Bridging Standards: Creating OMOP data via Fast Healthcare Interoperability Resources (FHIR) and Health Information Networks

Stephanie Hong MS, FAMIA[1], Thanaphop Na Nakhonphanom, MD, MMedSc, MS[2], Bryan Laraway, MS[2], Tanner Zhang MD, MS[1], Yvette Chen[1], MS, Richard Moffitt, PhD[3], Rob Schuff, MS[4], Tursynay Issabekova, MBA[2], Christopher Chute, MD, DrPH[1], Josh Lemieux, BA[4], Melissa Haendel, PhD[2], William Hogan MD, MS[5], Emily Pfaff, MS, PhD[2], Shahim Essaid, MD[2]

1. Johns Hopkins University School of Medicine, 2. University of North Carolina Chapel Hill, 3. Emory University, 4. OCHIN, 5. Medical College of Wisconsin

Background

The *All of Us* (AoU) Research Program is a landmark initiative to collect rich, multimodal health data from one million diverse participants across the United States, with a key component being the inclusion of electronic health record (EHR) data. While participating health organizations submit EHR data in the OMOP Common Data Model (CDM) format, many AoU participants have incomplete or missing records. To address this gap, the Center for Linkage and Acquisition of Data (CLAD) launched a pilot to supplement AoU's EHR data by leveraging participant-mediated requests through Health Information Networks and Exchanges (HINs/HIEs). These requests, based on prior participant authorization, require no additional action from participants. We utilized participant-mediated queries on HIN/HIE using Fast Healthcare Interoperability Resources (FHIR) API calls and compared them to source data submitted directly by the Health Provider Organizations (HPOs).

Methods

The CLAD pilot partnered with eHealth Exchange, a national Health Information Network (HIN) participating in the Trusted Exchange Framework and Common Agreement (TEFCA). Cedars Sinai was our pilot site for this phase of the project. Using Master Patient Index (MPI) identifiers for Cedars-Sinai participants, a set of FHIR queries were made through eHealth Exchange to retrieve structured clinical data in the form of FHIR Resource JSON responses, using participant-mediated consent mechanisms.

These responses were parsed to extract CodeableConcept elements representing clinical concepts such as conditions, procedures, lab results, medications, and demographics. A Spark-optimized transformation pipeline then performed both syntactic (format and structure alignment) and semantic (terminology mapping) harmonization. A dynamic lookup table was used to map extracted elements to standardized OMOP concepts using terminologies such as SNOMED CT, LOINC, and RxNorm. The harmonized data were loaded into OMOP CDM–compliant tables across multiple domains—including condition, procedure, measurement, drug exposure, device exposure, observation, and visit occurrence—supporting cross-site interoperability and standardized analytics.

