PEDSnet Data Science Training

Overview

- Web sessions

Data Model & Vocabularies	April 2019
Web Applications	May 2019
Data Request Fullfillment	June 2019
PEDSnet Standard R Framework	July 2019

- Hands-on workshop – August 01-02, 2019

- Use of network data and scientific projects

PEDSnet
A Pediatric Learning Health System

Web Sessions

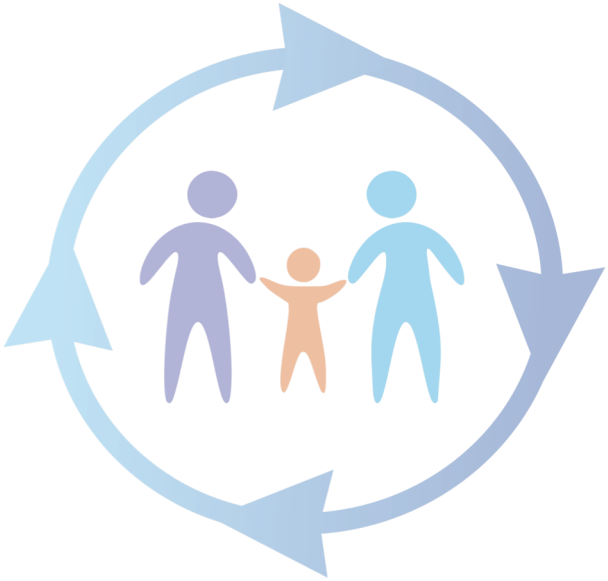
- Cover content areas central to using data for studies
  - Technical requirements
  - Process integration with DCC and PMO
  - Network technical conventions

- Sessions intended to be interactive – ask questions and make suggestions

- Pre-reading prior to each session
- Exercises distributed in the week after each session; 10 day turnaround

PEDSnet
A Pediatric Learning Health System

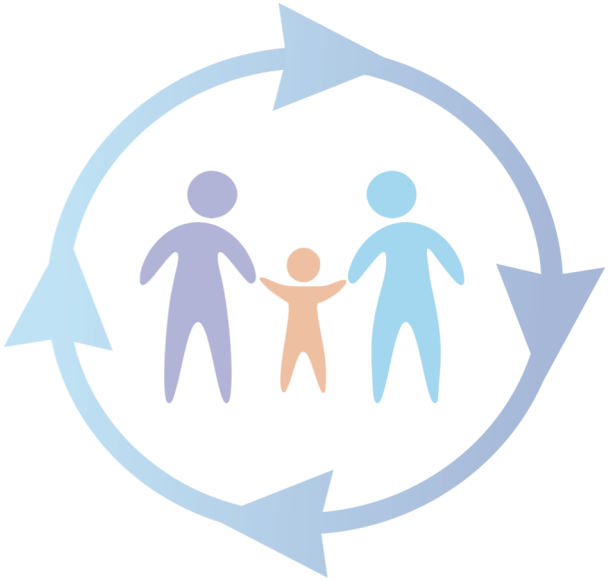
Session 1: Data Model



Agenda

- Introduction
- Vocabulary Tutorial
- Data Model Overview
- PEDSnet Domain Overviews
  - Conditions
  - Drugs
  - Procedures
  - Measurements
- PEDSnet Adaptations to Data Model
- Remaining Tables

Introduction

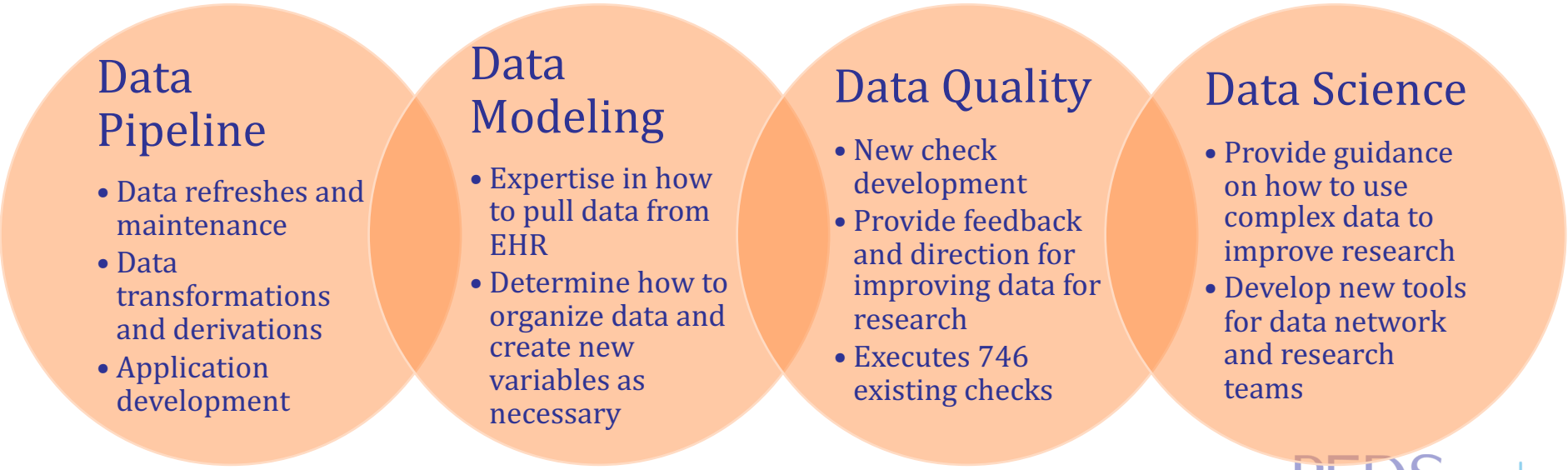


PEDSnet DCC

TEAM:

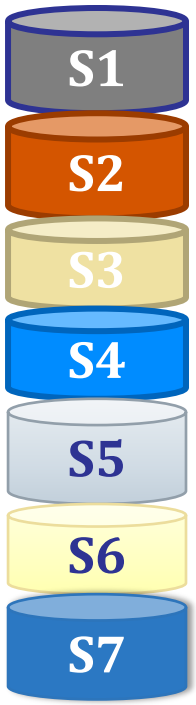
- Based at Children's Hospital of Philadelphia
- Composed of developers, programmers, coordinators, and data scientists

ROLE:



PEDSnet
A Pediatric Learning Health System

Data Flow



PEDSnet sites *(at various locations)*

Extract-transform-load (ETL) data quarterly

- Regular online meetings
- GitHub version control

- ETL specifications
- Data Quality

PEDSnet Common Data Model

PCORnet Common Data Model

Data Quality and Characterization

Query Fulfillment

Data Science

Data Modeling

Study Participation

PEDSnet
A Pediatric Learning Health System

Why Standardize Data?

- Different starting points
  - EHR v EDW
  - Existing v New
  - i2b2 v OMOP
- Different environments
  - Workflows
  - Data capture
  - Data semantics

PEDSnet
A Pediatric Learning Health System

The Goal:
Consistent CDM Data Semantics

- PEDSnet ETL Conventions
  - Github
  - Data Model WG
  - ETL Conventions document
- Comparative Data Characterization
  - Internal
  - External

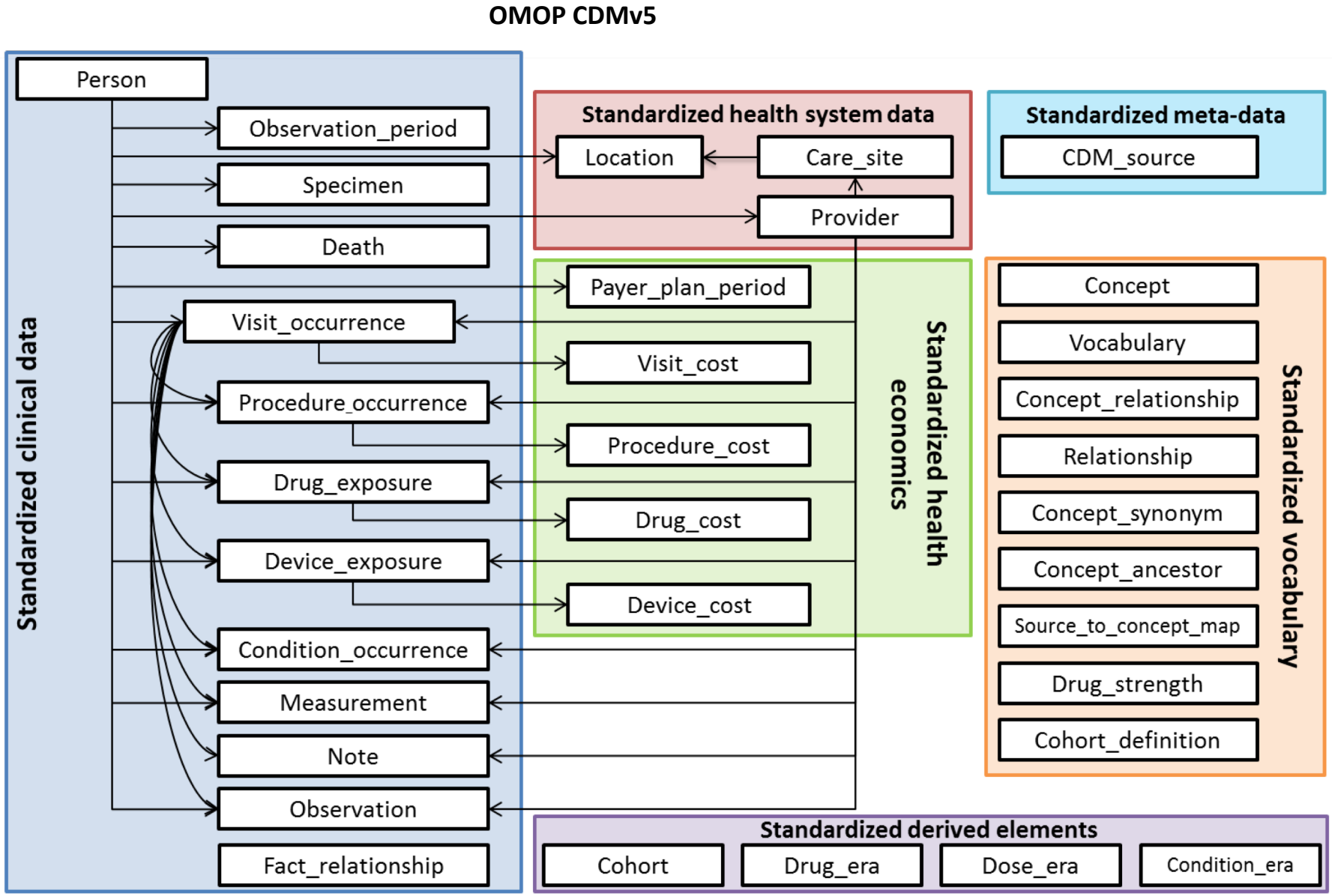
PEDSnet
A Pediatric Learning Health System

