OHDSI Scholarship

(Publications in 2022)

OHDSI Community Call

June 14, 2022 • 11 am ET

Upcoming OHDSI Community Calls

Date	Topic
June 21	10-Minute Tutorials
June 28	European Symposium Recap
July 5	NO MEETING
July 12	New Adopter Introductions and Q&A
July 19	Workgroup Updates
July 26	CDM Update Process

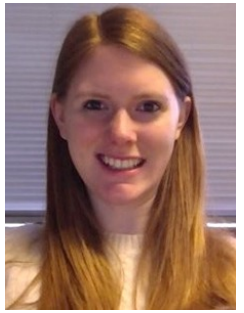
Upcoming OHDSI Community Calls

Date	Topic
June 21	10-Minute Tutorials
June 28	European Symposium Recap
July 5	NO MEETING
July 12	New Adopter Introductions and Q&A
July 19	Workgroup Updates
July 26	CDM Update Process

June 21: 10-Minute Tutorials

CAPR
Presenter: Martin Lavalee

Patient-Level Prediction
Presenter: Jenna Reps



PheValuator
Presenter: Joel Swerdel

Three Stages of The Journey

Where Have We Been?

Where Are We Now?

Where Are We Going?

OHDSI Shoutouts!

Congratulations to the team of **Rohan Khera, Martijn Schuemie, Yuan Lu, Anna Ostropolets, RuiJun Chen, George Hripcsak, Patrick Ryan, Harlan Krumholz, and Marc Suchard** on the publication of **Large-scale evidence generation and evaluation across a network of databases for type 2 diabetes mellitus (LEGEND-T2DM): a protocol for a series of multinational, real-world comparative cardiovascular effectiveness and safety studies** in BMJ Open.

Open access Protocol

BMJ Open

Large-scale evidence generation and evaluation across a network of databases for type 2 diabetes mellitus (LEGEND-T2DM): a protocol for a series of multinational, real-world comparative cardiovascular effectiveness and safety studies

Rohan Khera [1,2] Martijn J Schuemie [3,4] Yuan Lu [1,2] Anna Ostropolets [5] RuiJun Chen,[6] George Hripcsak,[5,7] Patrick B Ryan,[3,5] Harlan M Krumholz [1,2] Marc A Suchard [4,8,9,10]

To cite: Khera R, Schuemie MJ, Lu Y, *et al*. Large-scale evidence generation and evaluation across a network of databases for type 2 diabetes mellitus (LEGEND-T2DM): a protocol for a series of multinational, real-world comparative cardiovascular effectiveness and safety studies. *BMJ Open* 2022;**12**:e057977. doi:10.1136/bmjopen-2021-057977

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2021-057977).

Received 04 November 2021
Accepted 27 April 2022

ABSTRACT

Introduction Therapeutic options for type 2 diabetes mellitus (T2DM) have expanded over the last decade with the emergence of cardioprotective novel agents, but without such data for older drugs, leaving a critical gap in our understanding of the relative effects of T2DM agents on cardiovascular risk.

