

Using the Oral Health and Disease Ontology to Study Dental Outcomes in National Dental PBRN Practices

W. D. Duncan^{1,3*}, T.P. Thyvalikakath^{2,3}, Z. Siddiqui², M. LaPradd⁴, C. Wen³, J. Zheng³, A. Roberts³, D. Hood³, T. Schleyer⁴, A. Manimangalam², D. B. Rindal⁵, M. W. Jurkovich⁵, T. L. Shea⁵, D. Bogacz⁶, T. Yu⁷, J. L. Fellows⁸, V. V. Gordan⁹, G. H. Gilbert¹⁰, National Dental PBRN Collaborative Group¹⁰

¹Roswell Park Comprehensive Cancer Center, Buffalo, NY; ²Dental Informatics Core, Department of Cariology, Operative Dentistry & Dental Public Health, Indiana University School of Dentistry, Indianapolis, IN; ³Regenstrief Institute, Indianapolis, IN; ⁴Indiana University School of Medicine, Indianapolis, IN; ⁵Institute for Education and Research, HealthPartners, Minneapolis, MN; ⁶White Park Dental, Concord, NH ⁷Brighton Dental Health, Chandler, Arizona, ⁸Kaiser Permanente Center for Health Research, Portland, OR; ⁹University of Florida, Gainesville, FL; ¹⁰University of Alabama at Birmingham, Birmingham, AL

Keywords: *ontology, dental research, dental procedure, posterior composite restoration, root canal treatment*

Abstract

The use of electronic dental records (EDR) has grown rapidly over the past decade, but the development of methods to use EDR data for research and quality improvement is still in its infancy. In this study, we are investigating the feasibility of reusing semantically structured EDR data for research purposes. Our two use cases are to assess (1) longevity of posterior composite restorations (PCR) and (2) tooth loss following root canal treatment (RCT).

To answer these research questions, we recruited 99 National Dental PBRN¹ dental practices that had been using either Dentrix or Eaglesoft – two common EDR software systems – to record patient care for more than five years. The practices transferred the data to us by working with their EDR software vendors. After appropriate agreements were in place, the dental practice coordinated with the software vendor to have a copy of the EDR data extracted (**Figure 1**). The vendor,

on behalf of the practice, de-identified and transferred the study data to our research team. This workflow ensured that patient confidentiality is protected. In cases in which we may want to find out more information from a practice, protocols are in place that allow us to communicate using honest brokers.

Data from the practices are translated into the Web Ontology Language (OWL) [1] using terms from the Oral Health and Disease Ontology (OHD) [2], an ontology built to represent the diagnosis and treatment of oral conditions. **Figure 2** illustrates the workflow for the translation process. Instead of translating data all at once, we have established a translation pipeline in which we first extract data from a practice's EDR (using standard SQL) and save the data as text files. These text files undergo quality checks to ensure that the data have been extracted correctly and make sense. For example, we check that dental procedures on teeth have a tooth associated with the procedure. The text files are also loaded into a MySQL database. This allows us to more easily perform quality checks over data from multiple practices. The data are then translated into OWL and loaded into a GraphDB [3] triple store. Throughout the translation process, we regularly compared the data in the triple store to the extracted data.

