Preview of ATLAS 3.0

OHDSI Community Call
July 7, 2026 • 11 am ET

Upcoming Community Calls

Date	Topic
July 7	Preview of ATLAS 3.0
July 14	Workgroup Spotlight: Health Systems, NLP
July 21	Workgroup Spotlight: TBA
July 28	OHDSI Newcomer Introductions

Upcoming Europe Calls

Date	Topic
July 9	Zoom in on the Oh!-Factor
Aug. 13	CANCELLED: Summer Break
Sept. 10	National Node Focus: Germany
Oct. 8	National Node Focus: Italy
Nov. 10	National Node Focus: Norway
Dec. 10	National Node Focus: Portugal

Three Stages of The Journey

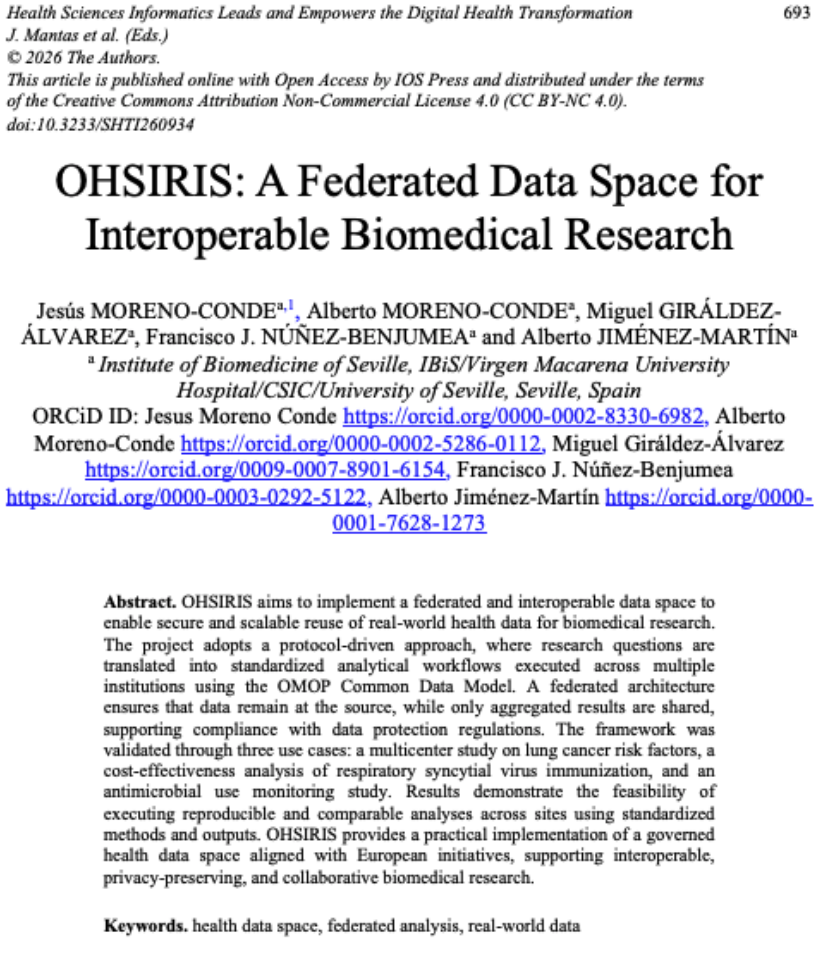
Where Have We Been?

Where Are We Now?

Where Are We Going?

OHDSI Shoutouts!

Congratulations to the team of **Jesús Moreno-Conde, Alberto Moreno-Conde, Miguel Giráldez-Álvarez, Francisco J. Núñez-Benjumea, and Alberto Jiménez-Martín** on the recent publication of **OHSIRIS: A Federated Data Space for Interoperable Biomedical Research** in *Volume 338 of Studies in Health Technology and Informatics: Health Sciences Informatics Leads and Empowers the Digital Health Transformation*.



Health Sciences Informatics Leads and Empowers the Digital Health Transformation 693
J. Mantas et al. (Eds.)

doi:10.3233/SHTI260934

OHSIRIS: A Federated Data Space for Interoperable Biomedical Research

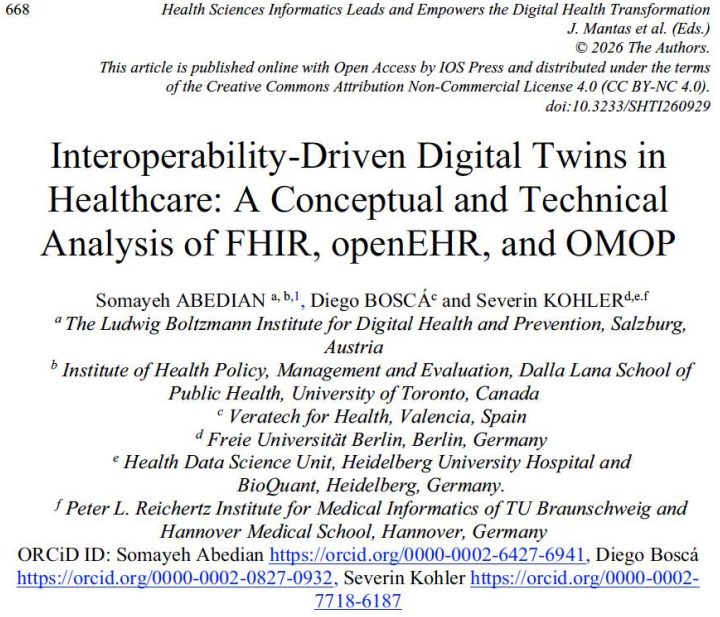
Jesús MORENO-CONDE[a,1], Alberto MORENO-CONDE[a], Miguel GIRÁLDEZ-ÁLVAREZ[a], Francisco J. NÚÑEZ-BENJUMEA[a] and Alberto JIMÉNEZ-MARTÍN[a]
[a] *Institute of Biomedicine of Seville, IBiS/Virgen Macarena University Hospital/CSIC/University of Seville, Seville, Spain*
ORCiD ID: Jesus Moreno Conde https://orcid.org/0000-0002-8330-6982, Alberto Moreno-Conde https://orcid.org/0000-0002-5286-0112, Miguel Giráldez-Álvarez https://orcid.org/0009-0007-8901-6154, Francisco J. Núñez-Benjumea https://orcid.org/0000-0003-0292-5122, Alberto Jiménez-Martín https://orcid.org/0000-0001-7628-1273