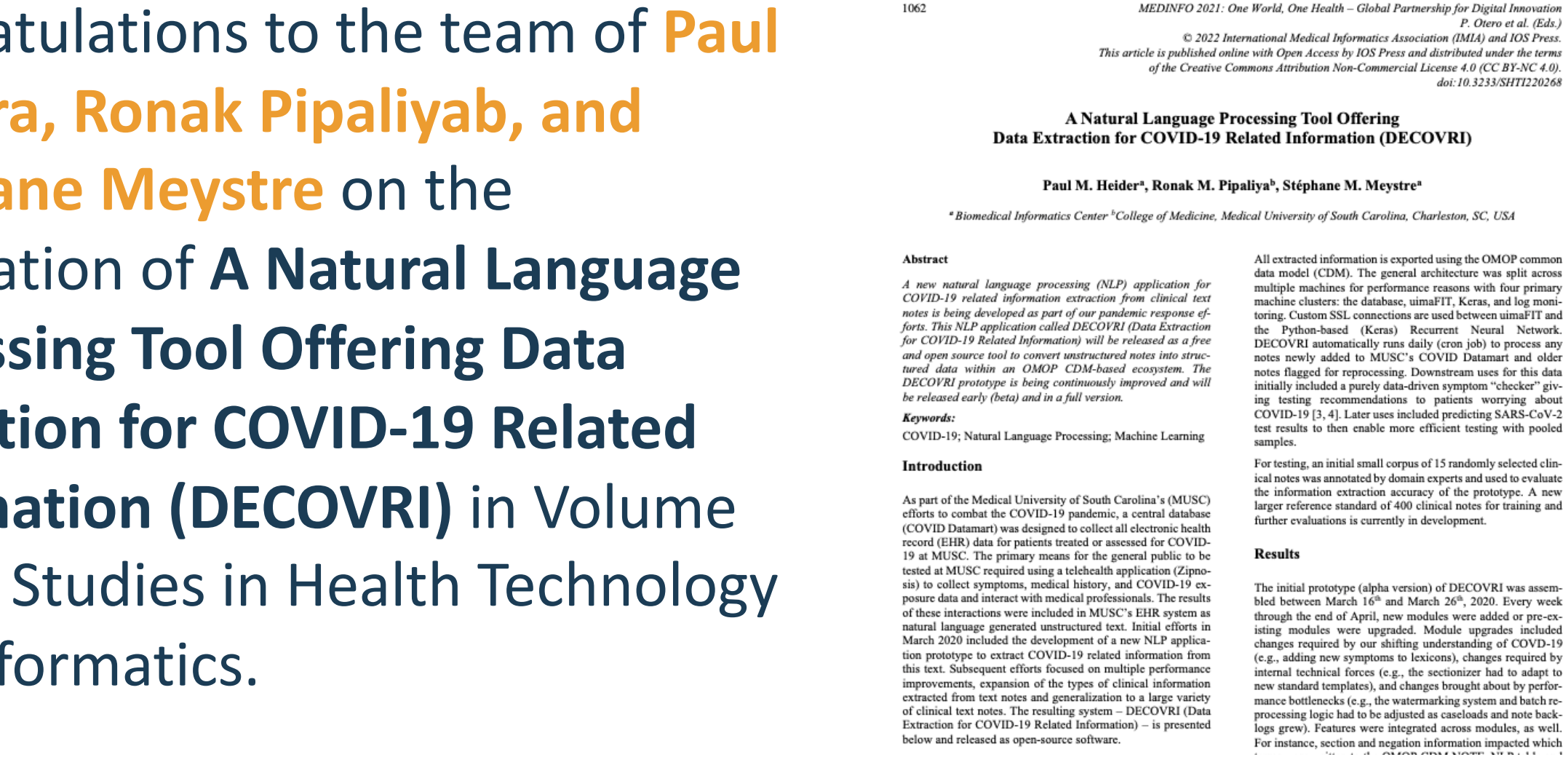
Methods and analysis The large-scale evidence generations across a network of databases for T2DM (LEGEND-T2DM) initiative is a series of systematic, large-scale, multinational, real-world comparative cardiovascular effectiveness and safety studies of all four major second-line anti-hyperglycaemic agents, including sodium–glucose co-transporter-2 inhibitor, glucagon-like peptide-1 receptor agonist, dipeptidyl peptidase-4 inhibitor and sulfonylureas. LEGEND-T2DM will leverage the Observational Health Data Sciences and Informatics (OHDSI) community that provides access to a global network of administrative claims and electronic health record data sources, representing 190 million patients in the USA and about 50 million internationally. LEGEND-T2DM will identify all adult, patients with T2DM who newly initiate a traditionally second-line T2DM agent. Using an active comparator, new-user cohort design, LEGEND-T2DM will execute all pairwise class-versus-class and drug-

STRENGTHS AND LIMITATIONS OF THIS STUDY

⇒ The proposal seeks to use health information encompassing millions of patients with type 2 diabetes mellitus (T2DM) in the multinational Observational Health Data Science and Informatics (OHDSI) community to determine real-world comparative effectiveness and safety of traditionally second-line T2DM agents.
⇒ The proposed set of studies will be comprehensive, with a systematic pairwise comparisons of all sodium–glucose co-transporter-2 inhibitor, glucagon-like peptide-1 receptor agonist, dipeptidyl peptidase-4 inhibitor and sulfonylurea agents at the drug, class and population subgroup level.
⇒ The studies will focus on a broad set of outcomes, including comprehensive measures of adverse cardiovascular events as well as secondary effectiveness and safety outcomes.
⇒ The studies use robust methods an observational, active-comparator, new-user cohort design with a systematic framework to address residual confounding, publication bias and p-hacking using data-driven, large-scale propensity adjustment

OHDSI Shoutouts!

Congratulations to the team of **Paul Heidera, Ronak Pipaliyab, and Stéphane Meystre** on the publication of **A Natural Language Processing Tool Offering Data Extraction for COVID-19 Related Information (DECOVRI)** in Volume 290 of Studies in Health Technology and Informatics.



MEDINFO 2021: One World, One Health – Global Partnership for Digital Innovation
P. Otero et al. (Eds.)

doi:10.3233/SHTI220268

A Natural Language Processing Tool Offering Data Extraction for COVID-19 Related Information (DECOVRI)

Paul M. Heider[a], Ronak M. Pipaliya[b], Stéphane M. Meystre[a]

[a] Biomedical Informatics Center [b] College of Medicine, Medical University of South Carolina, Charleston, SC, USA

Abstract

A new natural language processing (NLP) application for COVID-19 related information extraction from clinical text notes is being developed as part of our pandemic response efforts. This NLP application called DECOVRI (Data Extraction for COVID-19 Related Information) will be released as a free and open source tool to convert unstructured notes into structured data within an OMOP CDM-based ecosystem. The DECOVRI prototype is being continuously improved and will be released early (beta) and in a full version.

