

Large-Scale Implementation of the Medical Imaging Common Data Model at Johns Hopkins Medicine

Jen (Woo Yeon) Park, PhD¹, Teri Sippel Schmidt, MS¹, Blake Dewey, PhD², Paul Nagy, PhD¹

¹Biomedical Informatics and Data Science, School of Medicine, Johns Hopkins University, Baltimore, MD

²Department of Neurology, School of Medicine, Johns Hopkins University, Baltimore, MD

Background

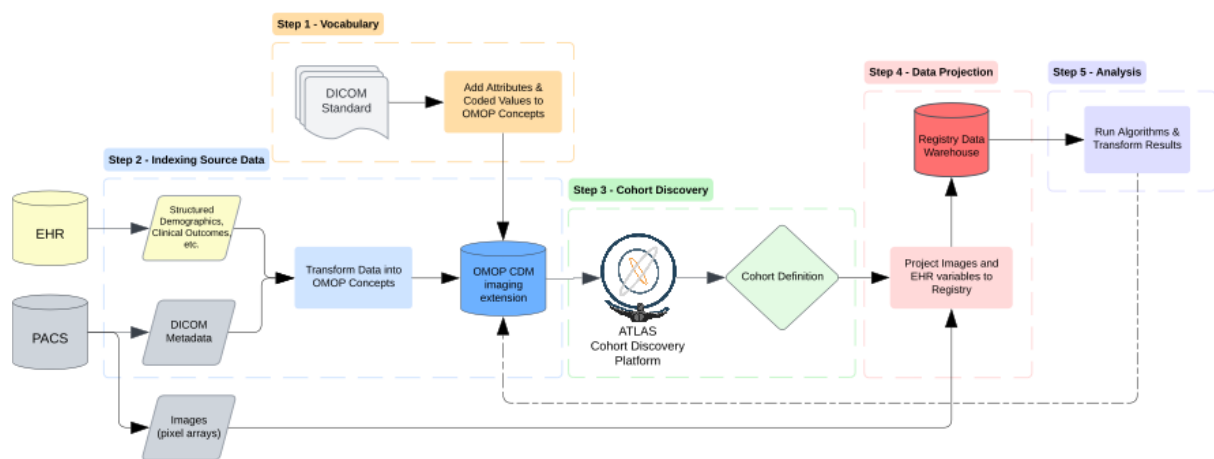
Medical imaging data captured in clinical Picture Archiving and Communication Systems remains structurally isolated from EHR-based research infrastructure, preventing reproducible multimodal observational research[1]. Since its proposal in 2024, the Medical Imaging Common Data Model (MI-CDM), an extension of the OMOP CDM introducing Image_occurrence and Image_feature tables, has been presented as a framework to bridge this gap[2]. Critics and collaborators alike have asked the same question: does it work in a real clinical environment? This abstract reports the first large-scale clinical implementation of MI-CDM, executed at Johns Hopkins Medicine (JHM) using MRI data from the Alzheimer's Disease Precision Medicine Center of Excellence registry.

Methods

DICOM metadata was extracted from the JHM Vendor Neutral Archive spanning 2016 to 2025, covering five scanner manufacturers (Siemens, GE, Philips, Toshiba, Hitachi). The MI-CDM pipeline encompassing vocabulary integration, structural extension, and metadata ETL was applied to the existing institutional OMOP CDM instance without modifying clinical data tables (Figure 1). Procedure linkage was established using Accession Number and DICOM Study Instance UID, evaluated comparatively. Sixteen MRI acquisition attributes were selectively extracted and loaded

into Image_occurrence and Measurement tables, with provenance preserved via measurement_event_id linked to Image_occurrence rather than duplicating rows in Image_feature. Data quality was evaluated across three dimensions including completeness by attribute and vendor, DICOM conformance of Code String attributes against DICOM Standard Part 3, and numeric plausibility using vendor-stratified distributions. To demonstrate reusability, the Alzheimer's disease cohort phenotype built in ATLAS for ADNI was re-executed against the JHM MI-CDM instance with a single procedural code adaptation[3].

Figure 1. The ETL pipeline diagram



Results

The ETL produced 210,173 imaging studies, 3,196,998 series, and approximately 41 million measurement values across 16 attributes for 81,899 patients. Accession Number linkage achieved a 99% match rate versus 78% for Study Instance UID, establishing a clear preferred pathway for clinical deployments. Eleven attributes including Echo Time, Repetition Time, Scanning Sequence, Sequence Variant, Flip Angle, and Slice Thickness achieved near-perfect completeness averaging 99% across all vendors (Figure 2). Research-relevant parameters were more variable, with Contrast/Bolus Agent averaging 15% completeness, Inversion Time 23%, and Trigger Time 3%, consistent with conditional recording requirements under the DICOM Standard. Body Part Examined was populated at 83% overall, ranging from 55% for GE to 92% for Toshiba, and 186 of 268 unique

values initially failed concept mapping, requiring study description fallback parsing. Most Code String attributes conformed to DICOM Standard defined terms at 100%. Scan Options was the notable exception with conformance ranging from 8% for GE to 100% for Hitachi, reflecting vendor non-standardization of a free-text-permissive field. Numeric plausibility was confirmed across all attributes within clinically expected ranges. Cohort regeneration produced 19,085 patients and 62,436 imaging series distributed plausibly across diagnostic group (CN > MCI > AD), age stratum, and sex (Table 1).