Abstract. OHSIRIS aims to implement a federated and interoperable data space to enable secure and scalable reuse of real-world health data for biomedical research. The project adopts a protocol-driven approach, where research questions are translated into standardized analytical workflows executed across multiple institutions using the OMOP Common Data Model. A federated architecture ensures that data remain at the source, while only aggregated results are shared, supporting compliance with data protection regulations. The framework was validated through three use cases: a multicenter study on lung cancer risk factors, a cost-effectiveness analysis of respiratory syncytial virus immunization, and an antimicrobial use monitoring study. Results demonstrate the feasibility of executing reproducible and comparable analyses across sites using standardized methods and outputs. OHSIRIS provides a practical implementation of a governed health data space aligned with European initiatives, supporting interoperable, privacy-preserving, and collaborative biomedical research.

Keywords. health data space, federated analysis, real-world data

OHDSI Shoutouts!

Congratulations to the team of **Somayeh Abedian, Diego Boscá, and Severin Kohler** on the recent publication of **Interoperability-Driven Digital Twins in Healthcare: A Conceptual and Technical Analysis of FHIR, openEHR, and OMOP** in *Volume 338 of Studies in Health Technology and Informatics: Health Sciences Informatics Leads and Empowers the Digital Health Transformation*.



doi:10.3233/SHTI260929

Interoperability-Driven Digital Twins in Healthcare: A Conceptual and Technical Analysis of FHIR, openEHR, and OMOP

Somayeh ABEDIAN [a,b,1], Diego BOSCÁ[c] and Severin KOHLER[d,e,f]
[a] *The Ludwig Boltzmann Institute for Digital Health and Prevention, Salzburg, Austria*
[b] *Institute of Health Policy, Management and Evaluation, Dalla Lana School of Public Health, University of Toronto, Canada*
[c] *Veratech for Health, Valencia, Spain*
[d] *Freie Universität Berlin, Berlin, Germany*
[e] *Health Data Science Unit, Heidelberg University Hospital and BioQuant, Heidelberg, Germany.*
[f] *Peter L. Reichertz Institute for Medical Informatics of TU Braunschweig and Hannover Medical School, Hannover, Germany*
ORCiD ID: Somayeh Abedian https://orcid.org/0000-0002-6427-6941, Diego Boscá https://orcid.org/0000-0002-0827-0932, Severin Kohler https://orcid.org/0000-0002-7718-6187

Abstract. Digital Twins (DTs) are gaining attention as a promising approach for next-generation healthcare systems. However, their real-world adoption remains limited, mainly due to fragmented data environments and a lack of effective interoperability across clinical systems. Although much of the current research focuses on modeling and advanced analytics, the role of interoperability as a structural enabler has received less attention. This paper examines interoperability as a core requirement for healthcare DTs. It considers how three commonly used standards, HL7 FHIR, openEHR, and the OHDSI OMOP Common Data Model, support different needs across the DT lifecycle, including data exchange, semantic representation, and data reuse for analysis. Each of these standards is effective in specific contexts. OMOP, for instance, is widely used for cohort studies, openEHR supports structured longitudinal records, and FHIR is effective for system integration and data exchange. However, developing more complete DT solutions requires coordinating these complementary capabilities. The paper takes a lifecycle- and role-based view, suggesting that DTs are better understood as the result of coordinated data infrastructures rather than standalone systems. This lifecycle framework provides actionable guidance for scalable DT architectures.

Keywords. Digital Twins, Healthcare Interoperability Standards, HL7 FHIR, openEHR, OMOP

OHDSI Shoutouts!

Congratulations to the team of **Falk Meyer-Eschenbach, Martin Vogel, Philipp Jacob, Severin Kohler, Louis Agha-Mir-Salim, Loreen Ruhm, Christof von Kalle, Elias Grünewald, Oliver Kumpf, and Felix Balzer** on the recent publication of **LLM-Assisted Clinical Data Harmonization: Combining Automated ETL Generation with Semantic Vocabulary Mapping for OMOP CDM** in *Volume 338 of Studies in Health Technology and Informatics: Health Sciences Informatics Leads and Empowers the Digital Health Transformation*.

