Building The OHDSI Evidence Network, Session II

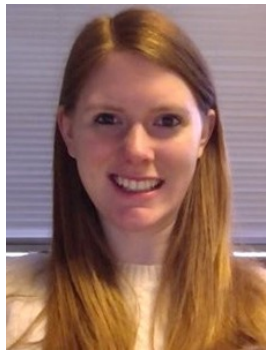
OHDSI Community Call

July 23, 2024 • 11 am ET

Upcoming Community Calls

Date	Topic
July 23	Building The OHDSI Evidence Network Sprint
July 30	Advances in Patient-Level Prediction
Aug. 6	Building The OHDSI Evidence Network Sprint
Aug. 13	Global Symposium Plenary and Tutorial Preview
Aug. 20	Building The OHDSI Evidence Network Sprint
Aug. 27	canceled due to ISPE 2024
Sept. 3	New Standardized Vocabularies Release
Sept. 10	Asia-Pacific Regional Updates

July 30: Patient-Level Prediction

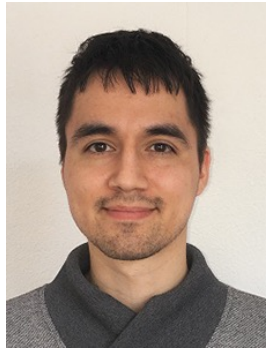


Jenna Reps

Johnson & Johnson

Chen Yanover

KI Research Institute



Henrik John

Erasmus MC

Alexander Saelmans

Erasmus MC

Egill Friðgeirsson

Erasmus MC

Ross Williams

Erasmus MC

Three Stages of The Journey

Where Have We Been?

Where Are We Now?

Where Are We Going?

OHDSI Shoutouts!

Congratulations to the team of **Tom Seinen, Jan Kors, Erik van Mulligen, and Peter Rijnbeek** on the publication of **Annotation-preserving machine translation of English corpora to validate Dutch clinical concept extraction tools** in *JAMIA*.

Journal of the American Medical Informatics Association, 2024, **31(8)**, 1725–1734
https://doi.org/10.1093/jamia/ocae159
Advance access publication 27 June 2024
Research and Applications

Research and Applications

Annotation-preserving machine translation of English corpora to validate Dutch clinical concept extraction tools

Tom M. Seinen, MSc*[,1], Jan A. Kors, PhD[1], Erik M. van Mulligen, PhD[1], Peter R. Rijnbeek, PhD[1]

[1]Department of Medical Informatics, Erasmus University Medical Center, 3015 GD Rotterdam, The Netherlands

***Corresponding author:** Tom M. Seinen, MSc, Department of Medical Informatics, Erasmus University Medical Center, Dr Molewaterplein 40, 3015 GD Rotterdam, The Netherlands (t.seinen@erasmusmc.nl)

Abstract

Objective: To explore the feasibility of validating Dutch concept extraction tools using annotated corpora translated from English, focusing on preserving annotations during translation and addressing the scarcity of non-English annotated clinical corpora.

Materials and Methods: Three annotated corpora were standardized and translated from English to Dutch using 2 machine translation services, Google Translate and OpenAI GPT-4, with annotations preserved through a proposed method of embedding annotations in the text before translation. The performance of 2 concept extraction tools, MedSpaCy and MedCAT, was assessed across the corpora in both Dutch and English.

Results: The translation process effectively generated Dutch annotated corpora and the concept extraction tools performed similarly in both English and Dutch. Although there were some differences in how annotations were preserved across translations, these did not affect extraction accuracy. Supervised MedCAT models consistently outperformed unsupervised models, whereas MedSpaCy demonstrated high recall but lower precision.

Discussion: Our validation of Dutch concept extraction tools on corpora translated from English was successful, highlighting the efficacy of our annotation preservation method and the potential for efficiently creating multilingual corpora. Further improvements and comparisons of annotation preservation techniques and strategies for corpus synthesis could lead to more efficient development of multilingual corpora and accurate non-English concept extraction tools.

Conclusion: This study has demonstrated that translated English corpora can be used to validate non-English concept extraction tools. The annotation preservation method used during translation proved effective, and future research can apply this corpus translation method to additional languages and clinical settings.

Key words: named entity recognition; clinical concept extraction; machine learning; natural language processing; text mining; corpus annotation.

