in Brazil and across the region, fostering integration, scientific collaboration and social impact. Thank you to **Valentina Martufi** and **Julio Oliveira** for leading this initiative.
- **New Workgroup: Jenna Reps** recently posted a call for developers with knowledge or interest in R or shiny or web development/design and UI to join the REVEAL Workgroup (Results Enhancements & Visualization for Evidence, Analysis & Learning). If you are interested, please respond to the forum post or sign up here.

Where Are We Now?

- **Clinical Definitions:** A critical aspect of the phenotype development and evaluation process is aligning on a clinical definition that provides a full specification for each of the indications, exposures, and outcomes of interest in our studies. But how good are we as a community at reaching that shared understanding? During the Aug. 4 community call, **Patrick Ryan** will lead us in examining the topic of clinical definitions by highlighting work in our community and through a series of interactive exercises.
- **ATLAS 3.0 Feedback: Peter Hoffmann** provided a preview of the ATLAS 3 demo site during the July 7 community call. The ATLAS workgroup is looking for feedback about the new site. The demo is available here (Db Authenticaton enabled: use ohdsi/ohdsi, and anonymous access will be enabled in a future update). Discussions will be held here. More details are included later in this newsletter. Thank you for collaborating on this critical open-source tool for our community.
- **Survey Requests: Jiwon Um**, a PhD student at Yonsei University in Korea, introduced a brief survey focusing on OHDSI in the Era of Medical AI. Medical AI is reshaping how observational research evidence is generated and used. The OHDSI community has spent more than a decade building infrastructure — OMOP CDM, standardized vocabularies, federated networks, open-source methodology — that may have a distinctive role in this transition. This survey asks how you, as a member of the community, see that role and the tensions surrounding it. Also, **Jiwoo Seo**, a PhD student at Florida State University, announced a study on the motivations that drive experts to continuously contribute to biomedical ontology communities like OHDSI. You can take part in this survey here.

Rebuilt for Performance: Explore the ATLAS 3.0 Preview and Share Your Feedback



ATLAS is a free, open-source web platform that empowers researchers to explore and analyze standardized healthcare data. By integrating with the OMOP Common Data Model, ATLAS enables users to search medical vocabularies, define specific patient cohorts, and examine population data through an intuitive web interface—all without needing direct database access or writing complex code.

The upcoming ATLAS 3.0 release completely modernizes the platform behind the scenes to deliver a faster, smoother user experience. On the surface, users will benefit from a refreshed interface with updated search tools, sharper interactive charts, and instant patient counts as you refine your study criteria. Under the hood, a new modular design makes the tool significantly more responsive and gives technical teams the flexibility to build custom extensions tailored to their organization's unique research needs.

As we prepare for the official rollout, we invite community members and partner organizations to explore the preview release, test its new features, and help shape its final release. You can check out the source code and deployment guides on GitHub or review the ATLAS 3.0 Preview slides for a deeper look at what's new. Please share your feedback, bug reports, and suggestions with the OHDSI development team on GitHub to help ensure ATLAS 3.0 delivers the best possible experience for researchers worldwide.

ATLAS 3.0 Preview

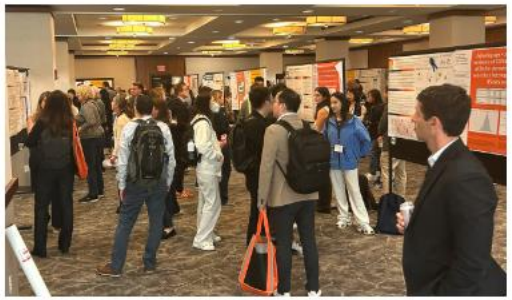
GitHub: ATLAS 3.0

Video: ATLAS 3.0 Presentation & Demo

My Journey: Maxim Moinat

In the latest installment of our "My Journey" series, Maxim Moinat (Researcher at the Medical Informatics Department of the Erasmus Medical Center) shares his love for turning "messy real-world data" into high-level evidence. Maxim discusses his pride in building open-source ETL tools used across the global OHDSI community, the warmth of the network's family-like culture, and the massive trajectory of the DARWIN EU initiative as it scales up parallel studies to impact international healthcare decision-making. *(If video does not appear, please click 'View this email in your browser')*

Global Showcase To Feature 160+ Submissions, Highlighting Wide Breadth of Research



The Collaborator Showcase remains one of the highlight events during any OHDSI symposium, and the one coming this October in New Jersey should be one of the most memorable. Following a record-breaking total of showcase submissions, our amazing Scientific Review committee of 41 volunteers (see graphic below) accepted more than 160 posters, software demos or lightning talks to share during the 2026 Global Symposium.

