Feasibility of Integrating DICOM Headers into the OMOP Medical Imaging CDM : A Pilot Study Using Chest CT Data

PRESENTER: Kyulee Jeon

INTRODUCTION

- Advancement in healthcare artificial intelligence demands integration of diverse data sources, yet OMOP CDM lacks support for imaging data
- While **Medical Image Workgroup** proposed **OMOP CDM extension for medical imaging** to incorporate Digital Imaging and Communication in Medicine (DICOM) into OMOP CDM, it remains theoretical without implementation
- This study implements **OMOP CDM Imaging Extension framework** using real-world imaging data to evaluate the feasibility of incorporating DICOM headers from chest CT scans into the standardized framework

METHODS

Step 1: Data Selection and Processing

- Extracted DICOM metadata (tag-value pairs) from randomly selected 200 chest CT studies of lung cancer patients (Severance Hospital, 2014-2023)
- Excluded metadata with text values >16 characters and data already harmonized in OMOP CDM (e.g., patient sex), while keeping only first instance of redundant series-level data

Step 2: Semantic Standardization

- Mapped all DICOM tags to temporary custom concept IDs from DICOM2OMOP github
- DICOM values were mapped to: (1) SNOMED-CT codes for predefined values in DICOM Standard, (2) custom concept IDs for other code string values, and (3) raw values stored in value_as_number or value_source_value