OMOP Common Data Model

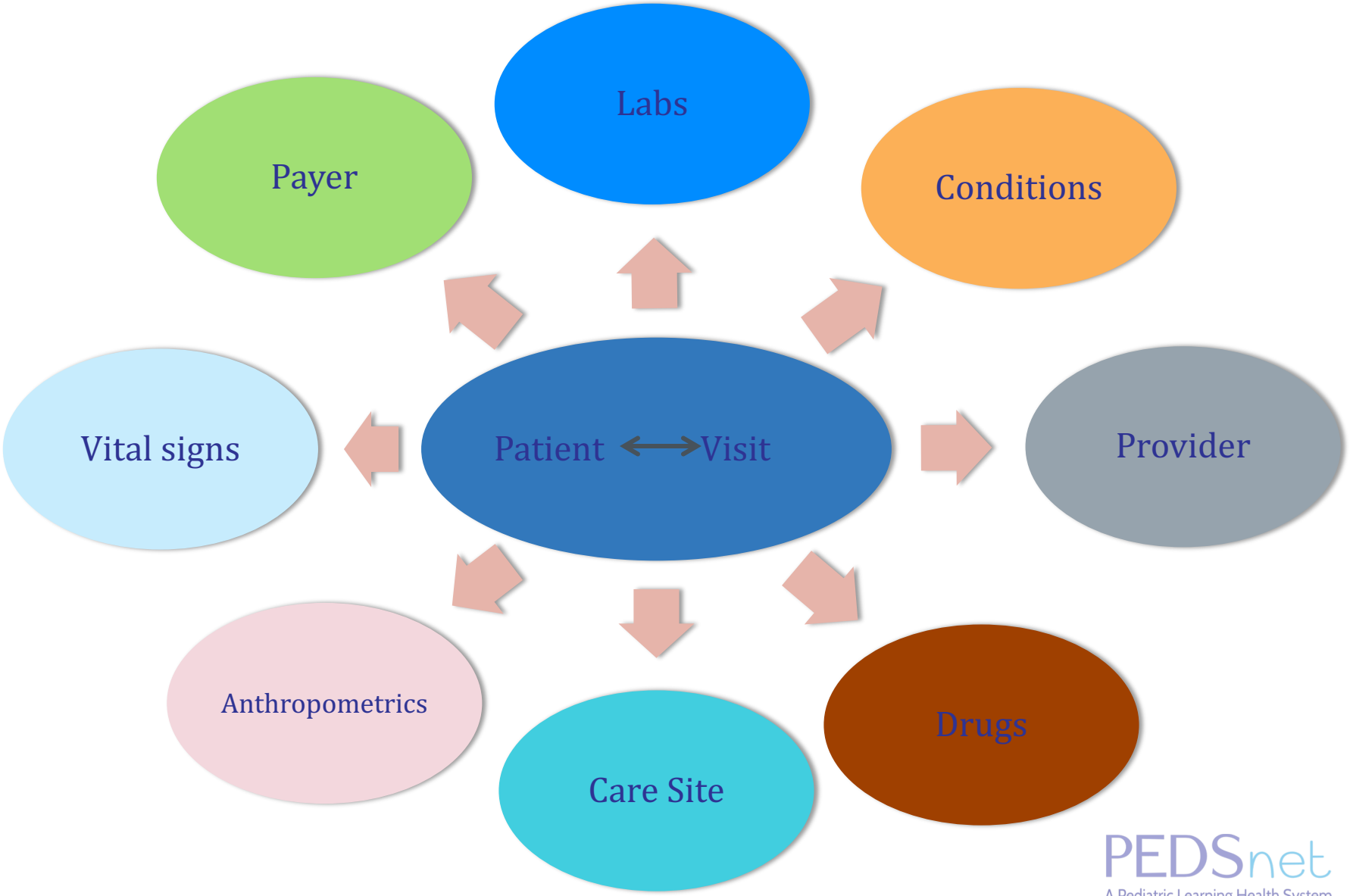
- **What?**
  - A standardized way to represent data structure (CDM) and content (vocabulary)
  - One model to accommodate data coming from disparate data sources
    - administrative claims, electronic health records
    - EHRs from both inpatient and outpatient settings
    - registries and longitudinal surveys
    - data sources both within and outside of US
- **Why?**
  - Enable standardization of structure and content to support a systematic and reproducible process to efficiently generate evidence
  - Support collaborative research both within and outside of US

PEDSnet
A Pediatric Learning Health System

The journey of the OMOP Common data model



OMOP CDM STRUCTURE



PEDSnet
A Pediatric Learning Health System

PEDSnet Data Model

- Adapted from OMOP (list below not comprehensive)
- Additional Domains:
  - adt_occurrence (ICU transfer)
  - measurement_organism (lab culture organism)
  - immunization
- Additional Fields:
  - gestational age
  - care site specialty
- Additional Valuesets:
  - head circumference
  - visit types

PEDSnet
A Pediatric Learning Health System

CDM Vocabulary

Introduction

- A medical code or idea is expressed via a unique identifier called a ***concept_id***

- concept_id's are expressed in every table --- conditions, drugs, procedures, etc

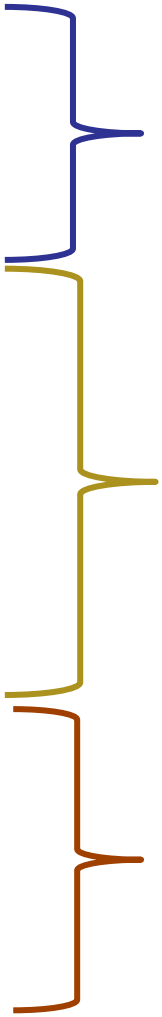
- Every individual demographic, drug, condition, procedure, etc. described by SNOMED, gender, ICD10, race, RxNorm, HCPT4, etc. is specified by its own concept_id in our database

Overview

TABLE	DESCRIPTION
concept	• dictionary of terms • provides definition of concept (e.g., concept_id 45932805 = `Croup`) • includes metadata about concept
concept_relationship	• describes how concepts may be related to each other • can be used to map similar concepts from different vocabularies
concept_ancestor	• shows hierarchy of standard ontologies • provides ability to quickly scan vocabularies/ontologies and pick terms of interest

concept Table

Field name	Example
concept_id (PK)	143193
concept_name	Asthma
concept_code	J45
vocabulary_id	ICD10
concept_class_id	ICD 10 Hierarchy
domain_id	Condition
standard_concept	[null]
valid_start_date	1990-05-01
valid_end_date	2099-12-31
invalid_reason	[null]



concept identifier and definition; concept_id is identifier in CDM; concept code is identifier in its own vocabulary

metadata about concept and how it is classified; a null standard concept means it is not used in the CDM tables

identifies whether concept is still in use

PEDSnet
A Pediatric Learning Health System

concept_relationship

concept_relationship table

concept_id_1	concept_id_2	relationship_id
45562457	45768910	Maps to

- *concept_id_1* and *concept_id_2* are related to each other
- *relationship_id* describes relationship of two concepts

concept table

concept_id	concept_code	concept_name	domain_id	vocabulary_id
45562457	J45.909	Unspecified asthma, uncomplicated	Condition	ICD10CM
45768910	707444001	Uncomplicated asthma	Condition	SNOMED