July Publications

Guo Y, Burn E, Jödicke AM, Mercadé-Besora N, Lopez-Güell K, Chen X, Du M, Li X, Kolde R, Oja M, Mosseveld M, Verhamme K, Politi J, Delmestri A, Man WY, Rijnbeek P, Prieto-Alhambra D, Català M. DrugUtilisation: An R Package to Support Drug Utilisation Research Using the OMOP Common Data Model. Pharmacoepidemiol Drug Saf. 2026 Jul;35(7):e70433. doi: 10.1002/pds.70433. PMID: 42448616; PMCID: PMC13368618.

Park J, Kim JH, Kim YS, Lim G, Song HJ, Rhee SY, Hwang S, Cho D, Kim M, Kim BY, Jeong CW, Kim DY, Park HA, Kim MH, Kim SH, Seo WW, Kim WJ, Suh YS, Park RW, Shin JY. GLP-1 Receptor Agonists for Weight Loss and Risk of Major Safety Outcomes: A Multicentre Cohort Study. Diabetes Obes Metab. 2026 Jul 6. doi: 10.1111/dom.71065. Epub ahead of print. PMID: 42410329.

Rueda M, Ramírez-Anguita JM, López-Sánchez V, Aguiló-Castillo S, López MEG, Labarga A, Mayer MÁ, Esteve JR, Gut IG. Enhancing semantic interoperability in precision medicine: converting OMOP CDM to Beacon v2 in the Spanish IMPaCT-Data project. BMC Med Inform Decis Mak. 2026 Jul 8. doi: 10.1186/s12911-026-03649-0. Epub ahead of print. PMID: 42420997.

Shubov M, Beernink M, Carus J, Wiederhold AJ; AI-CARE Consortium; Gundler C. Enhancing stratification for survival analyses across standardized data sources. BMC Med Inform Decis Mak. 2026 Jul 13;26(1):261. doi: 10.1186/s12911-026-03666-z. PMID: 42443871; PMCID: PMC13361719.

López-Sánchez I, Palomar-Cros A, Giuliodori A, Granés L, Pérez-Crespo L, Raventós B, Burn E, Barchuk A, Verbiest A, Eteve-Pitsaer C, Newby D, Rowlands EJ, Enerly E, Jadhav G, De Schutter H, Hsu JC, Evers J, Vrancic J, Peruchet-Noray L, Sung L, Kiss L, Van Swieten M, Šekerija M, Mosseveld M, Mabry PL, Kolde R, Patnoe S, You SC, Kim S, Reisberg S, Nguyen TKH, Lehtonen T, Myklebust TÅ, Bagyura Z, Golozar A, Roel E, Duarte-Salles T. 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 Jul 21;16(7):e119069. doi: 10.1136/bmjopen-2026-119069. PMID: 42481199; PMCID: PMC13410706.

Ozonze O, Okonor O, Dike U, Adedeji T. Enabling interoperability across disparate health data sources using Large Language Models. PLOS Digit Health. 2026 Jul 23;5(7):e0001511. doi: 10.1371/journal.pdig.0001511. PMID: 42490570; PMCID: PMC13395378.

Kim H, Kim M, Kim S, You SC. From study design to executable code: automating target trial emulation with large language models. JAMIA Open. 2026 Jul 9;9(4):ooag131. doi: 10.1093/jamiaopen/ooag131. PMID: 42428529; PMCID: PMC13348706.

Collaborator Spotlight: Masha Khitrun

OMOP CDM is often what people see first — the beautifully organized structure. But the vocabularies are what give that structure meaning — a universal language that enables collaborative research across various countries.

I like to think of it this way: if OMOP is the grammar of a shared scientific language, then vocabularies are the words. And just like any living language, they must evolve.

OHDSI
Observational Health Data Sciences and Informatics

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 23 at 10 pm ET.

