Asia-Pacific Regional Updates

OHDSI Community Call

Aug. 25, 2026 • 11 am ET

Upcoming Community Calls

Date	Topic
Aug. 25	Asia-Pacific Regional Updates
Sept. 1	Vocabulary Refresh, Summer 2026 Version
Sept. 8	Medical Imaging Workgroup Spotlight, CDM 5.5 Review
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 **Brenda Mbouamba Yankam, Fankoua Tchaptchet Luc Baudoin, François Anicet Onana Akoa, Pauline Andeso, Steve Bicko Cygu, Samuel Iddi, Michael Ochola, Tathagata Bhattacharjee, Mbele Onana Charles Lebon, Miranda Barasa, Agnes Kiragga, and Bertrand Hugo Mbatchou Ngahane** on the recent publication of **Harmonizing patient health data into the OMOP common data model: a case study using tuberculosis data from Douala General Hospital, Cameroon** in *Frontiers in Digital Health*.

frontiers in Digital Health

TYPE Original Research
PUBLISHED 03 August 2026
DOI 10.3389/fdgth.2026.1828475

OPEN ACCESS

EDITED BY
Zisis Kozlakidis,
International Agency For Research On Cancer (IARC), France

REVIEWED BY
Alladoumbaye Ngueilbaye,
Shenzhen University, China
Philip Quinlan,
University of Nottingham, United Kingdom

*CORRESPONDENCE
Brenda Mbouamba Yankam
brenda.yankammbouamba@ruhr-uni-bochum.de;
brenda.yankam@gmail.com

RECEIVED 11 March 2026
REVISED 04 May 2026
ACCEPTED 01 June 2026
PUBLISHED 03 August 2026

CITATION
Mbouamba Yankam B, Luc Baudoin FT, Onana Akoa FA, Andeso P, Cygu SB, Iddi S, Ochola M, Bhattacharjee T, Charles Lebon MO, Barasa M, Kiragga A and Mbatchou Ngahane BH (2026) Harmonizing patient health data into the OMOP common data model: a case study using tuberculosis data from Douala General Hospital, Cameroon. Front. Digit. Health 8:1828475. doi: 10.3389/fdgth.2026.1828475

Harmonizing patient health data into the OMOP common data model: a case study using tuberculosis data from Douala General Hospital, Cameroon

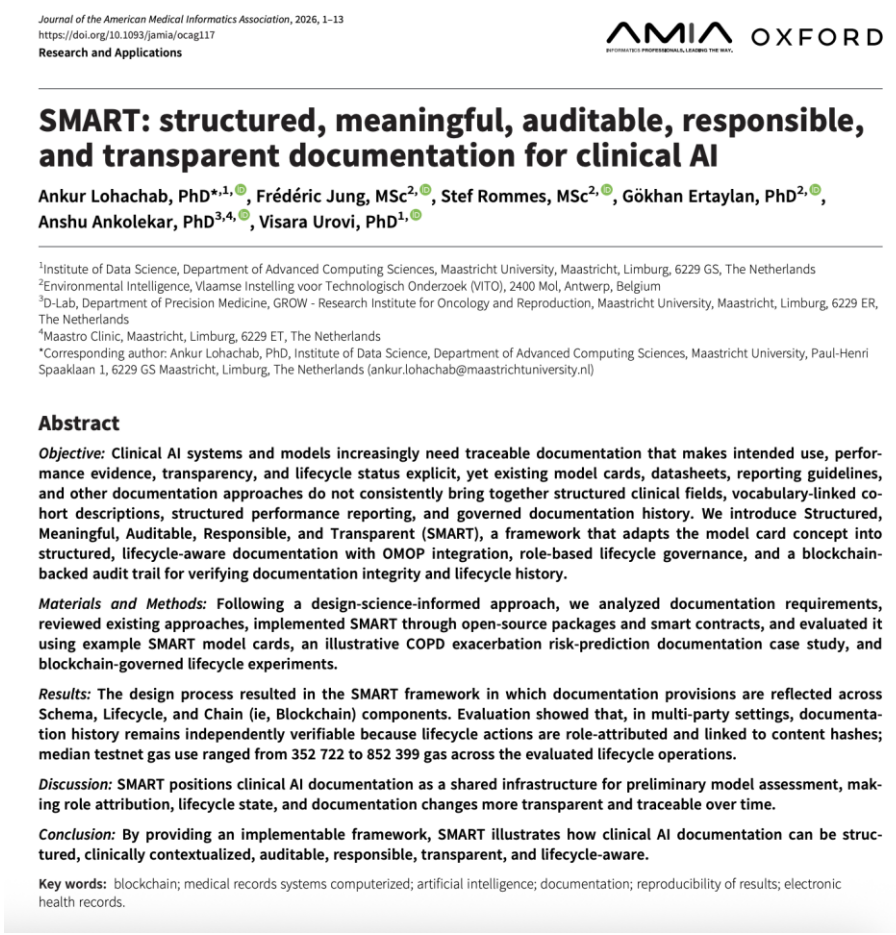
Brenda Mbouamba Yankam[1,2]*, Fankoua Tchaptchet Luc Baudoin[1], François Anicet Onana Akoa[1,3], Pauline Andeso[4], Steve Bicko Cygu[4], Samuel Iddi[4,5], Michael Ochola[4], Tathagata Bhattacharjee[6], Mbele Onana Charles Lebon[7], Miranda Barasa[4], Agnes Kiragga[4,8] and Bertrand Hugo Mbatchou Ngahane[1,9]

[1]Data Science Without Borders Project, Douala General Hospital, Douala, Cameroon, [2]Faculty of Mathematics, University of Bochum, Bochum, Germany, [3]School of Health Sciences, Catholic University of Central Africa, Yaounde, Cameroon, [4]African Population and Health Research Center (APHRC), Nairobi, Kenya, [5]Department of Statistics, University of Ghana, Accra, Ghana, [6]Department of Population Health, Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, University of London, London, United Kingdom, [7]Internal Medicine Department, Douala General Hospital, Douala, Cameroon, [8]Infectious Diseases Institute, College of Health Sciences, Makerere University, Kampala, Uganda, [9]Faculty of Medicine and Pharmaceutical Sciences, University of Douala, Douala, Cameroon

Introduction: Douala General Hospital is a tertiary healthcare facility in Cameroon that serves thousands of patients annually. Despite maintaining extensive patient records with potential for public health research, most data remain paper-based, limiting accessibility, reuse, and interoperability. In departments such as pulmonology, patient information is stored in heterogeneous, non-standardized formats, constraining its use for clinical research, care management, and decision-making. To address these challenges, we implemented a complete Extract, Transform, and Load (ETL) pipeline aligned with the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM) Version 5.4.
Objective: This study aimed to harmonise and integrate tuberculosis (TB) patient data into the OMOP CDM to improve data quality, interoperability, and reusability for research and clinical monitoring.

