Application of the R-CDM extension to capture metadata and features extracted from quantitative brain MRI and CT data

Jelle Praet[a], Jared Houghtaling[b], Frederic Jung[b], Steve De Backer[a], Jeroen Pinxten[a] and Dirk Smeets[a]

[a] Icometrix
[b] edenceHealth NV

Background:

The use of quantitative radiology data has long been a challenge within the Observational Health and Data Science Institute (OHDSI) community as the Observational Medical Outcomes Partnership (OMOP) common data model (CDM) is unable to capture all facets of radiological data. The research group of Rae Woong Park has recently addressed this limitation by proposing to extend the OMOP CDM with the R-CDM radiology tables for OMOP **[1]**. Icometrix and edenceHealth have joined forces, as part of the European Health Data and Evidence Network (EHDEN) project, to implement this R-CDM extension at Icometrix. Icometrix' proprietary software, icobrain, is a fully automated, deep-learning-based pipeline that provides standardised analytical measurements of brain magnetic resonance imaging (MRI) and computed tomography (CT) images for patients with multiple sclerosis (MS), dementia, traumatic brain injury, stroke and epilepsy. To make this standardised approach for quantitative radiological data available within OMOP, we built an ETL and combined it with a custom ontology and maintenance application, named *PROSA*, developed by edenceHealth **[2]**. AI-driven pipelines, like icobrain, generate massive amounts of quantitative data, which until recently was not available as radiolgoists had to manually and subjectively evaluate images. The adoption of the R-CDM extension meant we added 2 additional tables to the core OMOP CDM, in particular the 'Radiology Occurance' and 'Radiology Image' tables. These tables enable us to keep track of both hierarchical connections between, and the metadata about, radiological studies and their associated images for a given patient.

Methods:

We have created a structural mapping to link the DICOM metadata from icobrain to the R-CDM extension, with specific mappings shown below in **Tables 1 and 2.** Briefly, a python script first parses a repository of DICOM images and creates a data frame with one row per image and all associated metadata fields as columns. We integrated this script within our existing Extract, Transform, Load (ETL) processes, such that all OMOP CDM tables (both standard and extension) are properly filled per ETL execution. We are in the process of defining additional quality check queries to investigate the metadata contained in those tables, and to cross reference data in the standard OMOP tables with the supplemental extension tables.

Table 1. Source to target mappings for Radiology Occurrence Table

Source Field	Target Field	Req.	Type	Note
autogenerate	radiology_occ_id (PK)	Yes	Int	UID for each image session
PatientID	person_id (FK)	Yes	Int	FK to PERSON table, person who was imaged
SeriesDate	radiology_occ_date	No	Date	Date when the study was taken
SeriesDate + Time	radiology_occ_datetime	No	Datetime	Date and time when the study was taken
Modality	modality	No	VC(10)	DICOM file type
Manufacturer	manufacturer	No	VC(50)	Manufacturing company of imaging equipment
Custom Id	protocol_concept_id	No	Int	Custom concepts generated using unique protocols in dataset
Custom Name	protocol_source_value	No	VC(255)	Custom concept name
SeriesNumber	count_of_series	No	Int	Total series generated per study
ImagesInAcquisition	count_of_images	No	Int	Total images generated per study
ALL UNMAPPED	radiology_note	No	VC(Max)	Concatentate unmapped info into note

* Table inspired by content presented in **[1]**

Note that we needed to define custom concepts to adequately describe imaging protocols used to generate DICOM files, and we mapped those concepts and associated protocols to the *protocol_concept_id* field in the Radiology Occurrence table. Moreover, we concatenated the metadata not captured by the extension tables into the *radiology_note* field to ensure complete data coverage within the OMOP data set.

Results:

The DICOM parser we implemented was able to transform more than 100’000 DICOM files with both visual and meta information. Importantly, by using the radiology extension criteria, we are now able to generate custom cohorts for patients with MS, multiple sclerosis, brain injury, stroke and epilepsy, that account for the technical specifications of the imaging tools. This capability is critical, as it will lend insight into the effects of imaging protocol and manufacture bias, and it will also enable us to investigate and further train the algorithms supporting icobrain in its feature extraction and image quantification in the context of OMOP data.

Table 2. Source to target mappings for Radiology Image Table

Source Field	Target Field	Req	Type	Note
	radiology_image_id (PK)	Yes	Int	UID for each image
	radiology_occ_id (FK)	Yes	Int	UID for each image session
SeriesInstanceUID	radiology_series_id	Yes	Int	UID of each series
File_path	file_path	Yes	VC(255)	File path with image files
BodyPartExamined	body_part_source_value	No	VC(20)	Value indicating the photographed body part
Lookup concept using Laterality	laterality_concept_id	No	VC(20)	Image shooting direction (anatomical plane)
Lookup concept using SeriesDescription	series_type_concept_id	No	VC(20)	Value indicating the type of the series
Lookup concept using SeriesDescription	series_type_source_value	No	VC(20)	Additional source values describing the series
ImagesInAcquisition	series_total_number	No	Int	Number of images constituting each series
ImageNumber	series_serial_number	No	Int	Order of images within each series
Rows	image_resolution_rows	No	Int	Image resolution (# horiz. pixels)
Columns	image_resolution_columns	No	Int	Image resolution (# vert. pixels)
SliceThickness	CT_slice_thickness	No	Numeric	Thickness of CT image slide

* Table inspired by content presented in **[1]**

Conclusion:

Taken together, combining the quantitative output of the deep-learning based icobrain software with this newly built ETL enables us to bring quantitative brain imaging to OMOP at scale. This capability offers tremendous value to studies focusing on neurodegenerative diseases; from a clinical point of view, these studies often rely heavily on brain imaging data for diagnosis and prognosis.

References:

1. Park, C., You, S. C., Jeon, H., Jeong, C. W., Choi, J. W., & Park, R. W. (2022). Development and Validation of the Radiology Common Data Model (R-CDM) for the International Standardization of Medical Imaging Data. *Yonsei Medical Journal*, *63*, S74–S83. https://doi.org/10.3349/ymj.2022.63.S74
2. Houghtaling et al. (2023) Development of an OMOP Ontology Application – PROSA – for creation and maintenance of highly granular source concepts within the OMOP vocabulary structure. OHDSI Europe Symposium 2023
3. Belenkaya, R., Gurley, M. J., Golozar, A., Dymshyts, D., Miller, R. T., Williams, A. E., Ratwani, S., Siapos, A., Korsik, V., Warner, J., Scott Campbell, W., Rivera, D., Banokina, T., Modina, E., Bethusamy, S., Morgan Stewart, H., Patel, M., Chen, R., Falconer, T., … Reich, C. (2021). Extending the OMOP Common Data Model and Standardized

Vocabularies to Support Observational Cancer Research. *American Society of Clinical Oncology*, 12–20. https://doi.org/10.1200/CCI.20

4. Shin, S. J., You, S. C., Park, Y. R., Roh, J., Kim, J. H., Haam, S., Reich, C. G., Blacketer, C., Son, D. S., Oh, S., & Park, R. W. (2019). Genomic common data model for seamless interoperation of biomedical data in clinical practice: Retrospective study. *Journal of Medical Internet Research*, *21*(3). https://doi.org/10.2196/13249