



Future Collaboration Opportunities

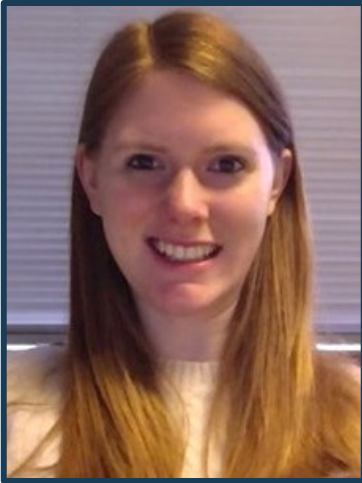
Breakouts: Methods Research | Data Standards |
Open-Source Development | Clinical Applications

OHDSI Community Call
Nov. 2, 2021 • 11 am ET



Nov. 9: Future Collaboration Opportunity Breakouts

Methods Research



Jenna Reps



Martijn Schuemie

Data Standards



Clair Blacketer

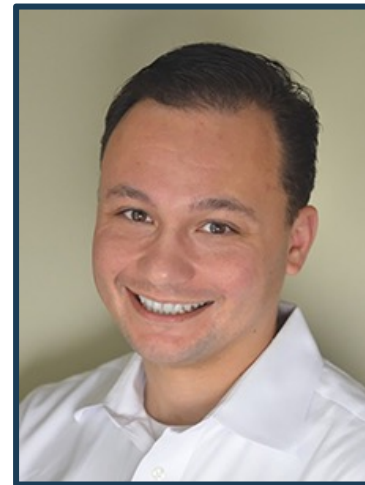


Maxim Moinat

Open-Source Development



Adam Black



Anthony Sena

Clinical Applications



Talita Duarte-Salles



Asieh Golozar



Remaining 2021 OHDSI Community Calls

Date	Topic
Nov. 2	Collaboration Opportunities: Methods Res., Data Standards, Open-Source, Clinical App.
Nov. 9	Demos: Tools for Adoption of OHDSI Data Standards
Nov. 16	Open Network Studies
Nov. 23	History of OHDSI
Nov. 30	Collaborator Showcase Presentations
Dec. 7	How Did We Do This Year? Final OKR Review
Dec. 14	Holiday-Themed Final Meeting Of 2021



Remaining 2021 OHDSI Community Calls

Date	Topic
Nov. 2	Collaboration Opportunities: Methods Res., Data Standards, Open-Source, Clinical App.
Nov. 9	Demos: Tools for Adoption of OHDSI Data Standards
Nov. 16	Open Network Studies
Nov. 23	History of OHDSI
Nov. 30	Collaborator Showcase Presentations
Dec. 7	How Did We Do This Year? Final OKR Review
Dec. 14	Holiday-Themed Final Meeting Of 2021



Nov. 9: Tools For Adoption Of OHDSI Data Standards



Clair Blacketer

Introduction To CDM 5.4



Michael Kallfelz

OHDSI Vocabularies



ETL Inspection Report



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	1 pm	Common Data Model
Tuesday	2 pm	Health Equity
Tuesday	3 pm	OMOP CDM Oncology – Outreach/Research Subgroup
Wednesday	2 am (4 pm KST)	Patient-Level Prediction/Population-Level Estimation (Eastern Hemi)
Wednesday	10 am	OMOP CDM Oncology – Development Subgroup
Wednesday	7 pm	Medical Imaging
Thursday	10 am	Medical Devices
Thursday	12 pm	Patient-Level Prediction/Population-Level Estimation (Western Hemi)
Thursday	1 pm	OMOP CDM Oncology – CDM/Vocabulary Subgroup
Friday	9 am	Education
Friday	9 am	Vaccine Vocabulary
Monday	8 am	Early-Stage Researchers (Europe/Western Hemisphere)
Monday	10 am	GIS-Geographic Information System
Tuesday	9 am	OMOP CDM Oncology – Genomic Subgroup

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



Get Access To Different Teams/WGs/Chapters



OHDSI

OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS

[Who We Are](#) [OHDSI Updates & News](#) [Standards](#) [Software Tools](#) [OHDSI Studies](#) [Book of OHDSI](#) [Resources](#) [New To OHDSI?](#)

[EHDSN Academy](#) [This Week In OHDSI](#) [2021 Global Symposium](#) [Events/Collaborations](#) [Collaborate in MSTeams](#) [Follow OHDSI](#)

Welcome to OHDSI!

The Observational Health Data Sciences and Informatics (or OHDSI, pronounced "Odyssey") program is a multi-stakeholder, interdisciplinary collaborative to bring out the value of health data through large-scale analytics. All our solutions are open-source.