Health Sciences Informatics Leads and Empowers the Digital Health Transformation 453
J. Mantas et al. (Eds.)

doi:10.3233/SHTI260883

LLM-Assisted Clinical Data Harmonization: Combining Automated ETL Generation with Semantic Vocabulary Mapping for OMOP CDM

Falk MEYER-ESCHENBACH[a,b,c,1,2*] and Martin VOGEL[c,2], Philipp JACOB[c], Severin KOHLER[e], Louis AGHA-MIR-SALIM[a], Loreen RUHM[b], Christof VON KALLE[b], Elias GRÜNEWALD[a], Oliver KUMPF[d] and Felix BALZER[a,c]
[a] *Institute of Medical Informatics, Charité - Universitätsmedizin Berlin, Berlin, Germany*
[b] *Clinical Study Center, Berlin Institute of Health, Charité, Germany*
[c] *Institute of Informatics, Humboldt-Universität zu Berlin, Berlin, Germany*
[d] *Department of Anesthesiology and Intensive Care Medicine, Charité, Germany*
[e] *Institute of Informatics, Freie Universität Berlin, Berlin, Germany*
ORCiD ID: *Falk Meyer-Eschenbach https://orcid.org/0009-0000-9813-0249, Martin Vogel https://orcid.org/0009-0003-5148-2900*

Abstract. Transforming clinical data into the OMOP Common Data Model requires structural schema mapping and semantic vocabulary harmonization, demanding both technical and clinical expertise. We investigated Large Language Models for both tasks using the eICU Collaborative Research Database. For structural transformation, context-enriched prompting with White Rabbit profiling reports guided Gemini 2.5 Pro to generate PostgreSQL Extract-Transform-Load (ETL) scripts, populating eight OMOP core tables after five debugging iterations. For semantic mapping, a hybrid approach combining vector similarity search with LLM refinement targeted Logical Observation Identifiers Names and Codes (LOINC) and Anatomical Therapeutic Chemical (ATC) codes, validated by clinical experts from Charité – Universitätsmedizin Berlin. Our approach achieved 93.9% precision for medications (n=148) and 78.5-96.8% for laboratory terms (n=158) depending on consensus criteria. LLM refinement raised medication precision from 60.8% to 93.9% and laboratory precision from 62.0–88.6% to 78.5–96.8%. Together, these phases reduce manual effort in OMOP transformation across three clinical domains, while expert supervision remains essential.

Keywords. OMOP Common Data Model, Large Language Models, Clinical Data Harmonization, Vocabulary Mapping, ETL

OHDSI Shoutouts!

Daniel Prieto-Alhambra announced as the first Oxford Professor of Regulatory Health Data Science at NDORMS

Share Share Share Share

Botnar Main

2 July 2026

The new professorship strengthens the department's leadership in health data science and real-world evidence and will specifically support regulatory frameworks and decision making to bring new safe, efficient and effective treatments to benefit patients and society.

FIND OUT MORE ABOUT

Our **Groups, Centres and Units**

Our **Research Themes**

Our **Clinical Trials**

Our **Programme Grants**

SIMILAR STORIES

NDORMS team wins Award for Excellence for transforming ulcerative colitis care

Mark Coles receives funding to tackle viral threats

Natural compound may help boost vaccine responses in older adults

Three Stages of The Journey

Where Have We Been?

Where Are We Now?

Where Are We Going?

Upcoming Workgroup Calls

Date	Time (ET)	Meeting
Wednesday	7 am	Medical Imaging
Wednesday	8 am	Psychiatry
Wednesday	9 am	Patient-Level Prediction
Wednesday	10 am	Common Data Model
Wednesday	2 pm	Canada Chapter
Wednesday	2 pm	Natural Language Processing
Wednesday	7 pm	Eyecare and Vision Research
Thursday	7 am	Europe Community Call
Thursday	10 am	Rare Diseases
Thursday	10 am	ATLAS/WebAPI
Thursday	10 am	GIS-Geographic Information System
Thursday	11:30 am	Surgery and Perioperative Medicine
Friday	9 am	Waveform
Friday	10 am	Transplant
Friday	11 am	Clinical Trials
Friday	11:30 am	Steering
Friday	11 pm	China Chapter
Tuesday	9 am	Oncology Genomic Subgroup
Tuesday	9 am	Data2Evidence

Clair Blacketer PhD Defense

Erasmus University Rotterdam

EN | NL Search

Education Research Engagement About the university Working at Campus Contact

PhD defence M.S. (Clair) Blacketer

Strengthening the Foundations of Real-World Evidence

On Wednesday 8 Juli 2026, M.S. Blacketer will defend the doctoral thesis titled: Strengthening the Foundations of Real-World Evidence

Erasmus MC

Promotor	Prof. dr. ir. P.R. Rijnbeek Erasmus MC
Co-promotor	Dr. M.J. Schuemie Erasmus MC
Date	Wednesday 8 Jul 2026, 13:00 - 14:30

July Newsletter is Available

OHDSI

The Journey Newsletter (July 2026)

Welcome to the July edition of the OHDSI newsletter, where we dive into the current state of generative AI research in our community and the cutting-edge tools currently in development. This issue is packed with essential updates, including our latest June publications and inspiring community spotlights on Melanie Philofsky and Sulev Reisberg. We also look ahead to the first LATAM Symposium and announce the dates for the APAC Symposium. Finally, anticipation is building for the upcoming Global Symposium (Oct. 20-22), where a unique Jam Session promises to bring new, fun collaboration opportunities to this year's event. #JoinTheJourney

Podcast: LLM Research, LATAM Growth

In the July 2026 On The Journey podcast, Patrick Ryan and Craig Sachson reflect on LLM research in the OHDSI community. They begin with reflections on Martijn Schuemie's overview of current research, challenges and opportunities in OHDSI, and then they look at some of the individual tool presentations that were shared throughout the month. They also look forward to the first LATAM Symposium and the importance of building new collaborations in South America. *(If video does not appear, please click 'view this email in your browser.)*

Community Updates

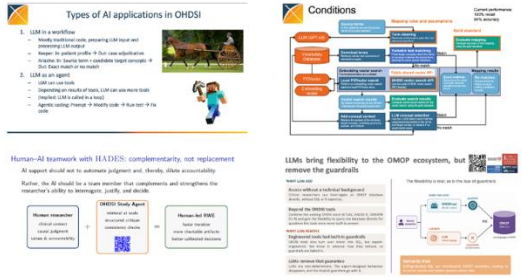
Where Have We Been?

