Intelligent Health Systems – From Technology to Data and Knowledge
E. Andrikopoulou et al. (Eds.)

© 2025 The Authors.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).
doi:10.3233/SHTI250432

OMOP-on-FHIR: Integrating the Clinical Data Through FHIR Bundle to OMOP CDM

Prabath JAYATHISSA[a,1], Lukas ROHATSCH[b], Stefan SAUERMANN[b], and Rada HUSSEIN[a]

[a]*Ludwig Boltzmann Institute for Digital Health and Prevention, Salzburg, Austria*
[b]*University of Applied Sciences/FH-Technikum Wien, Vienna, Austria*

ORCiD ID: Prabath Jayathissa https://orcid.org/0000-0003-3056-503X, Lukas ROHATSCH https://orcid.org/0000-0002-6168-2717, Stefan SAUERMANN https://orcid.org/0000-0003-0824-9989,Rada Hussein https://orcid.org/0000-0003-1257-4848

Abstract. The harmonization of the OMOP Common Data Model (CDM) with HL7 FHIR aims to enhance interoperability in clinical research by harmonizing diverse healthcare datasets. This process, referred to as OMOP-on-FHIR, leverages FHIR Bundles for real-time clinical data exchange and transforms these resources into OMOP CDM format using an ETL process. The ETL pipeline, facilitated by tools like XSLT, enables the extraction, transformation, and loading of data while maintaining semantic consistency. By bridging these two standards, OMOP-on-FHIR promotes the seamless exchange of data across clinical systems and research-oriented databases, supporting global health studies, advanced analytics, and personalized medicine. This methodology advances cross-border research by providing a standardized approach to data management and analysis, thereby improving healthcare outcomes.

Keywords. OMOP, ETL Pipeline, FHIR, Clinical Research, Healthcare Interoperability

1. Introduction

The globalization of healthcare and medical research is driving the need for conducting cross-border clinical studies that heavily rely on the ability to integrate and analyze data across diverse healthcare systems. However, A significant challenge in cross-border research is contributed by the fact that health data is stored and structured differently across different regions, where there is little standardization. Health systems make use of different data models; hence, it is difficult to share and analyze data on a uniform basis.

The OMOP CDM had been developed to offer a standard structure for observational healthcare data that enables researchers to conduct systematic analyses across different databases. In such a context, native healthcare data needs to be transformed into OMOP-CDM format through an ETL process. This step is critical to data harmonisation of the

[1] Corresponding Author: Prabath Jaythissa, Ludwig Boltzmann Institute of Digital Health and Prevention, Salzburg, Austria; E-mail: Prabath.Jayathissa@dhp.lbg.ac.at.

disparate datasets. Given the diversity of data formats, a generic and flexible ETL tool is required for simplifying and standardising this transformation, especially in large-scale and cross-border research projects. This concept can contribute to global health studies by guaranteeing interoperability and decreasing the integration complexity of healthcare data from various regions; clinical research can be advanced on a global scale [1]. Such clinical data interoperability between systems will help increase healthcare analytics and research. Two major standards enable the same: the OMOP CDM and HL7 FHIR. The OMOP-developed standard is one of the most used standards for standardising observational healthcare data and has been developed under OHDSI [2]. FHIR, under the banner of HL7 International, stands out as a lightweight protocol fit to exchange clinical data across health systems [3].

OMOP-on-FHIR Integration can blend the best of these two standards into an efficient data exchange. While OMOP CDM standardizes and harmonizes huge and various data, FHIR allows real-time health information exchange in structured FHIR Bundles. Integration with FHIR Bundles within the OMOP CDM allows interoperability between real-time clinical systems and research-oriented databases [3,4]. The Data Transformation Process bridges these two standards, FHIR Bundles-sets of related resources such as patients or observations extracted and transformed using XSLT into OMOP CDM format. Subsequently, it will cover the data into OMOP CDM, allowing the transformation process systematically to hold medical research, predictive analytics, and personalised medicine [4]. Healthcare organizations leverage OMOP-on-FHIR to overcome the management and analytics of large-scale clinical data. This allows for the development of data-driven medical research and treatment [5].

2. Methodology

The mapping of FHIR resources with the OMOP-CDM involves key steps to ensure data standardization, interoperability, and usability for both primary and secondary data purposes [4,6].

2.1. Data Extraction

Clinical data is extracted from source systems in the form of FHIR Bundles, which of organize related resources like patient records, observations, and encounters into a structured format. FHIR Bundles facilitate the handling of complex clinical data and ensure interoperability by conforming to international healthcare data standards [2].

2.2. Data Transformation

ETL transforms FHIR data into OMOP-CDM format. XSLT and other tools map the FHIR resources into OMOP CDM tables for conditions, measurements, and drug exposures and can keep semantic interoperability across the standards [3,5]. This use case integrates OMOP with FHIR and further expands clinical data interoperability among heterogeneous systems. The flexible structure of FHIR, along with the research-oriented data model of OMOP, enables the two bidirectional data exchange, thus enhancing both clinical use and secondary research. Furthermore, OMOP-on-FHIR enables fast data querying and system-to-system information exchange without losing in information integrity [5,8].