Introduction

Electronic health records (EHRs) have become an invaluable source of real-world data for observational research, offering insights into disease prevalence, patient outcomes, and treatment effectiveness.[1,2] While structured data, such as coded conditions, measurements, and prescriptions, are frequently used for analysis, a significant portion of valuable patient information remains locked within free text, such as nursing and physician notes.[3,4] The extraction of information from these unstructured data in a structured manner, such as standardized clinical concepts from the Unified Medical Language System (UMLS),[5] can greatly enhance observational

with the rise of real-world data utilization in observational clinical research across the multilingual continent of Europe,[17] as seen in initiatives like the European Medical Information Framework (EMIF),[18] the European Health Data & Evidence Network (EHDEN),[19] and the Data Analytics and Real World Interrogation Network (DARWIN EU).[20] Utilizing unstructured data in large-scale analyses within standardized frameworks, such as the Observational Medical Outcomes Partnership Common Data Model (OMOP CDM),[21,22] highlights the importance of reliable information extraction for different languages. Nevertheless, the landscape of concept extraction tools for relatively small

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	12 pm	Latin America
Wednesday	4 pm	Vulcan/OHDSI
Wednesday	7 pm	Medical Imaging
Thursday	9:30 am	Network Data Quality
Thursday	7 pm	Dentistry
Friday	9 am	Phenotype Development & Evaluation
Friday	10 am	GIS-Geographic Information System
Friday	11:30 am	Steering Group
Monday	9 am	Vaccine Vocabulary
Monday	10 am	CDM Survey Subgroup
Monday	10 am	Healthcare Systems Interest Group
Tuesday	12 pm	CDM Vocabulary Subgroup