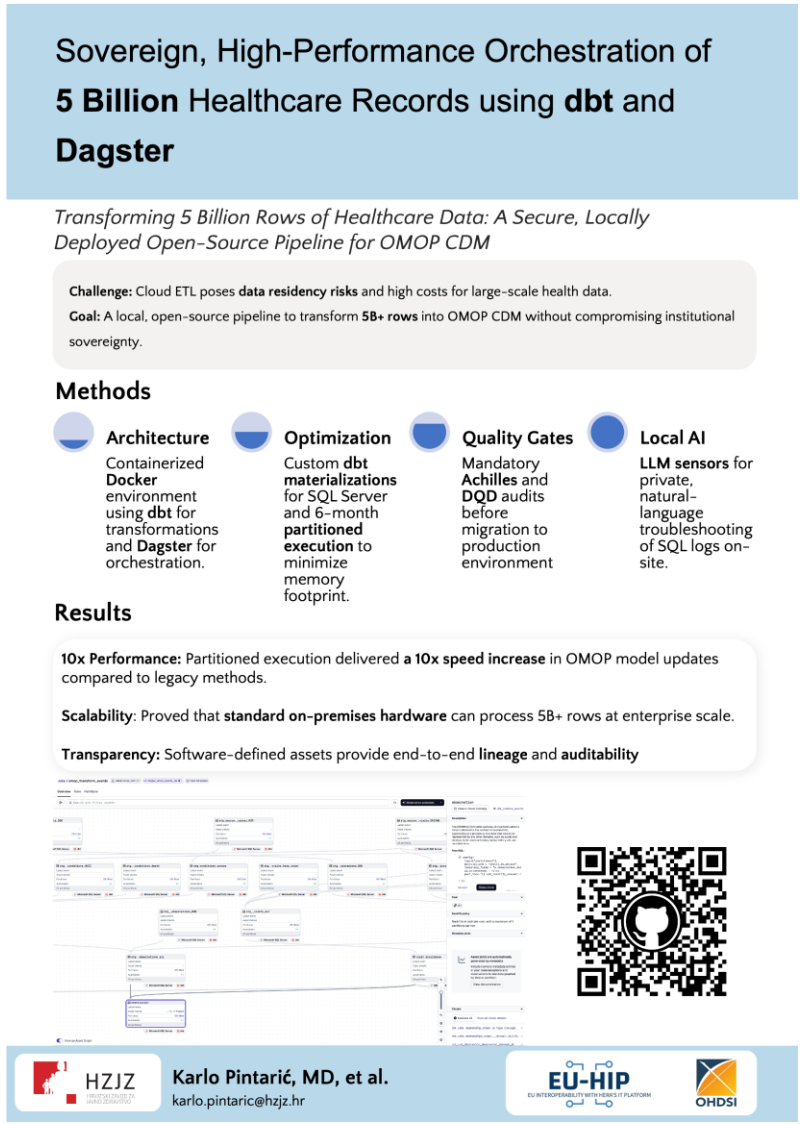
ohdsi.org/APAC2026

#OHDSISocialShowcase This Week

Monday

Transforming 5 Billion Rows of Healthcare Data: A Secure, Locally Deployed Open-Source Pipeline for OMOP CDM

(Karlo Pintarić, Marko Čavlina, Antea Jezidžić, Pero Ivanko)



#OHDSISocialShowcase This Week

Tuesday

Lateral specific concept mapping from ICD-10 to SNOMED CT: addressing information loss

(Vojtech Huser, Ilya Pyatin, Tatsiana Skuhareuskaya, Maria Khitrun, Sarah Seager, Sebastiaan van Sandijk, Polina Talapova, Oleg Zhuk)

Lateral specific concept mapping from ICD-10 to SNOMED CT: addressing information loss

Vojtech Huser[1], Ilya Pyatin[1], Tatsiana Skuhareuskaya[1], Maria Khitrun[1], Sarah Seager[1], Sebastiaan van Sandijk[1], Polina Talapova[2], Anna Ostropolets[3], Oleg Zhuk[1]
[1]EPAM, [2]SciForce, [3]Columbia University

BACKGROUND

Medical terminologies, such as SNOMED CT (Systematized Nomenclature of Medicine - Clinical Terms), play an important part in data harmonization. Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM) includes an important terminology layer in the model and uses the principle of non-standard terms being harmonized to standard terms. Any gaps and challenges in this principle and underlying terminologies may impact research use of the OMOP-shaped data.

Pre-coordination is a frequently adopted approach in medical terminologies. SNOMED International allows creation of pre-coordinated, lateral side specific disease concepts in SNOMED CT.[1] In current SNOMED CT version, this approach may not be fully implemented (i.e., not all expected concepts may currently exist). Another important terminology, ICD-10CM, which is a source data terminology, also uses pre-coordination to capture left or right location detail about a diagnosis.

