



# Cohort Selection of MIDRC Exams Using the LOINC RSNA Radiology Playbook

**Paul E. Kinahan, PhD**, Vice-Chair Radiology, Research, University of Washington

---

## Introduction

To develop artificial intelligence (AI) methods with improved repeatability and reproducibility, a recognized need is for large, publicly available, and curated imaging data sets. Notable public repositories of DICOM images exist, including the recent NIBIB-supported Medical Image and Data Resource Center (MIDRC). MIDRC currently has over 500,000 studies (exams) in process, of which over 180,000 have been processed and publicly available. In addition 20% of studies are sequestered for testing and validation studies. A challenge for users is to select appropriate cohorts using the highly variable Study Descriptions in the DICOM metadata supplied by the providing imaging centers.

## Hypothesis

We hypothesized that using a restricted subset of the Logical Observation Identifiers Names and Codes (LOINC) RSNA radiology playbook could be used for efficient cohort selection from the MIDRC collection. This would make use of the algorithmically-generated and unique LOINC Long Common Name as an adjunct to the DICOM Study Description.

## Methods

We used a restricted set, called the 'MIDRC-LOINC Mapping Table', from the ~10,000 Long Common Names from the LOINC RSNA radiology playbook. This set was selected in a hierarchical manner to balance the number of codes used versus the level of detail that is anticipated for cohort selection. A sample of 146,600 DR, CR, DX and CT exams containing 1,400 unique DICOM Study Descriptions in a highly-skewed long-tailed distribution was mapped with the MIDRC-LOINC Mapping Table.

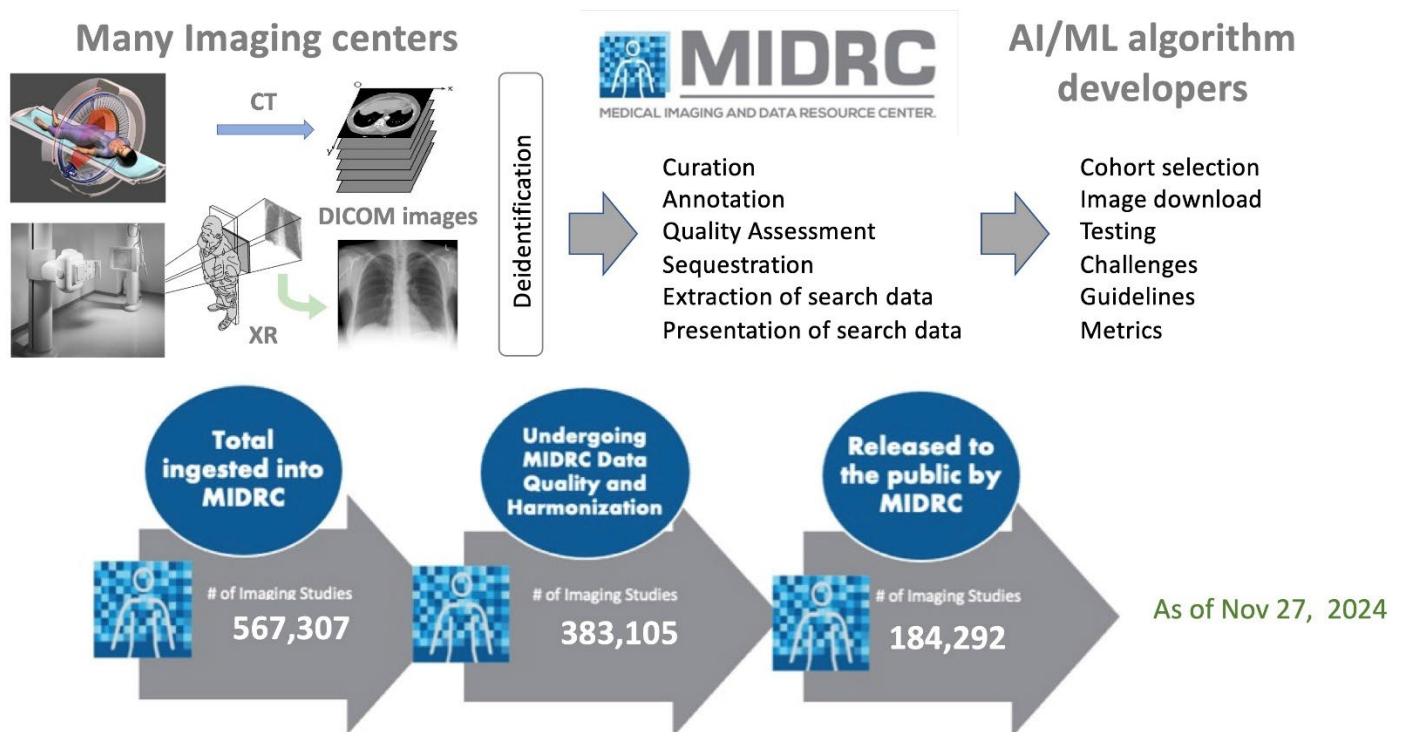
## Results

We were able to match over 97% of the Study Descriptions to 65 unique LOINC Long Common Names.