PEDSnet
A Pediatric Learning Health System

concept_ancestor

- Purpose: states the hierarchical relationships between concepts
- Indexed by a composite primary key: ancestor_concept_id and descendant_concept_id
  - The ancestor is the higher-level concept and the descendant is the lower-level (more specific) concept
  - E.g. We can look for descendants of “Uncomplicated asthma” (concept_code = 45768910) to find the concept_ids of lower-level Using the concept_id for asthma we found in concept_relationship as an ancestor_concept_id

concept_ancestor

concept_ancestor table

ancestor_concept_id	descendant_concept_id	max_levels_of_separation	min_levels_of_separation
45768910	42539549	1	1

concept table

concept_id	concept_code	concept_name	domain_id	vocabulary_id
45768910	707444001	Uncomplicated asthma	Condition	SNOMED
42539549	735588005	Uncomplicated allergic asthma	Condition	SNOMED

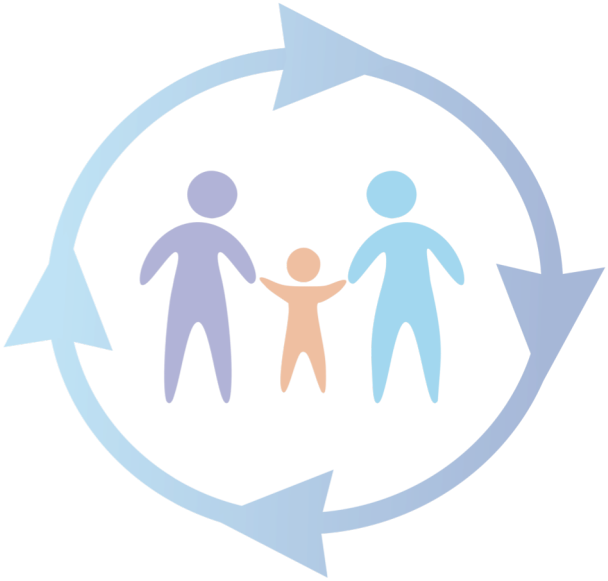
- *both of these codes be descendants the more general `asthma` code*

PEDSnet
A Pediatric Learning Health System

Standard CDM Vocabularies

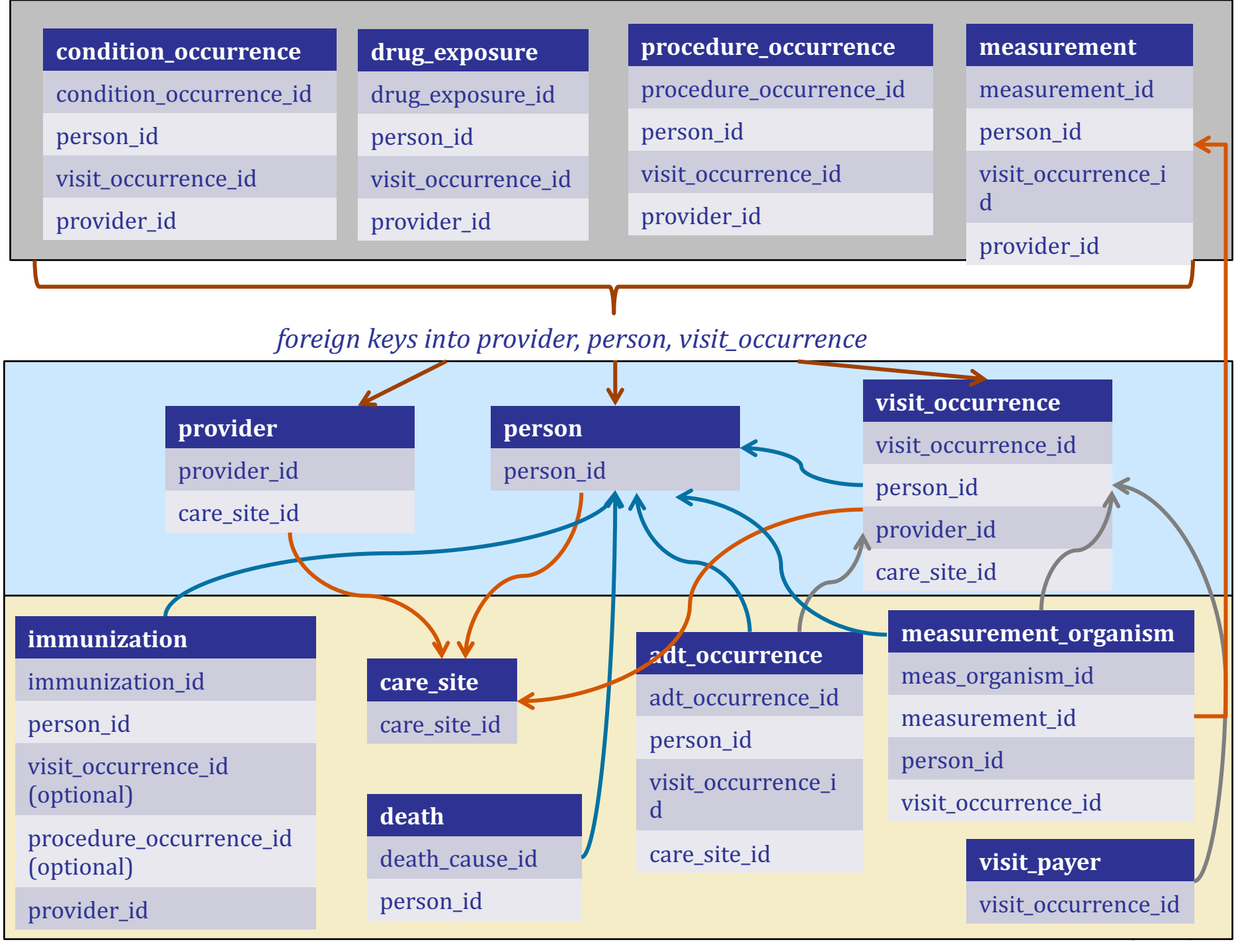
Table	Standardized Vocabularies	Field Name in CDM table
condition_occurrence	• SNOMED	condition_concept_id
drug_exposure	• RxNorm	drug_concept_id
procedure_occurrence	• CPT • ICD • SNOMED • HCPCS	procedure_concept_id
measurement	• LOINC • SNOMED	measurement_concept_id
immunization	• CVX	immunization_concept_id

Data Model



Overview

- The data model is captured in **person, fact, observation, and vocabulary tables** which join to one another based on unique identifiers
- Unique identifiers **(*primary keys)*** for each type of table are:
  - Person uses the **person_id** to distinguish unique patients
  - Fact tables use ***_occurrence_id** to distinguish unique instances
    - i.e. condition_occurrence_id captures unique instance of diagnosis
  - Observation captures examination facts using the **observation_id**
  - Vocabulary tables **use concept_id** to distinguish facts
    - e.g., a type of medication
- Primary key in one table can be a ***foreign key*** into another.
  - e.g., person_id is a foreign key in almost all tables


condition_occurrence
condition_occurrence_id
person_id
visit_occurrence_id
provider_id
drug_exposure
drug_exposure_id
person_id
visit_occurrence_id
provider_id
procedure_occurrence
procedure_occurrence_id
person_id
visit_occurrence_id
provider_id
measurement
measurement_id
person_id
visit_occurrence_id
provider_id
foreign keys into provider, person, visit_occurrence
provider
provider_id
care_site_id
person
person_id
visit_occurrence
visit_occurrence_id
person_id
provider_id
care_site_id
immunization
immunization_id
person_id
visit_occurrence_id (optional)
procedure_occurrence_id (optional)
provider_id
care_site
care_site_id
death
death_cause_id
person_id
adt_occurrence
adt_occurrence_id
person_id
visit_occurrence_id
care_site_id
measurement_organism
meas_organism_id
measurement_id
person_id
visit_occurrence_id
visit_payer
visit_occurrence_id

Data Model: Standard Fields

FIELDS	DESCRIPTION	EXAMPLE
(table name)*_id	typically the primary key of a table	*measurement_id* indicates one instance of a measurement; primary key of measurement table
(table name)_concept_id	primary variable for domain of interest	*condition_concept_id* is the primary variable for the condition table (e.g., condition_concept_id for croup)
*_type_concept_id	metadata about context during which fact occurred	*drug_type_concept_id* value set includes prescriptions, dispensing, inpatient order, and inpatient administration
*_start_date	date during which a particular fact occurred	*visit_start_date* indicate the start of the visit instance
*_source_value *_source_concept_id	value from a site's source data	*procedure_source_value* is the direct value of the name and code of source code; *procedure_source_concept_id* is the standardized concept of a site's source

PEDSnet
A Pediatric Learning Health System

person Table

Field name
person_id (PK)
provider_id (FK)
location_id (FK)
care_site_id (FK)
gender_concept_id (and source_value / source_concept_id)
race_concept_id (and source_value / source_concept_id)
ethnicity_concept_id (and source_value / source_concept_id)
year_of_birth
month_of_birth
day_of_birth
birth_datetime

primary and foreign keys

- provider_id is primary care provider or provider with most visits
- care_site is where patient received majority of care

demographic information of patient

birth day/date information

PEDSnet
A Pediatric Learning Health System

visit_occurrence Table

Field name*
visit_occurrence_id (PK)
person_id (FK)
provider_id (FK)
care_site_id (FK)
visit_concept_id
visit_type_concept_id
preceding_visit_occurrence_id
admitted_from_concept_id
discharge_to_concept_id
visit_start_date
visit_end_date
visit_start_datetime
visit_end_datetime

primary and foreign keys
- provider_id is provider in charge of visit

metadata regarding type of patient visit (e.g., visit_concept_id clarifies inpatient, outpatient, ED, etc)

visit date/time information

*source values and source concept id's not shown here

PEDSnet
A Pediatric Learning Health System

Clinical Tables

- Clinical tables capture patient-level clinical information
- Each row in each of the fact tables is a unique observation of the concept for the patient
- Clinical tables in the CDM:
  - diagnoses (condition_occurrence)
  - procedures (procedure_occurrence)
  - drug administrations (drug_exposure)
  - death (death)
  - immunizations (immunization)
  - labs, vitals, anthroprometrics (measurement)
  - lab cultures (measurement_organism)
  - ICU transfer during inpatient stay (adt_occurrence)