OHDSI Shoutouts!

Congratulations to the team of **Ankur Lohachab, Frédéric Jung, Stef Rommes, Gökhan Ertaylan, Anshu Ankolekar, and Visara Urovi** on the recent publication of **SMART: structured, meaningful, auditable, responsible, and transparent documentation for clinical AI** in *JAMIA*.



Journal of the American Medical Informatics Association, 2026, 1–13
https://doi.org/10.1093/jamia/ocag117
Research and Applications

AMIA OXFORD

SMART: structured, meaningful, auditable, responsible, and transparent documentation for clinical AI

Ankur Lohachab, PhD[*,1], Frédéric Jung, MSc[2], Stef Rommes, MSc[2], Gökhan Ertaylan, PhD[2], Anshu Ankolekar, PhD[3,4], Visara Urovi, PhD[1]

[1]Institute of Data Science, Department of Advanced Computing Sciences, Maastricht University, Maastricht, Limburg, 6229 GS, The Netherlands
[2]Environmental Intelligence, Vlaamse Instelling voor Technologisch Onderzoek (VITO), 2400 Mol, Antwerp, Belgium
[3]D-Lab, Department of Precision Medicine, GROW - Research Institute for Oncology and Reproduction, Maastricht University, Maastricht, Limburg, 6229 ER, The Netherlands
[4]Maastro Clinic, Maastricht, Limburg, 6229 ET, The Netherlands
*Corresponding author: Ankur Lohachab, PhD, Institute of Data Science, Department of Advanced Computing Sciences, Maastricht University, Paul-Henri Spaaklaan 1, 6229 GS Maastricht, Limburg, The Netherlands (ankur.lohachab@maastrichtuniversity.nl)

Abstract

Objective: **Clinical AI systems and models increasingly need traceable documentation that makes intended use, performance evidence, transparency, and lifecycle status explicit, yet existing model cards, datasheets, reporting guidelines, and other documentation approaches do not consistently bring together structured clinical fields, vocabulary-linked cohort descriptions, structured performance reporting, and governed documentation history. We introduce Structured, Meaningful, Auditable, Responsible, and Transparent (SMART), a framework that adapts the model card concept into structured, lifecycle-aware documentation with OMOP integration, role-based lifecycle governance, and a blockchain-backed audit trail for verifying documentation integrity and lifecycle history.**

Materials and Methods: **Following a design-science-informed approach, we analyzed documentation requirements, reviewed existing approaches, implemented SMART through open-source packages and smart contracts, and evaluated it using example SMART model cards, an illustrative COPD exacerbation risk-prediction documentation case study, and blockchain-governed lifecycle experiments.**

Results: **The design process resulted in the SMART framework in which documentation provisions are reflected across Schema, Lifecycle, and Chain (ie, Blockchain) components. Evaluation showed that, in multi-party settings, documentation history remains independently verifiable because lifecycle actions are role-attributed and linked to content hashes; median testnet gas use ranged from 352 722 to 852 399 gas across the evaluated lifecycle operations.**

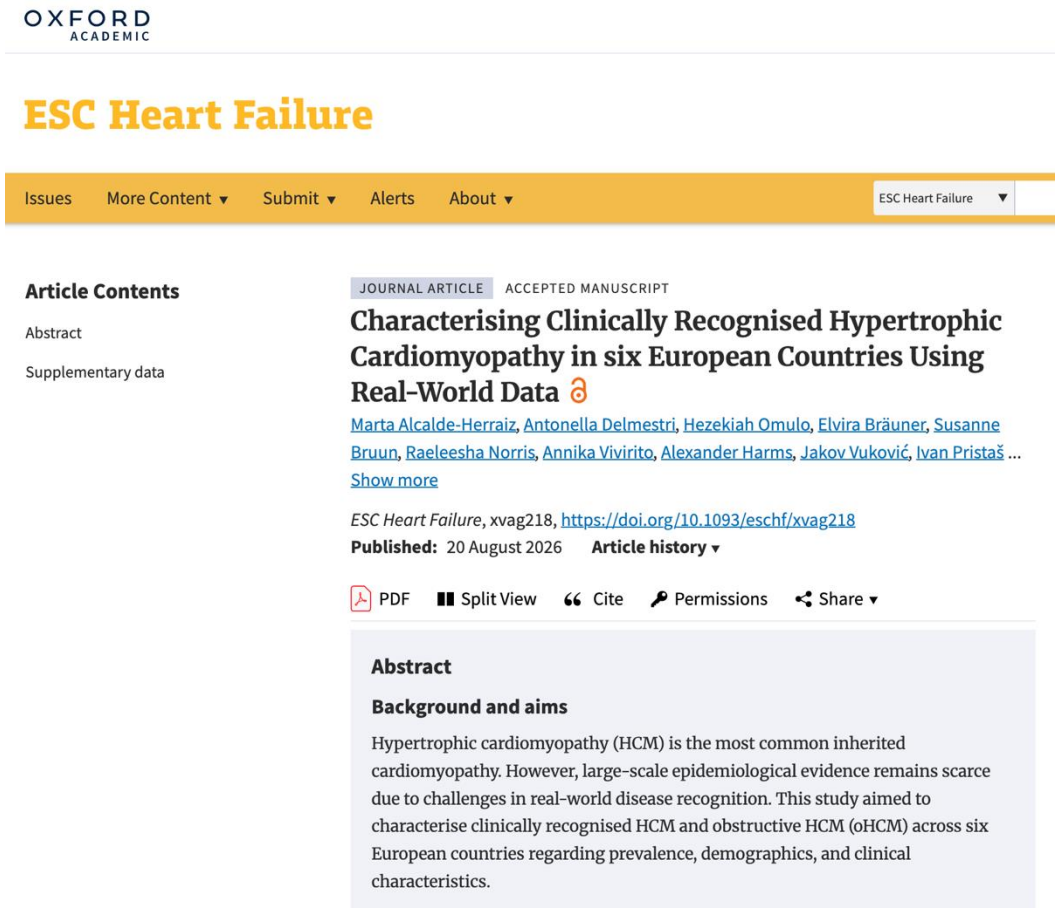
Discussion: **SMART positions clinical AI documentation as a shared infrastructure for preliminary model assessment, making role attribution, lifecycle state, and documentation changes more transparent and traceable over time.**

Conclusion: **By providing an implementable framework, SMART illustrates how clinical AI documentation can be structured, clinically contextualized, auditable, responsible, transparent, and lifecycle-aware.**

Key words: blockchain; medical records systems computerized; artificial intelligence; documentation; reproducibility of results; electronic health records.

OHDSI Shoutouts!