OHDSI2024 Conference Agenda

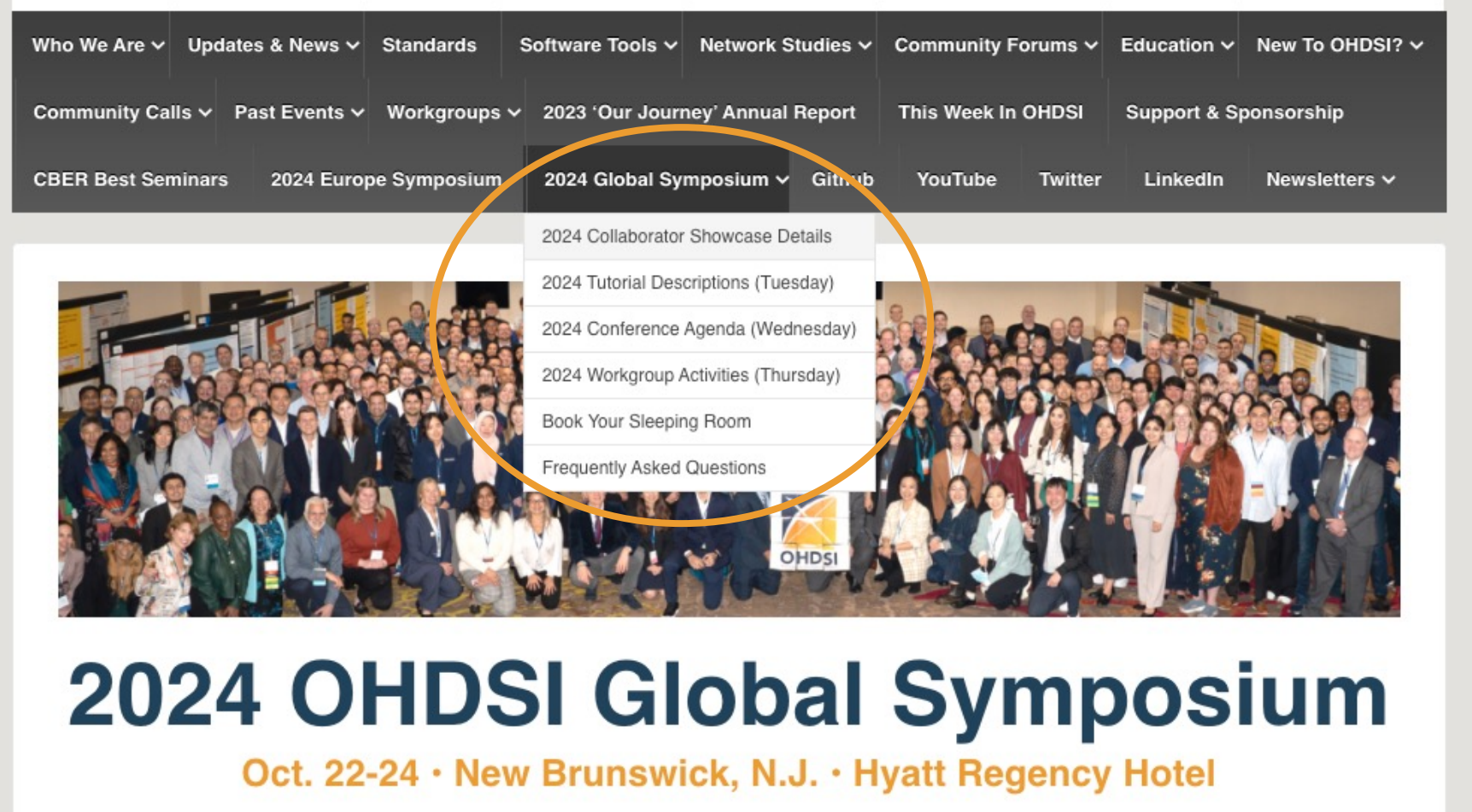
Agenda • Wednesday, Oct. 23

Time (ET)	Topic (Presenters)
7:30 - 8:30 am	**Registration and Lite Breakfast**
8:30 - 9:15 am	**State of the OHDSI Community** (George Hripcsak, Columbia Univ.)
9:15 - 10:15 am	**Plenary: Clinical Insights from LEGEND-T2DM** Introduction to LEGEND-T2DM (Moderator: Aline Pedroso, Brazil) Comparative Effectiveness of Second-line Antihyperglycemic Agents (Arya Aminorroaya, Yale Univ.) Effectiveness of First-line Antihyperglycemia Agents (Phyllis Thangaraj, Yale Univ.) Comparative Safety of SGLT2 for Risk of Diabetic Ketoacidosis (Hannah Yang/Evan Minty, Univ. of Calgary) Comparative Safety of GLP1-RA and the Risk of Thyroid Tumors (Daniel Morales, Univ. of Dundee)
10:15 - 10:35 am	**Networking Break**
10:35 - 11:20 am	**Plenary: Value Proposition for Participating in OHDSI Network Studies like LEGEND-T2DM** Introduction to OHDSI Evidence Network / Marketplace (Moderator: Clair Blacketer, Johnson & Johnson) Reflections from US Department of Veterans Affairs (Scott Duvall, VA) Reflections from SIDIAP (Spain) (Talita Duarte-Salles, IDIAP) Reflections from Taipei Medical University (Thanh-Phuc Phan, Taipei Medical Univ.) Reflections from a Global Commercial Data Provider (Sarah Seager, IQVIA)
11:20 am - 12 pm	**Plenary Q&A: Lessons Learned on LEGEND-T2DM Journey** (Moderator: Fan Bu, Univ. of Michigan; Panelists: LEGEND-T2DM co-authors)
12 - 12:45 pm	**Lunch**
12:45 - 1:30 pm	**Plenary Panel: JACC-OHDSI Partnership** (Moderators: Nicole Pratt, Univ. of South Australia/Marc Suchard, UCLA; Panelists: Harlan Krumholz, Yale Univ./Seng Chan You, Yonsei Univ./ Yuan Lu, Yale Univ.)
1:30 pm - 2 pm	**Plenary Activity: OHDSI Scavenger Hunt - Form Your Network Study Dream Team**
2 pm - 3 pm	**Collaborator Showcase: Posters and Software Demos**
3 pm - 4 pm	**Collaborator Showcase: Lightning Talks**
4 pm - 5 pm	**Collaborator Showcase: Posters and Software Demos**
5 pm - 6 pm	**Closing Talk & Titan Awards** (Patrick Ryan, Johnson & Johnson/Columbia Univ.)
6 pm - 7 pm	**Network Reception**

*agenda is subject to change

ohdsi.org/ohdsi2024

OHDSI2024 Homepage



ohdsi.org/ohdsi2024

#OHDSI2024 Registration Is Open!

Registration is OPEN for the 2024 OHDSI Global Symposium, which will be held **Oct. 22-24** at the **Hyatt Regency Hotel in New Brunswick, N.J., USA**.

Tuesday: Tutorials
Wednesday: Plenary/Showcase
Thursday: Workgroup Activities

ohdsi.org/OHDSI2024

The Center for Advanced Healthcare Research Informatics (CAHRI) at Tufts Medicine welcomes:

Melissa Haendel, PhD

Director of Precision Health & Translational Informatics and the Sarah Graham Kenan Distinguished Professor in the Department of Genetics at The University of North Carolina at Chapel Hill and co-founder of the Monarch Initiative and the National Covid Cohort Collaborative

‘Journeys across the translational divide: making healthcare and basic research data interoperable’

July 25, 2024, 11am-12pm EST
Virtually via Zoom

Please contact Marty Alvarez at malvarez2@tuftsmedicalcenter.org for calendar invite or questions.

TuftsMedicine
Tufts Medical Center

#OHDSISocialShowcase

MONDAY

External Validation of the Revised Cardiac Risk Index (RCRI) Clinical Prediction Model in Observational Health Care Databases

(Alexander Saelmans, Evan Minty, Peter Rijnbeek, Jenna Reps, Ross Williams)

External validation performance of the RCRI model suggests the need for an updated non-cardiac surgery postoperative risk model

External Validation of the Revised Cardiac Risk Index (RCRI) Clinical Prediction Model in Observational Health Care Databases

Background A more recent external validation of the model could inform clinicians better on the current performance of the model and give incentive to its improvement

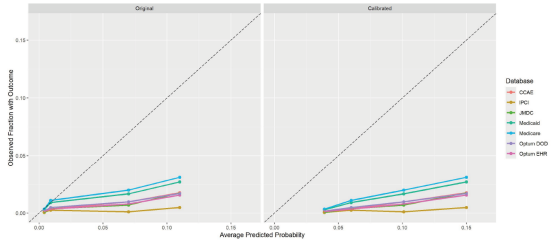
Result 1 Sample size, outcome and AUC of the RCRI model across different OMOP CDM databases*