Keywords:
COVID-19; Natural Language Processing; Machine Learning

Introduction

As part of the Medical University of South Carolina's (MUSC) efforts to combat the COVID-19 pandemic, a central database (COVID Datamart) was designed to collect all electronic health record (EHR) data for patients treated or assessed for COVID-19 at MUSC. The primary means for the general public to be tested at MUSC required using a telehealth application (Zipnosis) to collect symptoms, medical history, and COVID-19 exposure data and interact with medical professionals. The results of these interactions were included in MUSC's EHR system as natural language generated unstructured text. Initial efforts in March 2020 included the development of a new NLP application prototype to extract COVID-19 related information from this text. Subsequent efforts focused on multiple performance improvements, expansion of the types of clinical information extracted from text notes and generalization to a large variety of clinical text notes. The resulting system – DECOVRI (Data Extraction for COVID-19 Related Information) – is presented below and released as open-source software.

All extracted information is exported using the OMOP common data model (CDM). The general architecture was split across multiple machines for performance reasons with four primary machine clusters: the database, uimaFIT, Keras, and log monitoring. Custom SSL connections are used between uimaFIT and the Python-based (Keras) Recurrent Neural Network. DECOVRI automatically runs daily (cron job) to process any notes newly added to MUSC's COVID Datamart and older notes flagged for reprocessing. Downstream uses for this data initially included a purely data-driven symptom "checker" giving testing recommendations to patients worrying about COVID-19 [3, 4]. Later uses included predicting SARS-CoV-2 test results to then enable more efficient testing with pooled samples.

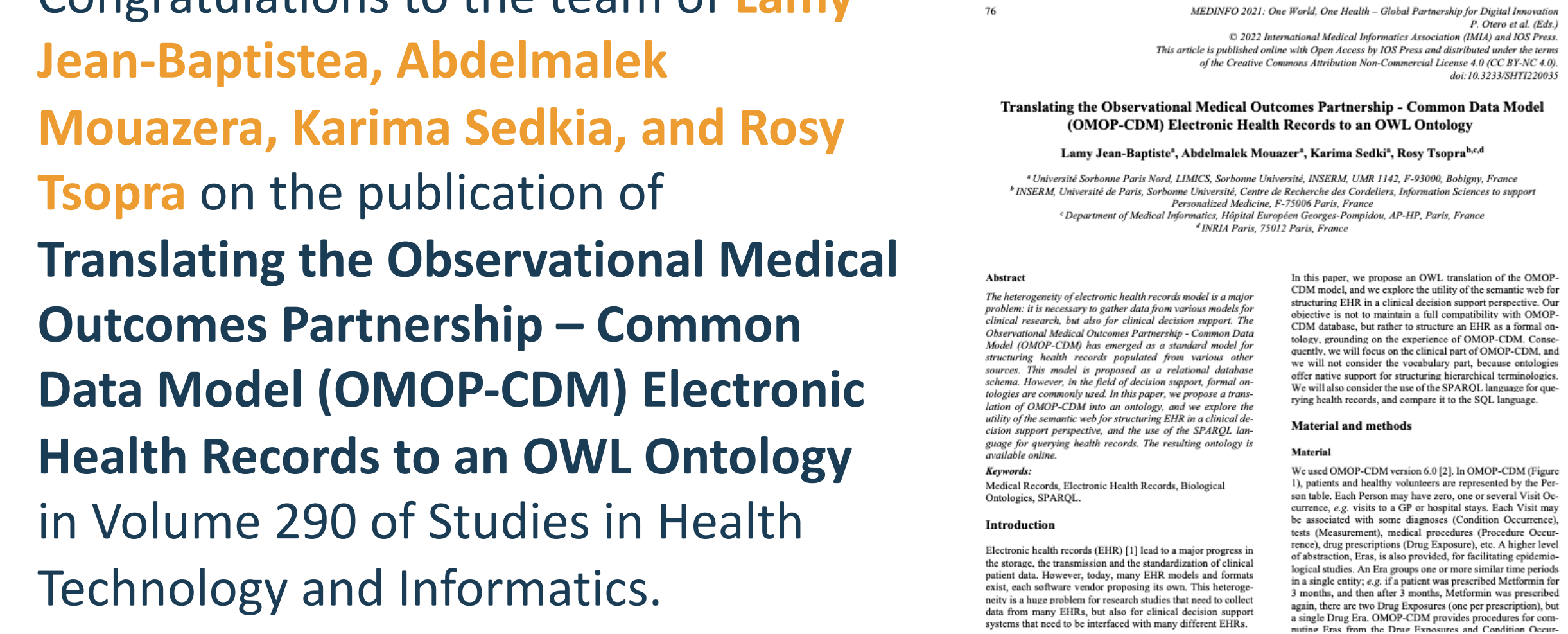
For testing, an initial small corpus of 15 randomly selected clinical notes was annotated by domain experts and used to evaluate the information extraction accuracy of the prototype. A new larger reference standard of 400 clinical notes for training and further evaluations is currently in development.

Results

The initial prototype (alpha version) of DECOVRI was assembled between March 16th and March 26th, 2020. Every week through the end of April, new modules were added or pre-existing modules were upgraded. Module upgrades included changes required by our shifting understanding of COVD-19 (e.g., adding new symptoms to lexicons), changes required by internal technical forces (e.g., the sectionizer had to adapt to new standard templates), and changes brought about by performance bottlenecks (e.g., the watermarking system and batch reprocessing logic had to be adjusted as caseloads and note backlogs grew). Features were integrated across modules, as well. For instance, section and negation information impacted which

OHDSI Shoutouts!