Current mapping of ICD10-CM concepts for diagnoses within OMOP model terminology layer (standardized vocabularies) often maps side specific diagnoses to standard concepts (from SNOMED CT terminology) that are not side specific (narrow to broad type of mapping). This leads to information loss and it is not optimal for some research questions. Our goal is submit new concept requests to SNOMED International (standard developing organization for SNOMED CT) for side specific condition concepts and to better quantify to what extent a mapping that "drops laterality" occurs. In this study, we focus on diagnoses in the ophthalmology domain.

MATERIALS

SNOMED International provides an Excel batch submission template. Each new proposed concept must be linked to a parent concept. A pilot submission of two concepts made by us was accepted by SNOMED and provides guidance for future (and larger) batch submission targeted by this study.

METHODS

We used published concept counts [2] to arrive at ICD-10CM ophthalmology conditions that include left or right laterality. We define DiseaseConceptWithLaterality (e.g., https://athena.ohdsi.org/search-terms/terms/ 37200408; Exudative age-related macular degeneration, left eye; H35.322) as concept that describes a disease and includes a specific left vs. right side location. Similarly, we define DiseaseConceptWithoutLaterality (e.g., https://athena. ohdsi.org/search-terms/terms/376966; Exudative age-related macular dege-neration; SNOMED CT concept code 414173003) as concept that describes a disease only and does not include left or right side as location. We used regular expressions to identify DiseaseConceptsWithLaterality in non-standard and standard concept subgroups.

We decided that the submission of new concepts will be split into batch A that will contain requests for new concepts where usage (count of events on one or multiple datasets) of the ICD10 term is over some percentile. In other words, batch A will target the most important diagnoses and avoid overloading SNOMED team with a large number of new concept requests. Remaining new concepts will be submitted in a possible later batch B. Results of the ophthalmology domain implementation will inform later phase 2 global scope expansion of the approach (for all diagnoses not limited to ophthalmology).

For batch submission, we generated new concepts with fully specified names (FSNs) using a script. We added 'of left eye' and 'of right eye' to the existing DiseaseConcept WithoutLaterality. This strategy was used in our pilot submission.

RESULTS

Pilot submission was made in October 2025 and appeared in production in December 2025 (SNOMED CT US Edition; facilitated by the monthly release cycle). Study repository at github.com/informaticsrepo/omop-lat contains result files referenced below. We found 3116 ICD10-CM source ophthalmology diagnosis concepts of type DiseaseConceptWithLaterality (see supplemental file s1.csv) mapped to 730 standard SNOMED CT concepts (file s2.csv). Study repository contains analogous results and supplemental files for phase 2 scope for all diagnoses, other expanded results and additional discussion points. For batch A, number of SNOMED CT parent concepts where lateral-specific terms will be requested is: 20 (for top 3 percentile), 34 (for top 5 percentile) and 132 (for top 20 percentile) (file s3.csv).

CONCLUSION

New SNOMED CT terms allow for preservation of laterality information contained in the source data and source concept (e.g., H35.322). In phase 2 context, we quantified that laterality is preserved in 17.2% and dropped in 92.8%.

DISCUSSION

There are several related considerations. First, OMOP model currently relies fully on pre-coordinated terms. Laterality addition is a poster-child example of possibly using a post-coordination approach. SNOMED CT expression language is a fully developed post-coordination framework. The HL7 FHIR standard also mostly works with pre-coordination approaches. Changes to OMOP model must be carefully considered (balanced against the virtue of relative model stability over time) and we concluded that large paradigm shift for *_concept_id columns (introducing post-coordination) is not something the OMOP research community may prefer as of 2026.

Second, solving laterality for diagnoses is related to solving related problem in procedural data. E.g., intraocular injection of anti-VEGF agents to the affected eye (left vs. right) would possibly have to be similarly addressed in procedural terminology. We provide some expanded discussion on this topic in study repository. We define analogous scopes of phase 3 (ophthalmology procedures) and phase 4 (general procedures) for procedures.

Third, we fully acknowledge that for the majority of research analyses, the laterality information is not needed. However, in the current SNOMED-CT it is clearly implemented in some diseases/symptoms and not implemented in others. It is a somewhat subjective decision to declare it important for some concepts and not for other concepts. Different researchers will place the implementation boundary differently.