Congratulations to the team of **Marta Alcalde-Herraiz, Antonella Delmestri, Hezekiah Omulo, Elvira Bräuner, Susanne Bruun, Raeleesha Norris, Annika Vivirito, Alexander Harms, Jakov Vuković, Ivan Pristaš, Anamaria Jurčević, Marko Čavlina, Antea Jezidžić, Pero Ivanko, Saeed Hayati, Nhung T H Trinh, Hedvig Nordeng, Talita Duarte-Salles, Anna Palomar-Cros, Agustina Giuliodori, Antonio Gómez-Outes, Patrick Vrijlandt, María Clara Restrepo-Méndez, Edward Burn, Albert Prats-Uribe, and Anna Saura-Lazaro** on the recent publication of **Characterising Clinically Recognised Hypertrophic Cardiomyopathy in six European Countries Using Real-World Data** in *ESC Heart Failure*.



OXFORD ACADEMIC

ESC Heart Failure

Issues | More Content | Submit | Alerts | About | ESC Heart Failure

Article Contents

Abstract

Supplementary data

JOURNAL ARTICLE ACCEPTED MANUSCRIPT

Characterising Clinically Recognised Hypertrophic Cardiomyopathy in six European Countries Using Real-World Data

Marta Alcalde-Herraiz, Antonella Delmestri, Hezekiah Omulo, Elvira Bräuner, Susanne Bruun, Raeleesha Norris, Annika Vivirito, Alexander Harms, Jakov Vuković, Ivan Pristaš ...

Show more

ESC Heart Failure, xvag218, https://doi.org/10.1093/eschf/xvag218

Published: 20 August 2026 **Article history**

PDF | Split View | Cite | Permissions | Share

Abstract

Background and aims

Hypertrophic cardiomyopathy (HCM) is the most common inherited cardiomyopathy. However, large-scale epidemiological evidence remains scarce due to challenges in real-world disease recognition. This study aimed to characterise clinically recognised HCM and obstructive HCM (oHCM) across six European countries regarding prevalence, demographics, and clinical characteristics.

OHDSI Shoutouts!

Congratulations to the team of **Rinat Gabbay-Benziv, Esther Maor-Sagie, Amos Stern, Inna Bleicher, Simon Shenhav, Enav Yefet, Yifat Wiener, Michal Kovo, and Maya Frank Wolf** on the recent publication of **Gestational diabetes and maternal morbidity in multifetal pregnancies: independent contributions of GDM and plurality** in *Diabetes Research and Clinical Practice*.

ELSEVIER

Diabetes Research and Clinical Practice
Available online 22 August 2026, 113517
In Press, Journal Pre-proof What's this?

Gestational diabetes and maternal morbidity in multifetal pregnancies: independent contributions of GDM and plurality

Rinat Gabbay-Benziv [a b], Esther Maor-Sagie [a b], Amos Stern [c d], Inna Bleicher [b e], Simon Shenhav [f g], Enav Yefet [h l], Yifat Wiener [i j], Michal Kovo [i j], Maya Frank Wolf [k l]

Show more

Add to Mendeley Share Cite

https://doi.org/10.1016/j.diabres.2026.113517 Get rights and content

Under a Creative Commons license Open access

Highlights

- Maternal morbidity was compared across four GDM-plurality groups.
- GDM was independently associated with hypertensive disorders.
- Multifetal gestation showed the strongest association with hypertensive disorders.
- No significant interaction between GDM and plurality was detected.
- Findings were derived from a national multicenter OMOP-based cohort.

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
Wednesday	12 pm	Latin America Chapter
Thursday	8 am	Medical Devices
Thursday	9 am	Phenotype Development & Evaluation
Thursday	10 am	Africa Chapter (ZOOM)
Thursday	10 am	GIS – Geographic Information System
Thursday	11 am	Perinatal & Repoductive Health
Friday	11 am	Clinical Trials
Friday	11:30 am	Steering
Monday	9 am	Vaccine Vocabulary
Tuesday	10 am	CDM Survey

Titan Nominations Are OPEN!

The OHDSI Titan Awards recognize those who have gone above and beyond to foster community engagement, lead research and development efforts and make significant contributions towards OHDSI's mission.

Individuals, teams or collaborating institutions can be nominated for an award, and you can make as many nominations as you would like.

Nominations will be accepted until 8pm ET Tuesday, Sept. 22, 2026.

ohdsi.org/titan-awards

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 Global Symposium Agenda

Start	End	Topic	Presenter/Lead
7:15 am	8:00 am	Newcomer Welcome	Paul Nagy/Marc Suchard
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

OHDSI 2026 Plenaries

Federated Learning Meets Negative Control Calibration: Toward Reliable Multi-Site Evidence Generation (**Yong Chen**, University of Pennsylvania)

Beyond the Defaults: How the OHDSI Community is Adapting, Extending, and Reimagining Its Tools (**Scott Duvall**, Purple Lab)

The role of national initiatives in supporting sustainability, collaboration, and growth of OHDSI (**Ed Burn**, Oxford University)

OHDSI 2026 Lightning Talks – Session 1

Using OWL Reasoning to Identify Structural Inconsistencies in the OMOP Standardized Vocabularies (**William Baumgartner**, University of Colorado Anschutz Medical Campus)

Comparative performance of the concurrent comparator design with existing vaccine safety surveillance approaches on real-world observational health data (**Shounak Chattopadhyay**, University of Virginia)

Making the most of What We Have: Regional Characterization of OHDSI Network Database (**Evelyn Goh**, National University of Singapore)

A Multi-Database, Retrospective, Long-term, Observational Safety Study of Pediatric Patients with Inflammatory Bowel Disease (**Kevin Haynes**, Johnson & Johnson)

Implementing Population Health Risk Analytics on a Standardized EHR Data Layer: Evaluation and Framework Development (**Haeun Lee**, Johns Hopkins University)

OHDSI 2026 Lightning Talks – Session 2

Evaluating Semantic Loss in EHR Terminology Mapping in Common and Rare Diseases (**Melanie Philofsky**, EPAM Systems)

EPISODE/EPISODE_EVENT-Based Representation of Microbiology Results and Antimicrobial Susceptibility Testing in OMOP CDM v5.4 (**Polina Talapova**, SciForce/Tufts CTSI)

UNO: Neural Pruning Enhances AI Reliability of Observational Data-based Findings (**Yue Wu**, University of Pennsylvania)

MOSAiC-H (**Qiaoshi Yang**, University of Pennsylvania)

Ontology-Governed LLM Search for Safe Drug Concept Mapping in Air-Gapped Healthcare Environments (**Seung Ju Yun**, Chonnam National University Hwasun Hospital)

2026 OHDSI Global Symposium

There are opportunities to be both a **sponsor** and an **exhibitor** at the Global Symposium.