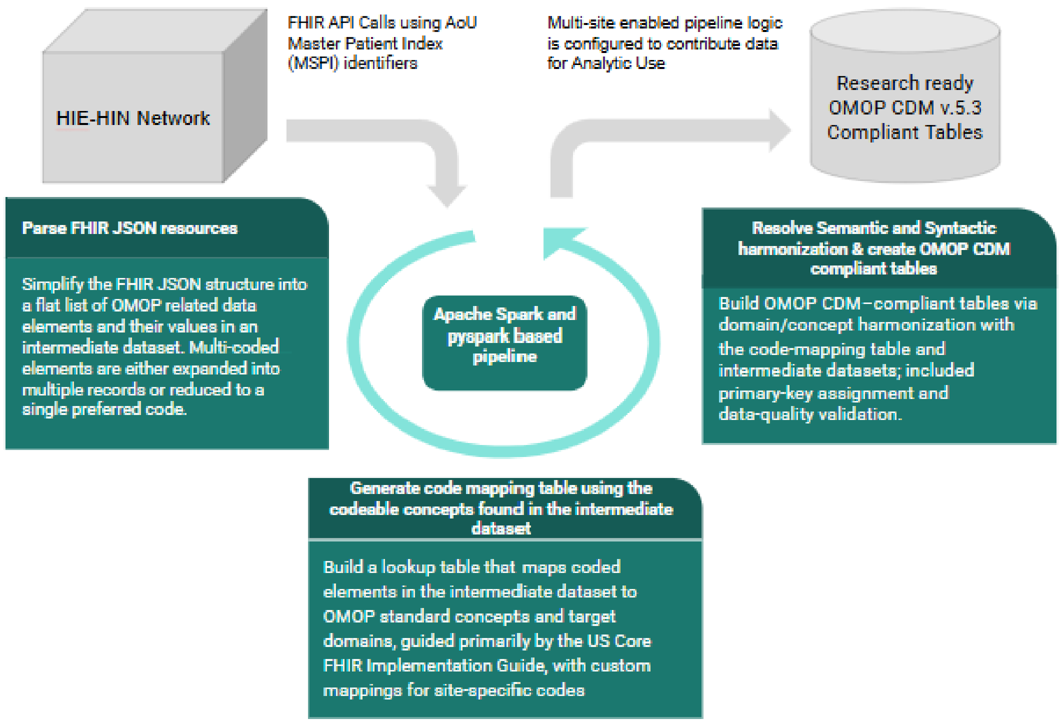

Figure 1. -FHIR to OMOP Transformation Architecture.

The FHIR-to-OMOP data transformation pipeline was built for scalability and efficient processing of high-volume FHIR API responses, consisting of 122 Python and 20 SQL transformation steps. It uses a dynamic lookup table to map US Core terminology from FHIR resources to OMOP standard concepts, with support for vocabularies such as SNOMED CT, LOINC, and RxNorm. A curated value-set mapping table addresses EPIC-specific custom codes, particularly for encounters and devices. The pipeline performs both semantic and syntactic harmonization, resolving structural differences and bridging mismatches between FHIR and OMOP vocabulary IDs. Designed for multi-site ingestion, the pipeline is configurable to accommodate additional FHIR payloads using a carbon copy logic—enabling a unified codebase to support new HPOs as data contributing partners while allowing site-specific customizations.

Results

As of April 4, 2025, over 10.5 million FHIR resource rows had been processed. A total of 1,443 unique data elements were mapped—covering 306 Epic-specific encounter codes and 21 device codes—with 49,493 unique CodeableConcept entries captured in the dynamic vocabulary mapping table.
This pilot demonstrated the feasibility and utility of using FHIR and health information network–based data exchange to scale EHR data acquisition and harmonization for precision medicine research.
As a result of running the transformation pipeline, the processed data were converted into the following OMOP CDM tables:

Transformed OMOP CDM Tables		
FHIR	OMOP	Rows
Patient	Person	7,047
Encounter	Visit_occurrence	591,054
Condition	Condition_occurrence	2,194,240
Procedure, Immunization	Procedure_occurrence	37,715
Medication, MedicationAdministration, Immunization	Drug_exposure	703,433
Device Related	Device_exposure	369
Observation	Measurement	4,057,184
Patient, Condition, Observation	Observation	3,670,804
Practitioner	Provider	53,186
Organization	Care_site	n/a
Supporting Reference Mapping Table		
mspi_to_omop_person_xwalk		7,118
code_map		49,493

Transformed OMOP CDM Tables	
FHIR_value_set_mapping_table	1,443

Table 1. FHIR Resource data transformation results.

Figure 2 illustrates the value of incorporating Cedars-Sinai FHIR-based submissions to enhance the completeness of clinical data. The FHIR feed significantly enriches visit data by contributing numerous additional records—particularly from 2023 onward—not captured in the DRC. Notably, the DRC payload had a data cutoff of October 2023, which highlights the FHIR feed's ability to extend longitudinal coverage beyond the existing dataset. Strong coverage is observed across all eight randomly selected patients, though patient-level discrepancies remain; for example, Patient 5 has pre-2022 visits in the FHIR data that are missing from the DRC. These results demonstrate that FHIR submissions meaningfully improve data completeness and highlight FHIR's utility as a scalable source of structured clinical information for network studies.