OHDSI has established an international network

2020 OHDSI Symposium

Our 2020 OHDSI Global Symposium brought together a global research community for 18 hours of open science, international collaboration and community fun. The day included research presentations from community members, panels that brought together leaders from major healthcare organizations, as well as network sessions, the annual collaborator

5. Select the workgroups you want to join (you can refer to the WIKI for work group objectives www.ohdsi.org/web/wiki/doku.php?id=projects:overview)

- ☐ ATLAS
- ☐ Clinical Trials
- ☐ Common Data Model
- ☐ Data Quality Dashboard Development
- ☐ Early-stage Researchers
- ☐ Education Work Group
- ☐ Electronic Health Record (EHR) ETL
- ☐ Geographic Information System (GIS)
- ☐ HADES Health Analytics Data-to-Evidence Suite
- ☐ Health Equity
- ☐ Latin America
- ☐ Medical Devices
- ☐ Natural Language Processing
- ☐ OHDSI APAC
- ☐ OHDSI APAC Steering Committee
- ☐ OHDSI Steering Committee
- ☐ Oncology
- ☐ Patient-Generated Health Data
- ☐ Pharmacovigilance Evidence Investigation

- ☐ Phenotype Development and Evaluation
- ☐ Population-Level Effect Estimation / Patient-Level Prediction
- ☐ Psychiatry
- ☐ Registry (formerly UK Biobank)
- ☐ Surgery and Perioperative Medicine
- ☐ Vaccine Safety
- ☐ Vaccine Vocabulary
- ☐ Women of OHDSI

6. Select the chapter(s) you want to join

- ☐ Africa
- ☐ Australia
- ☐ China
- ☐ Europe
- ☐ Japan
- ☐ Korea
- ☐ Singapore
- ☐ Taiwan

7. Select the studies you want to join

- ☐ HERA-Health Equity Research Assessment
- ☐ PIONEER for Prostate Cancer (study-a-thon ended)
- ☐ SCYLLA (SARS-Cov-2 Large-scale Longitudinal Analyses)

Get Access To Different Teams/WGs/Chapters



The screenshot shows the OHDSI website with the following elements:

- Header:** OHDSI OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS
- Navigation Bar:** Who We Are, OHDSI Updates & News, Standards, Software Tools, OHDSI Studies, Book of OHDSI, Resources, New To OHDSI?, EHDSN Academy, This Week In OHDSI, 2021 Global Symposium, Events/Collaborations, Collaborate in MTeams, Follow OHDSI.
- Main Content:** Welcome to OHDSI! The Observational Health Data Sciences and Informatics (or OHDSI, pronounced "oh-dsee") program is a multi-stakeholder collaborative to bring out the best through large-scale analytics. OHDSI has established...
- Annotations:**
 - A blue arrow points from the "Collaborate in MTeams" link in the navigation bar to the "Join Work groups, Chapters, and Studies Registration" section.
 - An orange circle highlights the "Join Work groups, Chapters, and Studies Registration" link in the navigation bar.
 - A blue arrow points from the "Join Work groups, Chapters, and Studies Registration" link in the navigation bar to the "Join Work groups, Chapters, and Studies Registration" section.
- Registration Section:** OHDSI MTeams Work groups, Chapters, and Studies Registration. OHDSI is using MTeams to further encourage active collaboration within the community. Within the OHDSI organization, there are separate teams for work groups, chapters, and studies, as well as OHDSI community activities (such as the OHDSI2020 Symposium). All teams are open to all collaborators. Below please indicate which Team you would like to join and the OHDSI coordinating center team will grant access.

5. Select the workgroups you want to join (you can refer to the WIKI for work group objectives www.ohdsi.org/web/wiki/doku.php?id=projects:overview)

- ☐ ATLAS
- ☐ Clinical Trials
- ☐ Common Data Model
- ☐ Data Quality Dashboard Development
- ☐ Early-stage Researchers
- ☐ Education Work Group
- ☐ Electronic Health Record (EHR) ETL
- ☐ Geographic Information System (GIS)
- ☐ HADES Health Analytics Data-to-Evidence Suite
- ☐ Health Equity
- ☐ Latin America
- ☐ Medical Devices
- ☐ Natural Language Processing
- ☐ OHDSI APAC
- ☐ OHDSI APAC Steering Committee
- ☐ OHDSI Steering Committee
- ☐ Oncology
- ☐ Patient-Generated Health Data
- ☐ Pharmacovigilance Evidence Investigation

- ☐ Phenotype Development and Evaluation
- ☐ Population-Level Effect Estimation / Patient-Level Prediction
- ☐ Psychiatry
- ☐ Registry (formerly UK Biobank)
- ☐ Surgery and Perioperative Medicine
- ☐ Vaccine Safety
- ☐ Vaccine Vocabulary
- ☐ Women of OHDSI

6. Select the chapter(s) you want to join

- ☐ Africa
- ☐ Australia
- ☐ China
- ☐ Europe
- ☐ Japan
- ☐ Korea
- ☐ Singapore
- ☐ Taiwan

7. Select the studies you want to join

- ☐ HERA-Health Equity Research Assessment
- ☐ PIONEER for Prostate Cancer (study-a-thon ended)
- ☐ SCYLLA (SARS-Cov-2 Large-scale Longitudinal Analyses)