- **LLM Research Presentations:** Our community spent June diving deep into the intersection of generative AI and observational research, kicking off the month with a comprehensive high-level overview of how Large Language Models (LLMs) are transforming our ecosystem and then showcasing several tools being developed in the community. More information and links are available later in this newsletter.
- **OHDSI Summer School at Columbia University:** The second-annual OHDSI Summer School in New York City was held in late June, and an international cohort of learners traveled the path from learning about observational research to performing their own network studies within one week.
- **Oxford Summer School in Real World Evidence:** Oxford University also welcomed a cohort of learners to discovery the possibilities using real-world data, OMOP, and the OHDSI way of generating reliable and robust real-world evidence. Thank you to the faculty at both Columbia and Oxford for welcoming new collaborators to our community.

Where Are We Now?

- **Agentic AI Call for Papers:** A new call for papers is open for a *Journal of Biomedical Informatics* special issue focusing on agentic AI for generating trustworthy Real-World Evidence (RWE). The issue seeks contributions on multi-agent collaboration and coordinated autonomy across the RWE lifecycle—such as data harmonization and causal analysis—while excluding single-task AI agents. Submissions are accepted from June 15 through November 15, 2026, with publication slated for September 2027.
- **OHDSI Newcomer Introductions:** The annual "Welcome OHDSI Newcomers" community call will be held July 28. 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. For scheduling purposes, please note your interest in taking part in this brief survey.
- **First LATAM Symposium:** Registration is still open for the first OHDSI Latin America Symposium, which will be held July 30-31 in Salvador, Brazil. The landmark event features a robust lineup of workshops covering an Introduction to OMOP, cohort building, ETL development, and opportunities for regional scientific collaboration and network studies.

The AI Wave in OHDSI: Accelerating LLM Innovation Through Community Collaboration



Our community spent June diving deep into the intersection of generative AI and observational research, kicking off the month with a comprehensive high-level overview of how Large Language Models (LLMs) are transforming our ecosystem (see video link below). Throughout the subsequent presentations and community discussions, several core themes emerged, showcasing how these advanced models can streamline the entire data lifecycle. From automating complex cohort definition and accelerating phenotype development to simplifying data harmonization and ETL pipeline design, the presentations highlighted a powerful shift toward embedding intelligent automation directly into the OMOP Common Data Model architecture.

The incredible variety of individual breakthroughs presented this month shows just how energized our network is by the potential of healthcare AI. To fully unlock the potential of these foundational models, our next great opportunity lies in bringing these exciting, innovative efforts together into unified, network-wide collaborations. By standardizing how LLMs interface with OHDSI vocabularies and ensuring these tools are reproducible across all data environments, we can scale these innovations globally through the same open, collective effort that has always driven our community forward.

It is important to remember that the insights and innovations spotlighted this past month represent just a small fraction of the AI momentum building across our global network. The upcoming OHDSI Global Symposium in October will feature a massive expansion of this research, diving far deeper into validated real-world evidence generation and the future of healthcare AI. If you want to be part of shaping the official roadmap for AI integration in observational health, make sure your registration is secured for New Brunswick!

LLM Research Presentations

Video: Overview of LLM Research in OHDSI

My Journey: Melanie Philofsky

In the latest installment of our "My Journey" series, Melanie Philofsky (Director for Clinical Informatics at Odysseus Data Services, an EPAM company) shares her profound transition from ICU nurse to clinical informaticist. Melanie discusses how her background at the bedside fuels her passion for transforming EHR data into the OMOP CDM format, leading crucial OHDSI working groups, and scaling up real-world evidence to impact tens of thousands of patient lives. *(If video does not appear, please click 'view this email in your browser.)*

Symposia Season Welcomes First LATAM Event, APAC Symposium in November

The second half of the year annually brings together an international collection of OHDSI Symposia, and for the first time, that will begin in South America. The inaugural OHDSI LATAM Symposium will be held July 30-31 in Salvador, Brazil, and it will bring together the public sector, industry and academia to advance health data standardization and strengthen an ethical, open and inclusive regional community. The symposium will work to accelerate OMOP Common Data Model adoption in Brazil and across the region, fostering integration, scientific collaboration and social impact.