*Recalibration did not influence discrimination

Result 2 Calibration plot of the original RCRI model across different OMOP CDM databases

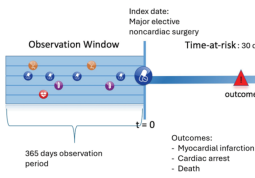
Result 3 Calibration plot of the calibrated RCRI model across different OMOP CDM databases

OMOP CDM Databases	Target cohort (n)	Outcome (%)	AUC
IPCI	68,202	0.0894	0.625
Medicaid	193,393	1.26	0.697
Medicare	757,956	1.38	0.699
Optum DOD	1,173,505	0.580	0.725
Optum EHR	281,389	0.404	0.747
CCAE	1,276,210	0.327	0.783
JMDC	38,795	0.291	0.795



Methods

1 Design of the RCRI model



2 Calculation of the RCRI score

RCRI predictors	Score
Elevated-risk surgery	1
History of ischemic heart disease	1
History of congestive heart failure	1
History of cerebrovascular disease	1
Pre-operative treatment with insulin	1
Pre-operative creatinine >2mg/dL/176.8µmol/L	1

RCRI score	Risk percentage	
	Original RCRI (1)	Calibrated RCRI (2)
0	0.4	3.9
1	0.9	6.0
2	7	10.1
≥3	11	15

References

1. Lee TH, et al. Derivation and prospective validation of a simple index for prediction of cardiac risk of major noncardiac surgery. Circulation. 1999;100(10):1043–9

2. Duceppe E, et al. Canadian cardiovascular society guidelines on perioperative cardiac risk assessment and management for patients who undergo noncardiac surgery. Can J Cardiol. 2017;33(1):17–32

Erasmus MC Health Data Science

Alexander Saelmans, Evan Minty, Peter Rijnbeek, Jenna Reps, Ross Williams

#OHDSISocialShowcase

TUESDAY

Conversion of the Papageorgiou General Hospital EHR to the OMOP Common Data Model

(Papapostolou Grigoris, Chytas Achilleas, Rekkas Alexandros, Bigaki Maria, Zeimpekis Demetrios, Dermentzoglou Lampros, Tortopidis George, Natsiavas Pantelis)

SNOMED overhaul and its impact on ETL and phenotyping

Masha Khitrun[1], Alexander Davydov[1], Oleg Zhuk[1]
[1]Odysseus Data Services Inc., Cambridge, MA

Background: Over the years, the Vocabulary team has been working on integrating SNOMED CT into the ecosystem of OHDSI Standardized vocabularies. However, due to its comprehensive structure, multiple adjustments to SNOMED vocabulary ETL logic and interventions on the content level have been necessary, leading to the accumulation of bugs and discrepancies over the years. The SNOMED load_stage script that integrates the SNOMED into the OMOP vocabularies, has grown larger and more complex than anticipated, resulting in significant delays of OHDSI releases and a time lag between the OMOP version of SNOMED and SNOMED sources.
We present the results of a comprehensive overhaul of SNOMED in OHDSI vocabularies. This overhaul included both technical changes to the load_stage, aimed at simplifying future releases, and content changes designed to optimize cohort creation and ETL process.

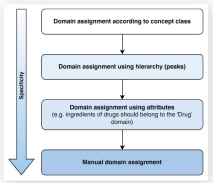
Methods: The vocabulary development follows the guiding principles outlined in the Book of OHDSI[1], ensuring adherence to established standards and practices within the OHDSI framework. Both developer[2] and end-user[3] documentation is maintained and made publicly accessible on GitHub, allowing for transparency and collaboration within the community.
To assess the impact of vocabulary changes on ETL processes, we conducted a comprehensive analysis leveraging completed and ongoing ETL projects. This analysis provided insights into the challenges posed by vocabulary modifications.
Recognizing the significant impact of vocabulary changes on ETL[4], we employ analytical methods tailored to mitigate these challenges. These methods[5], accompanied by our internal quality control approach[6] which includes the collection of vocabulary statistics, and a bunch of specific vocabulary checks[7] enable us to address the implications of vocabulary updates on ETL workflows and phenotyping proactively.

Table 1. Main changes implemented during the SNOMED overhaul and their estimated impact on ETL and Phenotyping

Changes	No. of concepts involved*	Impact on ETL	Impact on Phenotypes
Domain changes	24021	Low	High
De-standardization of concepts	3533	High	Low**
Split of pre-coordinated measurements and allergies	5357	Medium	High
Mapping of secondary neoplasms to Cancer Modifier	1042	High	High
Mapping following replacement links	9393	Low	Low
Retirement of the UK Drug Extension Module	464010	Possible	Low

* Overall number of SNOMED concepts: 1 084 286
**De-standardized concepts represent non-clinical events

Figure 1. Domain assignment algorithm in SNOMED



Results:

- **Domain assignment was improved** in its stability and consistency (Figure 1). As a result of this change, you may need to change the tables of interest (eg. querying **Condition_occurence instead of Observation**) in the process of cohort creation. However, **semantic "grey zones"** still exist, where the domain assignment is a matter of debate due to the **ambiguity of concept interpretation**. Domain flows (Table 2) in these grey zones were discovered using analytical methods[5], and domain improvement here **is a constant iterative process**.
- **Pre-coordinated SNOMED measurements** were moved **from the Condition to the Measurement** domain and splitted **in a post-coordinated way**. Thus, if you have previously used pre-coordinated SNOMED measurements, you should now start looking into the **Measurement** table where these concepts live as the Measurement / Value pairs.
- **Secondary neoplasms** were mapped to Standard concepts in **Cancer Modifier** vocabulary that belong to the **Measurement** domain. Thus, now you should look into the **Measurement** table to find the Secondary neoplasm concepts.
- In the course of the overhaul **110 SNOMED concepts in the Measurement domain**, mainly representing vital sign measurements, **were mapped to the Standard LOINC concepts.** These mappings are **erroneous** as SNOMED has a higher position in the hierarchy than LOINC, and **they may affect the hierarchy of measurements**. These issue is supposed to be solved in the course of the next release.
- We improved the **creation of 'Maps to'** relationships following the SNOMED sources' **replacement links**. As a result of this change, more concepts are now mapped to Standard (Figure 3), and the **number of events in cohorts may increase**.
- Concepts that belong to **Attribute, Location** (except countries), **Social Context** (except concepts that carry the semantics of relatives, religion, occupation), **Physical Force**, and **Physical Object** (except concepts in the Device domain) concept classes have been de-Standardized in the course of the overhaul.
- We have performed the **retirement of the UK Drug Extension** module aimed to declutter the vocabulary, with concepts deprecated and linked to their equivalents in the dm+d vocabulary.[8]

Figure 2. Number of concepts with mapping compared with the upgraded concepts

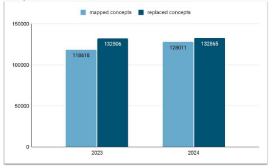


Table 2. Changes of domains for Standard SNOMED concepts over 2020-2024 years

2020	2021	2022	2023	2024	count
Condition	Condition	Condition	Condition	Observation	14216
Observation	Observation	Observation	Observation	Meas Value	2796
Observation	Observation	Measurement	Measurement	Measurement	1978
Observation	Observation	Observation	Observation	Measurement	1589
Procedure	Procedure	Procedure	Procedure	Observation	1308
Condition	Condition	Condition	Condition	Measurement	847
Observation	Observation	Observation	Language	Language	834
Observation	Observation	Observation	Observation	Procedure	614
Observation	Observation	Drug	Drug	Observation	500
Procedure	Procedure	Procedure	Procedure	Measurement	410
Condition	Condition	Observation	Observation	Observation	259

Conclusion:

The overhaul of the SNOMED vocabulary in OMOP has yielded significant improvements in ontology structure, cohort creation, and mapping efficiency. These enhancements contribute to more accurate data analysis, better research outcomes, and increased interoperability within the healthcare ecosystem. However, it is essential to consider their impact on ETL processes and phenotyping algorithms. The adjustments made to domain assignments and concept mappings may require updates to existing ETL workflows, and researchers should carefully review their phenotyping algorithms to ensure compatibility with the updated vocabulary structure and content.

References:

1. Observational Health Data Sciences and Informatics. The Book of OHDSI.
2. https://github.com/OHDSI/Vocabulary-v5.0/tree/master/SNOMED
3. https://github.com/OHDSI/Vocabulary-v5.0/wiki/Vocab.-SNOMED
4. https://forums.ohdsi.org/t/cpt-hierarchy-errors-lost-children-in-2023-and-changed-domains/18383
5. Dmitry Dymshyts, Frank DeFalco, Anthony Molinaro, Clair Blacketer. An Evaluation and maintenance of cohorts and concept sets in the OMOP Vocabulary Evolution. July 2023, Conference: OHDSI European Symposium 2023.
6. https://github.com/OHDSI/Vocabulary-v5.0/tree/master/working/packages/QA_TESTS
7. https://github.com/OHDSI/Vocabulary-v5.0/blob/master/working/manual_checks_after_generic_update.sql
8. https://github.com/OHDSI/Vocabulary-v5.0/releases/tag/v20240229_1709217174.000000

#OHDSISocialShowcase

WEDNESDAY

Automated OMOP-CDM pipeline for the new EBMT Registry

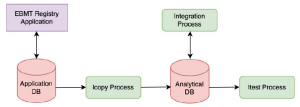
(Shirah Cashriel, Nir Assaraf, Ignacio Garcia, Fernando Cid, Freija Descamps)

Integrating a templated configurator into the ETL process

Automated OMOP-CDM pipeline for the new EBMT Registry

Background: The configurator is a web application solution that templates the process of converting the EBMT Registry's complex data-entry forms base into OMOP-CDM. This solution allows for new configurations to be easily entered into the application through the user interface and applied to the Extract-Transform-Load (ETL).

