Name:	Hamed Abedtash
Affiliation:	Indiana University School of Informatics and Computing
Email:	abedtash@iupui.edu
Presentation type:	Poster

CCD2OMOP: An Interoperable Extract-Transform-Load Package to Support the Implementation of OHDSI Software Tools Across Non-OMOP-based Electronic Health Records

Hamed Abedtash, PharmD[1], Jon Duke, MD, MSc[2]
[1]Indiana University School of Informatics and Computing, Indianapolis, IN; [2]Regenstrief Institute, Indianapolis, IN

Abstract

This paper introduces a solution to facilitate involvement of clinical institutions in large-scale observational data analytics using OHDSI software tools even with non-OMOP-based data warehouses. We developed an interoperable Extract-Transform-Load (ETL) Python Package that extracts patients' medical records from HL7 Continuity of Care Documents (CCD), transforms the data into OMOP Common Data Model, and loads the transformed information to an OMOP database for further processing and analytics. This package will support not only integrating and consolidating longitudinal data from diverse EHR systems, but also helps deploy OMOP-based predictive models for better clinical decision making at clinics.

Introduction

Observational Health Data Sciences and Informatics (OHDSI) community has developed numerous open-source software tools for large-scale analysis of observational medical data warehouses to support clinical decision making through generating real-world evidences (1), such as PatientLevelPrediction (2) and CohortMethod (3). The OHDSI community has recognized the importance of predictive analytics for better decision making and has developed PatientLevelPrediction R package to generate predictive models according to OMOP Common Data Model (CDM). The models can be used further to perform patient level prediction about the risk of health related events. The CohortMethod package is also meant to perform new-user cohort studies in an observational OMOP-based database.

OHDSI has made many other similar open-source packages available to epidemiology and data sciences researchers; however, implementing these tools such as deploying OMOP-based predictive models for decision support (e.g., real-time calculation of individual patient risk) is limited because 1) few centers have patient data in OMOP CDM format, and 2) even in such institutions it is unlikely that transactional patient data will be available in the OMOP CDM for real-time computation. Thus, we aimed to leverage an existing and commonly used standard for individual patient data representation–the HL7 Continuity of Care Document (CCD)–to develop an interoperable solution that will enable the myriad institutions that are not using the OMOP data model to benefit from OMOP-based predictive models in their clinical workflow.

CCD2OMOP Package

The CCD2OMOP is an ETL Python package that extracts a patient's record from a continuity of care document (CCD) in HL7 consolidated clinical document architecture (C-CDA) format, transforms the data into OMOP CDM format, and finally loads the transformed information into staging OMOP CDM tables to be used later for predictive model deployment. The package consists of five modules (Figure 1): CCD Parser, OMOP Mapper, Loader, OMOP CDM, and Database Connector. The current version of CCD2OMOP package renders CCD data for seven OMOP tables as shown in Table 1.

The CCD Parser module receives a single CCD file at a time and extracts the patient demographics, medications, conditions, laboratory test results, and observations. The extraction pipeline points to the template IDs specified by HL7 C-CDA. It also validates the pulled data whether data elements have the correct values for the destination OMOP table. After data is extracted, the OMOP Mapper module transforms the data into intermediate OMOP tables–which are instantiated from the OMOP CDM module–for further processing. The transformation process translates source codes into OMOP concept IDs (e.g., mapping an RxNorm code to OMOP) and extracts values where appropriate (e.g., year and month from CCD drug dispensing date). Next, the Loader module loads the transformed data from the

intermediate tables into an OMOP CDM database that can be accessible by the end user. Both the OMOP Mapper and Loader modules use the Database Connector module to lookup OMOP values and load transformed data into this database, respectively.

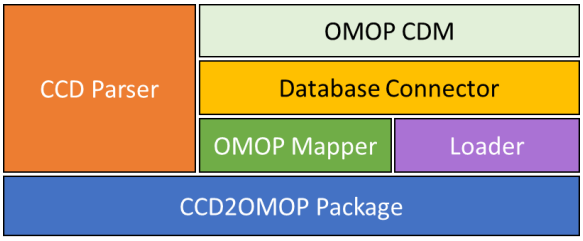

Figure 1. The architecture of CCD2OMOP package

Table 1. The CCD2OMOP package extracts, transforms, and loads data for seven OMOP tables.

OMOP Table	HL7 C-CDA Section
Person	Header
Death	Deceased Observation
Procedure Occurrence	Procedure Activity Act Procedure Activity Observation Procedure Activity Procedure
Drug Exposure	Medications Section Medications Section Immunizations Section Immunizations Section
Condition Exposure	Problem Section
Measurement	Results Section Vital Signs Section
Observation	Allergies and Intolerances Section

Conclusion

The developed CCD2OMOP package supports conversion of CCD documents into OMOP CDM format for a variety of purposes, including electronic health records (EHR) data integration, data consolidation, and deployment of OMOP-based predictive models both within and outside of OHDSI community. The package has shown to be a reliable and convenient method to transform EHR data to OMOP CDM that enables healthcare centers without OMOP data model in their data warehouse to benefit from OMOP-based predictive models. Once patient CCDs are parsed by CCD2OMOP, the transformed information can be used by OHDSI PatientLevelPrediction-produced models to make estimations about health related events. The transformed data also facilitates integrating and consolidating data from diverse data sources. If linked with clinical decision support systems, this package has the potential to help clinicians make the best decisions by providing real-time, personalized predictions about patients' health status.

References

1. Hripcsak G, Duke JD, Shah NH, Reich CG, Huser V, Schuemie MJ, et al. Observational Health Data Sciences and Informatics (OHDSI): Opportunities for Observational Researchers. Studies in health technology and informatics. 2015;216:574-8.
2. OHDSI. An R package for performing patient level prediction 2016 [6/22/2016]. Available from: https://github.com/OHDSI/PatientLevelPrediction.
3. OHDSI. CohortMethod 2016 [6/22/2016]. Available from: https://github.com/OHDSI/CohortMethod.