July Newsletter is Available

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

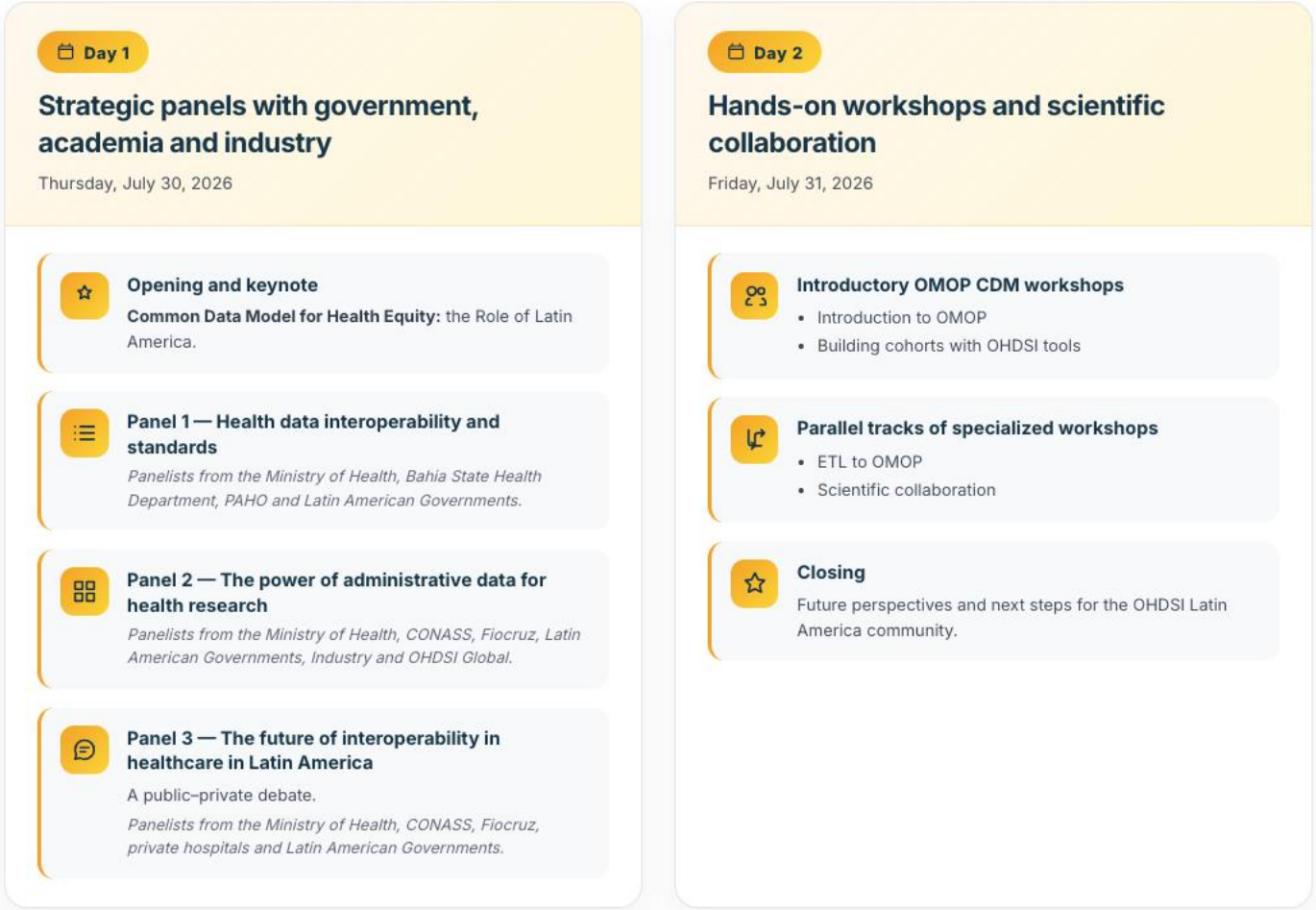
Whether you are an amateur hobbyist, a seasoned pro, an instrumentalist, or a vocalist, our very first OHDSI Jam Session is open to everyone. We'll do some free-form improvisation, but we will also vote on a few songs ahead of time so everyone has a chance to practice.

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!

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.



ohdsilatam.org

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

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 exporer, 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.framas 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

Standardizing the "Repositório Integrado de Conhecimento" at IPO Porto: Design and Implementation of an OMOP CDM ETL for Oncology Data

(Mariana Canelas-Pais, Renata Silva, Sofia Gomes, Tiago Taveira-Gomes, Rita Rb-Silva, Maria, José Bento, Teresa Garcia)

Standardizing the "Repositório Integrado de Conhecimento" at IPO Porto: Design and Implementation of an OMOP CDM ETL for Oncology Data

Mariana Canelas-Pais[1,2,3], Renata Silva[1], Sofia Gomes[1], Tiago Taveira-Gomes[2,4], Rita Rb-Silva[5], Maria José Bento[5,6,7], Teresa Garcia[5,6]

[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]SIGIL Scientific Enterprises, Dubai, United Arab Emirates
[5]Group of Epidemiology, Outcomes, Economics and Management in Oncology – Research Center, Porto Comprehensive Cancer Center (Porto.CCC) & RISE@CI-IPOP (Health Research Network), Portuguese Oncology Institute of Porto (IPO Porto), Porto, Portugal
[6]Department of Epidemiology, Portuguese Oncology Institute of Porto (IPO Porto), Porto, Portugal
[7]Population Studies Department. School of Medicine and Biomedical Sciences, ICBAS, University of Porto, Porto, Portugal.
mariana.pais@mtg-research.com

Background

IPO Porto is a national reference center for oncology care in Portugal. Its clinical registry, Repositório Integrado de Conhecimento (RIC), consolidates oncology data through a mix of automated ingestion and extensive manual entry by clinical experts. Despite its quality, coding and structural heterogeneity limit interoperability and secondary use[1]. This work presents the OMOP-CDM ETL design and planned next steps for extension[2].

Methods

Results

The first version cover the **PERSON**, **VISIT_OCCURRENCE**, **CONDITION_OCCURRENCE**, **MEASUREMENT**, **DEATH**, **CDM_SOURCE**, **OBSERVATION_PERIOD** domains. The ETL design decisions included:

- **ICD-O-3 Mapping:** Histology, behaviour, and topography concatenated for CONDITION_OCCURRENCE mapping.
- **Wide-to-Long Transformation:** Tumor characteristics were transformed into multiple longitudinal records across MEASUREMENT and CONDITION_OCCURRENCE.
- **Cancer Modifiers:** Staging and tumor attributes mapped to OMOP Cancer Modifier concepts in MEASUREMENT.
- **Temporal Proxies:** Tumor diagnosis date used as proxy when comorbidity onset dates were unavailable.

Figure 1. ETL design specification for the RIC.CaracteristicaTumorView source, detailing the transformation logic for mapping oncological data, TNM and ICD-O-3 classifications, into the OMOP Measurement and Condition_Occurrence tables.