The user enters the new configuration through the EBMT Registry's application user interface. The new configuration is saved to the application database. The nightly pipeline copies the relevant patient data and the configurations to the analytical database (icopy process), and uses the configurations during the ETL to map the patient data to OMOP-CDM format (integration process), and stored on the analytical database. The process is then tested for quality assurance (itest process).



Methods

1 Configuration data for each field that is recorded is stored in the Field table (1a). The Data Service Option table (1b) stores configuration data for fields with pre-fixed options. Patient data is stored in the Field Response table (1c).

Field

id	integrationConfig	eventTypeId	label
field1	{'references': {'omop_start_date':'field1'}, 'directs': {'concept_id': 4266367}}	event_type1	Date of diagnosis
field2	{'references': {'omop_start_date': field1}, 'data_sources': {'concept_id': data_service1}}	event_type1	Type of flu
field3	{'references': {'omop_start_date':field1, 'value_as_number':field3}, 'directs': {'concept_id': 4302666, 'unit_source_value': 'celsius'}}	event_type1	Body temperature

1a

Data Service Option

id	dataServiceId	value	dataConfig
option1	data_service1	1	{'standard_concept_id': 1111}
option2	data_service1	2	{'standard_concept_id': 2222}

1b

Field Response

id	patientId	patientEventId	fieldId	value
response1	patient1	patient_event1	field1	2008-08-08
response2	patient1	patient_event1	field2	1
response3	patient1	patient_event1	field3	38.7

1c

Stem

id	omop_start_date	concept_id	value_as_number	unit_source_value	patient_event_id	person_id
stem1	2008-08-08	426637			patient_event1	person1
stem2	2008-08-08	1111			patient_event1	person1
stem3	2008-08-08	4302666	38.7	celsius	patient_event1	person1

2 The ETL process first builds an intermediary STEM table with all the necessary data for each field. Using the configurations in the Field table, the process captures the date of each data point, even when it is not included in the Field Response.

3 The ETL process then joins the STEM and Concept tables on concept_ids to retrieve the domain and populates the associated clinical table.

Measurement

measurement_concept_id	measurement_date	value_as_number	unit_source_value	person_id	patient_event_id
4302666	2008-08-08	38.7	celsius	person1	patient_event1

Condition Occurrence

condition_concept_id	condition_start_date	person_id	patient_event_id
4266367	2008-08-08	person1	patient_event1
1111	2008-08-08	person1	patient_event1

Conclusion: Templating mapping configurations via user interface keeps the OMOP-CDM at the forefront when expanding data collection and allows for seamless expansion of the application database without needing to update the code, pipeline, or ETL process. As a result, this pipeline can be maintained by any non-technical team member that is familiar with OMOP-CDM.

Shirah Cashriel[1], Nir Assaraf[1], Ignacio Garcia[2], Fernando Cid[2], Freija Descamps[1]

[1]edenceHealth
[2]EBMT Registry

#OHDSISocialShowcase

THURSDAY

Incorporating Temporal Information from EHR Data in Clinical Prediction Modelling

(Estelle Lampel, Aniek Markus, Tom Seinen)

Incorporating Temporal Information from EHR Data in Clinical Prediction Modelling

Can you increase the predictive performance of models using binning methods or temporal weights?

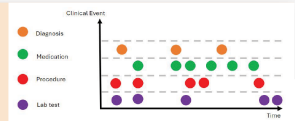
Background: Electronic Health Record (EHR) data is considered both to have a lot of potential for clinical prediction modelling, yet complex and challenging to model. These complexities have led most researchers to bin covariates across a whole observation period or parts of an observation period, ignoring temporal information present in the data.

Temporality of EHR data

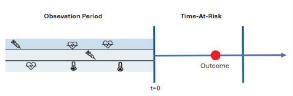
Sparsity
A patient goes to a clinician for a certain problem which is related to specific covariates, so not all covariates are recorded at every visit.

Irregular Intervals
Patients do not attend clinical visits regularly, meaning data is not sampled at regular intervals. But, increased visits may hint towards a patient's health worsening.

High-dimensionality
In each visit different combinations of features are measured. These features are measured at different levels of granularity, and multiple features can represent the same clinical concept.



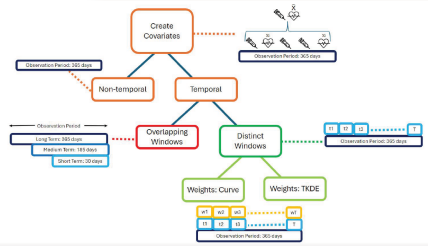
This is a tabular representation of EHR data, for a singular patient. The distance between clinical events are not regular.



Patient Level Prediction (hospital readmission)

Index date (t=0) (hospital admission)
The point at which one predicts if an outcome will occur.
Observation Period (one year prior admission)
Where predictors are observed, albeit asynchronously and inconsistently.
Time-At-Risk (one month post admission)
Post the index date, in which there is or is not a predicted outcome.