Fact Tables Example: Conditions

- Table: condition_occurrence
- Unique identifier (primary key): condition_occurrence_id
  - Each condition_occurrence_id has its own row in the table and is a unique diagnosis instance for the patient
- Other information in table:
  - Origin of the diagnosis code (e.g. where in the EHR this information can be found)
  - Person, provider, site, date for which the unique diagnosis was recorded
  - To determine the type of condition (e.g. scoliosis), link the condition_concept_id back to the concept_id in the concept table
    - while the condition_occurrence_id is unique for each instance of the condition, the condition_concept_id is shared for all of the patients with the same condition

Fact Tables Example: Procedures

- Table: procedure_occurrence
- Unique identifier (primary key): procedure_occurrence_id
  - Each procedure_occurrence_id has its own row in the table and is a unique procedure instance for the patient
- Other information in the table:
  - Information includes the associated concept_id for the procedure, the date of the procedure, and the unique identifier for the provider, and the type of procedure administered
  - To determine the type of procedure (e.g. tonsillectomy), link the procedure_concept_id back to the concept_id in the concept table
    - While the procedure_occurrence_id is unique for each instance of the condition, the procedure_concept_id is shared for all of the patients who underwent the same procedure and identifies the type of procedure

Observation Table

- The observation table includes information from examinations not captured by other domains, and in particular unstructured data

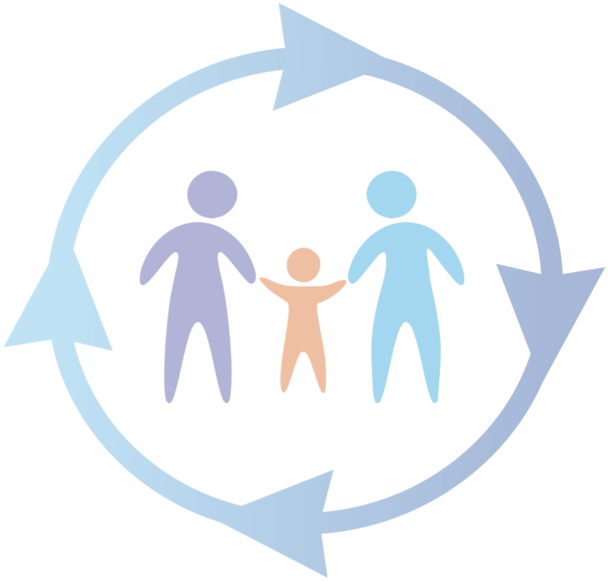
- It is indexed by the observation_id which is a unique identifier for an examination
  - I.e. An examination with an EHR could correspond to an entry in the observation table

- Information here includes the type of observation, source information, dates, and values from unstructured measurements

Vocabulary Tables

- Concept_ids are unique identifiers that describe a specific idea. Their meaning, domain (procedure, drug, etc.), and vocabulary (SNOMED, Gender, etc.) can be looked up in the concept table
  - I.e. concept_id ‘9201’ represents ‘Inpatient visit'
- Concept_ids can relate to one another (i.e. the same condition in both SNOMED and ICD10) and can be linked via the concept_relationship table
- They can have a hierarchy (i.e. ICD10 codes that get more and more specific) that is outline in the concept_ancestor table

Conditions



PEDSnet `condition_occurrence` Table

- All codes entered as conditions or able to be mapped back to ICD or SNOMED codes
- Standardized to **SNOMED-CT**
- Diagnoses in CDM should be **clinician-based** (not billing)
  - Conditions entered as Intelligent Medical Objects (IMO)
  - Mapped to SNOMED-CT
- Huge breadth of conditions captured

PEDSnet `condition_occurrence` Table

Top conditions in PEDSnet

condition_concept_id	concept_name	count(*)
443364	**Patient encounter status**	6031400
4132649	**Well child visit**	2880079
432526	Needs influenza immunization	1687441
257011	Acute upper respiratory infection	1344623
75860	Constipation	1282864
4008161	**Well child**	1188398

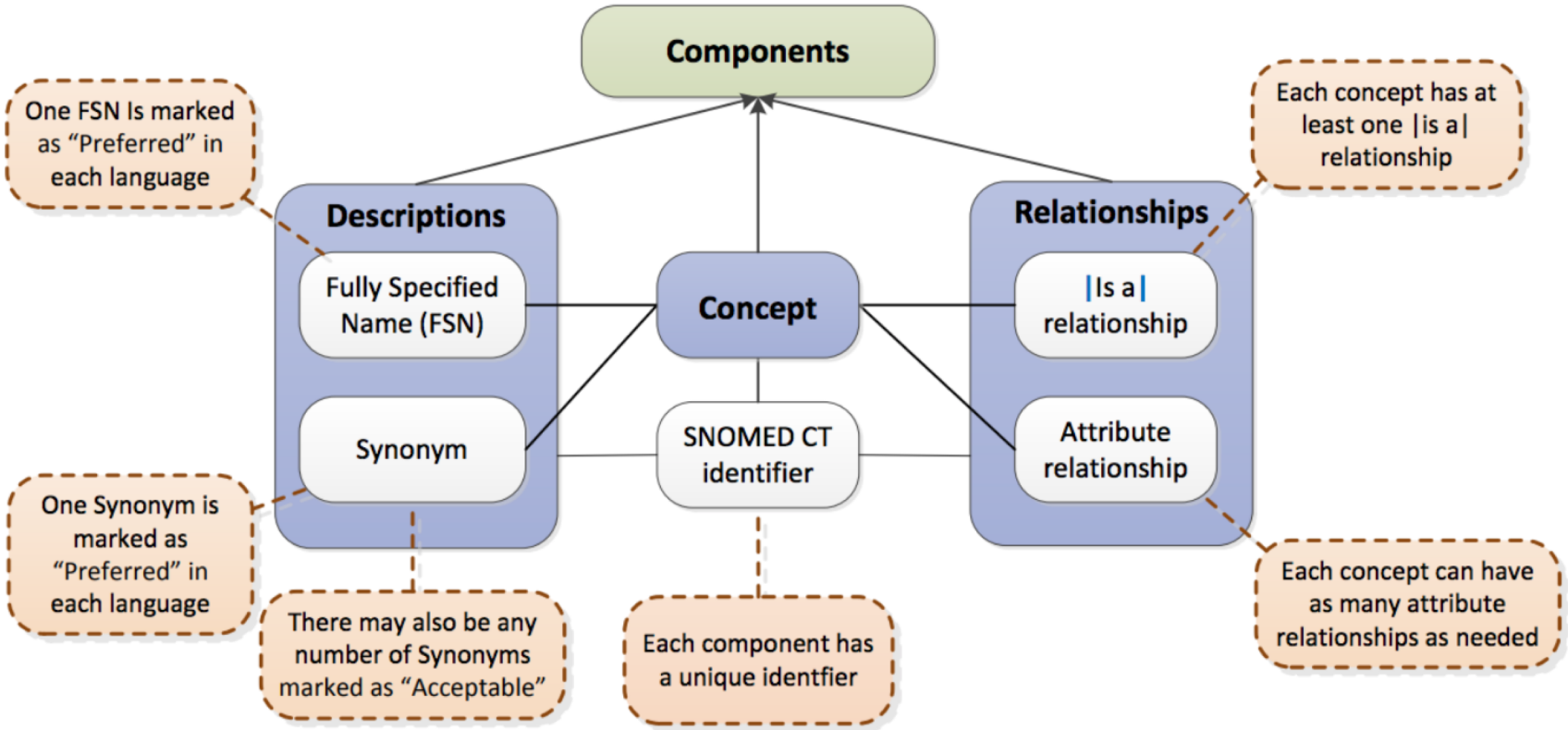
PEDSnet
A Pediatric Learning Health System

SNOMED-CT Overview

- The **Systematized Nomenclature of Medicine – Clinical Terms,** or SNOMED-CT, is an international clinical healthcare terminology designed for use in electronic health record (EHR) systems.
- 4 primary components:
  - Concepts – numerical codes to identify clinical terms, arranged in a hierarchy
  - Descriptions – the text used to describe a Concept
  - Relationships – connect Concepts with Related meanings.
  - Reference Sets – used to group Concepts or Descriptions into sets.
- The 6-18 digit concept SNOMED-CT codes provide a standard for the unambiguous identification and easy exchange of data between EHRs.
- SNOMED-CT has over **300,000 concept codes** covering not only diagnoses, but also classes of healthcare data like clinical findings, procedures, observations, body structures, organisms.