Conclusion

This work demonstrates the feasibility of transforming a curated oncology registry into OMOP CDM through a formally specified ETL process. Future iterations will expand the analytical scope by capturing systemic therapies, surgical procedures, and critical clinical performance indicators like ECOG

References
1. Ajmal A, Boussou O, Brash J, Cheeseman S, Banduge PG, Gomes AL, et al. Establishing standards: harmonising coding principles for a minimal cancer dataset in the OMOP Common Data Model. ESMO Real World Data and Digital Oncology [Internet]. 2025 Sept;3(100079):100079. Available from: https://dx.doi.org/10.1016/j.esmorw.2025.100079
2. Observational Health Data Sciences. Informatics. The Book of OHDSI [Internet]. 2021 [cited 2026 Feb 2]. Available from: https://ohdsi.github.io/TheBookOfOhdsi/

#OHDSISocialShowcase This Week

Tuesday

Standardising feasibility assessments: a proposal from the IHI-GREG project

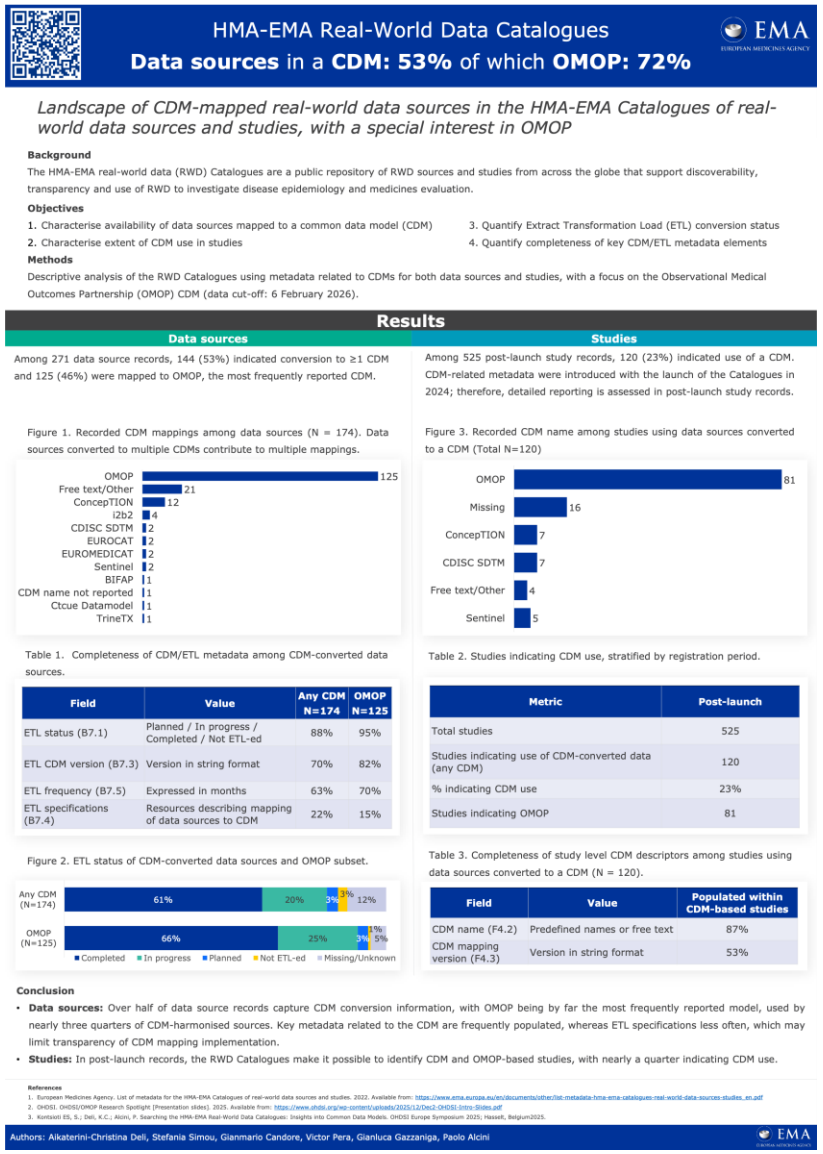
(Lucía A. Carrasco-Ribelles, Lucia Bellas, Annika Jödicke, Daniel Prieto-Alhambra)

#OHDSISocialShowcase This Week

Wednesday

The landscape of OMOP-mapped real-world data sources in the HMA-EMA Catalogues of real-world data sources and studies

(Aikaterini-Christina Deli, Stefania Simou, Gianmario Candore, Victor Pera, Gianluca Gazzaniga, Paolo Alcini)



HMA-EMA Real-World Data Catalogues

Data sources in a **CDM: 53%** of which **OMOP: 72%**

EMA

Landscape of CDM-mapped real-world data sources in the HMA-EMA Catalogues of real-world data sources and studies, with a special interest in OMOP

Background

The HMA-EMA real-world data (RWD) Catalogues are a public repository of RWD sources and studies from across the globe that support discoverability, transparency and use of RWD to investigate disease epidemiology and medicines evaluation.

Objectives

1. Characterise availability of data sources mapped to a common data model (CDM)
2. Characterise extent of CDM use in studies
3. Quantify Extract Transformation Load (ETL) conversion status
4. Quantify completeness of key CDM/ETL metadata elements

Methods

Descriptive analysis of the RWD Catalogues using metadata related to CDMs for both data sources and studies, with a focus on the Observational Medical Outcomes Partnership (OMOP) CDM (data cut-off: 6 February 2026).

Results

Data sources

Among 271 data source records, 144 (53%) indicated conversion to ≥1 CDM and 125 (46%) were mapped to OMOP, the most frequently reported CDM.