Fourth, OMOP vocabularies as of January 2026 consist of 447430 non-standard condition concepts and 167045 standard concepts (with SNOMED CT accounting for 105324 concepts and ICD-O-3 for 56858). Addition of several thousands of pre-coordinated terms within SNOMED CT is unlikely to disrupt existing approaches, tools or software packages. OMOP tools and model infrastructure for concept relationships can easily accommodate such an expansion. For users not interested in lateral side location detail, they can continue using existing concepts (with subsumed concepts included) as before.

REFERENCES

1. van Berkum MM, Rallins MC, Spackman KA. Choosing Sides. Assigning Laterality as an Attribute in SNOMED CT. Proc AMIA Symp. 2002:1184.
2. Datasets (Center for Clinical Observational Investigations) https://lhncbc.nlm.nih.gov/CCOI/datasets/datasets.html [accessed Dec 3, 2025]

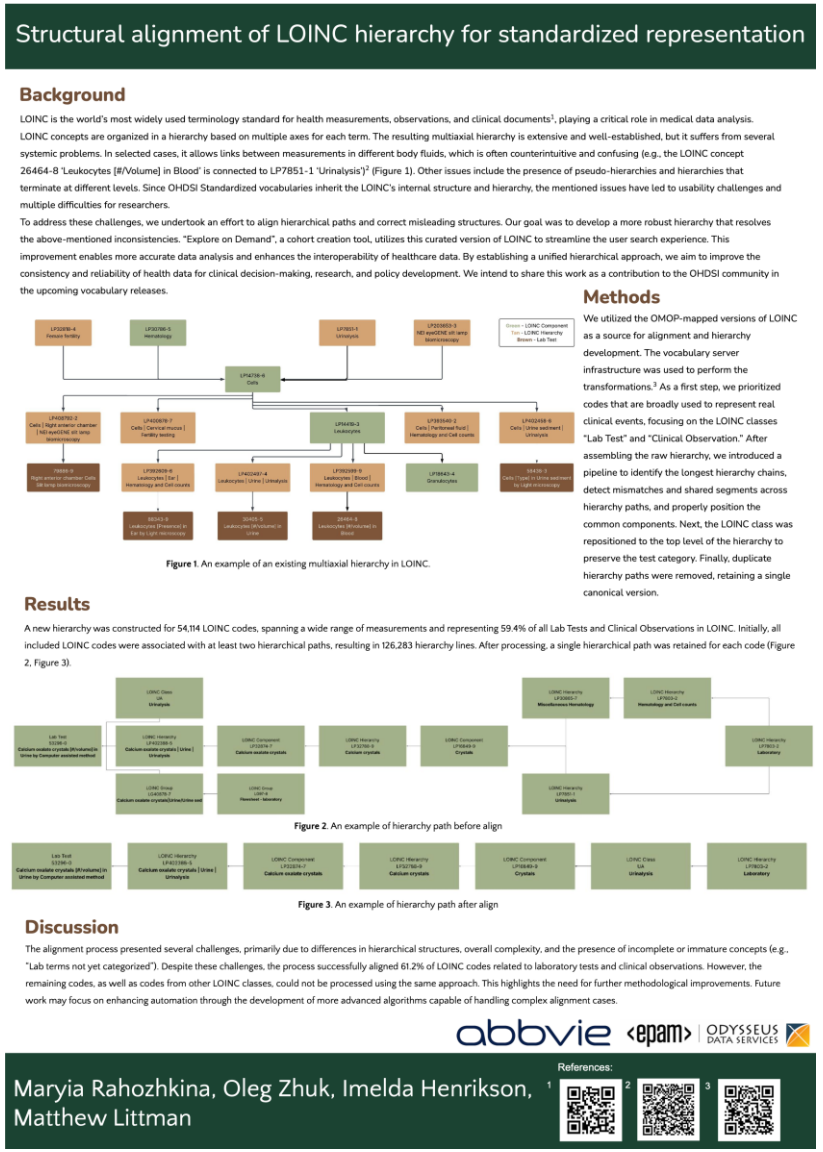
<epam> | ODYSSEUS DATA SERVICES

#OHDSISocialShowcase This Week

Wednesday

Structural alignment of LOINC hierarchy for standardized representation

(Maryia Rahozhkina, Oleg Zhuk, Imelda Henrikson, Matthew Littman)



Structural alignment of LOINC hierarchy for standardized representation

Background

LOINC is the world's most widely used terminology standard for health measurements, observations, and clinical documents[1], playing a critical role in medical data analysis. LOINC concepts are organized in a hierarchy based on multiple axes for each term. The resulting multiaxial hierarchy is extensive and well-established, but it suffers from several systemic problems. In selected cases, it allows links between measurements in different body fluids, which is often counterintuitive and confusing (e.g., the LOINC concept 26464-8 'Leukocytes [#/Volume] in Blood' is connected to LP7851-1 'Urinalysis')[2] (Figure 1). Other issues include the presence of pseudo-hierarchies and hierarchies that terminate at different levels. Since OHDSI Standardized vocabularies inherit the LOINC's internal structure and hierarchy, the mentioned issues have led to usability challenges and multiple difficulties for researchers.