2021 APAC Symposium • Nov. 18

Nov. 18 (APAC Time Zone)	Time (Korea time)	Contents	Speaker(s)
Morning	9:00 – 9:25 am	OHDSI State of the Community	George Hripcsak/Patrick Ryan
	9:25 – 9:50 am	OHDSI APAC State of the Community	Mui Van Zandt
	9:50 – 10:00 am	Energy Break	
	10:00 – 10:25 am	EHDEN	Peter Rijnbeek
	10:25 – 10:50 am	FHIR and OHDSI Collaboration	Christian Reich
	10:50 – 11:00 am	Energy Break	
	11:00 - 12:30 pm	APAC Chapter Visions for 2022	Chapter Leads
Lunch Break	12:30 – 13:00 pm		
Afternoon (in GatherTown)	13:00 – 14:00 pm	Workgroup Sessions (Medical Image, FHIR, CDM Tables)	
	14:00 – 15:00 pm	Collaboration Showcase	
	15:00 – 16:00 pm	APAC Study Sessions	

www.ohdsi.org/apac



New OMOP Modeling and OHDSI Tools Course

Profissionais
ensinando
a prática

Com OMOP você faz
pesquisas clínicas com
o MUNDO todo!

CURSO
**OMOP/
OHDSI**

Modelagem OMOP e
Ferramentas OHDSI

Inscrições para o curso
até 3 de novembro

Início das aulas 4/11

Realização: 

Apoio: 





Latest Edition Of The Journey Newsletter

Community Updates

Where Have We Been

- Recordings, materials and other resources from the first workshop between HL7 and OHDSI [are now available here](#). Five new subgroups have been formed to advance the use cases/projects: [Data Model Harmonization](#), [Digital Quality Measurement](#), [Oncology](#), [Registries and Applications](#), and [Terminologies](#). The FHIR and OMOP workgroup has returned, so if you would like to join this collaborative effort, [please sign up here](#).
- **Patrick Ryan** recently led a Halloween-themed interactive demonstration of how you can use the OHDSI tools to quickly generate insights from your OMOP CDM. This demo [is now available for anybody to watch](#).

Where Are We Now

- Nov. 1 marks the official start date of the LEGEND Initiative study around type 2 diabetes. This announcement follows the release of the [research protocol on MedRxiv](#) and a [presentation on the Initiative](#) by the LEGEND team. More on the LEGEND Initiative is available later in this newsletter.
- As announced recently, CDM v5.4 has been released and is available to the OHDSI Community. CDM workgroup leader **Clair Blacketer** will provide a live demo on the new version during the Nov. 9 community call.
- We are now in our second month of the #OHDSISocialShowcase, which highlights all of the global research shared during the 2021 Global Symposium. Each weekday, one presentation will be highlighted on both our [Twitter](#) and [LinkedIn](#) feeds. Please share with your networks to spread the word about our global efforts!

Where Are We Going

- The 2021 Asia-Pacific (APAC) Symposium will be held virtually Nov. 18 (Korea time), and will feature both a series of talks and an afternoon of collaborative activities, including workgroup meetings, a collaborator showcase and study sessions. You can [register for the APAC Symposium now](#), and more information on the event is available later in this newsletter.
- The upcoming month of community calls will feature breakouts for collaboration opportunities, presentations on open network studies within the community, a discussion on the history of OHDSI, and plenty more. Check out the full November schedule below; our meetings are held Tuesdays at 11 am ET, and everybody is invited. If you missed any, all presentations are recorded and posted to [our community calls homepage](#).

The Journey Newsletter (November 2021)

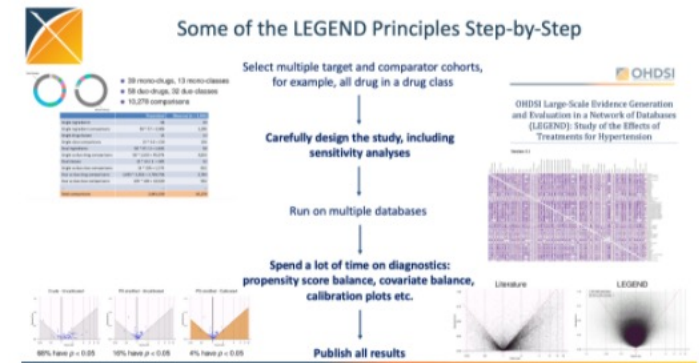
The LEGEND (Large-scale Evidence Generation and Evaluation across a Network of Databases) Initiative has already resulted in several published studies in high-impact journals, and the team has a new focus in store. You can learn more about this work, look ahead to the upcoming Asia-Pacific Symposium, and get more community updates in this edition of The Journey.

[#JoinTheJourney](#)

Monthly Update Podcast



The LEGEND Initiative — What Is It, How Did It Impact Hypertension, And What Comes Next?