Source: http://www.hl7.org/documentcenter/public_temp_85220E30-1C23-BA17-0C411958FBD32C64/calendarofevents/himss/2012/SNOMED%20CT%20and%20HL7%20--%20Bringing%20Standards%20Together.pdf

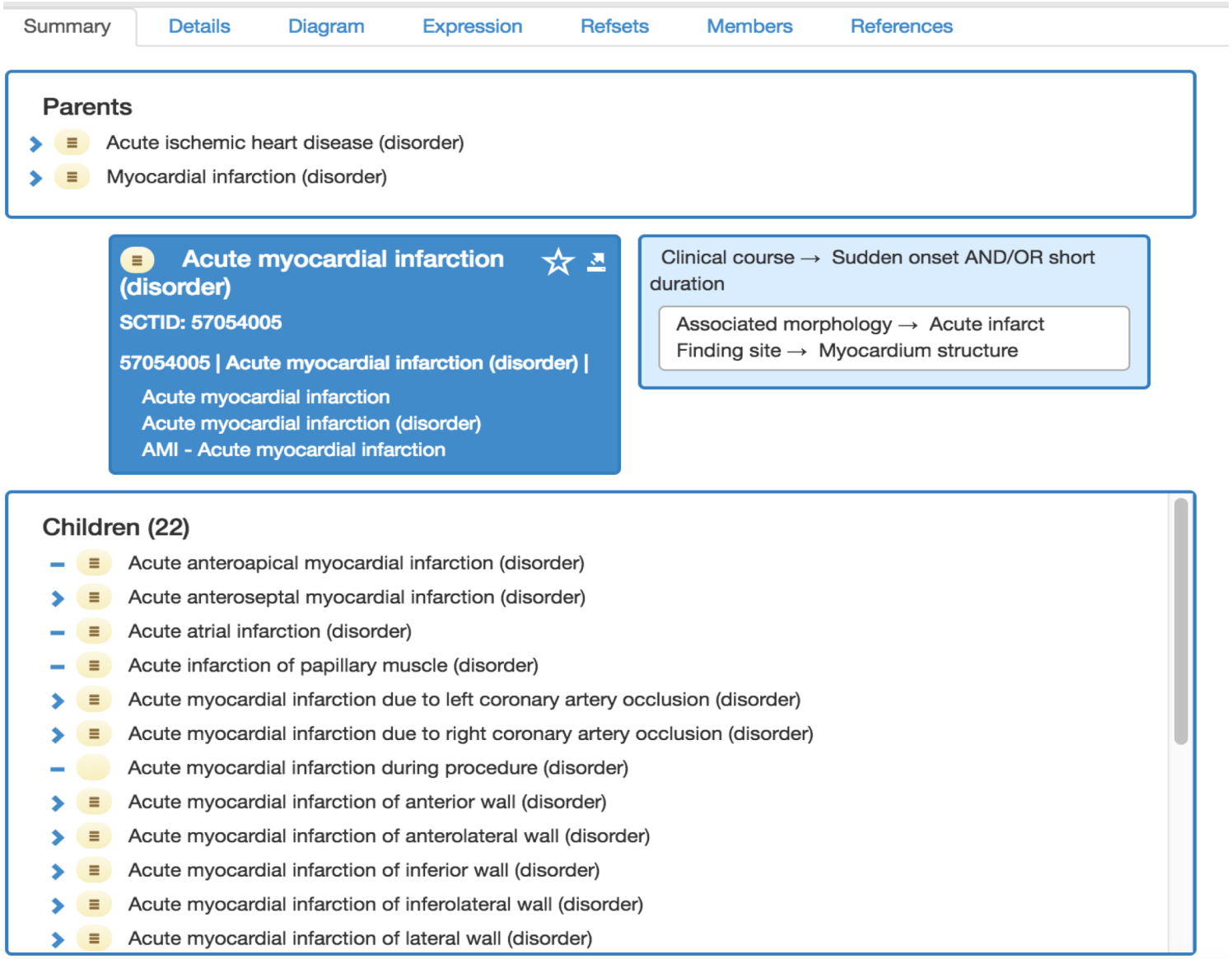
SNOMED-CT Overview



Source: www.snomed.org/starterguide.pdf

PEDSnet
A Pediatric Learning Health System

SNOMED-CT Browser



Source: http://browser.ihtsdotools.org/

PEDSnet
A Pediatric Learning Health System

SNOMED vs ICD-9 Comparison

SNOMED-CT	ICD-9

Structure:
• Reference terminology/ontology
  • relationships between concepts
  • collection of terms organized by meaning
	Structure:
	• Classification system
	  • Arrangement of terms into classes or groups based on common characteristics
	  • no relationships between groups

Intended Use: • Capture of clinical entities • Clinical encounters (problem lists)	Intended Use: • Billing • Origins to track morbidity and mortality reporting

SNOMED-CT vs ICD-9: Brain Tumor Example

ICD-9 CODE	COUNT IN DATABASE
Benign neoplasm of brain	449
Benign neoplasm of cerebral meninges	30
Benign neoplasm of cranial nerve	81
Benign neoplasm of nervous system	<11
Benign neoplasm of spinal cord	44
Benign neoplasm of spinal meninges	<11
Malignant neoplasm of brain	831
Malignant neoplasm of cerebral meninges	11
Malignant neoplasm of cerebrum	69
Malignant neoplasm of frontal lobe	30
Malignant neoplasm of occipital lobe	<11
Malignant neoplasm of parietal lobe	17
Malignant neoplasm of spinal meninges	<11
Malignant neoplasm of temporal lobe	25
Overlapping malignant neoplasm of brain and other parts of the central nervous system	77
Primary malignant neoplasm of brain stem	86
Primary malignant neoplasm of cerebellum	190
Primary malignant neoplasm of cerebral ventricle	34
Primary malignant neoplasm of cranial nerve	84
Primary malignant neoplasm of nervous system	14
Primary malignant neoplasm of spinal cord	39

SNOMED CODE	COUNT IN DATABASE
Anaplastic astrocytoma of brain	32
Anaplastic astrocytoma of spinal cord	<11
Astrocytoma of brain	144
Astrocytoma of retina	<11
Astrocytoma of spinal cord	<11
Ependymoma	114
Ependymoma of brain	16
Ependymoma of spinal cord	<11
Glioblastoma multiforme	42
Glioblastoma multiforme of brain	<11
Glioblastoma multiforme of spinal cord	<11
History of astrocytoma of brain	<11
Medulloblastoma	174
Medulloblastoma of cerebellum	65
Primitive neuroectodermal tumor	81
Supratentorial primitive neuroectodermal tumor	<11

PEDSnet
A Pediatric Learning Health System

PEDSnet `condition_occurrence` Table

Field name
condition_occurrence_id (PK)
person_id (FK)
visit_occurrence_id (FK)
provider_id (FK)
condition_concept_id (and source value / source_concept_id)
condition_type_concept_id
condition_start_date (and datetime)
condition_end_date (and datetime)
poa_concept_id
stop_reason
condition_status_concept_id (source value)

primary and foreign keys

specifies diagnosis and type (inpatient, outpatient, ED header position, or problem list)

dates of condition start/end for instance of condition

optional condition metadata fields

PEDSnet
A Pediatric Learning Health System

Codeset Creation

- Investigator usually starts with ICD codeset
- Can begin with ICD mappings and then search SNOMED separately

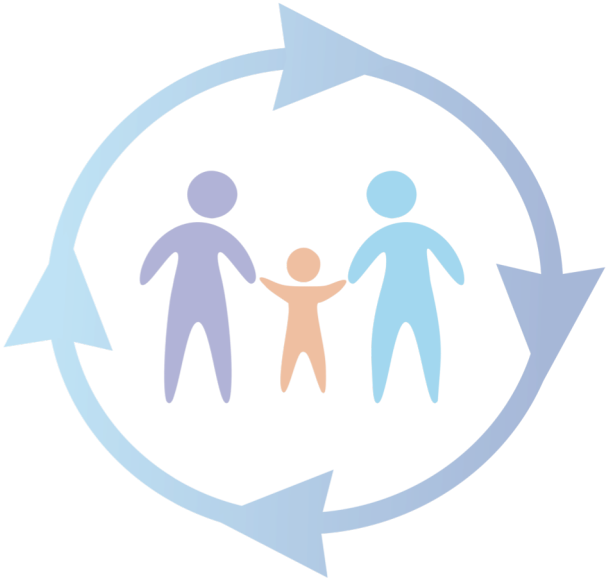
ICD codes for muscular dystrophy and crosswalked codes

icd_code	icd_name	snomed_code	snomed_name
359	Muscular dystrophies and other myopathies	129565002	Disorder of muscle
359.0	Congenital hereditary muscular dystrophy	111501005	Congenital hereditary muscular dystrophy
359.1	Hereditary progressive muscular dystrophy	193225000	Hereditary progressive muscular dystrophy
359.2	Myotonic disorders	193237003	Myotonic disorder
359.21	Myotonic muscular dystrophy	77956009	Steinert myotonic dystrophy syndrome
359.22	Myotonia congenita	57938005	Congenital myotonia, autosomal dominant form
359.23	Myotonic chondrodystrophy	29145002	Schwartz-Jampel syndrome
359.24	Drug- induced myotonia	193237003	Myotonic disorder
359.29	Other specified myotonic disorder	193237003	Myotonic disorder
359.3	Periodic paralysis	198030008	Periodic paralysis
359.4	Toxic myopathy	66952001	Toxic myopathy
359.5	Myopathy in endocrine diseases classified elsewhere	57958006	Endocrine myopathy
359.6	Symptomatic inflammatory myopathy in diseases classified elsewhere	193246009	Symptomatic inflammatory myopathy associated with another disorder
359.7	Inflammatory and immune myopathies, NEC	128496001	Inflammatory disorder of muscle
359.71	Inclusion body myositis	72315009	Inclusion body myositis
359.79	Other inflammatory and immune myopathies, NEC	129565002	Disorder of muscle
359.8	Other myopathies	443819006	Critical illness myopathy
359.81	Critical illness myopathy	443819006	Critical illness myopathy
359.89	Other myopathies	129565002	Disorder of muscle
359.9	Myopathy, unspecified	75047002	Disorder of skeletal muscle