Methods



Binning Strategies

One Window
The observation period acts as one window

Overlapping Window
One or more windows that are anchored at the index date.

Distinct Window
T distinct windows over the observation period.

Weighting Strategies

Knowledge Based
Temporal discounting based on covariate. Non-chronic related covariates are discounted faster than chronic related covariates.

Probabilistic Weights
A probability density function is fitted to each covariate, this gives the probability of a covariate occurring. These probabilities are used as weights.

Learned Weights
Coefficients estimated through a regression are normalised and combined with reciprocal temporal discounting.

Estelle Lampel
Supervisor(s):
Aniek Markus and Tom Seinen

Health Data Science
Erasmus MC

#OHDSISocialShowcase

FRIDAY

Defining international approaches for the detection of emergent metastasis and the classification of location of metastasis from hospital EHR

(Stelios Theophanous, Sue Cheeseman, Elin Hallan Naderi, Elisabeth Ross, Anne-Lore Bynens, Prabash Galgane Banduge, Petros Kalendralis, Aiara Lobo Gomes, Piers Mahon)

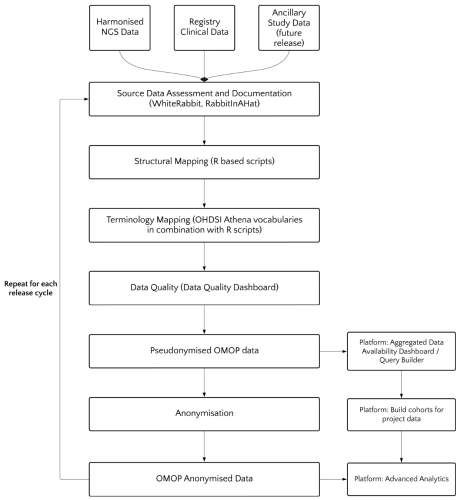
WAYFIND-R

Empowering research requires **seamless** delivery of **high quality data.** The WAYFIND-R® platform enables automation and **accelerates insights generation** from primary data collection to research-ready data.

Title: Empowering research with seamless data flow and research-ready, anonymised data in OMOP CDM: Learnings from the design of WAYFIND-R, a global precision oncology registry and research platform

Background: WAYFIND-R is a global precision oncology registry (NCT04529122) and has the aim to advance science and provide the scientific community worldwide with access to real-world data, enabling epidemiological and clinical research, and collaborations across research groups. The WAYFIND-R® Data Sharing and Collaboration Platform enables researchers to access anonymised clinico-genomic data from the registry transformed to the OMOP CDM within a secure research environment.

Methods



Outcomes

- Researchers able to access quality OMOP anonymised aggregate and individual-level data from registry
- Automation of concept mapping simplifies review process on an ongoing basis
- Defined data quality processes allow for routine checking of data and actions for each data release cycle. For example queries raised within clinical database
- ETL automation ensures consistency and reproducible output
- Changes to the structure registry clinical database require ongoing assessments to determine impact to data and platform
- Implementation of cohort definition within platform