The LEGEND (Large-scale Evidence Generation and Evaluation across a Network of Databases) Initiative applies high-level analytics to perform observational research on hundreds of millions of patient records within OHDSI's international database network.

LEGEND principles have been applied to studying the effects of treatments for depression, hypertension, and COVID-19, and are being applied to Type 2 diabetes. The clinical impact of LEGEND has already been observed, with important evidence that promotes better health decisions published in *Lancet*, *JAMA Internal Medicine*, and *Hypertension*.

During the Oct. 19 OHDSI Community Call, members of the LEGEND team provided a comprehensive presentation around this initiative, its work around hypertension, and plans around Type 2 diabetes.

[Watch The LEGEND Initiative Presentation](#)

[Find Study Protocol, Study Package, Presentation Slides](#)



Latest Edition Of The Journey Newsletter



OHDSI

OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS

Who We Are ▾	OHDSI Updates & News ▾	Standards	Software Tools	OHDSI Studies ▾	Book of OHDSI ▾	Resources ▾	New To OHDSI? ▾
EHDEN Academy ▾	This Week In OHDSI ▾	2021 Global Symposium ▾	Events/Collaborations ▾	Join OHDSI In MSTEams/Pick A Workgroup ▾			
NEW: Our Journey – Where The OHDSI Community Has Been, And Where We Are Going				Follow OHDSI/Newsletters ▾			

Newsletter	>	Subscribe
OHDSI on Twitter		November 2021
OHDSI on LinkedIn		October 2021
OHDSI on YouTube		SPECIAL: OHDSI2021 Symposium Preview
		September 2021
		August 2021
		July 2021
		June 2021
		May 2021
		April 2021
		March 2021
		February 2021

Welcome to OHDSI!

The Observational Health Data Sciences and Informatics (or OHDSI, pronounced "Odyssey") program is a multi-stakeholder, interdisciplinary collaborative to bring out the value of health data through large-scale analytics. All our solutions are open-source.

OHDSI has established an international network of researchers and observational health

20

The 2021 OHDSI Global Symposium presentations on OHDSI COVID-19 Pandemic, as well as to Reliable Evidence. The main the State of the Community Pre Collaborator Showcase, and a Closing Ceremony that focused work through the perspective of



#OHDSISocialShowcase This Week



Development of an ETL Process for Bulk and Incremental Load of German Patient Data into OMOP CDM Using FHIR

Elisa Henke¹, Yuan Peng², Ines Reinecke³, Michèle Zoch³, Martin Sedlmayr⁴

¹Institute of Medical Informatics and Biometry, Carl Gustav Carus Faculty of Medicine, Technische Universität Dresden, Germany

This work is part of the project MIRACUM, funded by the German Ministry of Education and Research (FKZ 01ZZ1801A/L).



Background

Motivation:

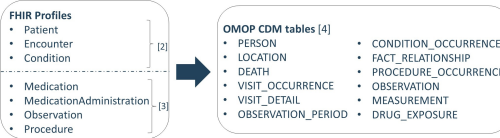
- The Use Case 'Alerting in Care – IT Support for Patient Recruitment' [1] in MIRACUM (Medical Informatics in Research and Care in University Medicine) aims to develop a Clinical Trials Recruitment Support System (CTRSS).
- This system suggests patients for clinical trials based on data in the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM).

Objective:

- To provide data for the CTRSS we need to design and develop an ETL (Extract-Transform-Load) process for filling OMOP CDM using Fast Healthcare Interoperability Resources (FHIR) profiles from MI-I and MIRACUM as data source.
- The ETL process has to support an initial (bulk) load as well as near real time or at least once a day updates (incremental load) of the data in OMOP CDM, to enable quick recruitment.

Methods

Semantic Mapping

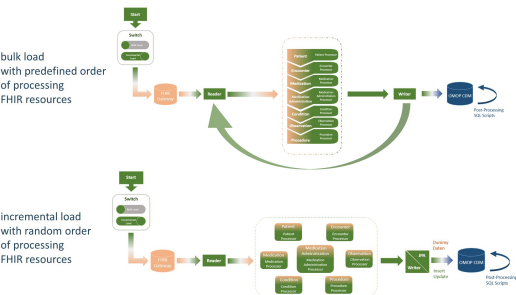


SpringBatch Framework [5]



Results

We have designed and implemented an ETL process which transforms MI-I and MIRACUM FHIR resources to OMOP CDM. This ETL process uses a switch to select whether the ETL process is executed as bulk load or as incremental load.