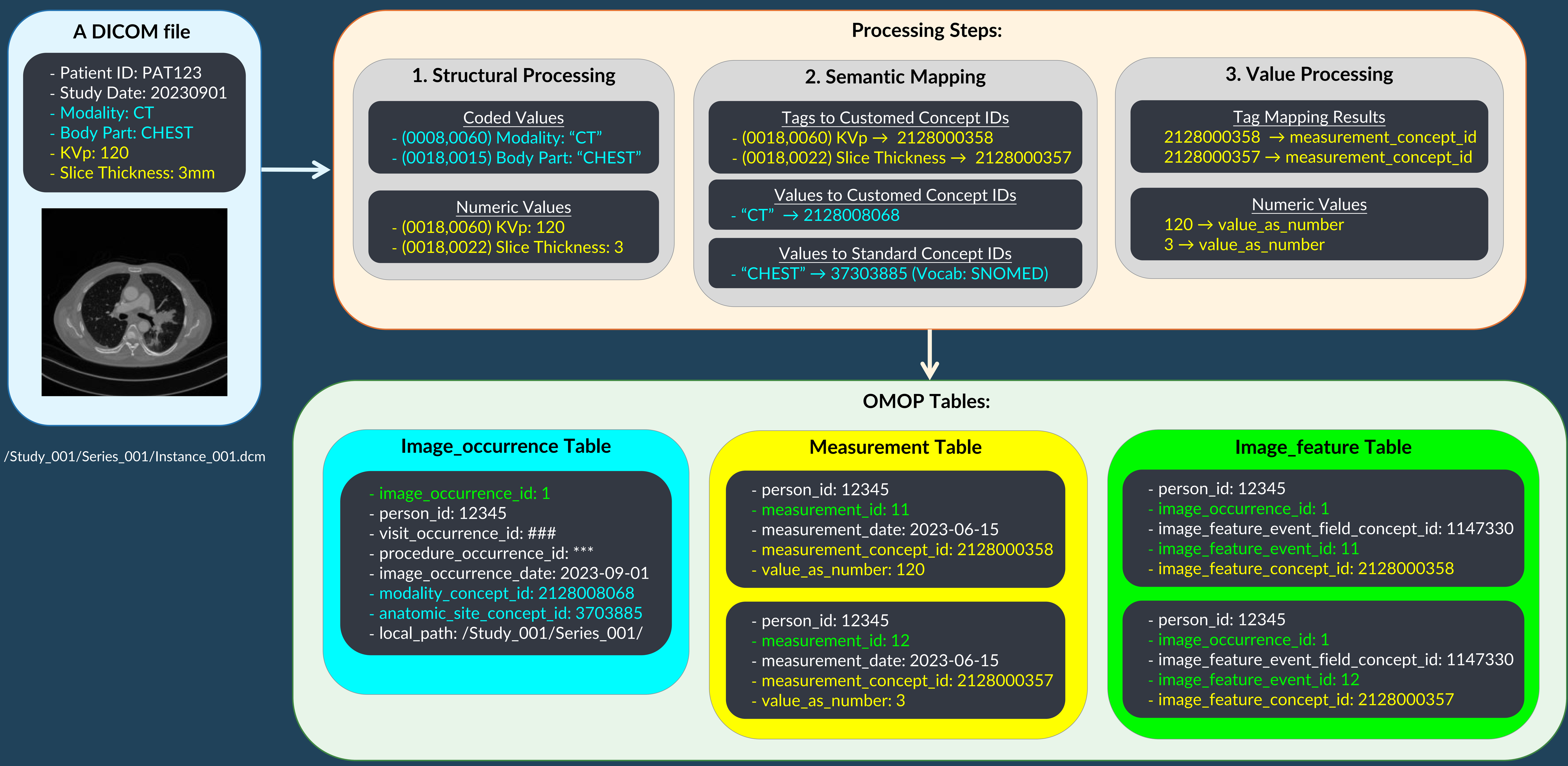
Step 3: CDM Extension Implementation

- Built **Image_occurrence table** for imaging events and **Measurement table** with DICOM metadata (mapping tags to measurement_concept_id and storing values as value_source_value, value_as_number, or value_concept_id)
- Created **Image_feature table** to link measurements with source imaging for traceability

RESULTS

- The comprehensive dataset included 200 CT studies (196 patients) comprising 1,576 series and 225,289 DICOM files (115.5 GiB), successfully integrated into three tables: Image_occurrence (1,576 rows), Measurement (1,622,381 rows), and Image_feature (1,622,381 rows)

Bridging the Gap: Real-World Implementation of OMOP CDM Imaging Extension

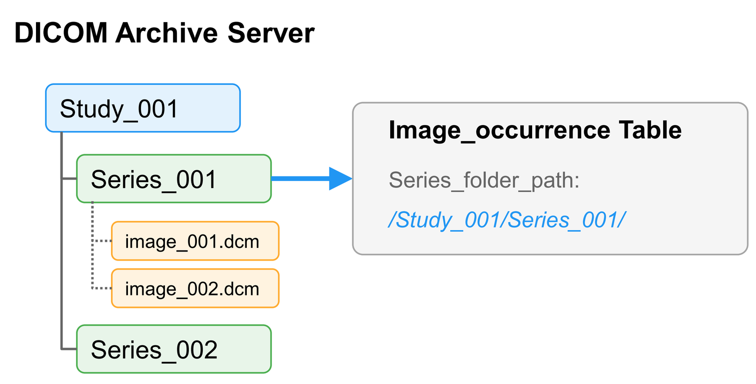


Example of DICOM metadata:

Tag	Attribute Name	Value (Example)
(0008,0020)	Study Date	20230901
(0008,0060)	Modality	CT
(0008,0070)	Manufacturer	Siemens Healthineers
(0008,1090)	Manufacturer's Model Name	SOMATOM Force
(0018,0015)	Body Part Examined	CHEST
(0018,0022)	Scan Options	HELICAL MODE
(0018,0050)	Slice Thickness	3
(0018,0060)	KVP	120
(0018,1041)	Contrast/Bolus Volume	0
(0018,9345)	CTDIvol	4.13998747826087 (mGy)
(0028,0004)	Photometric Interpretation	MONOCHROME2

Hierarchical Structure of DICOM:

- In DICOM, a single imaging study can contain multiple series of images acquired with different protocols and parameters.
- This framework processes imaging data at the series-level as its basic unit.



Why It Matters for OHDSI:

- This implementation opens doors for incorporating imaging biomarkers into observational studies, enabling new research questions that combine imaging and clinical data at scale.

Next Steps:

- Map imaging vocabularies (e.g., DICOM, RadLex) to OMOP concepts
- Define essential DICOM tags for observational research
- Handle long free-text metadata such as Protocol Name, Series description

Kyulee Jeon[1,2], Woo Yeon Park[3], Teri Sippel Schmidt[3], Paul Nagy[3], Seng Chan You[1,2]

Contact:
- Kyulee Jeon (jklee320@yuhs.ac)
- Seng Chan You (chandryou@yuhs.ac)

YONSEI UNIVERSITY
JOHNS HOPKINS SCHOOL of MEDICINE
OHDSI