

Assessing the Implementation of CDIS to Extract FHIR Data and its Transformation to OMOP: FHIR-OMOP-REDCap-CDIS-EHR (FORCE) Pipeline

Lina Sulieman, PhD¹, Alex Cheng, PhD¹, Michelle Jones, MEd¹, Adam Lewis, MS¹,
Francesco Delacqua, MS¹, Karthik Natarajan, PhD², Paul Harris, PhD¹

¹Vanderbilt University Medical Center, Nashville, TN; ²Columbia University Medical Center, New York, New York

Introduction

The Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM) is widely used to support standardized observational research across institutions. By harmonizing data structure and vocabulary, OMOP enables reproducible analyses and facilitates multi-site collaboration. National initiatives, including the All of Us Research Program and the National COVID Cohort Collaborative (N3C), have implemented OMOP to support large-scale data harmonization^{1,2}. However, creating and maintaining OMOP repositories remains challenging, particularly for institutions with limited technical resources. In addition, many Extract, Transform, and Load (ETL) workflows are tightly coupled to local electronic health record (EHR) systems, limiting portability and reuse across organizations. Clinical Data Interoperability Services (CDIS), an advanced REDCap feature, enables the retrieval of selected EHR data through Fast Healthcare Interoperability Resources (FHIR) web services across EHR systems such as Epic and Cerner³. In this pilot study, we evaluate the feasibility of the FHIR-OMOP-REDCap-CDIS-EHR (FORCE) pipeline, an ETL pipeline that uses CDIS to extract EHR data through FHIR and transform it into the OMOP CDM.

Methods

From Vanderbilt University Medical Center (VUMC) and Kansas University Medical Center (KUMC), we randomly selected 100 patients who met criteria designed to ensure sufficient structured EHR data for ETL evaluation. Eligible patients had at least three outpatient visits, at least one inpatient visit, at least five distinct medications prescribed across three or more days, at least two procedures, at least five measurements recorded across three or more days, and at least five unique conditions recorded across three or more days. For each patient, we used Clinical Data Interoperability Services (CDIS) to extract structured EHR data and store the output as FHIR JSON files. We then developed a transformation script to convert CDIS-extracted FHIR resources into OMOP CDM tables. To map CDIS-extracted concepts to OMOP-standard concepts, we used code-based matching supplemented by exact code and matching, as well as identifying similar concepts using MedEmbed-small-v0.1. Because both VUMC and KUMC maintain institutional OMOP EHR instances, we compared the number of transformed rows generated by the FORCE pipeline with the corresponding row counts in the existing OMOP databases.

Results

The CDIS FHIR-to-OMOP transformation pipeline demonstrated meaningful concordance with each site's native OMOP instance across several clinical domains at VUMC and KUMC. Agreement was strongest for Drug Exposure, where FHIR-derived and native record counts were closely aligned at both sites (VUMC: 17,356 vs. 19,294; KUMC: 18,961 vs. 16,443), as Table 1 shows. Condition Occurrence and Measurements—the two highest-volume clinical domains—were also represented in both pipelines, with the FHIR-to-OMOP approach recovering a substantial share of records at KUMC (Condition Occurrence: 22,251 vs. 32,789; Measurements: 60,527 vs. 87,082). Concordance varied by domain and site, however, and several gaps were evident. The FHIR pipeline captured a smaller proportion of Procedures records at both sites (VUMC: 2,047 vs. 9,267; KUMC: 2,831 vs. 5,631) and was most divergent for Observation at VUMC (1,225 vs. 17,021), while at KUMC the same domain exceeded native counts (20,447 vs. 14,986). Visit Occurrence showed the opposite pattern at VUMC, where FHIR-derived records surpassed the native instance (16,508 vs. 8,293). Overall, these results suggest that the CDIS FHIR-to-OMOP transformation achieves reasonable cross-domain concordance with established ETL processes, while highlighting domain- and site-specific differences—most notably in Procedures and Observation—that warrant further harmonization. In VUMC data, CDIS FHIR-to-OMOP provided 26 dental findings, 45 genomics, 222 allergies, and 4315 reasons for visits. In KUMC, FHIR-to-OMOP provided 265 allergies and 13231 reasons for visits. Genomics and dental findings extraction did not contain any data that might indicate some minor issue with access permissions for the user who performed the extraction.