Congratulations to the team of **Lamy Jean-Baptistea, Abdelmalek Mouazera, Karima Sedkia, and Rosy Tsopra** on the publication of **Translating the Observational Medical Outcomes Partnership – Common Data Model (OMOP-CDM) Electronic Health Records to an OWL Ontology** in Volume 290 of Studies in Health Technology and Informatics.



76

MEDINFO 2021: One World, One Health – Global Partnership for Digital Innovation
P. Otero et al. (Eds.)

doi:10.3233/SHTI220035

Translating the Observational Medical Outcomes Partnership - Common Data Model (OMOP-CDM) Electronic Health Records to an OWL Ontology

Lamy Jean-Baptiste[a], Abdelmalek Mouazer[a], Karima Sedki[a], Rosy Tsopra[b,c,d]

[a] Université Sorbonne Paris Nord, LIMICS, Sorbonne Université, INSERM, UMR 1142, F-93000, Bobigny, France
[b] INSERM, Université de Paris, Sorbonne Université, Centre de Recherche des Cordeliers, Information Sciences to support Personalized Medicine, F-75006 Paris, France
[c] Department of Medical Informatics, Hôpital Européen Georges-Pompidou, AP-HP, Paris, France
[d] INRIA Paris, 75012 Paris, France

Abstract

The heterogeneity of electronic health records model is a major problem: it is necessary to gather data from various models for clinical research, but also for clinical decision support. The Observational Medical Outcomes Partnership - Common Data Model (OMOP-CDM) has emerged as a standard model for structuring health records populated from various other sources. This model is proposed as a relational database schema. However, in the field of decision support, formal ontologies are commonly used. In this paper, we propose a translation of OMOP-CDM into an ontology, and we explore the utility of the semantic web for structuring EHR in a clinical decision support perspective, and the use of the SPARQL language for querying health records. The resulting ontology is available online.

Keywords:
Medical Records, Electronic Health Records, Biological Ontologies, SPARQL.

Introduction

Electronic health records (EHR) [1] lead to a major progress in the storage, the transmission and the standardization of clinical patient data. However, today, many EHR models and formats exist, each software vendor proposing its own. This heterogeneity is a huge problem for research studies that need to collect data from many EHRs, but also for clinical decision support systems that need to be interfaced with many different EHRs.

In this paper, we propose an OWL translation of the OMOP-CDM model, and we explore the utility of the semantic web for structuring EHR in a clinical decision support perspective. Our objective is not to maintain a full compatibility with OMOP-CDM database, but rather to structure an EHR as a formal ontology, grounding on the experience of OMOP-CDM. Consequently, we will focus on the clinical part of OMOP-CDM, and we will not consider the vocabulary part, because ontologies offer native support for structuring hierarchical terminologies. We will also consider the use of the SPARQL language for querying health records, and compare it to the SQL language.