To address these challenges, we undertook an effort to align hierarchical paths and correct misleading structures. Our goal was to develop a more robust hierarchy that resolves the above-mentioned inconsistencies. "Explore on Demand", a cohort creation tool, utilizes this curated version of LOINC to streamline the user search experience. This improvement enables more accurate data analysis and enhances the interoperability of healthcare data. By establishing a unified hierarchical approach, we aim to improve the consistency and reliability of health data for clinical decision-making, research, and policy development. We intend to share this work as a contribution to the OHDSI community in the upcoming vocabulary releases.

Methods

We utilized the OMOP-mapped versions of LOINC as a source for alignment and hierarchy development. The vocabulary server infrastructure was used to perform the transformations.[3] As a first step, we prioritized codes that are broadly used to represent real clinical events, focusing on the LOINC classes "Lab Test" and "Clinical Observation." After assembling the raw hierarchy, we introduced a pipeline to identify the longest hierarchy chains, detect mismatches and shared segments across hierarchy paths, and properly position the common components. Next, the LOINC class was repositioned to the top level of the hierarchy to preserve the test category. Finally, duplicate hierarchy paths were removed, retaining a single canonical version.

Figure 1. An example of an existing multiaxial hierarchy in LOINC.

Results

A new hierarchy was constructed for 54,114 LOINC codes, spanning a wide range of measurements and representing 59.4% of all Lab Tests and Clinical Observations in LOINC. Initially, all included LOINC codes were associated with at least two hierarchical paths, resulting in 126,283 hierarchy lines. After processing, a single hierarchical path was retained for each code (Figure 2, Figure 3).

Figure 2. An example of hierarchy path before align

Figure 3. An example of hierarchy path after align

Discussion

The alignment process presented several challenges, primarily due to differences in hierarchical structures, overall complexity, and the presence of incomplete or immature concepts (e.g., "Lab terms not yet categorized"). Despite these challenges, the process successfully aligned 61.2% of LOINC codes related to laboratory tests and clinical observations. However, the remaining codes, as well as codes from other LOINC classes, could not be processed using the same approach. This highlights the need for further methodological improvements. Future work may focus on enhancing automation through the development of more advanced algorithms capable of handling complex alignment cases.

