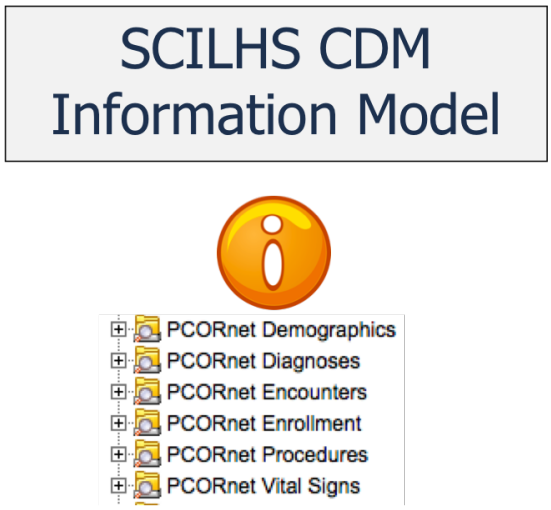
SCILHS PCORnet Common Data Model Ontology Home



SCILHS i2b2 PCORnet Common Data Model Ontology

The Scalable Collaborative Infrastructure for a Learning Health System (SCILHS, pronounced "skills") is a network of 10 health centers across the United States that will cover over 8 million patients. SCILHS is a Clinical Data Research Network (CDRN) in the Patient-Centered Outcomes Research Institute's PCORnet.

SCILHS (with much help from others) has developed an i2b2 information model that represents the PCORnet Common Data Model (CDM). This information model consists of an i2b2 ontology/terminology and a process for mapping local data elements to the ontology without changing the underlying imported data. This approach highlights i2b2's ability to separate data model from both information model and the underlying data format.

By conforming to this ontology, SCILHS sites can query across the network with an interoperable data model using SHRINE, and we provide a script to programmatically generate data marts in the PCORnet data model format. This will enable detailed analysis through the PCORnet Distributed Research Network (DRN). Likewise, if PCORNet Queries were rewritten to run against the i2b2 schema, they could run reliably at every site using the PCORNet CDM i2b2 ontology.

The ontology is live at https://www.i2b2.org/webclient/. Change the username to pcori, leave the password as demouser, and click 'Login'. The existing demo data has been mapped to the PCORI ontology, so most queries involving demographics, diagnoses, procedures, and enrollment will work. This did not require any changes to the demo data, only to the information model (ontology). Adding new demo data for the remaining sections of the ontology is on our roadmap.

This release, v2.1, represents several rounds of improvement over our initial release, and it includes much of CDM v1-v3. It has been vetted by ten of our sites - the mapping has been completed and it is successfully being used for inter- and intra- network querying. This release corresponds to the PCORnet CDM v3.0, though it is presently missing the death and PRO tables.

- Jeffrey Klann, PhD

About the ontology

The core ontology was originally generated from Dan Connolly's code to generate ontologies from the PCORnet CDM v1 spec, and edited significantly based on site feedback. Note that some clarifications on the meaning of various elements appear in the tooltips. Many terminology trees were added, as described below. The ontology has been modified to do as much as possible out-of-the-box. By default, the ontology is mapped to relevant sections in the i2b2 metadata, and some elements are computed automatically (such as number of patients enrolled). Sites can adapt the ontology to their local data through a mapping process described in the documentation, and no changes to actual data are necessary.

What does the ontology cover?

- Demographics, v2.1.1: Similar to the i2b2demodata, with a more granular age tree to support pediatric age queries.
- Diagnoses, v2.1.2: Includes ICD-10 2015AA and ICD-9 2014AA, generated by Lori Phillips from ontologies on BioPortal.
- Procedures, v2.1.2: Includes ICD-9 2015AA, ICD-10-PCS 2015AA, Partners' RPDR CMS-DRG tree, Beth Israel's MS-DRG tree (version unknown), and Nathan Wilson's HCPCS tree
- Encounters, v2.1.2: Includes PCORnet-specific additions, and a tree of 3-digit zip codes derived from the i2b2 demo data
- Enrollment, v2.1: To allow sites to specify which patients have "complete longitudinal data" for selected date ranges.
- Vitals, v2.1: A simple vitals implementation including height, weight, blood pressure, and smoking status.
- Labs, v2.1: Selected high-priority lab categories (70) and their LOINC codes, based on LOINC v236
- Medications, v2.2: RxNorm codes organized by NDF-RT (VA Drug Class), based on RxNorm's November 2015 release

What's not in the ontology?

- Death, death cause, and patient-reported outcomes will be included in a future release.
- SNOMED Clinical Findings v1.2 are available on BioPortal, but they are not included in this version of the ontology. We suspect SNOMED is not presently used, and the language from the coordinating center indicates CPT and HCPCS are preferred over LOINC for procedures at present: "Only billed procedures should be included in the PROCEDURE table. The ORDER concept may be incorporated into future phases of the CDM." We plan to include this eventually for completeness.
- CPT is not included because it is not freely available. We have an integrated CPT tree that has retired codes reconciled across sites. If you have a CPT license, you can contact us to get it.

Download the ontology

To download, please view the README at our new GitHub-hosted site!