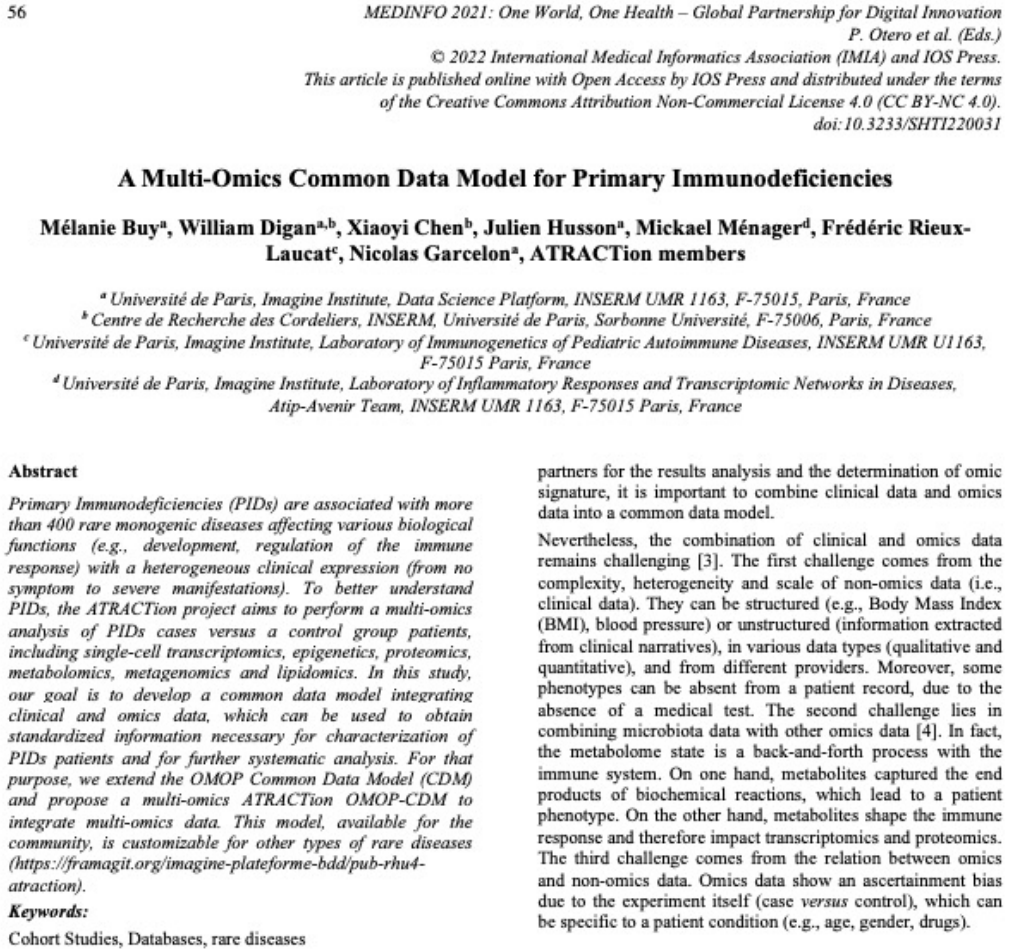
Material and methods

Material

We used OMOP-CDM version 6.0 [2]. In OMOP-CDM (Figure 1), patients and healthy volunteers are represented by the Person table. Each Person may have zero, one or several Visit Occurrence, *e.g.* visits to a GP or hospital stays. Each Visit may be associated with some diagnoses (Condition Occurrence), tests (Measurement), medical procedures (Procedure Occurrence), drug prescriptions (Drug Exposure), etc. A higher level of abstraction, Eras, is also provided, for facilitating epidemiological studies. An Era groups one or more similar time periods in a single entity; *e.g.* if a patient was prescribed Metformin for 3 months, and then after 3 months, Metformin was prescribed again, there are two Drug Exposures (one per prescription), but a single Drug Era. OMOP-CDM provides procedures for computing Eras from the Drug Exposures and Condition Occur-

OHDSI Shoutouts!

Congratulations to the team of **Mélanie Buy, William Digan, Xiaoyi Chen, Julien Husson, Mickael Ménager, Frédéric Rieux Laucat, Nicolas Garcelon, and ATRACTion members** on the publication of **A Multi-Omics Common Data Model for Primary Immunodeficiencies** in Volume 290 of Studies in Health Technology and Informatics.



MEDINFO 2021: One World, One Health – Global Partnership for Digital Innovation
P. Otero et al. (Eds.)

doi:10.3233/SHTI220031

A Multi-Omics Common Data Model for Primary Immunodeficiencies

Mélanie Buy[a], William Digan[a,b], Xiaoyi Chen[b], Julien Husson[a], Mickael Ménager[d], Frédéric Rieux-Laucat[c], Nicolas Garcelon[a], ATRACTion members

[a] Université de Paris, Imagine Institute, Data Science Platform, INSERM UMR 1163, F-75015, Paris, France
[b] Centre de Recherche des Cordeliers, INSERM, Université de Paris, Sorbonne Université, F-75006, Paris, France
[c] Université de Paris, Imagine Institute, Laboratory of Immunogenetics of Pediatric Autoimmune Diseases, INSERM UMR U1163, F-75015 Paris, France
[d] Université de Paris, Imagine Institute, Laboratory of Inflammatory Responses and Transcriptomic Networks in Diseases, Atip-Avenir Team, INSERM UMR 1163, F-75015 Paris, France

Abstract

Primary Immunodeficiencies (PIDs) are associated with more than 400 rare monogenic diseases affecting various biological functions (e.g., development, regulation of the immune response) with a heterogeneous clinical expression (from no symptom to severe manifestations). To better understand PIDs, the ATRACTion project aims to perform a multi-omics analysis of PIDs cases versus a control group patients, including single-cell transcriptomics, epigenetics, proteomics, metabolomics, metagenomics and lipidomics. In this study, our goal is to develop a common data model integrating clinical and omics data, which can be used to obtain standardized information necessary for characterization of PIDs patients and for further systematic analysis. For that purpose, we extend the OMOP Common Data Model (CDM) and propose a multi-omics ATRACTion OMOP-CDM to integrate multi-omics data. This model, available for the community, is customizable for other types of rare diseases (https://framagit.org/imagine-plateforme-bdd/pub-rhu4-atraction).

Keywords:
Cohort Studies, Databases, rare diseases

partners for the results analysis and the determination of omic signature, it is important to combine clinical data and omics data into a common data model.

Nevertheless, the combination of clinical and omics data remains challenging [3]. The first challenge comes from the complexity, heterogeneity and scale of non-omics data (i.e., clinical data). They can be structured (e.g., Body Mass Index (BMI), blood pressure) or unstructured (information extracted from clinical narratives), in various data types (qualitative and quantitative), and from different providers. Moreover, some phenotypes can be absent from a patient record, due to the absence of a medical test. The second challenge lies in combining microbiota data with other omics data [4]. In fact, the metabolome state is a back-and-forth process with the immune system. On one hand, metabolites captured the end products of biochemical reactions, which lead to a patient phenotype. On the other hand, metabolites shape the immune response and therefore impact transcriptomics and proteomics. The third challenge comes from the relation between omics and non-omics data. Omics data show an ascertainment bias due to the experiment itself (case *versus* control), which can be specific to a patient condition (e.g., age, gender, drugs).

OHDSI Shoutouts!

Any shoutouts from the community? Please share and help promote and celebrate OHDSI work!