- 94 SNOMED concepts for muscular dystrophy (using descendants from muscular dystrophy code)
- **Selected** codes below

snomed_code	omop_concept_id	snomed_name
719985001	36714666	Autosomal dominant limb girdle muscular dystrophy type 1A
725907002	37111246	Autosomal recessive muscular dystrophy due to Torsin-1A-interacting protein 1 deficiency
387732009	4308668	Becker muscular dystrophy
240078009	4344283	Benign congenital muscular dystrophy with finger flexion contractures
240076008	4347193	Benign scapuloperoneal muscular dystrophy
240072005	4345564	Benign scapuloperoneal muscular dystrophy with cardiomyopathy
111501005	78149	Congenital hereditary muscular dystrophy
240059009	4344280	Congenital muscular dystrophy
58795000	4243215	Distal muscular dystrophy
76670001	4296473	Duchenne muscular dystrophy
702343002	45765411	Early onset myopathy with fatal cardiomyopathy
240063002	4347186	Eichsfeld type congenital muscular dystrophy
240053005	4347184	Hereditary myopathy limited to females
240069003	4347191	Late onset proximal muscular dystrophy with dysarthria
240050008	4344278	Manifesting female carrier of X-linked muscular dystrophy
129596006	4030587	Menopausal muscular dystrophy syndrome
111503008	4006462	Merosin deficient congenital muscular dystrophy
111505001	4008571	Muscle-eye-brain disease, congenital muscular dystrophy
73297009	4247802	Muscular dystrophy
77097004	4297463	Oculopharyngeal muscular dystrophy
193227008	4102501	Pelvic muscular dystrophy
240058001	4344606	Reunion-Indiana Amish type muscular dystrophy
240074006	4347192	Scapulohumeral muscular dystrophy
277373000	4172671	Severe childhood autosomal recessive muscular dystrophy
77956009	79804	Steinert myotonic dystrophy syndrome
698846009	44782849	Tibial muscular dystrophy
240062007	4344608	Ullrich congenital muscular dystrophy
111504002	4009773	Walker-Warburg congenital muscular dystrophy
240060004	4344607	Western type of congenital muscular dystrophy

Drugs



PEDSnet `drug_exposure` Table

- Standardized to **RxNorm**
- Types of drug exposures captured:
  - prescriptions (outpatient and inpatient)
  - dispensing (when available)
  - inpatient ordering (pcornet requirement)
  - inpatient administration
- fields with variable data quality because of source data ambiguity
  - e.g., frequency, dose, quantity

RxNorm Overview

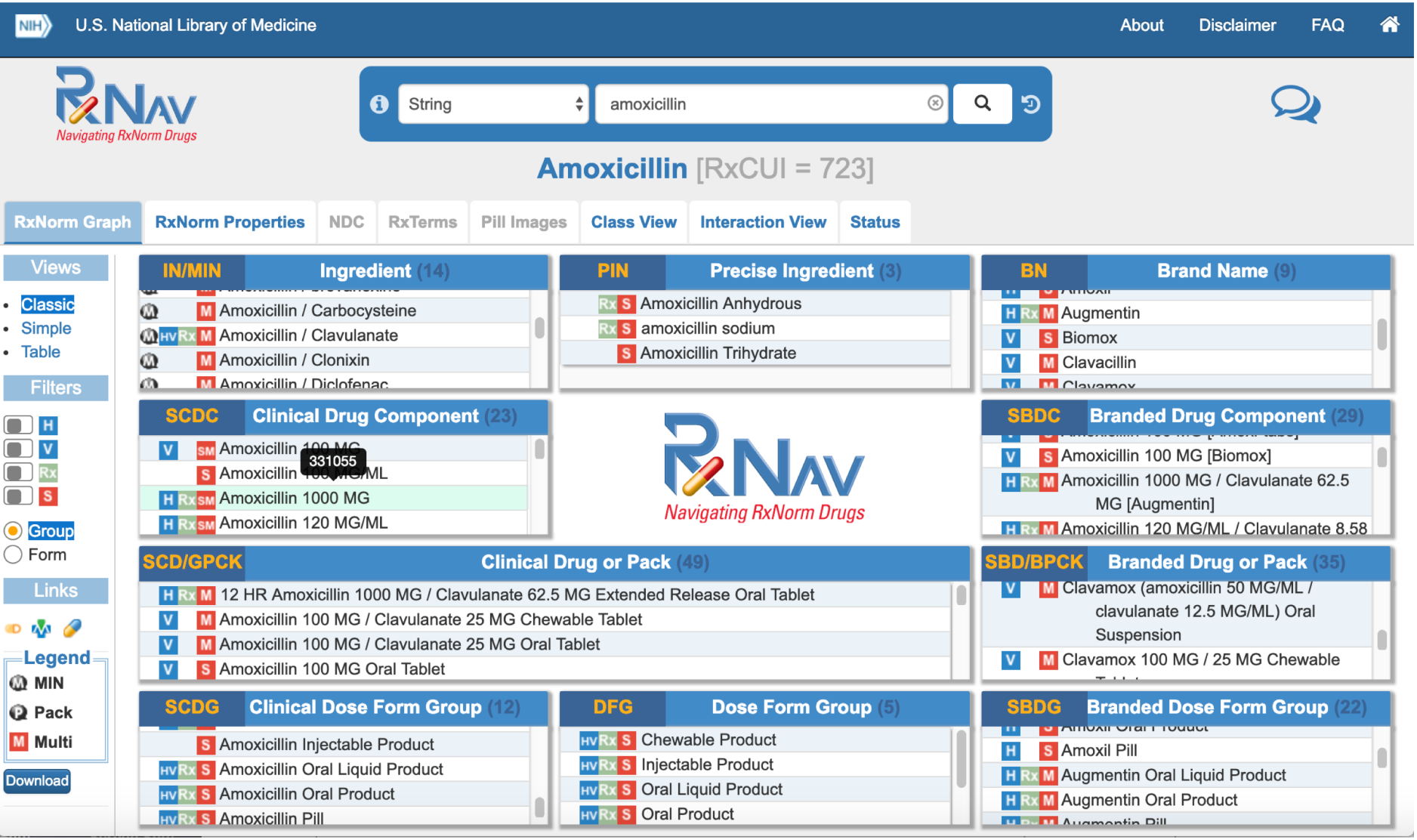
- RxNorm links different levels of specificity using relationships, e.g.:
  - 'Naproxen 250 MG Oral Tablet' has_dose_form 'Oral Tablet'
  - 'Naproxen' ingredient_of 'Naproxen 250 MG'
  - 'Naproxen 250 MG Oral Tablet' isa 'Naproxen oral Tablet'
  - 'Naproxen Pills' has_ingredient 'Naproxen'

Source: https://www.nlm.nih.gov/research/umls/rxnorm/overview.html

RxNorm Hierarchy

Term Type	Description	Example
Ingredient (IN)	Ingredient	Amoxicillin
Precise Ingredient (PIN)	Most PIN are salt or isomer forms	Amoxicillin Trihhydrate
Multiple Ingredient (MIN)	Two or more ingredients appearing together	Amoxicillin / Clavulanate
Semantic Clinical Drug Component (SCDC)	Ingredient + Strength	Amoxicillin 100 mg
Semantic Clinical Drug Form (SCDF)	Ingredient + Dose Form	Amoxiciillin Oral Solution
Semantic Clinical Dose Group (SCDG)	Ingredient + Dose Form Group	Amoxicillin Oral Product
Semantic Clinical Drug (SCD)	Ingredient + Strength + Dose Form	Amoxicillin 100 mg Oral Tablet
Brand Name (BN)	Brand Name	Augmentin
Semantic Branded Drug Component (SBDC)	Ingredient + Strength + Brand Name	Amoxicillin 1000 mg / clavulanate 62.5 mg [Augmentin]
Semantic Branded Dose Form Group (SBDG)	Brand Name + Dose Form Group	Augmentin Oral Product
Semantic Branded Drug (SBD)	Ingredient + Strength + Dose Form + Brand Name	Augmentin 250-mg Oral Tablet