Conclusions

- The developed ETL process can transform and load data from FHIR into OMOP CDM as bulk load or incremental load.
- Thus, patient data can be updated to enable rapid recruitment with the CTRSS based on OMOP CDM.
- In the future, it is our aim to:
 - use meta data from FHIR and OMOP CDM to automate the ETL process
 - update the ETL process to new versions of the FHIR profiles from MI-I

References:
[1] Reinecke I, Gulden C, Kümmel M, Nassirian A, Blasini R, Sedlmayr M. Design for a Modular Clinical Trial Recruitment Support System Based on FHIR and OMOP. Stud Health Technol Inform. 2020 16;270:158-162.
[2] Medical Informatics Initiative Germany. Basismodule des Kern Datensatzes der MI-I. Available from: <https://www.medizin-informative.de/de/basismodule-deskern-datensatzes-der-mi-i>.
[3] MIRACUM. MIRACUM Core Implementation Guide – Table of Contents. 2020. Available from: <https://fhir.miracum.org/core/toc.html>.
[4] Observational Health Data Sciences and Informatics. OMOP CDM v5.3.1. Available from: <https://ohdsi.github.io/CommonDataModel/cdm531.html>.
[5] Ward L, Syer D, Risberg T, Kasanicky R, Garrette D, Lund W, Minella M, Schaefer C, Hillert G, Renfro G, Bryant J, Hassine M B. Spring Batch – Reference Documentation. 2021. Available from: <https://docs.spring.io/spring-batch/docs/current/reference/html/index.html>.

Contact: Elisa.Henke@ukdd.de

MONDAY

Development of an ETL Process for Bulk and Incremental Load of German Patient Data into OMOP CDM Using FHIR

Authors: Elisa Henke, Yuan Peng, Ines Reinecke, Michele Zoch, Martin Sedlmayr



#OHDSISocialShowcase This Week



OHDSI

OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS

Evaluating the performance of Austin's standardized difference heuristic in observational cohort studies with varying sample size

Mitch Conover¹, Azza Shoaibi¹, Joshua Ide¹, Martijn Schuemie¹

OHDSI PLE/PLP Working Group
September 3, 2021

¹All authors are employees of Janssen Research & Development

Evaluating the performance of Austin's standardized difference heuristic in observational cohort studies with varying sample size

Authors: Mitchell Conover (presenter), Azza Shoaibi, Joshua Ide, Martijn Schuemie

TUESDAY



#OHDSI SocialShowcase This Week

Assessing the Technical and Operational Feasibility of a Next-generation Data Capture and Processing Approach—An Extension of OMOP-based Data Translation Developed by the Center for International Blood and Marrow Transplant Research (CIBMTR)

INTRODUCTION

The Center for International Blood and Marrow Transplant Research (CIBMTR) is a research collaboration between the National Marrow Donor Program (NMDP) (Be The Match) and the Medical College of Wisconsin (MCW). The CIBMTR collaborates with the global scientific community to advance hematopoietic cell transplantation (HCT) and cellular therapy worldwide to increase survival and enrich the quality of life for transplant patients. The CIBMTR collects outcomes data from approximately 400 cellular therapy centers worldwide, including for all allogeneic HCTs and 80% of autologous HCTs performed in the US.

The CIBMTR aims to reduce the administrative burden on transplant centers, in the first phase, CIBMTR focused on reducing manual data capture by beginning to transmit data directly from participating sites' source databases formatted in HL7 Fast Healthcare Interoperability Resources (FHIR) (Figure 1).

Figure 1. DTI FHIR Transmission

METHODS

The traditional forms-based approach can potentially result in unnecessary effort throughout data capture and management.

A domain-based workflow was implemented to mitigate or eliminate many of these inefficiencies, creating a more flexible data collection approach, a significantly streamlined data processing pipeline, and expedient analysis and sharing options.

The tool's user interface optimizes hybrid manual and automatic data entry.

Combining abstracted and extracted data in a centralized OMOP CDM, the CIBMTR significantly streamlines its data processing pipeline and creates the foundation for expanded analysis and sharing options

Use your phone to download the full paper

DATA TRANSFORMATION
Data. Knowledge. Lives Saved.

The next generation system assesses what data exists from the EMR extracts and combines them together with the manually abstracted data. (Figure 2)

Figure 2. Next Generation Transmission

As the chart abstractor enters information, the system intelligently sequences questions based on multiple criteria:

1. Determination of variable collection requirements, considering factors like the primary disease and the time of the transplant (e.g., pre-HCT, post-HCT, extended follow-up milestones)
2. Optimal time to validate assumptions and derivations based on real-time confidence and risk scoring (e.g., selecting the right lab panel for a target date when it may likely be inferable but not fully disambiguated)
3. Impact on downstream variable collection (e.g., position in the scenario tree, degree of use in subsequent derivation logic)
4. Simplification of process for abstractor (e.g., treatment chronology vs grouping similar data types)
5. Site and abstractor preferences

CONCLUSION

Upcoming work will build on this foundation by developing a collaborative research environment featuring prototype stakeholder portals, analysis tools, and links to additional datasets such as genomics, patient-reported outcomes, reimbursement claims, and manufacturing data.

Kristina Bloomquist¹, Matt Prestegard¹, Ben Smith², Trent Peterson²
¹ CIBMTR, Minneapolis, MN;
² Principia Health Sciences, Cary, NC.