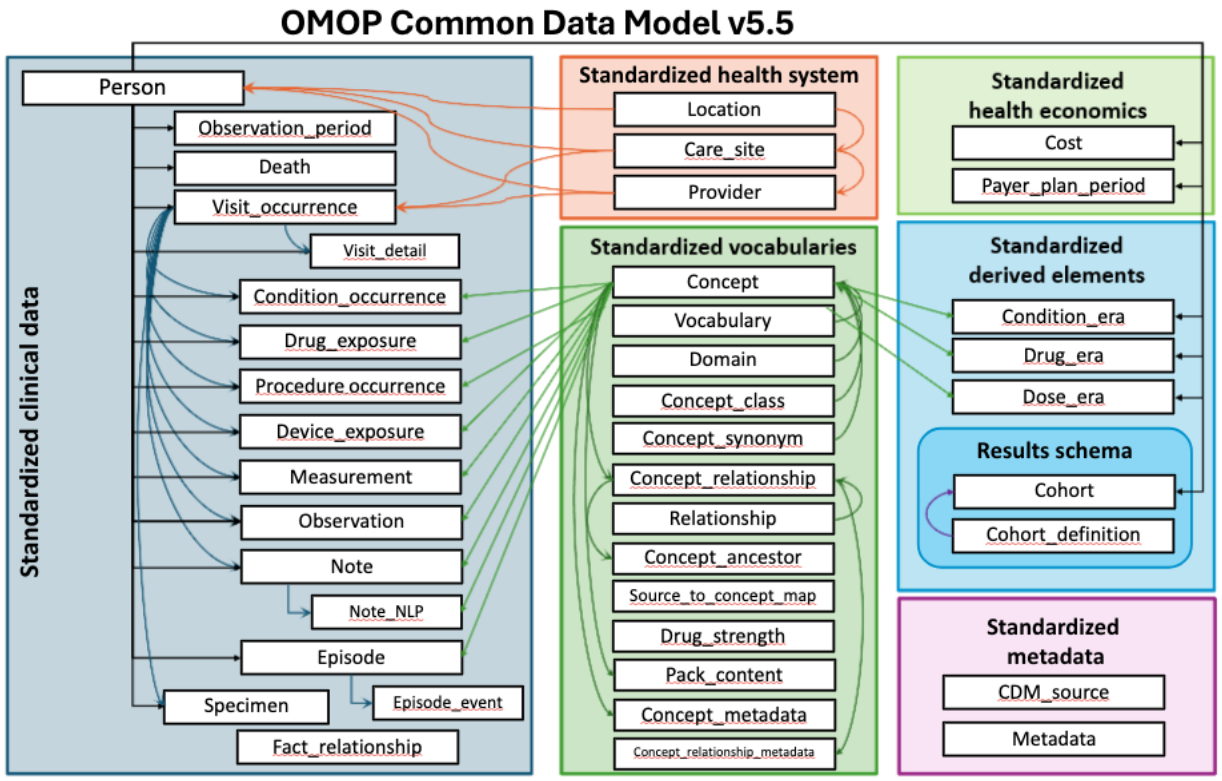
Please reach out to **symposium@OHDSI.org** for more information.

ohdsi.org/OHDSI2026

Collaborate on OMOP CDM v5.5

The Common Data Model working group has been planning CDM v5.5 over the last 18 months and there is now a development version to share. Please take a look, review, and test in your environment if you are able.

This version will be released alongside the August 2026 vocabulary and will support community needs and continued vocabulary updates.



github.com/OHDSI/CommonDataModel/tree/develop

EUROPE

You've built your OMOP CDM. Now let's make it shine!

Take your OMOP implementation to the next level

- 2.5-day hands-on expert training (max. 30 participants)
- Work with Data Quality Dashboard, ATLAS, CohortDiagnostics
- Learn study design, validation, execution
- Optional: bring your own ETL + OMOP CDM

What to expect

- Practical OMOP expertise
- Concrete CDM improvements
- Skills for evidence generation

Practical

🗓️ 7–9 September 2026
📍 Rotterdam

FALCON Symposium & Bladder Cancer Studyathon

Sep 28 – Oct 1, 2026 Dunden, Antwerp, Belgium

Sep 28: Symposium — Public afternoon with oncologists, data scientists, regulators, and hospital leaders. Short talks and a network drink.

Sep 29 – Oct 1: Studyathon — Three hands-on days: cohort building, analytics, and abstract drafting.

Study related materials:

- All study related materials are here: GDE2025 - Bladder cancer treatment | OHDSI | Microsoft Teams
- GitHub repo: FALCON Bladder Git Repo

Register now

Details, agenda updates, venue information, and registration are available on the FALCON website.

www.falcon-network.org

OMOP School: Oct. 13-15 + Oct. 16 (optional)

The OMOP School in Uppsala, Sweden will be held Oct. 13-15, with an extra optional day on Oct. 16.

This will feature a 3+1 Day OMOP CDM Bootcamp with hands-on training and workshop that turns your data harmonization vision into reality.

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 is open! All details can be found on the event homepage.

ohdsi.org/APAC2026

Job Openings

Associate Director, Data Science (Real World Data)

Apply

Boston, MA; New York, NY

About the Position

As Associate Director of RWD Intelligence at Formation Bio, you will lead the strategy and execution of our real-world data (RWD) capabilities, building the data foundations that power drug acquisition, clinical development, and portfolio decision-making. You will own the end-to-end lifecycle of RWD: sourcing, procurement, ingestion, harmonization, quality assurance, and delivery of analysis-ready datasets to downstream consumers across Product, Data Science, Clinical Development, and Business Development.

This role sits at the intersection of data engineering, data science, and drug development. You will build and maintain scalable data infrastructure (pipelines, data models, lakes/marts) while ensuring semantic interoperability across heterogeneous data sources through ontology-driven harmonization frameworks such as OMOP. You will also manage vendor relationships and data procurement, evaluating and integrating new data assets as the portfolio evolves. The ideal candidate combines deep RWD domain expertise with strong data fluency, enabling Formation Bio to treat real-world evidence as a first-class strategic asset.