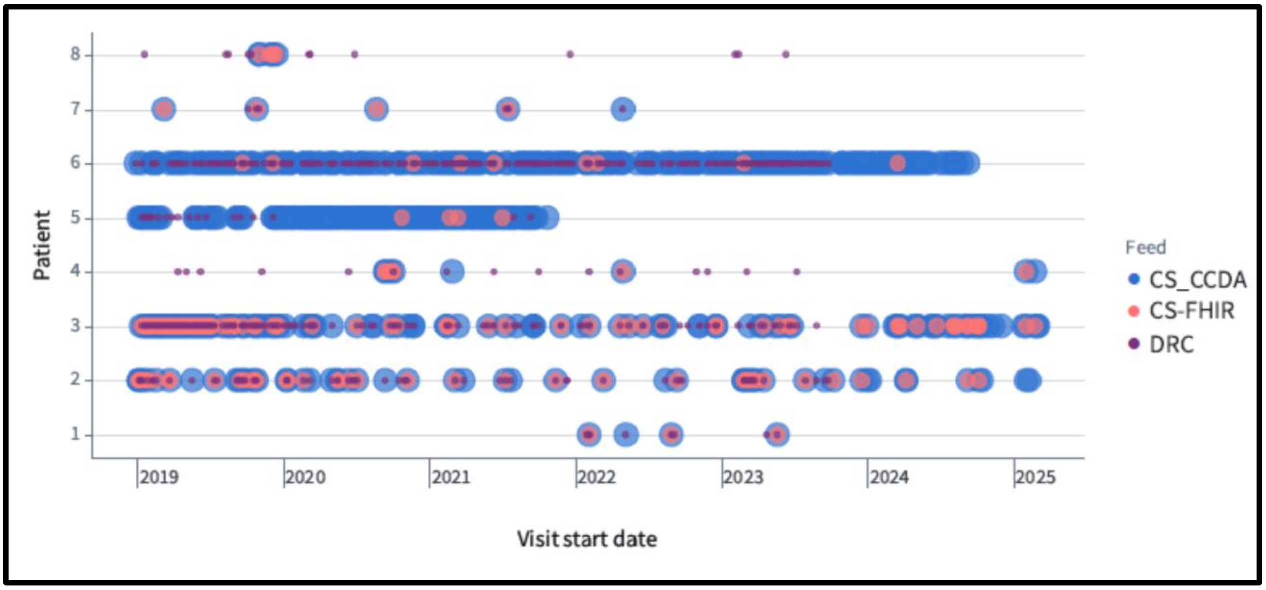

Figure 2. Example data temporal comparison across different HIN/HIE queries versus direct submission. Cedars Sinai FHIR and CCDA responses are shown versus the direct submission to AoU. The visit timeline for 8 Random Patients is shown, illustrating the complementarity of the content that is retrieved via the different mechanisms.

Despite the overall high quality of the integrated data, several issues were identified during the transformation process. A small number of duplicate MSPI entries (0.01%) indicated the need for improved deduplication strategies. A major challenge was the high rate of unmapped records resulting from Epic-specific codes, particularly within Encounter (45.8%), Procedure (51.6%), and Device (93.5%)

resources. Additionally, gaps in code system recognition led to unmapped MedicationRequest (4.1%) and Immunization (0.07%) records. These issues highlight the ongoing need for early vocabulary harmonization and robust maintenance of terminology mapping tools to ensure comprehensive and accurate OMOP CDM integration.

Conclusion

As of April 4, 2025, over 10.5 million FHIR resource rows had been processed, with 51,022 data elements mapped—covering 306 EPIC-specific encounter codes and 21 device codes—49,490 CodeableConcept entries captured in the dynamic vocabulary mapping table. This pilot assessed the feasibility and utility of using FHIR and network-based data exchange to scale EHR data acquisition and harmonization for precision medicine research.

The transformed data were successfully integrated into the OMOP CDM, populating millions of records across key clinical domains, including over 2.1 million condition occurrences and 4 million measurements. Supporting reference tables facilitated concept mapping and person linkage. These results highlight the scalability and effectiveness of FHIR-based data acquisition and transformation in enriching EHR completeness for All of Us participants, demonstrating the broader value of leveraging national health data infrastructure and interoperability standards for large-scale biomedical research.

Reference

1. US Core Implementation Guide (IG) - https://www.hl7.org/fhir/us/core/; Defines a minimum set of data elements and profiles (based on USCDI) required for nationwide exchange of patient data in the U.S, Based On: HL7 FHIR R4 (4.0.1)
2. USCDI - https://www.healthit.gov/isa/united-states-core-data-interoperability-uscdi ; Source for what data classes and elements are required for US Core IG, current version USCDI v5 (as of 2025)
3. HL7 FHIR R4 Specification -https://hl7.org/fhir/R4/; specifies full base FHIR resource definitions, RESTful API details, data types and structure definitions, terminology and value set support
4. FHIR mapping language - https://build.fhir.org/mapping-language.html
https://www.healthit.gov/isa/united-states-core-data-interoperability-uscd
5. HL7 FHIR IG Registry- https://registry.fhir.org/
6. Information Blocking Actors - https://www.healthit.gov/sites/default/files/2024-04/IB_Actors_Fact_Sheet_508_0.pdf

Acknowledgement

This work was supported by NIH All of Us Research Program Award No. OT2 OD036113-01. Please contact info@cladteam.io for future collaboration work.