CIBMTR
CENTER FOR INTERNATIONAL BLOOD & MARROW TRANSPLANT RESEARCH

PRINCIPIA
HEALTH SCIENCES

OHDSI

Assessing the Technical and Operational Feasibility of a Next-generation Data Capture and Processing Approach An Extension of OMOP-based Data Translation Developed by the Center for International Blood and Marrow Transplant Research (CIBMTR)

Authors: Ben Smith, Trent Peterson, Kristina Bloomquist



#OHDSISocialShowcase This Week

Mapping UK Biobank to the OMOP CDM: challenges and solutions using the Delphyne ETL framework

Sofia Bazakou¹*, Maxim Moinat¹, Alessia Peviani¹, Anne van Winzum¹, Stefan Payralbe¹, Vaclav Papez², Spiros Denaxas²

¹ The Hyve, Utrecht, The Netherlands * Contact: sofia@thehyve.nl / +31 (0)30 7009713

² University College London, London, United Kingdom



Background

UK Biobank¹ (UKB) is a large-scale registry containing medical and genetic data from 500,000 consented participants from the UK's general population, aged between 40 and 69 years (Figure 1). UKB is an extraordinary resource for human health research, accessible to approved research initiatives worldwide.

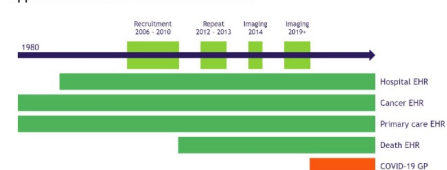


Figure 1. UKB data structure and timeline. The data include baseline assessments (light green), such as surveys, samples, and imaging, electronic health records (EHR) from different sources (dark green), and information on COVID-19 testing (red). Picture adapted from Spiros Denaxas, Professor of Biomedical Informatics, Institute of Health Informatics, University College London.

As part of the European Health Data Evidence Network² (EHDEN), The Hyve was contracted by University College London (UCL) to map the UKB data to the OMOP CDM v5.3. The main goal of the collaboration was to make the dataset available for research related to the COVID-19 pandemic. The Hyve implemented the data conversion pipeline, while UCL provided the source data expertise.

The UKB data conversion effort came with several challenges:

- ETL development without direct access to the data.
- Mapping of free-text and non-standard ontologies.
- Large heterogeneity of source terms amongst data providers.
- Conversion of a large wide format table to long format.
- Working with an evolving data source.

Methods

The Hyve overcame the lack of direct access to the UKB data by adopting a collaborative Agile-based development approach with UCL (Figure 2). This process made it possible to develop the ETL code relying entirely on **synthetic data**.

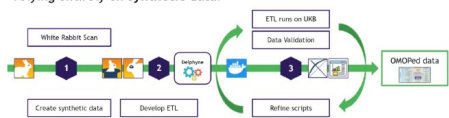


Figure 2. Collaborative Agile development approach using existing OHDSI tools and Delphyne. Step 1: UCL provided a White Rabbit scan report of the UKB data, which The Hyve used to generate synthetic data for ETL development. Step 2: The Hyve performed syntactic and semantic mappings with Rabbit in a Hat and Usagi, respectively, and developed the ETL with Delphyne. Step 3: UCL ran the ETL (locally deployed with Docker) on a UKB data subset, and executed quality checks with Achilles and the Data Quality Dashboard. The Hyve refined the ETL based on feedback before a new iteration. Finally, UCL ran the ETL on the full UKB dataset to produce the complete "OMOPed" data.

Our UKB mapping workflow made use of existing tools from the OHDSI suite and Delphyne^{3,4}, a specialized ETL framework for mapping data to the OMOP CDM, developed internally by The Hyve. Delphyne was particularly helpful in tackling data heterogeneity between UKB healthcare providers, and the mapping of the wide format *baseline* table (500,000x9,000), which both required specific handling logic. Our development approach relied heavily on feedback, which Delphyne helped to provide through detailed logging and summary reports. It also made trivial to extend the CDM model with custom provenance fields, enabling more informative data quality assessments. Finally, Delphyne automated several tasks, such as CDM table creation, vocabulary loading, and mapping of non-standard to standard ontologies, saving time whenever a UKB data or vocabulary update was available. Overall, Delphyne allowed us to build a highly specialized ETL, for maximum mapping coverage and quality.

Conclusion

UKB is an incredible resource for healthcare research. Given its size and complexity, mapping the data to the OMOP CDM model came with several challenges. A powerful and flexible ETL framework such as Delphyne was invaluable in carrying out the conversion effort. Together with existing open-source tools from the OHDSI suite, Delphyne allowed us to perform the conversion without direct access to the UKB data, and to deliver a **high-coverage mapping** (Figure 3). The mapping of UKB data to the OMOP CDM will in turn enable future research, including studies on COVID-19, to build upon our efforts.