Have a study published? Please send to sachson@ohdsi.org so we can share during this call and on our social channels.
Let's work together to promote the collaborative work happening in OHDSI!

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	Common Data Model Vocabulary Subgroup
Tuesday	3 pm	OMOP CDM Oncology Outreach/Research Subgroup
Wednesday	7 am	Medical Imaging
Wednesday	9 am	FHIR and OMOP Data Model Harmonization Subgroup (ZOOM)
Wednesday	9 am	Africa Chapter
Wednesday	11 am	Open-Source Community
Wednesday	12 pm	Health Equity Journal Club
Wednesday	12 pm	FHIR and OMOP Terminologies Subgroup (ZOOM)
Thursday	12 pm	FHIR and OMOP Oncology Subgroup
Thursday	12 pm	HADES
Friday	9 am	GIS – Geographic Information System
Friday	9 am	Education
Monday	10 am	Healthcare Systems Interest Group

www.ohdsi.org/upcoming-working-group-calls

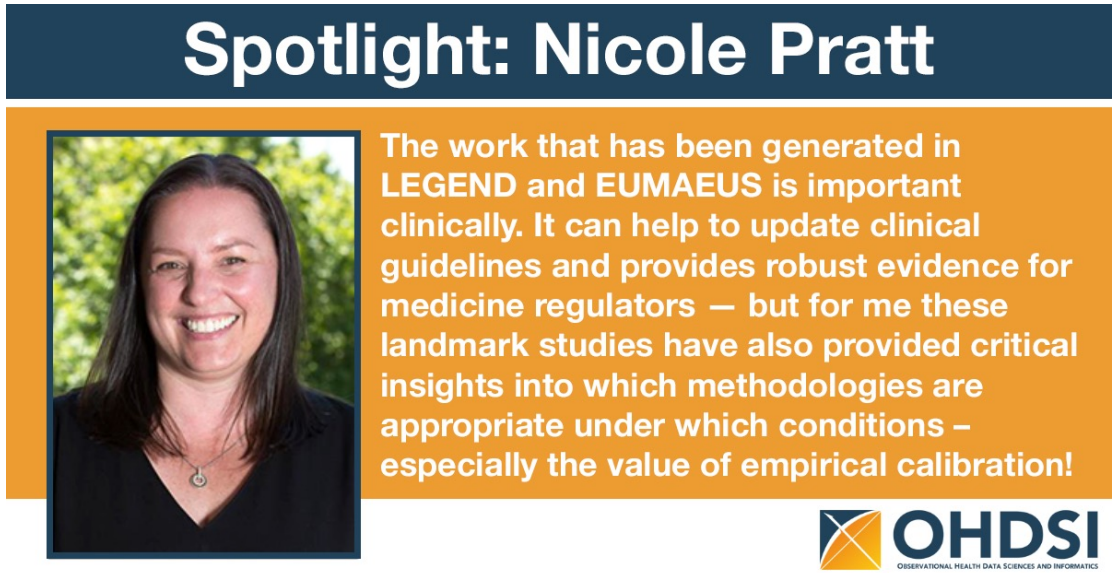
#OHDSI2022 Collaborator Showcase

We are now **10 days away** from the submission deadline for the 2022 OHDSI Global Symposium. All submissions for poster presentations, software demos and/or lightning talks are due no later than 8pm (EST) on **Friday, June 24**.