abbvie epam ODYSSEUS DATA SERVICES

Maryia Rahozhkina, Oleg Zhuk, Imelda Henrikson, Matthew Littman

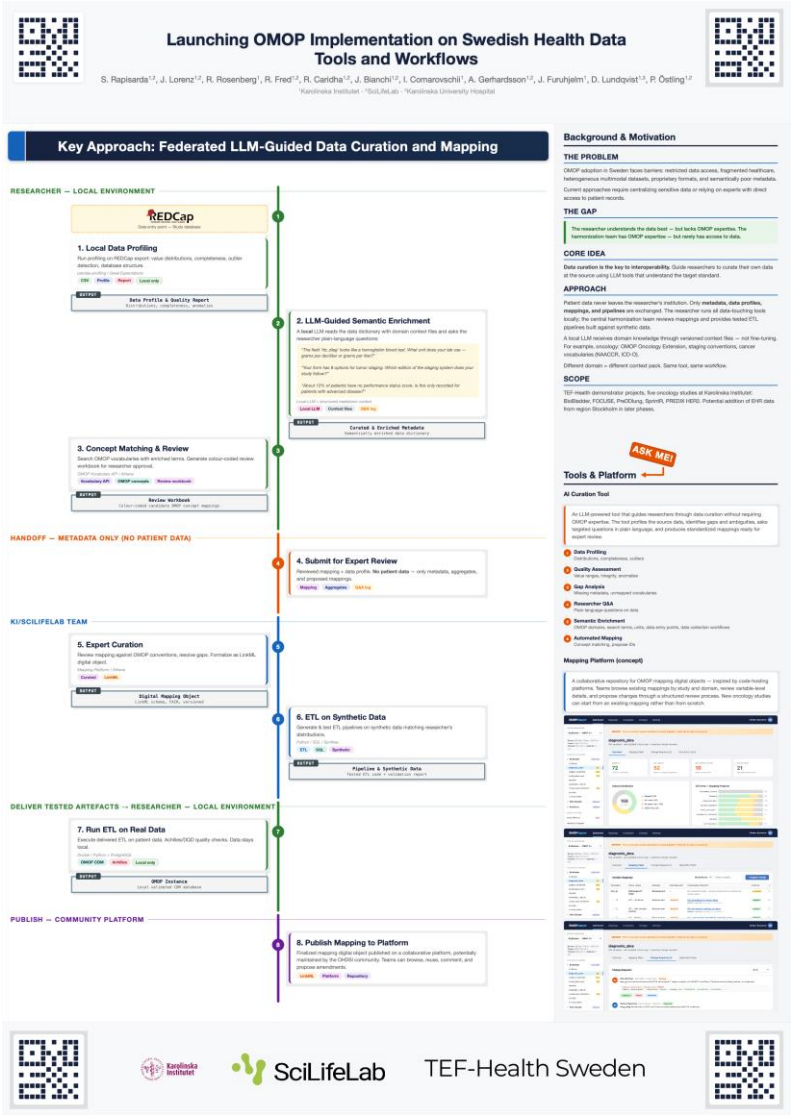
References:

#OHDSISocialShowcase This Week

Thursday

Launching OMOP Implementation on Swedish Health data – Tools and Workflows

(Stefano Rapisarda, Jan Lorenz, Rebecka Rosenberg, Rikard Fred, Rozina Caridha, Julie Bianchi, Igori Comarovschii, Andreas Gerhardsson, Johanna Furuhjelm, Daniel Lundqvist, Päivi Östling)

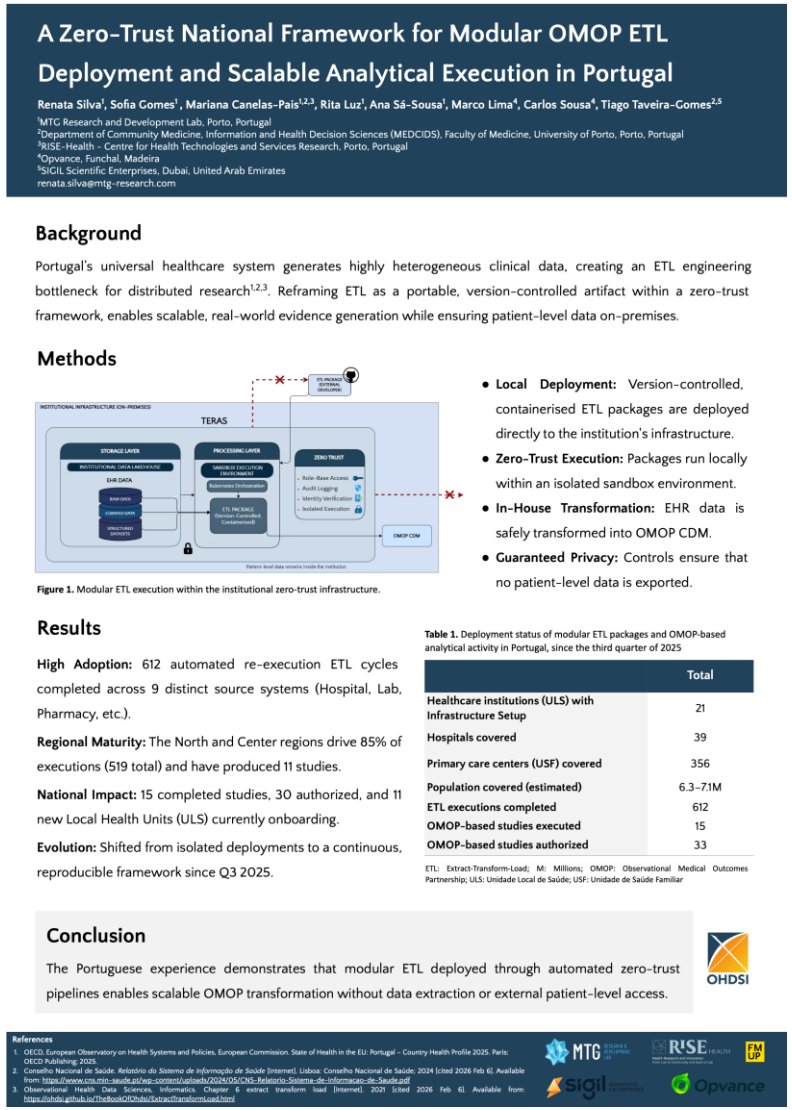


#OHDSISocialShowcase This Week

Friday

A Zero-Trust National Framework for Modular OMOP ETL Deployment and Scalable Analytical Execution in Portugal