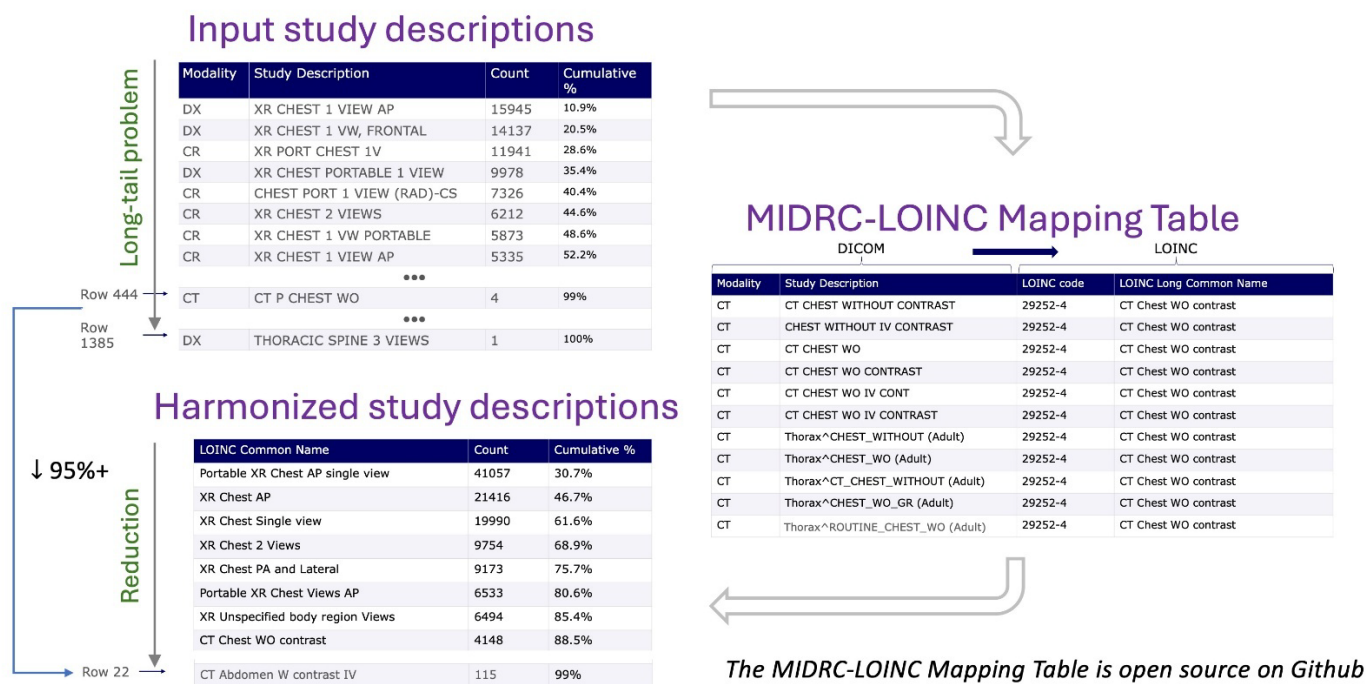
## Conclusion

Using DICOM metadata, most incoming imaging data to MIDRC can be mapped to a restricted set of LOINC Long Common Names that are suitable for cohort selection from DICOM exams pooled from multiple imaging centers. The MIDRC-LOINC Mapping Table and a companion LOINC code attribute table are being regularly updated and are publicly available on GitHub.

Figure(s)



**Figure 1.** The primary goal of MIDRC is building and expanding its artificial intelligence-ready data commons to fuel machine intelligence research, coupled with standardization, curation, and compliance with ethical responsibilities to honor patients' privacy. In order to leverage existing infrastructure, MIDRC is a high-quality data commons adhering to the FAIR principles (Findable, Accessible, Interoperable and Re-usable).



**Figure 2.** Process flow for mapping DICOM Study Descriptions to Logical Observation Identifiers Names and Codes (LOINC) Long Common Names for cohort selection. The 95% reduction of the number of terms (Long-tail problem) is indicated on the left-hand side.

## Keywords

Artificial Intelligence/Machine Learning; Imaging Research; Standards & Interoperability