Responsibilities

- Lead the RWD Intelligence function within Data Science, owning data strategy, sourcing, and delivery of analysis-ready datasets
- Architect and maintain the supporting infrastructure (pipelines, ingestion workflows, data models, lakes/marts) across EHR/EMR, claims, registries, and genomics-linked cohorts
- Drive adoption and extension of harmonization frameworks (e.g., OMOP CDM) across heterogeneous data sources, leveraging AI/ML tools for entity resolution, ontology mapping, data quality monitoring, and automated harmonization
- Manage RWD vendor relationships end-to-end: evaluate providers, negotiate data use agreements, broker new partnerships, and integrate acquired datasets into the platform
- Partner with Data Science, Clinical Development, Business Development, and Engineering teams to define RWD use cases (trial feasibility, synthetic control arms, epidemiology, label expansion) and productize ad hoc pipelines into scalable, production-grade systems
- Foster a culture of data quality rigor, documentation, and reproducibility across all RWD assets

Senior Data Scientist - Real World Data

Apply

New York, NY; Boston, MA; San Francisco, CA; Raleigh-Durham, NC

About the Position

We're looking for a Senior Data Scientist to join the Scientific Data Intelligence (SDI) team at Formation Bio to help transform Real World Data (RWD)—spanning electronic health records, claims, and other longitudinal patient data sources—into structured, analytics-ready assets. In this role, you'll be partnering closely with our Data Science team not only to model and transform data, but also to actively analyze it: answering research questions, generating evidence, and supporting scientific decision-making across our drug portfolio.

This position sits at the intersection of healthcare data engineering, real-world evidence analysis, and generative AI. While a strong foundation in building reliable, scalable pipelines is essential, you'll be equally expected to roll up your sleeves and work directly with the data—constructing cohorts, running analyses, and translating findings into actionable insights for scientific and business stakeholders.

The ideal candidate is a hybrid of data engineer and applied scientist: someone who can build the infrastructure and then use it, with familiarity in RWD study design, GenAI fluency (e.g., LLM-based entity extraction, summarization, classification), and strong technical expertise with modern data tooling. You'll play a key role in shaping how real-world patient data becomes discoverable, structured, and impactful across the organization.

Responsibilities

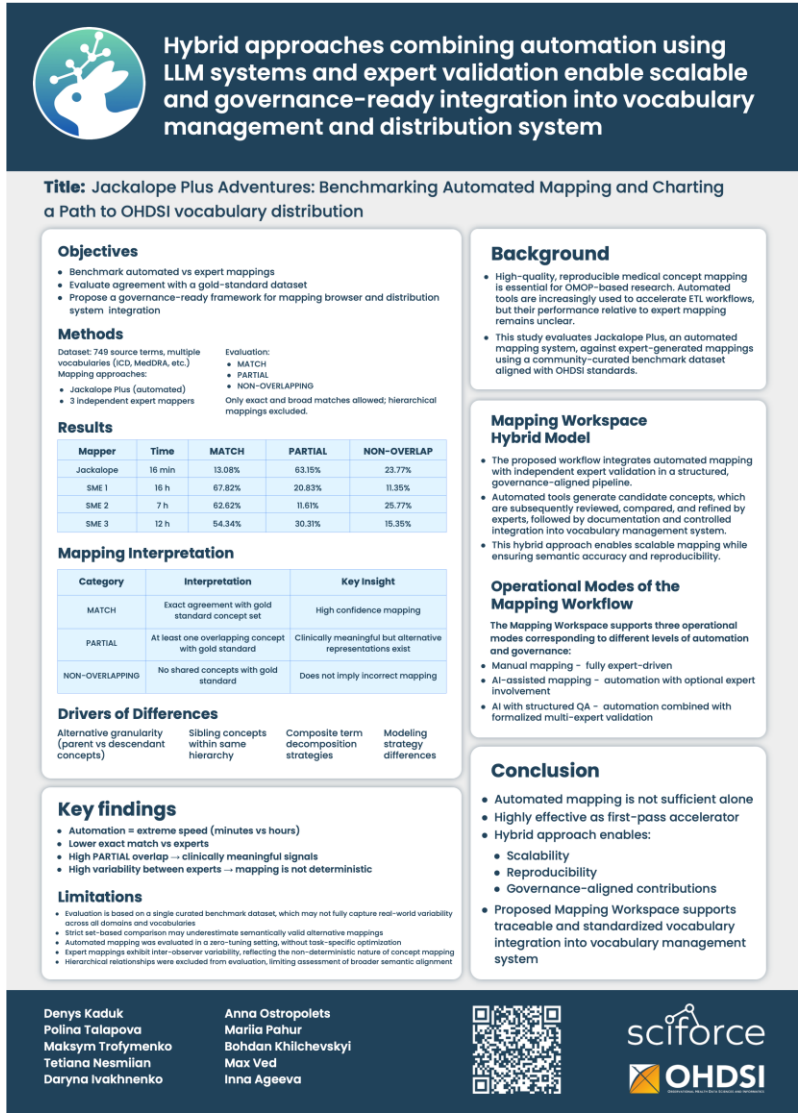
- Model and transform raw EHR and claims data into clean, canonical, and analytics-ready datasets using SQL, Python, and clinical standards like OMOP.
- Build and manage scalable data pipelines using Dagster for orchestration, dbt for transformation, and Snowflake as the primary compute and storage engine.
- Conduct hands-on RWD analyses to answer scientific and strategic research questions—including disease epidemiology, treatment patterns, patient journey characterization, and comparative effectiveness.
- Partner with Data Scientists and clinical leads to design and execute observational studies, translating scientific questions into well-structured, reproducible analyses.
- Implement data validation, completeness, and observability frameworks to ensure real-world datasets are accurate, comprehensive, and trustworthy for downstream research and product use.
- Apply Generative AI techniques within transformation and analysis layers to accelerate data structuring and insight generation.
- Communicate findings clearly to both technical and non-technical stakeholders, including summaries for portfolio teams and leadership.

#OHDSISocialShowcase This Week

Monday

Jackalope Plus Adventures: Benchmarking Automated Mapping and Charting a Path to OHDSI Vocabulary management system

(Denys Kaduk, Polina Talapova, Maksym Trofymenko, Tetiana Nesmiian, Daryna Ivakhnenko, Anna Ostropolets, Bohdan Khilchevskyi, Max Ved, Inna Ageeva)