Figure 1. Recorded CDM mappings among data sources (N = 174). Data sources converted to multiple CDMs contribute to multiple mappings.

Table 1. Completeness of CDM/ETL metadata among CDM-converted data sources.

Field	Value	Any CDM N=174	OMOP N=125
ETL status (B7.1)	Planned / In progress / Completed / Not ETL-ed	88%	95%
ETL CDM version (B7.3)	Version in string format	70%	82%
ETL frequency (B7.5)	Expressed in months	63%	70%
ETL specifications (B7.4)	Resources describing mapping of data sources to CDM	22%	15%

Figure 2. ETL status of CDM-converted data sources and OMOP subset.

Studies

Among 525 post-launch study records, 120 (23%) indicated use of a CDM. CDM-related metadata were introduced with the launch of the Catalogues in 2024; therefore, detailed reporting is assessed in post-launch study records.

Figure 3. Recorded CDM name among studies using data sources converted to a CDM (Total N=120)

Table 2. Studies indicating CDM use, stratified by registration period.

Metric	Post-launch
Total studies	525
Studies indicating use of CDM-converted data (any CDM)	120
% indicating CDM use	23%
Studies indicating OMOP	81

Table 3. Completeness of study level CDM descriptors among studies using data sources converted to a CDM (N = 120).

Field	Value	Populated within CDM-based studies
CDM name (F4.2)	Predefined names or free text	87%
CDM mapping version (F4.3)	Version in string format	53%

Conclusion

- **Data sources:** Over half of data source records capture CDM conversion information, with OMOP being by far the most frequently reported model, used by nearly three quarters of CDM-harmonised sources. Key metadata related to the CDM are frequently populated, whereas ETL specifications less often, which may limit transparency of CDM mapping implementation.
- **Studies:** In post-launch records, the RWD Catalogues make it possible to identify CDM and OMOP-based studies, with nearly a quarter indicating CDM use.

References

1. European Medicines Agency. List of metadata for the HMA-EMA Catalogues of real-world data sources and studies. 2022. Available from: https://www.ema.europa.eu/en/documents/other/list-metadata-hma-ema-catalogues-real-world-data-sources-studies_en.pdf
2. OHDSI. OHDSI/OMOP Research Spotlight [Presentation slides]. 2025. Available from: https://www.ohdsi.org/wp-content/uploads/2025/12/Dec2-OHDSI-Intro-Slides.pdf
3. Kontsioti ES, S.; Deli, K.C.; Alcini, P. Searching the HMA-EMA Real-World Data Catalogues: Insights into Common Data Models. OHDSI Europe Symposium 2025; Hasselt, Belgium2025.

Authors: Aikaterini-Christina Deli, Stefania Simou, Gianmario Candore, Victor Pera, Gianluca Gazzaniga, Paolo Alcini

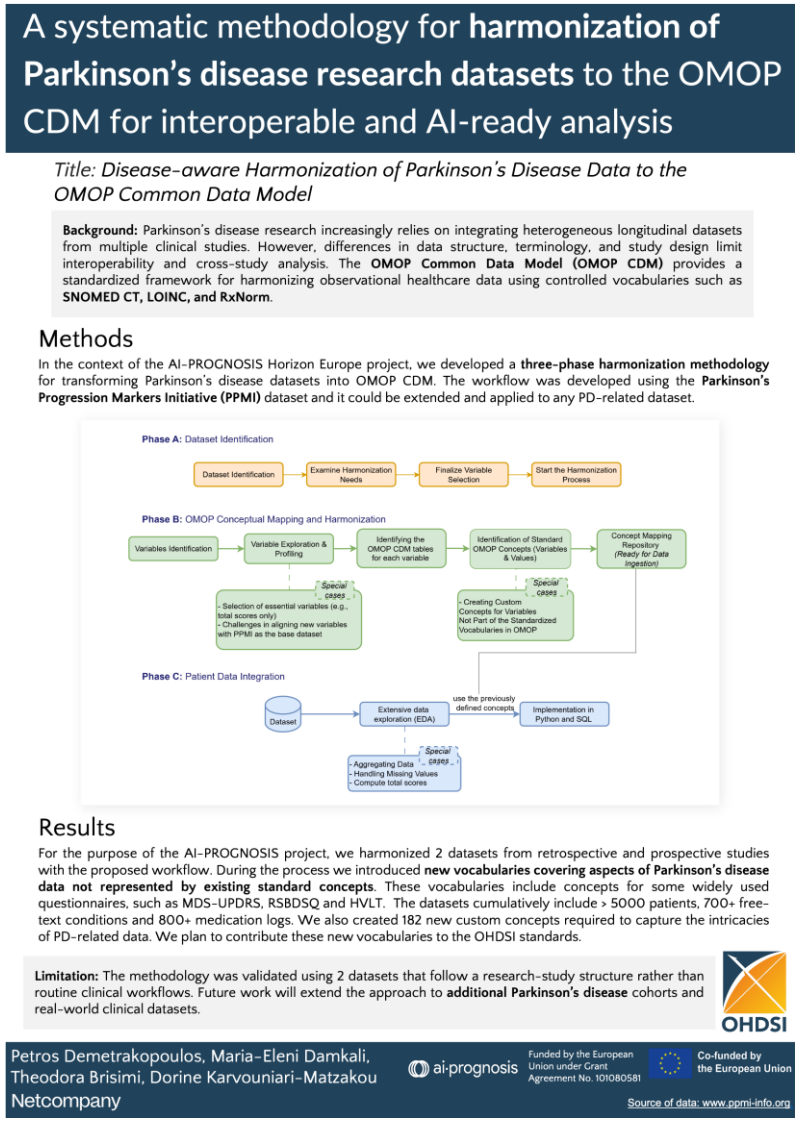
EMA

#OHDSISocialShowcase This Week

Thursday

Disease-aware Harmonization of Parkinson's Disease Research Data to the OMOP Common Data Model: A Systematic Methodology from AI-PROGNOSIS