www.ohdsi.org/ohdsi2022collaboratorshowcase

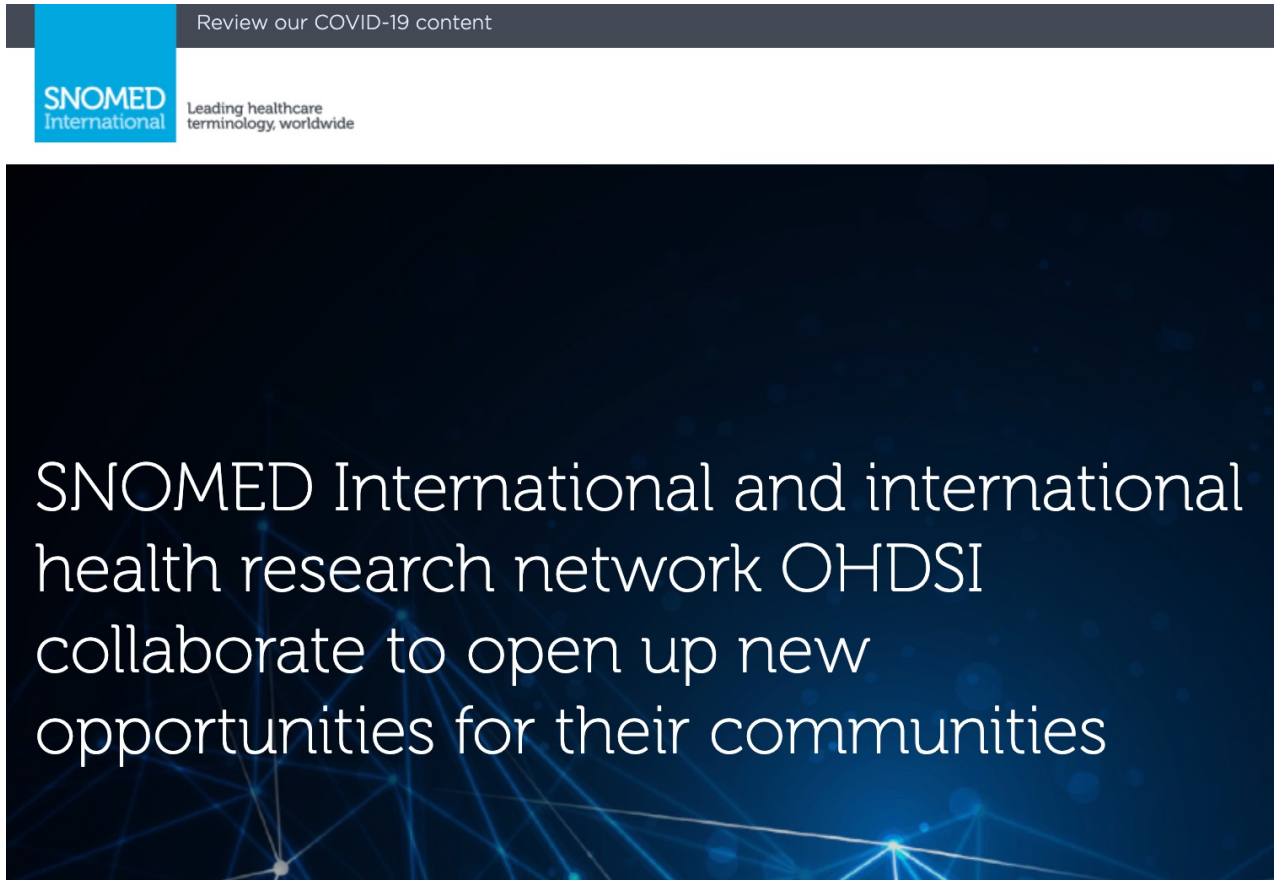
Collaborator Spotlight: Nicole Pratt

The latest edition of the Collaborator Spotlight features **Nicole Pratt,** whose interest in the effectiveness and safety of medicine use has led her to collaborate on network initiatives like **LEGEND** and **EUMAEUS**. She has also played an important role in the continued growth of the **APAC community**.



SNOMED, OHDSI Formalize Relationship

OHDSI and **SNOMED International** have formalized their long-time relationship with a **five-year collaborative agreement** that will benefit both of their user communities.



2022 European Symposium

www.ohdsi-europe.org/symposium-2022

Job Opening

Professor **Peter Rijnbeek** announced an opening for an epidemiologist to work with his team at Erasmus MC.

This position will be responsible for all aspects of observational research including protocol writing, input in the statistical analysis plan, study execution, interpretation of results and report/manuscript writing.

The application deadline is July 8, 2022.

Epidemiologist

Published	Deadline	Location
9 Jun	7 Jul	Rotterdam

JOB DESCRIPTION

This research will be performed in close collaboration with the Observational Health Data Sciences and Informatics (OHDSI) initiative, which is a global, multi-stakeholder, interdisciplinary collaborative to bring out the value of health data through large-scale analytics , and the EU-sponsored European Health Data and Evidence Network (EHDEN) which develop frameworks to generate reliable real-world evidence.

In your function as Epidemiologist you will be responsible for all aspects of observational research including protocol writing, input in the statistical analysis plan, study execution, interpretation of results and report/manuscript writing.

Job Opening

There is a new opening for a Postdoctoral Data Scientist within Dani Prieto-Alhambra's team at the University of Oxford.

This person would be involved with the work happening around both DARWIN EU and EHDEN.

The application deadline is June 27, and more information and the application link will be posted on the community calls page.

UNIVERSITY OF OXFORD

Postdoctoral Data Scientist

Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, Botnar Research Centre, Windmill Road, Oxford

We have an exciting opportunity for Postdoctoral Data Scientist to join a Pharmaco- and Device epidemiology research group led by Professor Daniel Prieto-Alhambra at the Botnar Research Centre, NDORMS, University of Oxford. The NDORMS Pharmaco- and Device epidemiology research group is involved in a number of national and international studies exploring the conditions of use (adherence, compliance, off and on-label use) of a number of licensed drugs, devices, and vaccines for the prevention and treatment of human disease in 'real world' (routine practice) conditions.

As a Postdoctoral Data Scientist you will develop analysis plans, protocols, ethical (and similar panel) submissions, governance and regulatory submissions as required for ongoing and future studies. You will generate and analyse OMOP-mapped real world health data assets, adapt existing and develop new research methodologies and materials. You will carry out collaborative research projects with colleagues in partner institutions and report research findings in the form of conference abstracts at national and international conferences.

You will hold a Doctoral (or be near completion) degree in informatics/information technology, engineering, statistics, biostatistics, mathematics, health data sciences or a related field. Demonstrable advanced skills and expertise in R programming, advanced skills in programming in Python, SQL, or similar languages and ability to work well within multi-disciplinary teams and independently are essential. Experience in propensity score/s, instrumental variable/s, and/or other methods to adjust for confounding for indication in pharmaco-epidemiological studies, experience in prediction modelling and good track record of peer reviewed scientific publications are desirable.