Data Loading: The transformed data is loaded into an OMOP-compliant database for research and advanced analytics such as population health studies or chronic disease surveillance [6,9]. Validation: Transformed data undergo validation checks for accuracy, consistency, and completeness. High-quality output data are necessary not only for primary clinical uses but also for secondary use, such as research applications [7,10]. In this way, health organizations can enjoy the dual benefit of FHIR and OMOP in ensuring that real-time clinical data becomes standardized, and interoperable not only in use within healthcare settings but also in research environments.

3. Results

Indeed, through the process of implementing OMOP-ETL, there has been the potential to standardize data coming from different regions and institutions in OMOP-CDM format. Improvement of research across borders through facilitation of pooling and analysis of data. Global studies like COVID-19 can be enabled on integrated data coming from a variety of sources.

3.1. Mapping to OMOP Domain

There are two major phases involved in mapping to OMOP CDM, namely Profiling and Syntactic Mapping and Semantic Mapping. Structuring: Data profiling and syntactic mapping to the OMOP CDM schema are done at this stage. The syntactic mapping defines how tables, fields, and values from the source could be translated into the target model of OMOP. Patient identifiers, data fields, values, units, and dates are structured in a format acceptable to the OMOP CDM. Semantic Mapping: Meaning assurance across Systems. Once data is syntactically aligned, semantic mapping describes the associations among the different data elements of the medical terms or codes in such a way that their meaning is preserved upon integration into the OMOP CDM. The consistency and interoperability of the clinical concepts presented in the CDM are thereby guaranteed. These two mapping processes complete the necessary standardization for the usability of heterogeneous healthcare data in large-scale research within the OMOP framework.

3.2. Extending OMOP Vocabularies

The extension of the OMOP CDM and standardized vocabularies is a very important feature for enabling observational research, especially in very specialized domains such as oncology. The OMOP CDM is extended to carry cancer-specific information fields in tumour stages, diagnosis codes, and treatment regimens, among others, to support observational cancer research. The extension is realized by vocabulary-driven ETL integrating cancer registry data. These activities help the data maintain interoperability across heterogeneous datasets with clinically accurate semantics. This enhancement allows researchers now to apply standard vocabularies in the analysis of outcomes and treatments of cancer more efficiently, increasing the model's applicability in oncology, among other specialized areas.

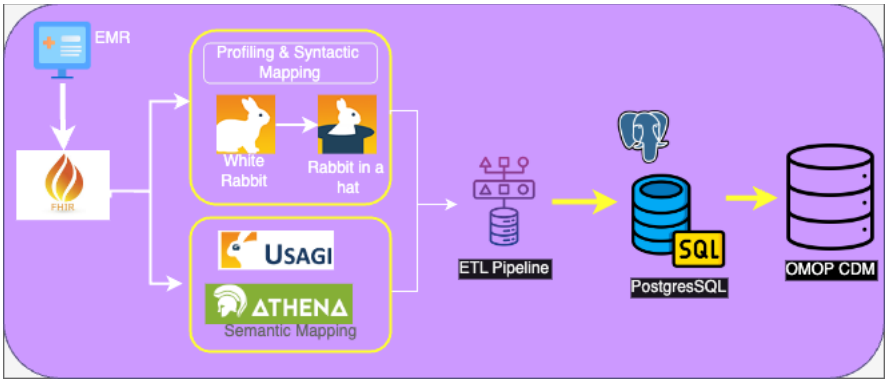

Figure 1. The architecture of FHIR to OMOP CDM ETL Process

3.3. The architecture of FHIR to OMOP CDM ETL Process

The architecture outlines a streamlined workflow to map data from EMRs to OMOP-CDM. It begins with extracting EMR data, which is processed according to the FHIR standard through profiling and syntactic mapping, leveraging tools such as White Rabbit and Rabbit in a Hat to align the data structure with the required analytical format. Semantic alignment is achieved using tools like USAGI and ATHENA, ensuring consistent interpretation of data elements across systems. The data then flows through an ETL pipeline for further transformation and is subsequently stored in a PostgreSQL database, formatted as SQL to enable efficient structured queries. The final output is structured data conforming to the OMOP-CDM standard, facilitating standardized healthcare data analysis and interoperability. The final output is structured data that adheres to the OMOP CDM, enabling standardized analysis of healthcare data.