Table 1. The number of rows generated by FORCE compared to institutional OMOP database in two institutions: Vanderbilt University Medical Center (VUMC) and Kansas University Medical Center (KUMC)

OMOP data type	CDIS source	VUMC		KUMC	
		CDIS FHIR→OMOP	VUMC OMOP	CDIS FHIR→OMOP	KUMC OMOP
Conditions Occurrence	Diagnosis, problem list, medical history, infections	11454	23490	22,251	32,789
Drug Exposure	Medications, immunizations	17356	19294	18,961	16,443
Procedures	Procedures	2047	9267	2,831	5,631
Measurements	Lab, vital signs	85568	202469	60,527	87,082
Visit Occurrence	Encounters, appointments, scheduled surgeries	16508	8293	18,427	35,947
Observation	Social history, Coverage, core characteristic	1225	17021	20,447	14,986

Discussion and Conclusion

This pilot study demonstrates the feasibility of using the FORCE pipeline to extract structured EHR data through CDIS/FHIR and transform that data into OMOP-compatible tables across two institutions. FORCE generated records across all evaluated OMOP domains, including conditions, drug exposures, procedures, measurements, visits, and observations, supporting the potential of CDIS as a reusable interoperability layer for OMOP ETL development.

Row count differences between FORCE -generated OMOP records and existing institutional OMOP instances varied by site and domain. These differences likely reflect variation in FHIR resource availability, CDIS extraction scope, and site-specific OMOP ETL logic. For example, FORCE produced higher counts for selected domains (VUMC visit occurrence; KUMC drug exposure and observation), while producing lower counts for conditions, procedures, and measurements; the largest discrepancy was VUMC observation, at roughly 7% of the native count (1,225 vs. 17,021). Upon reviewing the observation at VUMC, we noticed that Observation table included Consult visits that might be a better fit in the Observation table. These findings suggest that CDIS-based extraction can capture meaningful structured EHR data. However, because these comparisons are based on record counts rather than concept- or patient-level matching, higher counts cannot be interpreted as greater completeness and may indicate differences in data granularity. Additional evaluation is therefore needed to understand domain-specific completeness and mapping accuracy. CDIS FHIR was able to extract data that the current OMOP terminology may not yet support, such as dental findings, reason for visit, and genomics. Those data types can be important for researchers and making them available can support their research.

Overall, FORCE offers a promising approach for reducing the technical burden of OMOP implementation by leveraging existing REDCap CDIS functionality and FHIR-based EHR interoperability. Future work will focus on expanding the number of evaluated patients and sites, assessing patient-level and concept-level concordance with institutional OMOP repositories, improving vocabulary mapping, and evaluating the portability of the pipeline across additional EHR environments. This work represents an initial step toward a scalable and portable CDIS–FHIR–OMOP ETL framework for institutions seeking to participate in OMOP-based observational research.

References

1. The “All of Us” Research Program. *New England Journal of Medicine*. 2019;381(7):668–76. doi:10.1056/NEJMSr1809937
2. Haendel MA, Chute CG, Bennett TD, et al. The National COVID Cohort Collaborative (N3C): rationale, design, infrastructure, and deployment. *J Am Med Inform Assoc*. 2021 Mar;28(3):427-443. doi:10.1093/jamia/ocaa196.

3. Cheng AC, Duda SN, Taylor R, et al. REDCap on FHIR: Clinical Data Interoperability Services. *J Biomed Inform.* 2021;121:103871. doi:10.1016/j.jbi.2021.103871.