

Pilot Study of Dental CBCT Dataset and OMOP Medical Imaging Extension

Julie Cha¹, Vidhya Venkateswaran¹

¹National Institute of Dental and Craniofacial Research

Background

This abstract describes a pilot study to assess the feasibility of applying the OMOP Medical Imaging Extension, proposed by Park et al.¹, which aims to integrate medical imaging data within the OMOP CDM. Our approach involves mapping DICOM metadata and its annotation data along with the creation of two new tables – `image_occurrence` and `image_feature` for dental CBCT images. The project explored the data quality of dental DICOM files, the availability of dental vocabulary within OMOP standard concepts, and a potential AI imaging use case.

Methods

The ToothFairy dataset is a publicly available dataset of 443 dental CBCT DICOM files, with annotations for the inferior alveolar nerve. Exploratory data analysis of the DICOM attributes evaluated data completeness and value distributions.² The files were filtered to only include CBCT files and exclude localizer and secondary reconstructions.^{3,4,5,6} (Figure 1)

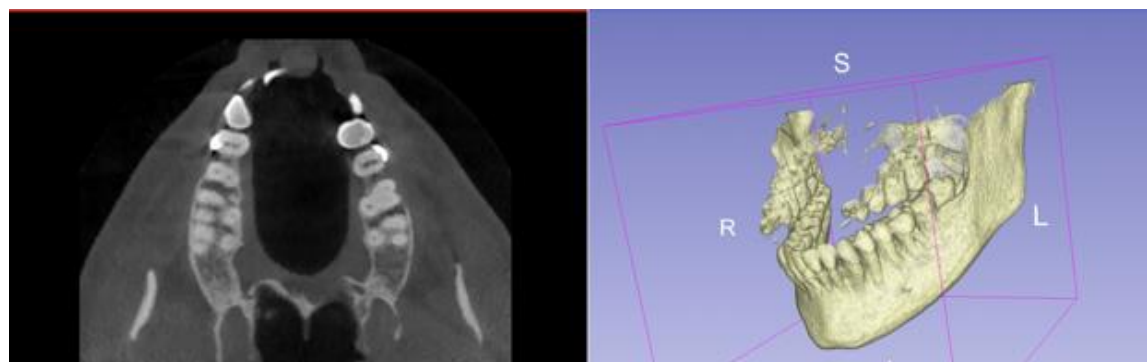


Figure 1. ToothFairy Dataset - dental CBCT DICOM files

The ToothFairy data was then mapped to the `Image_occurrence`, `Procedure_occurrence`, `Image_feature`, and `Measurement` tables of the Medical Imaging Extension. The AI imaging use case explored deriving the presence or absence of individual teeth using the ToothSeg model^{7,8} for tooth instance level segmentation and labeling. Model outputs were stored in the `Image_feature` table.

Results

The ToothFairy dataset showed generally good completeness of DICOM attributes. Missing tags included Body Part Examined, Institution, Series Description, and Provider, while there was 100% fill of features of interest such as Manufacturer, KVP, Exposure, and X-ray Tube Current. Vocabulary challenges were

uncovered during mapping of the data to the Medical Imaging Extension tables. Many imaging acquisition parameters did not have existing entries in the OMOP CDM standard vocabulary and therefore required custom concept IDs. Some common dental terms were found to have either ambiguous or missing mappings. For e.g., for `anatomic_site_concept_id` in the `image_feature` table, the SNOMED concept 4281237 refers to “jaw region structure”, when the area of interest was more specifically the maxilla and mandible. Similarly, `modality_concept_id` does not allow for CBCT of teeth and jaws to be specified and instead is folded under CT of the head or Cone Beam CT Teeth. This is not an appropriate record because dental CBCTs can extend from single tooth to segmental to whole head and neck. Thus, dental CBCT needs additional granularity to record the specific type of CBCT.

For the AI imaging use case, the ToothSeg model applied to the ToothFairy dataset showed generally good segmentation with zero shot inference but showed some inaccuracies in tooth numbering compared with the ground truth. (Figure 2)



Figure 2. ToothSeg model - tooth instance level segmentation and labeling

OMOP Imaging_Feature Table Fields	Corresponding Concept Ids
image_feature_event_field_concept_id	1147330 refers to the MEASUREMENT table (with further linkage to dental-specific terminology)
image_feature_concept_id	Custom ID generated based on DICOM tag (not in OMOP)
image_feature_type_concept_id	Custom ID to record AI algorithm-derived feature based on DICOM (not in OMOP)
image_finding_concept_id	4201530 refers to the entire tooth (SNOMED)
value_as_concept_id	4150446 refers to tooth present, 4122603 refers to tooth absent (SNOMED)
anatomic_site_concept_id	4281237 refers to jaw region structure (SNOMED)

Table 1. Potential mappings for tooth AI-identified presence or absence in a dental CBCT using the imaging extension

Conclusion

This pilot study found that dental CBCT DICOM data can be mapped to the OMOP Medical Imaging Extension (Table 1), with caveats and modifications. The ToothFairy dataset had good DICOM attribute data quality, suggesting a good fit with the proposed extension. Vocabulary gaps were found for both imaging acquisition parameters and dental specific concepts. The vocabulary to support imaging data in the Medical Imaging Extension and dental vocabulary within the OMOP CDM are both still under development.

References

1. Park WY, Jeon K, Schmidt TS, Kondylakis H, Alkasab T, Dewey BE, 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.
2. Ji H, Kim S, Sunwoo L, Jang S, Lee HY, Yoo S. Integrating clinical data and medical imaging in lung cancer: feasibility study using the Observational Medical Outcomes Partnership Common Data Model Extension. *JMIR Med Inform*. 2024;12:e59187. doi:10.2196/59187.
3. ToothFairy [Internet]. Modena: University of Modena and Reggio Emilia; [cited 2026 Jun 2]. Available from: <https://ditto.ing.unimore.it/toothfairy/>
4. Bolelli F, Lumetti L, Vinayahalingam S, Di Bartolomeo M, Pellacani A, Marchesini K, et al. Segmenting the inferior alveolar canal in CBCTs volumes: the ToothFairy Challenge. *IEEE Trans Med Imaging*. 2024.
5. Lumetti L, Pipoli V, Bolelli F, Ficarra E, Grana C. Enhancing patch-based learning for the segmentation of the mandibular canal. *IEEE Access*. 2024.
6. Cipriano M, Allegretti S, Bolelli F, Pollastri F, Grana C. Improving segmentation of the inferior alveolar nerve through deep label propagation. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Piscataway: IEEE; 2022. p. 21105-21114.
7. van Nistelrooij N, Kramer L, Kempers S, Beyer M, Bolelli F, Xi T, et al. ToothSeg: robust tooth instance segmentation and numbering in CBCT using deep learning and self-correction. *IEEE J Biomed Health Inform*. 2026;PP. doi:10.1109/JBHI.2025.3650444.
8. ToothSeg [Internet]. GitHub; [cited 2026 Jun 2]. Available from: <https://github.com/MIC-DKFZ/ToothSeg>