RxNav

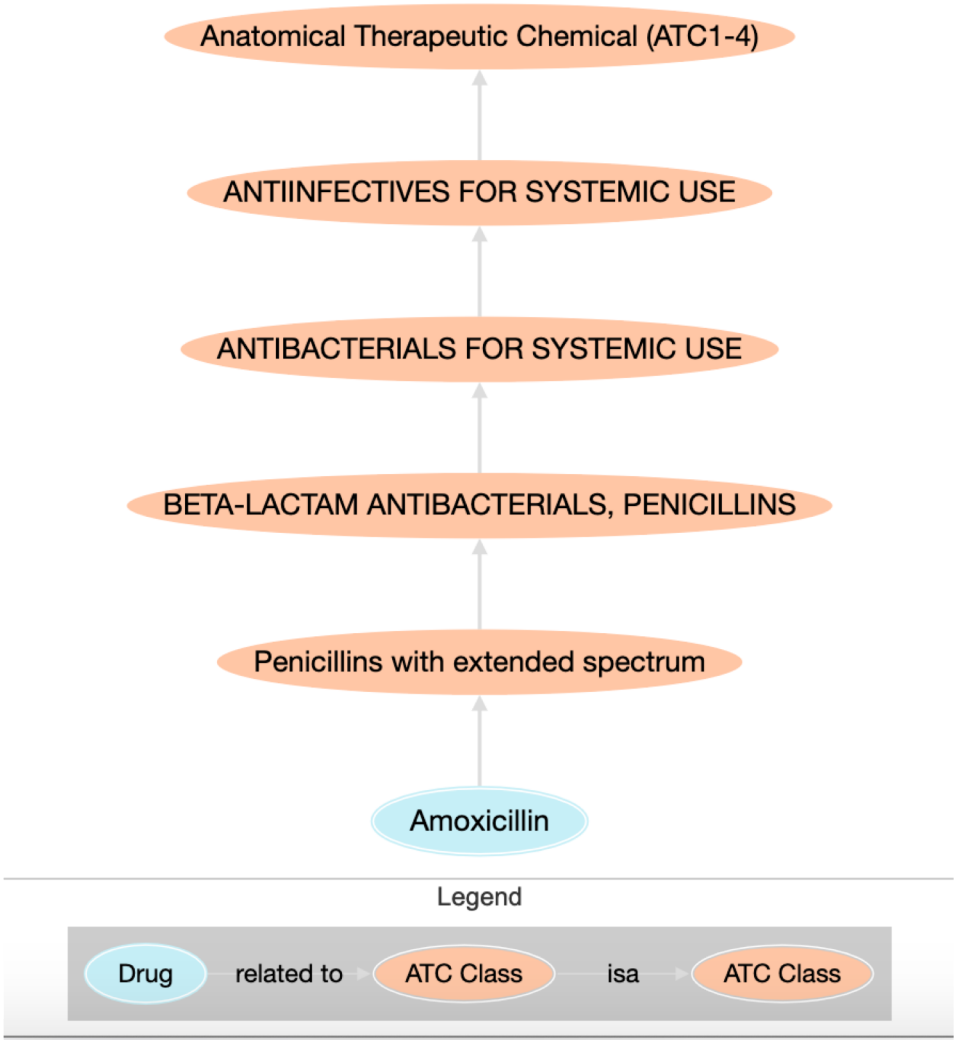


Source: https://mor.nlm.nih.gov/RxNav/

PEDSnet
A Pediatric Learning Health System

RxNav

- Includes ATC, Mesh, VA, and other classification systems



PEDSnet
A Pediatric Learning Health System

PEDSnet `drug_exposure` Table

Field name
drug_exposure_id (PK)
person_id (FK)
provider_id (FK)
visit_occurrence_id (FK)
drug_concept_id (and source value / source_concept_id)
drug_type_concept_id
drug_exposure_start_date / end_date (and datetime)
drug_exposure_order_date / end date (and datetime)
effective_drug_dose
refills
quantity
days_supply
frequency
route_concept_id (and source value)
dose_unit_concept_id (and source value)
stop_reason
lot_number
sig
dispense_as_written_concept_id

primary and foreign keys

specifies drug and type (prescription, inpatient order, inpatient admin, dispensed)

dates of order and start

drug instance metadata

optional fields

PEDSnet
A Pediatric Learning Health System

PEDSnet `drug_exposure` Table

Field name
drug_exposure_id (PK)
person_id (FK)
provider_id (PK)
visit_occurrence_id (PK)
drug_concept_id (and source value / source_concept_id)
drug_type_concept_id
drug_exposure_start_date / end_date (and datetime)
drug_exposure_order_date / end date (and datetime)
effective_drug_dose
refills
quantity
days_supply
frequency
route_concept_id (and source value)
dose_unit_concept_id (and source value)
stop_reason
lot_number
sig
dispense_as_written_concept_id

- Order of specificity:
  - BPCK
  - GPCK ~ 87% of drug data
  - SBD
  - SCD
  - SBDF ~ 5% of drug data
  - SCDF
  - MIN
  - SBDC
  - SCDC ~ 8% of drug data
  - PIN
  - IN

PEDSnet
A Pediatric Learning Health System

Codeset for Studies

- Number of drugs increase with specificity
- Prednisolone as example
  - Ingredient: 1
  - Clinical Dose Group: 40
  - Clinical Drug Form: 213
  - Clinical Drug: 423

Source: https://www.nlm.nih.gov/research/umls/rxnorm/overview.html

Codesets for Studies: Prednisolone

- Recommend review at Clinical Dose Group or Form

RxNorm Term Type	Examples
Ingredient	• Prednisolone
Clinical Dose Group	• Prednisolone Pill • Prednisolone Ophthalmic Product • Prednisone Topical Product • Prednisolone Rectal Product • Prednisolone Oral Liquid Product
Clinical Drug Form	• Prednisolone Rectal Foam • Prednisolone Topical Cream • Prednisone Oral Tablet
Clinical Drug	• prednisolone 0.001 MG/MG Topical Ointment • prednisolone 5 MG Delayed Release Oral Tablet

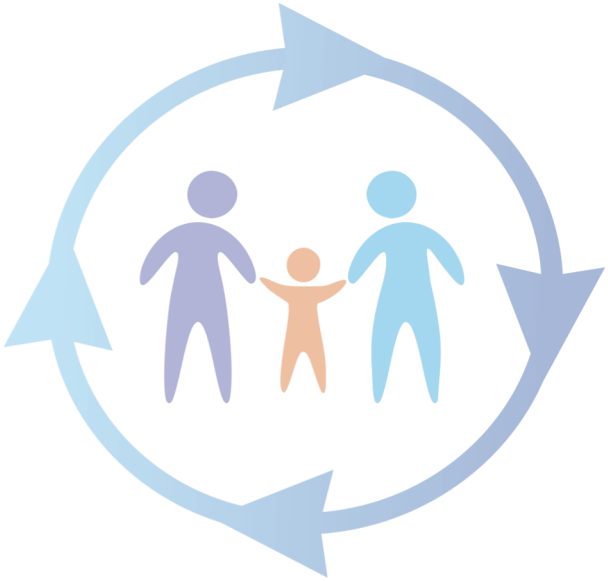
→ Too broad: can not specify route of administration

→ Likely the right level of specificity

→ Too specific: too many to sort through

PEDSnet
A Pediatric Learning Health System

Procedures



PEDSnet `procedure_occurrence` Table

- Any action other than medications performed by the health system for a patient
  - Office interventions
  - Laboratory testing
  - Diagnostic testing
  - Surgery
  - Administrative/billing process

- Should be **completed orders**

PEDSnet
A Pediatric Learning Health System

Procedure Vocabularies

- CPT (Current Procedural Terminology)
  - Most common for “smaller” procedures, including labs, testing, and office procedures
- HCPCS (HealthCare Common Procedural Coding System)
  - Additional codes with semantics similar to CPT
  - Particularly common for ED care
- ICD Procedure
  - Used by some institutions for “larger” procedures, particularly surgery
- SNOMED-CT
  - Typically found as standard form of pseudo-diagnosis codes

Finding Procedures

- `procedure_occurrence`
  - Orders and billing for non-medication activity
- `measurement`
  - Results of procedures that generate discrete data
- `condition_occurrence`
  - Pseudo-diagnoses generated for billing purposes (*e.g.* office procedures, infusions)

PEDSnet `procedure_occurrence` Table

Field name
procedure_occurrence_id
person_id
provider_id
visit_occurrence_id
procedure_concept_id (source value / source_concept_id)
modifier_concept_id (source vaule)
procedure_type_concept_id
quantity
procedure_date
procedure_datetime

primary and foreign keys

specifies procedure and info about procedure

- type = primary / secondary
- modifier = more info about procedure (e.g., bilateral)
- quantity = number ordered

date and datetime of order

PEDSnet
A Pediatric Learning Health System

Codeset Construction

- Procedure terminologies do not have formal hierarchy
- Informal semantic nesting by code prefix

- Need to account for aggregation
  - Lab panels
  - “Bundling” of interventions

- Don’t forget measurement and pseudo-condition records (PEDSnet differs from OHDSI here)

Codeset Construction

- Can still use concept_ancestor table to search for descendant / ancestor codes

Descendants for CPT code 1020217 *(Computed tomographic (CT) colonography, diagnostic, including image postprocessing*)