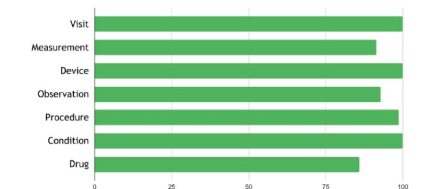


Figure 3. Percentage of UKB source codes mapped to a standard OMOP concept per domain, by record frequency. We achieved a near full mapping coverage for the Visit, Device and Condition domains (>99%), and the lowest mapping coverage for the Drug domain (86%). Note that for the baseline data mapping, we converted a subset of the original variables.

References

- [1] <https://www.ukbiobank.ac.uk/>
- [2] <https://ehden.eu/>
- [3] <https://www.thehyve.nl/cases/mapping-uk-biobank-to-omop-using-delphyne>
- [4] <https://delphyne.readthedocs.io/en/latest/>



Mapping UK Biobank to the OMOP CDM: challenges and solutions using the delphyne ETL framework

Authors: Sofia Bazakou, Maxim Moinat, Alessia Peviani, Anne van Winzum, Stefan Payralbe, Vaclav Papez, Spiros Denaxas

THURSDAY



@OHDSI

www.ohdsi.org

#JoinTheJourney



ohdsi



#OHDSISocialShowcase This Week



“Perseus”: Design and run your own ETL to CDM

Anton Ivanov, MS¹, Samus Sergey, MS¹, Alexander Efimov, MS¹, Maxim Draschinsky, MS¹, Dolotova Maria, MS¹, Matvey Chudakov, MS¹, Karpushina Anna, MS¹, Semenov Artem, MS¹
¹Arcadia Inc., St. Petersburg, Russia

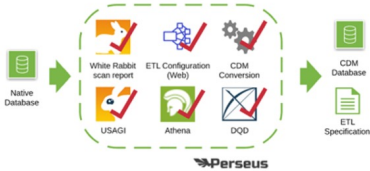


Background

There are several opensource ETL tools designed for CDM and available for OHDSI community. Tools for code mapping and quality control provide a good starting point to design ETL process for getting CDM but ETL implementation still requires technical skills in programming languages and platforms. Presented solution addresses this need and provides a visual designer for ETL implementation targeting CDM.

Methods

“Perseus” is combination of Web application for ETL configuration and engine for: conversion of native data into CDM, data quality check, code mapping, generating ETL document and vocabulary search.



“Perseus” has instruments for all major steps of creating an ETL.

Step 1: Scan source data

- ✓ Use White Rabbit scan report
- ✓ or scan the source data directly

Step 2: Create the Code Mappings

- ✓ Visually map raw data to CDM
- ✓ Embedded set of transformations
- ✓ Embedded set of lookups
- ✓ Combine source tables
- ✓ Combine fields into groups
- ✓ Configure conditional mappings
- ✓ Automatic era creation
- ✓ Automatic domain switching

Step 3: Design and run ETL



Step 4: Quality Control

- ✓ Integrated Data Quality Dashboard

Conclusions

“Perseus” reduces the time spent on manual processes such as writing code and mapping source data to target systems. The configuring and running ETLs with “Perseus” makes it repeatable, shareable and saves time.

References:

1. Perseus - <https://github.com/SoftwareCountry/Perseus>
2. White Rabbit, and Rabbit-in-a-Hat <https://github.com/SoftwareCountry/WhiteRabbit>
3. Usagi - <https://github.com/OHDSI/Usagi>
4. Achilles - <https://github.com/OHDSI/Achilles>
5. DQD - <https://github.com/SoftwareCountry/DataQualityDashboard>
6. CDM Builder - <https://github.com/SoftwareCountry/ETL-CDMBuilder>



Contact: info@softwarecountry.com

FRIDAY

Perseus Design and run your own ETL to CDM
Authors: Anton Ivanov, Samus Sergey, Alexander Efimov, Maxim Draschinsky



Three Stages of The Journey

Where Have We Been?

Where Are We Now?

Where Are We Going?





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?





OHDSI Shoutouts!



Congratulations to **Antoine Lamer, Osama Abou-Arab, Alexandre Bourgeois, Adrien Parrot, Benjamin Popoff, Jean-Baptiste Beuscart, Benoît Tavernier, and Mouhamed Djahoum Moussa** for the publication of **“Transforming Anesthesia Data Into the Observational Medical Outcomes Partnership Common Data Model: Development and Usability Study”** in the Journal of Medical Internet Research.

[Journal of Medical Internet Research](#) [Journal Information](#) [Browse Journal](#) [Subscribe](#)

Published on 29.10.2021 in Vol 23, No 10 (2021): October

Preprints (earlier versions) of this paper are available at <https://preprints.jmir.org/preprint/29259>, first published April 26, 2021.



