

6-11-2020

## Maintenance of ONC Terminology for i2b2 Metadata

James Campbell

Alfred J. Anzalone

Jay Pedersen

James C. McClay

Follow this and additional works at: [https://digitalcommons.unmc.edu/com\\_emerg\\_pres](https://digitalcommons.unmc.edu/com_emerg_pres)



Part of the **Bioinformatics Commons**

---

# Maintenance of ONC Terminology for i2b2 Metadata

James Campbell, MD<sup>1</sup>, Jerrod Anzalone, MS<sup>2</sup>, Jay Pedersen, MA<sup>3</sup>, James McClay, MD, MS<sup>4</sup>

1. Department of Internal Medicine, University of Nebraska Medical Center, Omaha NE 68198-3332
2. Department of Neurological Sciences, University of Nebraska Medical Center, Omaha NE 68198-8440
3. Department of Pathology/Microbiology, University of Nebraska Medical Center, Omaha NE 68198-5900
4. Department of Emergency Medicine, University of Nebraska Medical Center, Omaha NE 68198-1150

## Problem

ONC terminologies are constantly adding new content and deactivating existing codes. The University of Nebraska Medical Center (UNMC) deploys three primary code sets that require regular updating to support research: SNOMED CT, RXNORM / NDC, and LOINC. A problem across the i2b2 community is keeping these terminologies up-to-date and loading them into i2b2 for timely analysis of EHR data. We have developed tool kits for rapid deployment of SNOMED CT metadata and will be extending the work to RXNORM/NDC and LOINC.

## Objectives

Our primary objective is to keep our i2b2 ontology build current with standards development organization terminology releases. SNOMED CT is an ontological model that relies on polyhierarchies, which are challenging to deploy in a clinical research data warehouse reliant upon relational database structures. By building an architecture using queryable transitive closure tables that can be loaded directly from SDO terminology releases, we create an i2b2 metadata table capable of supporting advanced querying of patient data.

## Methodology

We developed an approach to rapidly respond to the problem using transitive closure tables as a component of i2b2 metadata. Transitive closure tables include every concept in a hierarchy in a relationship with all of its hierarchical descendants. Transitive closure tables support i2b2 folder searching and also the build of all other CRC tables. We created a SQL procedure to produce those tables quickly as needed whenever we receive a new release of SNOMED CT. We are now working to implement this tooling and architecture for RXNORM and LOINC.

## SNOMED CT Transitive Closure Build

Transitive closure tables are built from the terminology by traversing and listing all the IS\_A relationships which relate concepts within the ontology. Each SUPERTYPE\_SCTID reference is accompanied by one row for each unique SUBTYPE\_SCTID which is a subtype of that concept. The GENERATION counts the minimum number of relationship traversals between the two concepts in the SNOMED CT ontology.

Domain Description: Contains SNOMED CT reference codes and terms with lists of dependent (subtype) concepts for use in query and creating SNOMED CT value sets.

## SNOMED CT Table Specification

Field Name	Data Type (SQL)	Value Set	Definition and Comments	Source
SUPERTYPE_SCTID	code (bigint)	SNOMEDCT	The SNOMED CT concept identifier of the node which is the supertype (parent) in this relationship between SNOMED CT concepts. Format 19N.	IHTSDO
SUPERTYPE_CNCP	char(40)	SNOMEDCT	The context free reference code for the SNOMED CT concept that is the supertype concept in this relationship, of the format 'SNOMEDCT:NNNNNN'. This is the reference as it is stored in the i2b2 table OBSERVATION_FACT.	IHTSDO
SUPERTYPE_TERM	char(2048)	n/a	The Fully specified name of the SNOMED CT parent concept	IHTSDO
SUBTYPE_SCTID	code (bigint)	SNOMEDCT	The SNOMED CT concept identifier of the node which is the subtype (child, grandchild, great-grandchild...) in this relationship between SNOMED CT concepts. Format 19 N.	IHTSDO
CONCEPT_CD	char(40)	SNOMED CT	The context free reference code for the SNOMED CT concept that is the subtype concept in this relationship, of the format 'SNOMEDCT:NNNNNN'. This is the reference as it is stored in the i2b2 table OBSERVATION_FACT and used for searching by i2b2.	IHTSDO
SUBTYPE_TERM	char(2048)	n/a	The Fully specified name of the SNOMED CT child concept.	IHTSDO
GENERATION	integer	n/a	How many IS_A relationships separate the two concepts within SNOMED CT. (The shortest path length from the SUPERTYPE concept to the SUBTYPE concept in the SNOMED CT hierarchy)	IHTSDO

## SNOMEDCT\_TRC to i2b2 Metadata

Transitive closure tables contain all necessary data to construct the SNOMED CT hierarchy. A second SQL procedure processes the table into an i2b2 metadata table strictly adherent to i2b2 formatting, but employing folder searching using equivalence searching into the transitive closure table.

## Pneumococcal Pneumonia Exemplar

SUPERTYPE_SCTID	SUPERTYPE_CNCP	SUPERTYPE_TERM	SUBTYPE_SCTID	SUBTYPE_CNCP	SUBTYPE_TERM	GENERATION
233607000	0	Pneumococcal pneumonia (disorder)	233607000	SNOMEDCT:233607000	Pneumococcal pneumonia (disorder)	0
233607000	0	Pneumococcal pneumonia (disorder)	10625711000119	1009105	Bronchopneumonia caused by Streptococcus pneumoniae (disorder)	1
233607000	0	Pneumococcal pneumonia (disorder)	420787001	SNOMEDCT:420787001	Pneumococcal pneumonia associated with acquired immunodeficiency syndrome (disorder)	1
233607000	0	Pneumococcal pneumonia (disorder)	266350000	SNOMEDCT:266350000	Pneumococcal lobar pneumonia (disorder)	1

## Conclusions

Developing methods for implementing rapidly evolving medical ontologies as i2b2 metadata requires tool kits and procedures supporting efficient processing of complex data into useful i2b2 artifacts. In light of the recent invasion of COVID-19 on the world health scene, the ability to respond and deploy changed terminology content is imperative to support the mission of i2b2 as instrument of interoperable research in a changing world.

## Acknowledgements

The project is supported by the National Institute of General Medical Sciences, 1 U54 GM115458, which funds the Great Plains IDeA-CTR Network and PCORnet 2.0 Clinical Research Network (CRN) Infrastructure; PCRF Contract #1247 (PI: Waitman). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH or PCORI.