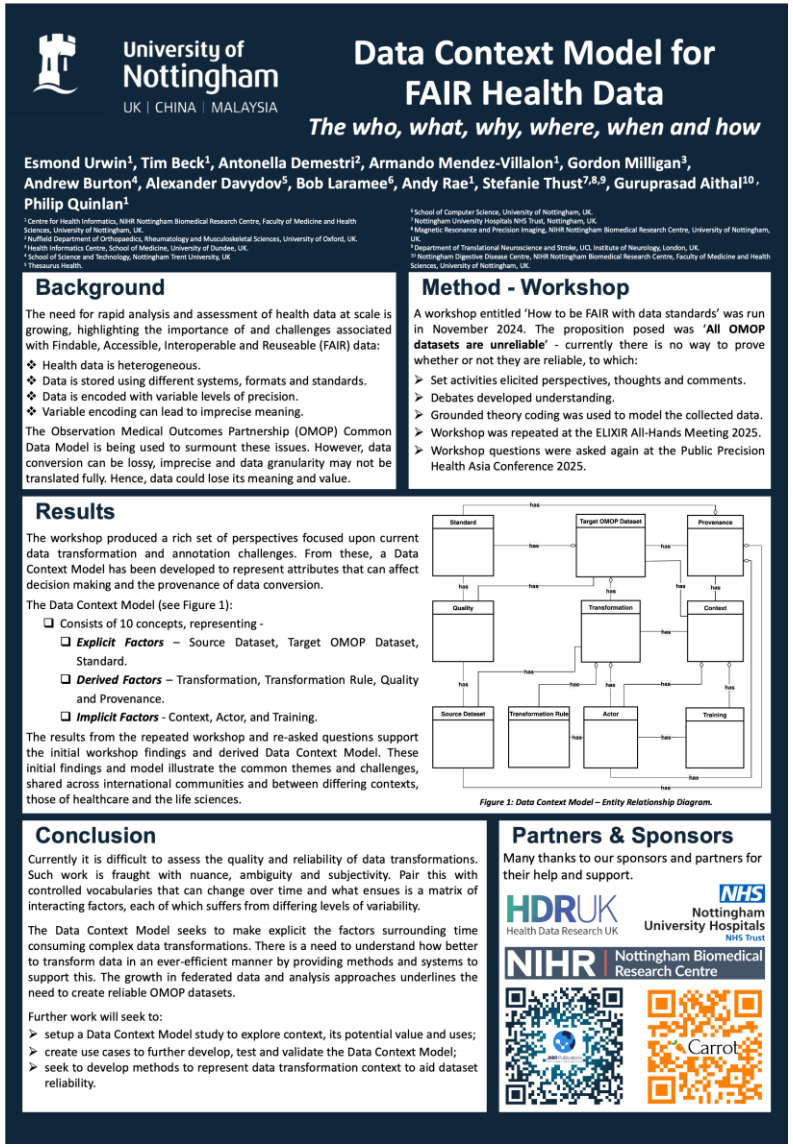


#OHDSISocialShowcase This Week

Tuesday

Data Context Model for FAIR Health Data - The who, what, why, where, when and how

(Esmond Urwin, Tim Beck, Antonella Delmestri, Armando Mendez-Villalon, Gordon Milligan, Andrew Burton, Andy Rae, Stefanie Thust, Philip Quinlan)



University of Nottingham
UK | CHINA | MALAYSIA

Data Context Model for FAIR Health Data

The who, what, why, where, when and how

Esmond Urwin[1], Tim Beck[1], Antonella Demestri[2], Armando Mendez-Villalon[1], Gordon Milligan[3], Andrew Burton[4], Alexander Davydov[5], Bob Laramee[6], Andy Rae[1], Stefanie Thust[7,8,9], Guruprasad Aithal[10], Philip Quinlan[1]

[1] Centre for Health Informatics, NIHR Nottingham Biomedical Research Centre, Faculty of Medicine and Health Sciences, University of Nottingham, UK.
[2] Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, UK.
[3] Health Informatics Centre, School of Medicine, University of Dundee, UK.
[4] School of Science and Technology, Nottingham Trent University, UK
[5] Thesaurus Health.
[6] School of Computer Science, University of Nottingham, UK.
[7] Nottingham University Hospitals NHS Trust, Nottingham, UK.
[8] Magnetic Resonance and Precision Imaging, NIHR Nottingham Biomedical Research Centre, University of Nottingham, UK.
[9] Department of Translational Neuroscience and Stroke, UCL Institute of Neurology, London, UK.
[10] Nottingham Digestive Disease Centre, NIHR Nottingham Biomedical Research Centre, Faculty of Medicine and Health Sciences, University of Nottingham, UK.

Background

The need for rapid analysis and assessment of health data at scale is growing, highlighting the importance of and challenges associated with Findable, Accessible, Interoperable and Reuseable (FAIR) data:

- Health data is heterogeneous.
- Data is stored using different systems, formats and standards.
- Data is encoded with variable levels of precision.
- Variable encoding can lead to imprecise meaning.

The Observation Medical Outcomes Partnership (OMOP) Common Data Model is being used to surmount these issues. However, data conversion can be lossy, imprecise and data granularity may not be translated fully. Hence, data could lose its meaning and value.

Method - Workshop

A workshop entitled 'How to be FAIR with data standards' was run in November 2024. The proposition posed was '**All OMOP datasets are unreliable**' - currently there is no way to prove whether or not they are reliable, to which:

- Set activities elicited perspectives, thoughts and comments.
- Debates developed understanding.
- Grounded theory coding was used to model the collected data.
- Workshop was repeated at the ELIXIR All-Hands Meeting 2025.
- Workshop questions were asked again at the Public Precision Health Asia Conference 2025.

Results

The workshop produced a rich set of perspectives focused upon current data transformation and annotation challenges. From these, a Data Context Model has been developed to represent attributes that can affect decision making and the provenance of data conversion.

The Data Context Model (see Figure 1):

- Consists of 10 concepts, representing -
  - ***Explicit Factors*** – Source Dataset, Target OMOP Dataset, Standard.
  - ***Derived Factors*** – Transformation, Transformation Rule, Quality and Provenance.
  - ***Implicit Factors*** - Context, Actor, and Training.

The results from the repeated workshop and re-asked questions support the initial workshop findings and derived Data Context Model. These initial findings and model illustrate the common themes and challenges, shared across international communities and between differing contexts, those of healthcare and the life sciences.

Figure 1: Data Context Model – Entity Relationship Diagram.

Conclusion

Currently it is difficult to assess the quality and reliability of data transformations. Such work is fraught with nuance, ambiguity and subjectivity. Pair this with controlled vocabularies that can change over time and what ensues is a matrix of interacting factors, each of which suffers from differing levels of variability.

The Data Context Model seeks to make explicit the factors surrounding time consuming complex data transformations. There is a need to understand how better to transform data in an ever-efficient manner by providing methods and systems to support this. The growth in federated data and analysis approaches underlines the need to create reliable OMOP datasets.

Further work will seek to:

- setup a Data Context Model study to explore context, its potential value and uses;
- create use cases to further develop, test and validate the Data Context Model;
- seek to develop methods to represent data transformation context to aid dataset reliability.

Partners & Sponsors

Many thanks to our sponsors and partners for their help and support.