4. Discussion and Conclusions

The application of an OMOP-ETL tool in cross-border research significantly reduces barriers caused by incompatible data models. This paves the way for global collaboration, enabling researchers to leverage diverse datasets to address critical health challenges. The development of a generic OMOP-ETL tool aims to standardize and harmonize various data sources, ensuring consistency that supports more robust, collaborative studies to advance healthcare outcomes worldwide. By standardizing data sources, the tool facilitates efficient cross-border clinical research while enhancing researchers' ability to conduct impactful studies. Best practices within the ETL process ensure the conversion of data is both efficient and accurate. Collaboration between data and CDM experts plays a crucial role in designing the ETL process, with tools like White Rabbit and Rabbit-in-a-Hat providing essential support for analyzing source data and mapping it effectively to the CDM. This is particularly important to medical professionals when the source data uses coding systems different from the OMOP vocabulary. Tools like Usagi help to suggest mappings for textual similarities, although the task is very complex. On the other hand, the technical team will have to implement the ETL, using for this purpose the most suitable programming language, considering the competencies of the working team in either SQL or Java. Finally, quality control is of paramount

importance, with the thorough review of design documents and code against source and target data comparison, and the use of Achilles to monitor data quality. This will enable timely resolution of possible issues. Thus, an overall approach can be taken toward high-quality conversion to OMOP-CDM. Equally important is the development and use of the OMOP-ETL tool for minimal irrelevant hurdles created by these different models. Such cooperation in research allows collaboration across borders. The standardization of the diverse data henceforward allows the researchers to study health challenges much more effectively than in the past. Best practices of ETL studies would involve collaboration between data experts and medical experts in using such tools as White Rabbit, Rabbit-in-a-Hat, and Usagi to simplify mapping and conversion. This shall be ensured through comprehensive reviews and Achilles-like tools that shall make integrations of data into OMOP CDM accurate and efficient, moving the world ahead in healthcare research.

References

[1] Espinoza J, Sikder S, Lulejian A, Levine B. Development of an OpenMRS-OMOP ETL tool to support informatics research and collaboration in LMICs. Comput Methods Programs Biomed Update. 2023;4:100119. doi:10.1016/j.cmpbup.2023.100119.

[2] Gulden C, Macho P, Reinecke I, Strantz C, Prokosch HU, Blasini R. recruIT: A cloud-native clinical trial recruitment support system based on HL7 FHIR and the Observational Medical Outcomes Partnership Common Data Model (OMOP CDM). Comput Biol Med. 2024;174:108411. doi:10.1016/j.compbiomed.2024.108411.

[3] Belenkaya R, Gurley MJ, Extending the OMOP Common Data Model and Standardized Vocabularies to Support Observational Cancer Research. JCO Clin Cancer Inform. 2021 Jan;5:12-20. doi: 10.1200/CCI.20.00079. PMID: 33411620; PMCID: PMC8140810.

[4] Papez V, Moinat M, Voss EA, Bazakou S, Van Winzum A, Peviani A, Transforming and evaluating the UK Biobank to the OMOP Common Data Model for COVID-19 research and beyond. J Am Med Inform Assoc. 2022 Dec 13;30(1):103-111. doi: 10.1093/jamia/ocac203..

[5] Reich C, Ostropolets A, Ryan P, Rijnbeek P, , Hripcsak G. OHDSI Standardized Vocabularies-a large-scale centralized reference ontology for international data harmonization. J Am Med Inform Assoc. 2024 Feb 16;31(3):583-590. doi: 10.1093/jamia/ocad247. PMID: 38175665; PMCID: PMC10873827.

[6] Hussein R, Balaur I, Burmann A, Cwiek-Kupczy´ nska H, G, Jayathissa P, Katsch F,´ Kremer A, Stamm T, Muehlendyck C, Gribbon P. Getting ready for the European Health Data Space (EHDS): IDERHA's plan to align with the latest EHDS requirements for the secondary use of health data. Open Res Eur. 2024 Jul 30;4:160. doi: 10.12688/openreseurope.18179.1. PMID: 39185338; PMCID: PMC11342032.

[7] Jathissa P, Rohatsch L, Sauermann S, Hussein R. OMOP-on-FHIR: A FHIR server development to facilitate data interaction with the OMOP-CDM and FHIR for PGHD. Stud Health Technol Inform. 2024;316:157-158. doi:10.3233/SHTI240367.

[8] Rinaldi E, Thun S. From OpenEHR to FHIR and OMOP Data Model for Microbiology Findings. Stud Health Technol Inform. 2021 May 27;281:402-406. doi: 10.3233/SHTI210189. PMID: 34042774.

[9] Essaid S, Andre J, Brooks IM, Hohman KH, Hull M, Jackson SL, Kahn MG, Kraus EM, Mandadi N, Martinez AK, Mui JY, Zambarano B, Soares A. MENDS-on-FHIR: leveraging the OMOP common data model and FHIR standards for national chronic disease surveillance. JAMIA Open.2024;7(2). doi:10.1093/jamiaopen/ooae045.

[10] Jayathissa P, Hewapathirana R. HAPI-FHIR Server Implementation to Enhancing Interoperability among Primary Care Health Information Systems in Sri Lanka: Review of the Technical Use Case. Eur Mod Stud J. 2024;7(6):225-41.