Opportunities for Collaboration

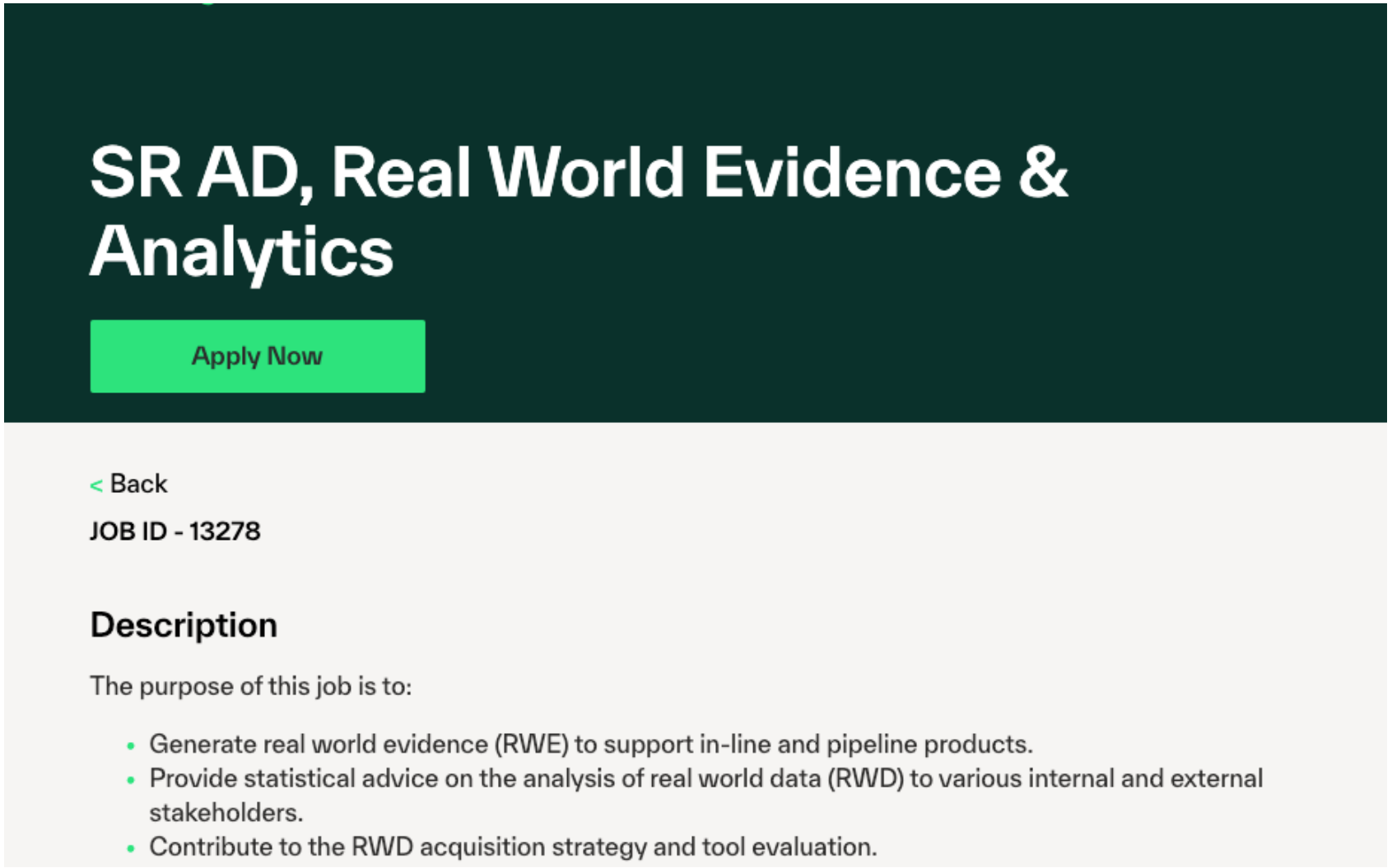
- Gene and biomarker ontologies
- Oncology extension
- Registry studies - our learnings

Acknowledgements:
We thank the patients and their families who take part in WAYFIND-R, as well as the staff, research coordinators, and investigators at each participating institution.

Tom Stone[1], Yuri Pyatkin[2], Ana Ferro[1], Dimitar Toshev[2]
[1] Roche Products Limited, Welwyn Garden City, UK; [2] F. Hoffmann-La Roche Ltd, Basel, Switzerland

Roche

Opening: Sr AD, Real World Evidence & Analytics Boehringer Ingelheim



Opening: Lead Director, RWE Distributed Research CVS Health

Lead Director, RWE Distributed Research

Apply

Hybrid

PA - Blue Bell
IL - Northbrook
CT - Hartford
RI - Woonsocket
AZ - Scottsdale

View All 7 Locations

Full time

Posted 6 Days Ago

R0268183

Bring your heart to CVS Health. Every one of us at CVS Health shares a single, clear purpose: Bringing our heart to every moment of your health. This purpose guides our commitment to deliver enhanced human-centric health care for a rapidly changing world. Anchored in our brand — with heart at its center — our purpose sends a personal message that how we deliver our services is just as important as what we deliver.

Our Heart At Work Behaviors™ support this purpose. We want everyone who works at CVS Health to feel empowered by the role they play in transforming our culture and accelerating our ability to innovate and deliver solutions to make health care more personal, convenient and affordable.

About us

CVS Health®

Our Work Experience is the combination of everything that's unique about us: our culture, our core values, our company meetings, our commitment to sustainability, our recognition programs, but most importantly, it's our people. Our

Read More

Openings: Postdoctoral Fellow, Johns Hopkins Univ.

JOHNS HOPKINS
BLOOMBERG SCHOOL
of PUBLIC HEALTH

PHARMACOEPIDEMIOLOGY POST-DOCTORAL TRAINING PROGRAM

Co-Directors: Caleb Alexander, MD, MS and Jodi Segal, MD, MPH

The **Pharmacoepidemiology Training Program** at the Johns Hopkins Bloomberg School of Public Health (BSPH) is currently **seeking to support postdoctoral fellows.** All supported trainees work with core faculty on existing or newly developed research projects on pharmacoepidemiology, so as to optimize the safe and effective use of medicines to treat heart, lung and blood diseases in the United States.

Deadline for applications: rolling

Where Are We Going?

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

Three Stages of The Journey

Where Have We Been?

Where Are We Now?

Where Are We Going?

July 23: Building The Evidence Network, Session II

Clair Blacketer

Director, Epidemiology Analytics,
Janssen Research & Development, Inc.

Paul Nagy

Deputy Director, Johns Hopkins Medicine Technology Innovation Center
Director of Education, Biomedical Informatics and Data Science Graduate Training Programs

- Technical description on what it takes to join the Evidence Network
- Live demo
- Language to support the IRB process
- More!

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