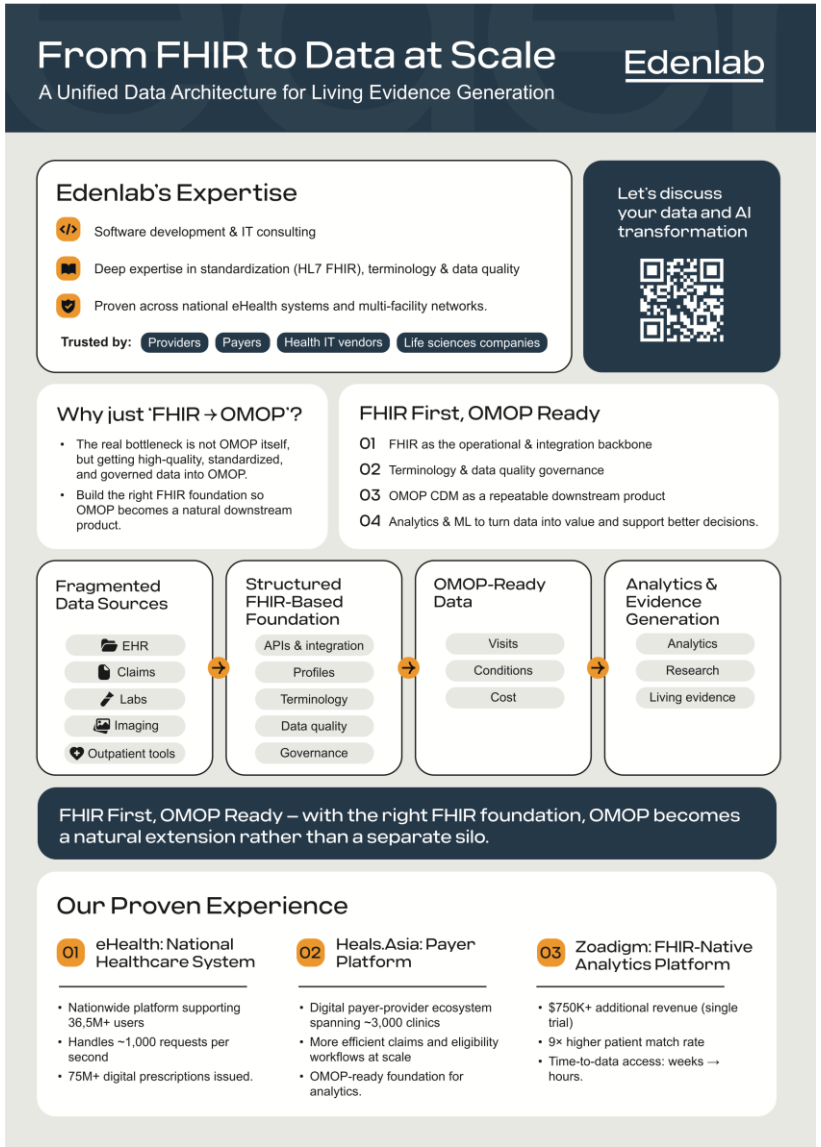
HDRUK Health Data Research UK
NHS Nottingham University Hospitals NHS Trust
NIHR | Nottingham Biomedical Research Centre
Carrot

#OHDSISocialShowcase This Week

Wednesday

From FHIR to working with data at scale: a unified data architecture for living evidence generation

(Jan Vekemans, Kseniia NiKolaienko)

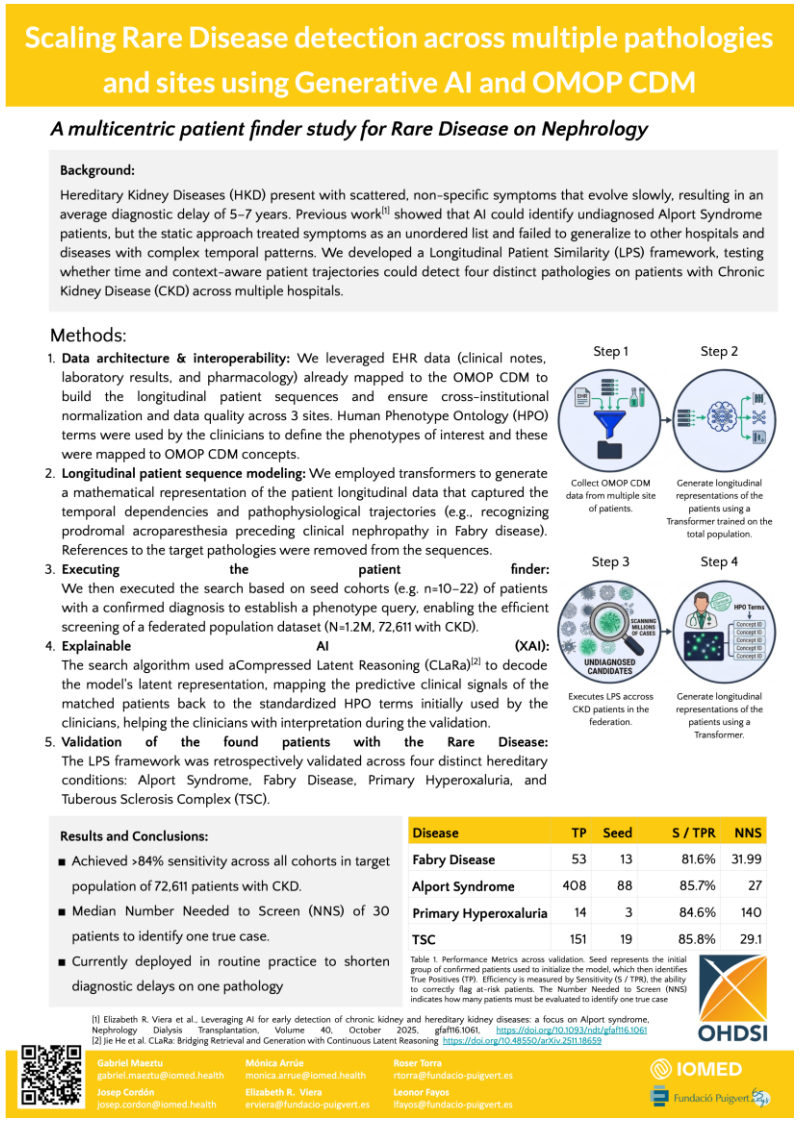


#OHDSISocialShowcase This Week

Thursday

Scaling Rare Disease Detection Across Multiple Pathologies using Generative AI and OMOP CDM

(Gabriel de Maeztu, Josep Cordón, Mónica Arrúe)



Scaling Rare Disease detection across multiple pathologies and sites using Generative AI and OMOP CDM

A multicentric patient finder study for Rare Disease on Nephrology

Background:

Hereditary Kidney Diseases (HKD) present with scattered, non-specific symptoms that evolve slowly, resulting in an average diagnostic delay of 5–7 years. Previous work[1] showed that AI could identify undiagnosed Alport Syndrome patients, but the static approach treated symptoms as an unordered list and failed to generalize to other hospitals and diseases with complex temporal patterns. We developed a Longitudinal Patient Similarity (LPS) framework, testing whether time and context-aware patient trajectories could detect four distinct pathologies on patients with Chronic Kidney Disease (CKD) across multiple hospitals.