(Petros Demetrakopoulos, Maria-Eleni Damkali, Theodora Brisimi, Dorine Karvouniari-Matzakou)

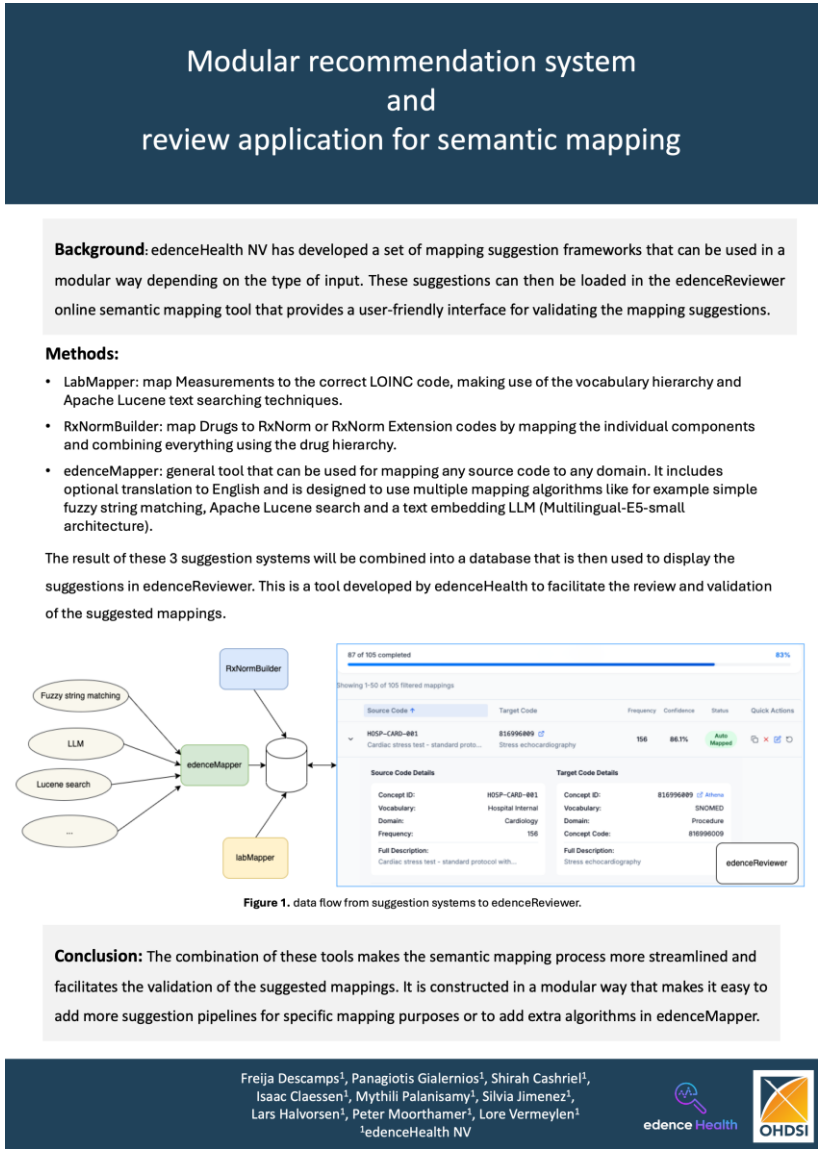


#OHDSISocialShowcase This Week

Friday

Modular recommendation system and review application for semantic mapping

(Freija Descamps, Panagiotis Gialernios, Shirah Cashriel, Isaac Claessen, Mythili Palanisamy, Silvia Jimenez, Lars Halvorsen, Peter Moorthamer, Lore Vermeylen)



Modular recommendation system
and
review application for semantic mapping

Background: edenceHealth NV has developed a set of mapping suggestion frameworks that can be used in a modular way depending on the type of input. These suggestions can then be loaded in the edenceReviewer online semantic mapping tool that provides a user-friendly interface for validating the mapping suggestions.

Methods:

- LabMapper: map Measurements to the correct LOINC code, making use of the vocabulary hierarchy and Apache Lucene text searching techniques.
- RxNormBuilder: map Drugs to RxNorm or RxNorm Extension codes by mapping the individual components and combining everything using the drug hierarchy.
- edenceMapper: general tool that can be used for mapping any source code to any domain. It includes optional translation to English and is designed to use multiple mapping algorithms like for example simple fuzzy string matching, Apache Lucene search and a text embedding LLM (Multilingual-E5-small architecture).

The result of these 3 suggestion systems will be combined into a database that is then used to display the suggestions in edenceReviewer. This is a tool developed by edenceHealth to facilitate the review and validation of the suggested mappings.

Figure 1. data flow from suggestion systems to edenceReviewer.

Conclusion: The combination of these tools makes the semantic mapping process more streamlined and facilitates the validation of the suggested mappings. It is constructed in a modular way that makes it easy to add more suggestion pipelines for specific mapping purposes or to add extra algorithms in edenceMapper.

Freija Descamps[1], Panagiotis Gialernios[1], Shirah Cashriel[1], Isaac Claessen[1], Mythili Palanisamy[1], Silvia Jimenez[1], Lars Halvorsen[1], Peter Moorthamer[1], Lore Vermeylen[1]
[1]edenceHealth NV

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