Figure 2. The data completion rate by attribute and manufacturers

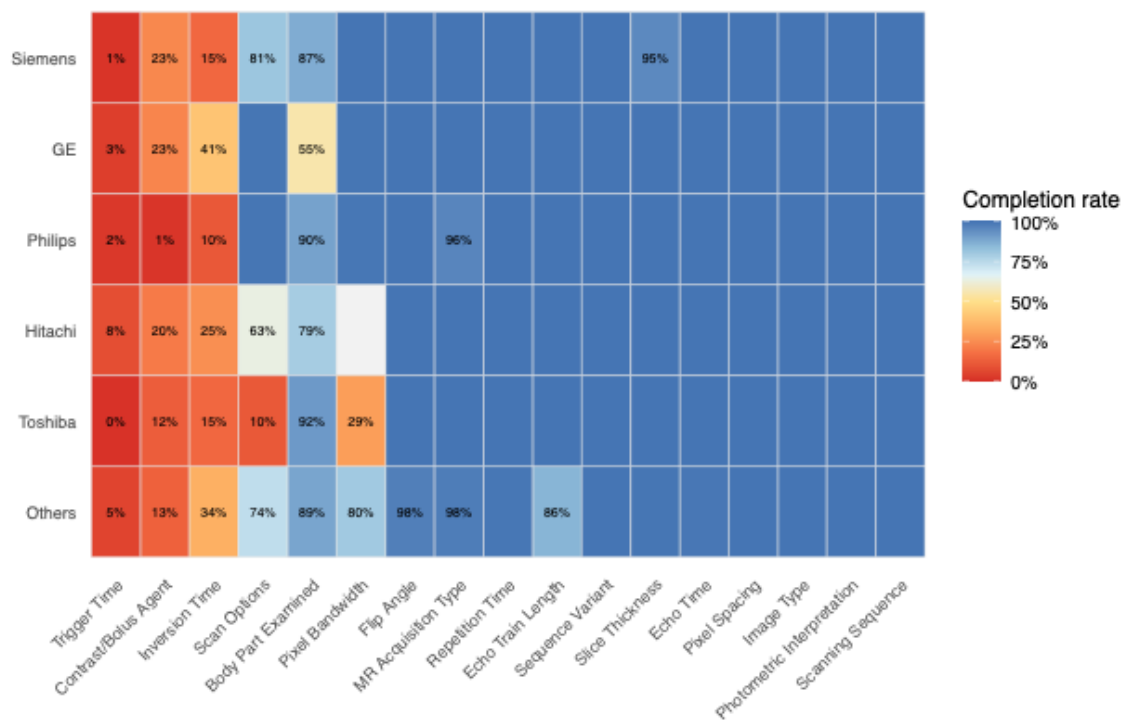


Table 1. JHM Cohort by Diagnostic Group, Age stratum, and Sex

Sex	Age group	CN Patient (Series)	MCI Patients (Series)	AD Patients (Series)
Male	55-65	3,042 (10,678)	141 (337)	40 (102)
Male	66-79	3,380 (10,960)	320 (895)	231 (603)
Male	80-95	964 (2,544)	121 (292)	144 (341)
Female	55-65	4,183 (16,748)	133 (378)	61 (152)
Female	66-79	4,104 (13,912)	306 (802)	343 (914)
Female	80-95	1,153 (2,973)	110 (260)	205 (545)

Conclusions

The MI-CDM is operationally viable at clinical scale. Over 200,000 real-world MRI studies were successfully standardized within an existing OMOP CDM instance, confirming that the extension transfers from a controlled research registry to a living clinical environment without structural modification. Data quality findings characterize the known gap between DICOM as specified and DICOM as practiced across vendors, providing a reference baseline for future implementations. The ETL convention established here, covering linkage strategy, attribute selection, SOP class scoping, and PHI handling, offers a replicable deployment pattern that reduces the adoption barrier for other institutions. Cohort regeneration with minimal modification demonstrates that the infrastructure supports portable, computable research definitions in practice. The MI-CDM has moved from theoretical framework to operational foundation.

References

1. Jeon K, Park WY, Kahn CE Jr, Nagy P, You SC, Yoon SH. Advancing Medical Imaging Research Through Standardization: The Path to Rapid Development, Rigorous Validation, and Robust Reproducibility. *Invest Radiol.* 2025;60(1):1-10. doi:10.1097/RLI.0000000000001106
2. Park WY, Jeon K, Schmidt TS, et al. Development of Medical Imaging Data Standardization for Imaging-Based Observational Research: OMOP Common Data Model Extension. *J Imaging Inform Med.* 2024;37(2):899-908. doi:10.1007/s10278-024-00982-6
3. Park WY, Sippel Schmidt T, Salvador G, et al. Breaking data silos: incorporating the DICOM imaging standard into the OMOP CDM to enable multimodal research. *J Am Med Inform Assoc.* 2025;32(10):1533-1541. doi:10.1093/jamia/ocaf091