Methods:

1. **Data architecture & interoperability:** We leveraged EHR data (clinical notes, laboratory results, and pharmacology) already mapped to the OMOP CDM to build the longitudinal patient sequences and ensure cross-institutional normalization and data quality across 3 sites. Human Phenotype Ontology (HPO) terms were used by the clinicians to define the phenotypes of interest and these were mapped to OMOP CDM concepts.
2. **Longitudinal patient sequence modeling:** We employed transformers to generate a mathematical representation of the patient longitudinal data that captured the temporal dependencies and pathophysiological trajectories (e.g., recognizing prodromal acroparesthesia preceding clinical nephropathy in Fabry disease). References to the target pathologies were removed from the sequences.
3. **Executing the patient finder:** We then executed the search based on seed cohorts (e.g. n=10–22) of patients with a confirmed diagnosis to establish a phenotype query, enabling the efficient screening of a federated population dataset (N=1.2M, 72,611 with CKD).
4. **Explainable AI (XAI):** The search algorithm used aCompressed Latent Reasoning (CLaRa)[2] to decode the model's latent representation, mapping the predictive clinical signals of the matched patients back to the standardized HPO terms initially used by the clinicians, helping the clinicians with interpretation during the validation.
5. **Validation of the found patients with the Rare Disease:** The LPS framework was retrospectively validated across four distinct hereditary conditions: Alport Syndrome, Fabry Disease, Primary Hyperoxaluria, and Tuberous Sclerosis Complex (TSC).

Results and Conclusions:

- Achieved >84% sensitivity across all cohorts in target population of 72,611 patients with CKD.
- Median Number Needed to Screen (NNS) of 30 patients to identify one true case.
- Currently deployed in routine practice to shorten diagnostic delays on one pathology

Disease	TP	Seed	S / TPR	NNS
Fabry Disease	53	13	81.6%	31.99
Alport Syndrome	408	88	85.7%	27
Primary Hyperoxaluria	14	3	84.6%	140
TSC	151	19	85.8%	29.1

Table 1. Performance Metrics across validation. Seed represents the initial group of confirmed patients used to initialize the model, which then identifies True Positives (TP). Efficiency is measured by Sensitivity (S / TPR), the ability to correctly flag at-risk patients. The Number Needed to Screen (NNS) indicates how many patients must be evaluated to identify one true case

[1] Elizabeth R. Viera et al., Leveraging AI for early detection of chronic kidney and hereditary kidney diseases: a focus on Alport syndrome, Nephrology Dialysis Transplantation, Volume 40, October 2025, gfaf116.1061. https://doi.org/10.1093/ndt/gfaf116.1061
[2] Jie He et al. CLaRa: Bridging Retrieval and Generation with Continuous Latent Reasoning https://doi.org/10.48550/arXiv.2511.18659

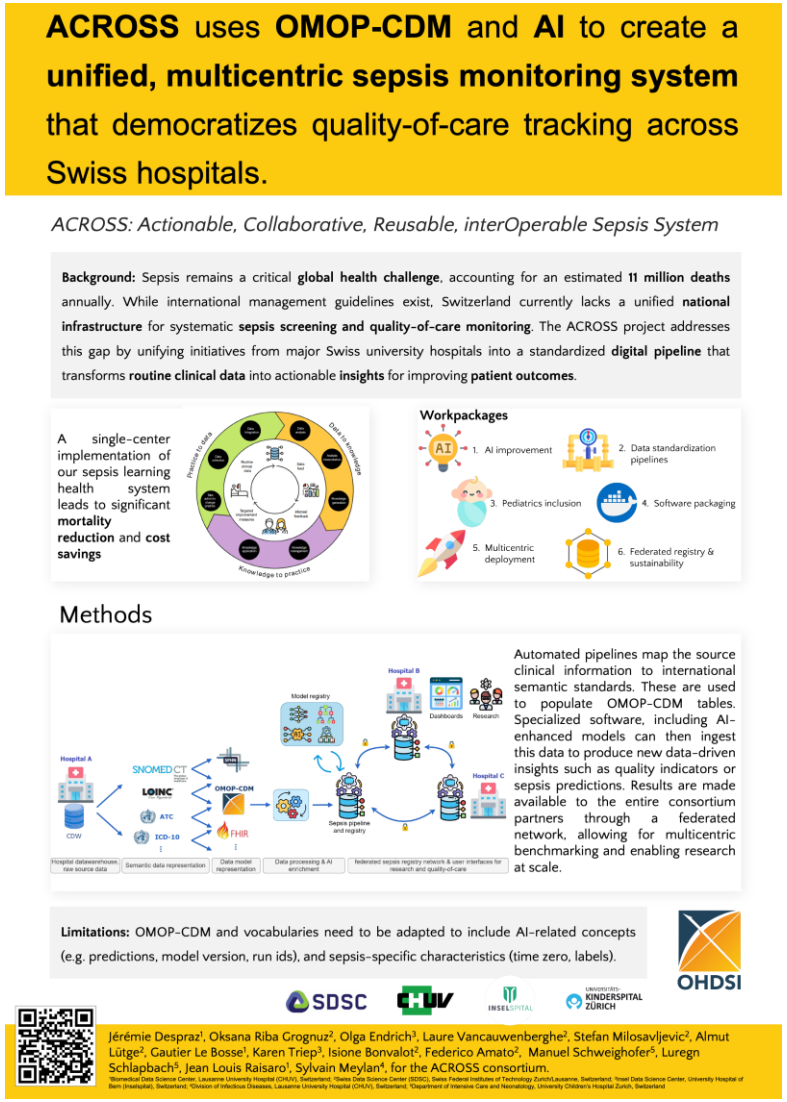
Gabriel Maeztu gabriel.maeztu@iomed.health
Mónica Arrúe monica.arrue@iomed.health
Roser Torra rtorra@fundacio-puigvert.es
Josep Cordón josep.cordon@iomed.health
Elizabeth R. Viera erviera@fundacio-puigvert.es
Leonor Fayos lfayos@fundacio-puigvert.es

#OHDSISocialShowcase This Week

Friday

ACROSS: Actionable, Collaborative, Reusable, interOperable Sepsis System

(Jérémie Despraz, Oksana Riba Grognuz, Olga Endrich, Laure Vancauwenberghe, Stefan Milosavljevic, Almut Lütge, Gautier Le Bosse, Karen Triep, Isione Bonvalot, Federico Amato, Manuel Schweighofer, Luregn Schlapbach, Jean Louis Raisaro, Sylvain Meylan)



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