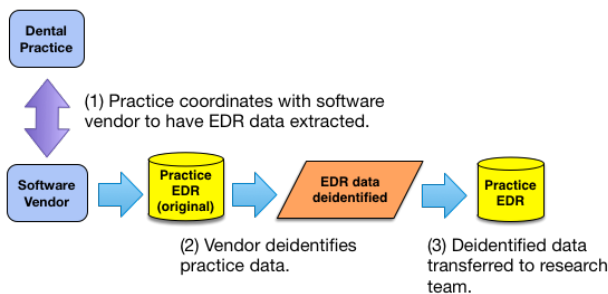


Figure 1: Extraction and transfer of practice data

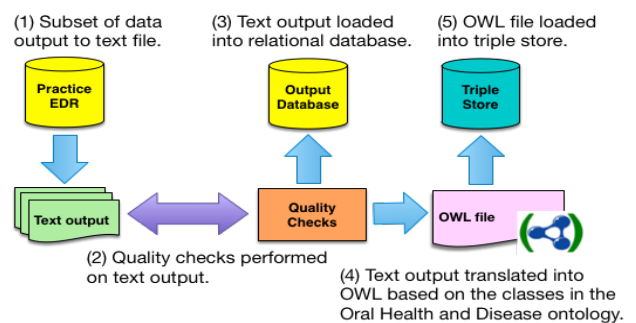


Figure 2: Workflow for translating data to OWL

¹ <http://www.NationalDentalPBRN.org>

* corresponding author

The triple store is configured to use GraphDB’s OWL2-RL automated reasoner. Although this has quite an impact on load time (approximately 20 hours to load data), we leverage the reasoning power to classify individuals as instances of defined classes of interest to dental researchers. For example, we represent a tooth restoration procedure as having a restored tooth as its output.

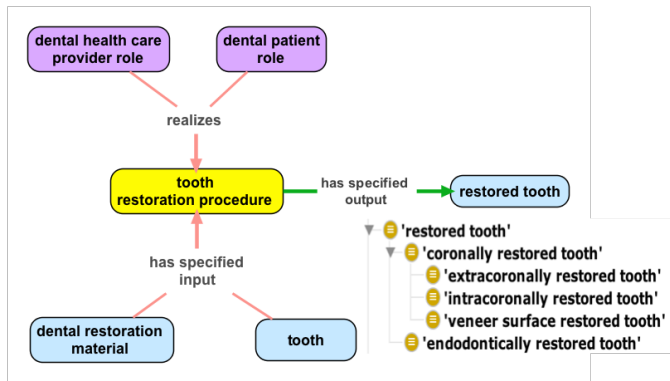


Figure 3: OHD representation for tooth restoration procedures

While it is possible to query for restored teeth using the pattern depicted in in **Figure 3**, we specify that the class ‘restored tooth’ is equivalent to:

Tooth **and** (‘has part’ **some** ‘dental restoration material’)

This allows us to more easily query the triple store for restored teeth and associated subtypes.

Presently, our triple store holds **1,160,388,319** triples. **Table 1** summarizes the number of unique patients, teeth, and dental procedures that we represent using the OHD.

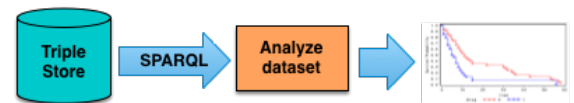
Table 1: Number of instances in dental triple store

Type	Number of Instances
patient	346,494
female patient	186,949
male patient	149,743
gender not known	9,802
tooth	1,488,174
restored tooth	1,320,294
PCR procedure	1,199,708
RCT procedure	75,108

Our comparative analysis of the triple store dataset and the dataset generated from MySQL database indicates a difference

in the number of unique patients, procedures and other data types. This difference occurred because the dataset from MySQL database did not include those records with a missing procedure for a specific tooth, procedure codes that involved multiple teeth, and with a missing gender.

Figure 4 illustrates our planned workflow to analyze outcomes for PCR and RCT procedures. Data are extracted using the SPARQL [4] query language, and saved to a text file. The text file is then analyzed using Statistical Analysis Software (SAS) to assess the longevity of PCRs (i.e., how long does a PCR last before another restoration is necessary that involves one or more of the same tooth surfaces) and tooth-specific tooth loss rates following an RCT on a specific tooth. To test the accuracy of our results, we will perform a comparable analysis using traditional relational database methods.



(1) SPARQL used to extract dataset from triple store. (2) Extracted dataset undergoes statistical analysis.

Figure 4: Workflow for analyzing data.

A significant contribution of this study is that it lays the groundwork for making quality improvement a part of dental practice. The methods developed for this study can be incorporated into developing tools that permit clinicians to analyze the data in their EDR for selected quality measures, implement appropriate interventions (if necessary), and repeat the analyses at a later date to determine the outcomes of the intervention.

ACKNOWLEDGMENT

This work is supported in part by NIDCR grant U19-DE-22516. Opinions and assertions contained herein are those of the authors and are not to be construed as necessarily representing the views of the respective organizations or the National Institutes of Health.

REFERENCES

[1] <http://www.w3.org/TR/owl2-overview>
 [2] <https://github.com/oral-health-and-disease-ontologies>
 [3] <http://graphdb.ontotext.com>
 [4] <http://www.w3.org/TR/rdf-sparql-query>