This is a full-time fixed-term appointment for 2 years.

The closing date for this position is 12 noon on 27 June 2022. You will be required to upload a CV and supporting statement as part of your online application

Contact Person :	HR Team, NDORMS	**Vacancy ID :**	158193
Contact Phone :		**Closing Date & Time :**	27-Jun-2022 12:00
Pay Scale :	STANDARD GRADE 7	**Contact Email :**	hr@ndorms.ox.ac.uk
Salary (£) :	Grade 7: £33,309 - £40,927 p.a.		

Where Are We Going?

Any other announcements of upcoming work, events, deadlines, etc?

Three Stages of The Journey

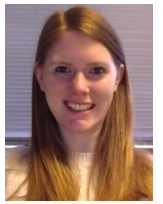
Where Have We Been?

Where Are We Now?

Where Are We Going?

June 14: OHDSI Publications in 2022

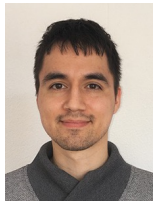
Learning patient-level prediction models across multiple healthcare databases: evaluation of ensembles for increasing model transportability
Presenter: Jenna Reps



Analysis of Dual Combination Therapies Used in Treatment of Hypertension in a Multinational Cohort
Presenter: Yuan Lu

Factors Influencing Background Incidence Rate Calculation: Systematic Empirical Evaluation Across an International Network of Observational Databases
Presenter: Anna Ostropolets

Logistic regression models for patient-level prediction based on massive observational data: Do we need all data?
Presenter: Henrik John

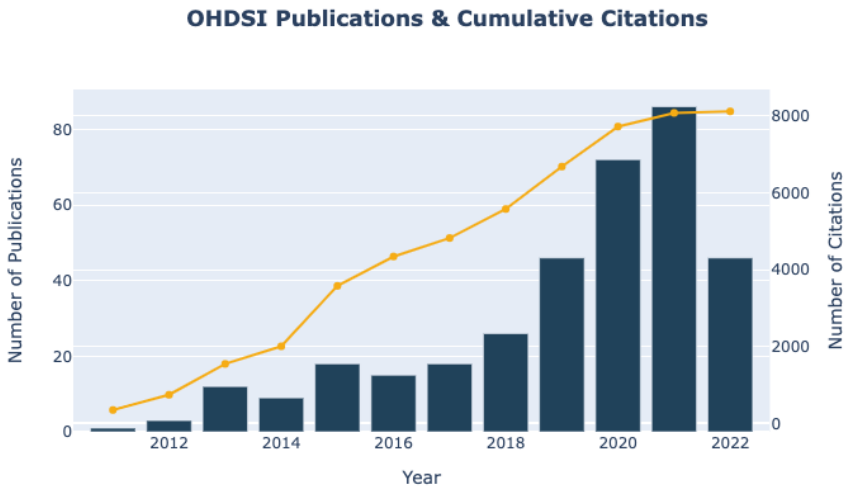


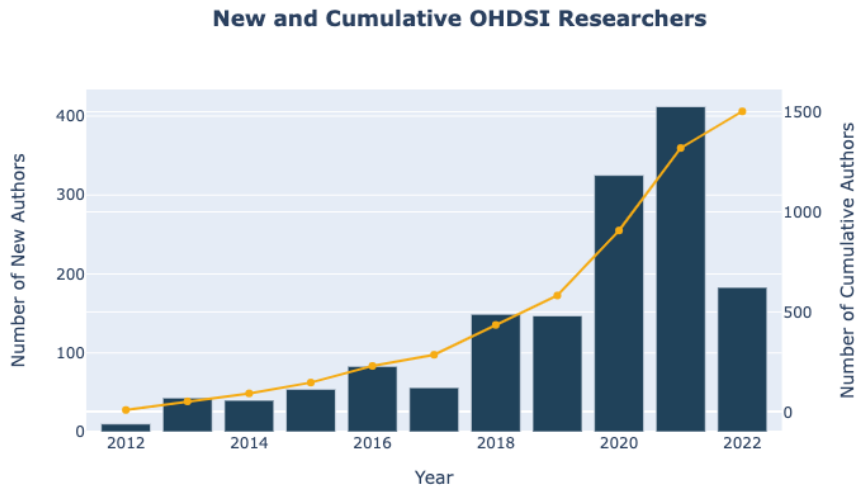
Prior-Preconditioned Conjugate Gradient Method for Accelerated Gibbs Sampling in "Large n, Large p" Bayesian Sparse Regression
Presenter: Aki Nishimura

Publication Dashboard

OHDSI Community Dashboard Dashboards

Publication Analysis





PubMed ID	Creation Date	Authors	Publication	Journal	MeSH Terms	Citation Count
filter data...						
35685531	2022/06/10	Pareen Vora, Henry Morgan Stewart, Beth Russell, + 2 authors, Gunnar Brobert	Time Trends and Treatment Pathways in Prescribing Individual Oral Anticoagulants in Patients with Nonvalvular Atrial Fibrillation: An Observational Study of More than Three Million Patients from Europe and the United States.	International journal of clinical practice	Not Yet Available	0