(Renata Silva, Sofia Gomes, Mariana Canelas-Pais, Rita Luz, Ana Sá-Sousa, Marco Lima, Carlos Sousa, Tiago Taveira-Gomes)



A Zero-Trust National Framework for Modular OMOP ETL Deployment and Scalable Analytical Execution in Portugal

Renata Silva[1], Sofia Gomes[1], Mariana Canelas-Pais[1,2,3], Rita Luz[1], Ana Sá-Sousa[1], Marco Lima[4], Carlos Sousa[4], Tiago Taveira-Gomes[2,5]
[1]MTG Research and Development Lab, Porto, Portugal
[2]Department of Community Medicine, Information and Health Decision Sciences (MEDCIDS), Faculty of Medicine, University of Porto, Porto, Portugal
[3]RISE-Health – Centre for Health Technologies and Services Research, Porto, Portugal
[4]Opvance, Funchal, Madeira
[5]SIGIL Scientific Enterprises, Dubai, United Arab Emirates
renata.silva@mtg-research.com

Background

Portugal's universal healthcare system generates highly heterogeneous clinical data, creating an ETL engineering bottleneck for distributed research[1,2,3]. Reframing ETL as a portable, version-controlled artifact within a zero-trust framework, enables scalable, real-world evidence generation while ensuring patient-level data on-premises.

Methods

Figure 1. Modular ETL execution within the institutional zero-trust infrastructure.

- **Local Deployment:** Version-controlled, containerised ETL packages are deployed directly to the institution's infrastructure.
- **Zero-Trust Execution:** Packages run locally within an isolated sandbox environment.
- **In-House Transformation:** EHR data is safely transformed into OMOP CDM.
- **Guaranteed Privacy:** Controls ensure that no patient-level data is exported.

Results

High Adoption: 612 automated re-execution ETL cycles completed across 9 distinct source systems (Hospital, Lab, Pharmacy, etc.).

Regional Maturity: The North and Center regions drive 85% of executions (519 total) and have produced 11 studies.

National Impact: 15 completed studies, 30 authorized, and 11 new Local Health Units (ULS) currently onboarding.

Evolution: Shifted from isolated deployments to a continuous, reproducible framework since Q3 2025.

Table 1. Deployment status of modular ETL packages and OMOP-based analytical activity in Portugal, since the third quarter of 2025

	Total
Healthcare institutions (ULS) with Infrastructure Setup	21
Hospitals covered	39
Primary care centers (USF) covered	356
Population covered (estimated)	6.3–7.1M
ETL executions completed	612
OMOP-based studies executed	15
OMOP-based studies authorized	33

ETL: Extract-Transform-Load; M: Millions; OMOP: Observational Medical Outcomes Partnership; ULS: Unidade Local de Saúde; USF: Unidade de Saúde Familiar

Conclusion

The Portuguese experience demonstrates that modular ETL deployed through automated zero-trust pipelines enables scalable OMOP transformation without data extraction or external patient-level access.

References
1. OECD, European Observatory on Health Systems and Policies, European Commission. State of Health in the EU: Portugal – Country Health Profile 2025. Paris: OECD Publishing; 2025.
2. Conselho Nacional de Saúde. *Relatório do Sistema de Informação de Saúde* [Internet]. Lisboa: Conselho Nacional de Saúde; 2024 [cited 2026 Feb 6]. Available from: https://www.cns.min-saude.pt/wp-content/uploads/2024/05/CNS-Relatorio-Sistema-de-Informacao-de-Saude.pdf
3. Observational Health Data Sciences, Informatics. Chapter 6 extract transform load [Internet]. 2021 [cited 2026 Feb 6]. Available from: https://ohdsi.github.io/TheBookOfOhdsi/ExtractTransformLoad.html

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?

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

Find your workgroup.

Fuel our mission.

ohdsi.org/workgroups