Transforming Anesthesia Data Into the Observational Medical Outcomes Partnership Common Data Model: Development and Usability Study

Antoine Lamer ^{1,2,3} ; Osama Abou-Arab ⁴ ; Alexandre Bourgeois ⁵ ; Adrien Parrot ² ; Benjamin Popoff ⁶ ; Jean-Baptiste Beuscart ¹ ; Benoît Tavernier ^{1,7} ; Mouhamed Djahoum Moussa ⁷

Article	Authors	Cited by	Tweetations	Metrics
<ul style="list-style-type: none">AbstractIntroductionMethodsResultsDiscussionReferencesAbbreviationsCopyright	<h4>Abstract</h4> <p>Background: Electronic health records (EHRs, such as those created by an anesthesia management system) generate a large amount of data that can notably be reused for clinical audits and scientific research. The sharing of these data and tools is generally affected by the lack of system interoperability. To overcome these issues, Observational Health Data Sciences and Informatics (OHDSI) developed the Observational Medical Outcomes Partnership (OMOP) common data model (CDM) to standardize EHR data and promote large-scale observational and longitudinal research. Anesthesia data have not previously been mapped into the OMOP CDM.</p>			



OHDSI Shoutouts!



Congratulations to **William Wood, Peter Marks, Robert Plovnick, Kathleen Hewitt, Donna Neuberg, Sam Walters, Brendan Dolan, Emily Tucker, Charles Abrams, Alexis Thompson, Kenneth Anderson, Paul Kluetz, Ann Farrell, Donna Rivera, Matthew Gertzog, and Gregory Pappas** for the publication of **“ASH Research Collaborative: A Real-World Data Infrastructure to Support Real-World Evidence Development and Learning Healthcare Systems in Hematology”** in Blood Advances.



ASH Research Collaborative: A Real-World Data Infrastructure to Support Real-World Evidence Development and Learning Healthcare Systems in Hematology

Tracking no: ADV-2021-005902R1

William Wood (University of North Carolina at Chapel Hill, United States) Peter Marks (Food and Drug Administration, United States) Robert Plovnick (American Society of Hematology, United States) Kathleen Hewitt (ASH Research Collaborative, United States) Donna Neuberg (Dana-Farber Cancer Institute, United States) Sam Walters (Breakthrough Healthcare, United States) Brendan Dolan (The University of Wisconsin School of Medicine and Public Health, United States) Emily Tucker (ASH Research Collaborative, United States) Charles Abrams (University of Pennsylvania School of Medicine, United States) Alexis Thompson (Ann & Robert H. Lurie Children's Hospital of Chicago/Northwestern Univ, United States) Kenneth Anderson (Dana-Farber Cancer Institute, United States) Paul Kluetz (Oncology Center of Excellence, Food and Drug Administration, United States) Ann Farrell (US FDA, United States) Donna Rivera (Oncology Center of Excellence, Food and Drug Administration, United States) Matthew Gertzog (ASH Research Collaborative, United States) Gregory Pappas (Food and Drug Administration, United States)

Abstract:

The ASH Research Collaborative is a nonprofit organization established through the American Society of Hematology's commitment to patients with hematologic conditions and the science that informs clinical care and future therapies. The ASH Research Collaborative houses two major initiatives: 1) the Data Hub and 2) the Clinical Trials Network (CTN). The Data Hub is a program for hematologic diseases in which networks of clinical care delivery sites are developed in specific disease areas, with individual patient data contributed through electronic health record (EHR) integration, direct data entry through electronic data capture, and external data sources. Disease-specific data models are constructed so that data can be assembled into analytic datasets and used to enhance clinical care through dashboards and other mechanisms. Initial models have been built in multiple myeloma and sickle cell disease using the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM) and Fast Healthcare Interoperability Resources (FHIR) standards. The Data Hub also provides a framework for the development of disease-specific Learning Communities and the testing of healthcare delivery strategies. The ASH Research Collaborative CTN is a clinical trials accelerator that creates efficiencies in the execution of multicenter clinical trials and has been initially developed for sickle cell disease. Both components are operational, with the Data Hub actively aggregating source data and the CTN reviewing study candidates. This manuscript describes processes involved in developing core features of the ASH Research Collaborative to inform the stakeholder community in preparation for expansion to additional disease areas.



OHDSI Shoutouts!



AMIA
@AMIAinformatics

Congratulations to those new [#ACMI](#) Fellows in person with us in San Diego! [#AMIA2021](#)



Patrick Ryan

Inductee

ACMI Fellows Class of 2021
Induction Ceremony
(ACMI Fellows Only)

11:30 am PT • Sapphire M



COLUMBIA

COLUMBIA UNIVERSITY
DEPARTMENT OF
BIOMEDICAL INFORMATICS



@OHDSI

www.ohdsi.org

[#JoinTheJourney](#)



ohdsi



OHDSI Shoutouts!

A video frame showing J. Marc Overhage, a man with glasses and a blue shirt, speaking. The background is a grey curtain.

J. Marc Overhage
Chief Medical Informatics Officer • Anthem, Inc.

The OHDSI logo, consisting of a stylized yellow and blue icon followed by the text "OHDSI" and "OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS" in smaller text below it.