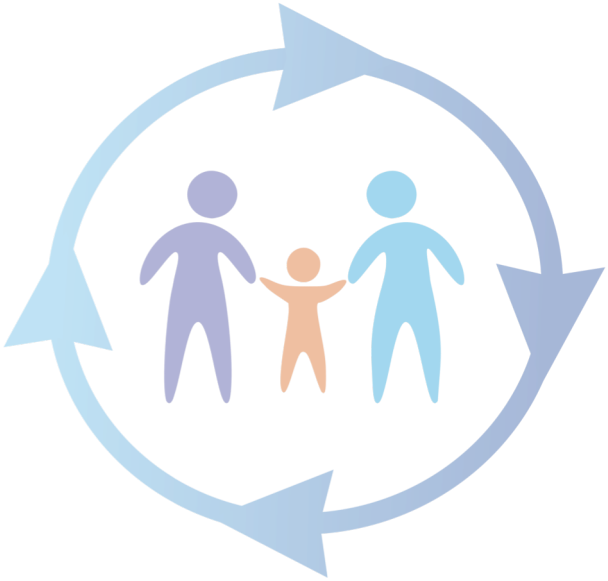
CPT code	CPT term
×74262	Computed tomographic (CT) colonography, diagnostic, including image postprocessing; with contrast material(s) including non-contrast images, if performed
74261	Computed tomographic (CT) colonography, diagnostic, including image postprocessing; without contrast material

Codeset Construction

Descendants for **ICD10PCS** code D82 *(×Radiation Therapy, Eye, Stereotactic Radiosurgery)*

ICD10PCS code	CPT term
×D820J	×Radiation Therapy @ Eye @ Stereotactic Radiosurgery @ Eye @ Stereotactic Gamma Beam Radiosurgery
D820DZZ	×Stereotactic Other Photon Radiosurgery of Eye
D820DZ	Radiation Therapy @ Eye @ Stereotactic Radiosurgery @ Eye @ Stereotactic Other Photon Radiosurgery @ None
D820HZ	Radiation Therapy @ Eye @ Stereotactic Radiosurgery @ Eye @ Stereotactic Particulate Radiosurgery @ None
D820D	Radiation Therapy @ Eye @ Stereotactic Radiosurgery @ Eye @ Stereotactic Other Photon Radiosurgery
D820H	Radiation Therapy @ Eye @ Stereotactic Radiosurgery @ Eye @ Stereotactic Particulate Radiosurgery
D820HZZ	Stereotactic Particulate Radiosurgery of Eye
D820JZZ	Stereotactic Gamma Beam Radiosurgery of Eye
D820	Radiation Therapy @ Eye @ Stereotactic Radiosurgery @ Eye
D820JZ	Radiation Therapy @ Eye @ Stereotactic Radiosurgery @ Eye @ Stereotactic Gamma Beam Radiosurgery @ None

Measurement



PEDSnet `measurement` Table

- Discrete results of procedures
  - Laboratory testing
  - Diagnostic testing (*e.g.* PFTs)
  - Anthropometrics
    - DCC derives z-scores
  - Vital signs

- ETL conventions contains details of vital signs (those included and PEDSnet standards)

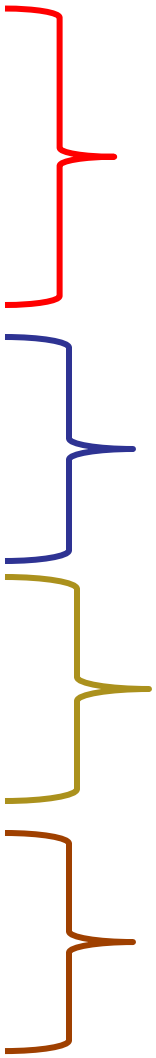
PEDSnet
A Pediatric Learning Health System

Finding Measurements

- `measurement`
  - Results of procedures that generate discrete data
  - Internal division into `measurement_labs`, `measurement_anthro`, and `measurement_vitals` – generally not to be used
- `observation`
  - Small number of measurements, typically qualitative
- `procedure_occurrence`
  - Orders for measurements – for when utilization is the endpoint rather than values

PEDSnet `measurement` Table (I)

Field Name
measurement_id (PK)
person_id (FK)
visit_occurrence_id (FK)
provider_id (FK)
measurement_concept_id (source_value and source_concept_id)
measurement_type_concept_id
measurement_date (datetime)
measurement_order_date (datetime)
measurement_result_date (datetime)
priority_concept_id (source value)
specimen_concept_id
specimen_source_value



primary and foreign keys

specifies measurement and type (e.g., lab, vital sign, etc)

lab dates

lab metadata; priority of lab (e.g., routine) and analyte

PEDSnet
A Pediatric Learning Health System

PEDSnet `measurement` Table (II)

operator_concept_id
value_as_number
value_as_concept_id
value_source_value
unit_concept_id (source_value)
range_low (source_value)
range_low_source_value
range_low_operator_concept_id
range_high (source_value)
range_high_operator_concept_id

result of measurement

range for calibration (labs)

PEDSnet
A Pediatric Learning Health System

Measurement Vocabularies

- LOINC (Logical Observation Identifiers Names & Codes)
  - Most common, especially for laboratory testing
  - LOINC code often incorporates specimen and/or units, but coverage is not complete, so do not rely on this
- SNOMED-CT
  - May be used for measurements lacking LOINC codes
- PEDSnet
  - Small number of additions for derived data

Codeset Construction

- Measurement terminologies do not have formal hierarchy
- PEDSnet has standard lab codes; variably adopted
- LOINC is piloting a hierarchy; it is not yet stable or built into CDM vocabularies
- Records occur at component level, not panel

- Coding is a work in progress

Codeset Construction

- Simple lab such as blood glucose has 60 values

concept_code character varying (256)	concept_id integer	concept_name character varying (512)
32779-1	3046616	Glucose-6-Phosphatase [Enzymatic activity/volume] in Red Blood Cells
49926-9	3029025	Glucose-6-Phosphate dehydrogenase [Enzymatic activity/mass] in Red Blood Cells by WHO met...
39480-9	3038515	Glucose [Moles/volume] in Venous blood
14770-2	3017703	Fasting glucose [Moles/volume] in Capillary blood by Glucometer
51596-5	3040151	Glucose [Moles/volume] in Capillary blood
11047-8	3002217	Glucose phosphate isomerase [Enzymatic activity/volume] in Red Blood Cells
62851-1	40765583	PhenX - fasting plasma glucose for diabetes screening - blood draw protocol
40858-3	3038962	Glucose [Mass/volume] in Capillary blood --baseline
2340-8	3011424	Glucose [Mass/volume] in Blood by Automated test strip
6689-4	3034530	Glucose [Mass/volume] in Blood --2 hours post meal
14743-9	3001501	Glucose [Moles/volume] in Capillary blood by Glucometer
71593-8	42869437	Glucose-6-Phosphate dehydrogenase [Enzymatic activity/substance] in Red Blood Cells
53553-4	3039720	Glucose mean value [Moles/volume] in Blood Estimated from glycated hemoglobin
59813-6	40762874	Glucose [Moles/volume] in Capillary blood by Glucometer --7 AM specimen
49266-0	3050115	Glucose-6-Phosphate dehydrogenase phenotype [Interpretation] in Red Blood Cells Narrative
44050-3	3046146	Glucose phosphate isomerase [Enzymatic activity/mass] in Red Blood Cells
49925-1	3031932	Glucose-6-Phosphate dehydrogenase [Enzymatic activity/mass] in Red Blood Cells by Glock an...
1557-8	3037187	Fasting glucose [Mass/volume] in Venous blood
41604-0	3035250	Fasting glucose [Mass/volume] in Capillary blood by Glucometer

PEDSnet Standards and Adaptations

PEDSnet Specific Standards

- PEDSnet data teams have adapted OMOP data model to meet complex data needs of system
- Adaptations of certain fields documented in ETL specifications
- Goal is to assist with debugging and data quality issues when they arise

PEDSnet Specific Standards

STANDARD	EXAMPLES
Expanded value sets	• *visit_concept_id* includes visit types not specified in omop; separate clinician face to face from imaging and lab visits • *care_site.specialty_concept_id* missing in omop; provides information about the specialty of the care site of visit
New tables to fit study needs	• *adt_occurrence* for identification and metadata of ICU transfer visits • *immunization* provides all immunization data available • *measurement_organism* for blood cultures
Creating separate tables to search for site-specific codes	• *procedure_occurrence_source_value* lists all standard concepts with unique source values • available for almost all fact tables to search for anomalies and completion of codesets
Partitioning	• *measurement* partitioned by vitals, labs, and anthroprometrics to increase efficiency and query execution time
Data Derivations	• *measurement_anthro* includes derived height/weight/BMI z-scores and BMI values

PEDSnet Specific Standards: Source Values

Condition source_value and source_concept_id

Please use the following logic to populate the `condition_concept_id`, `condition_source_concept_id` and `condition_source_value` based on what is available in your source system:

You have in your source system	condition_concept_id	condition_source_concept_id	condition_source_value
Any diagnosis that was captured as a term or name (e.g. IMO to SNOMED)	Corresponding SNOMED concept id	Corresponding concept for site diagnosis captured (must correspond to ICD9/ICD10 concept mapping)	Diagnosis Name "|" IMO Code "|" Diagnosis Code
Any diagnosis that was captured directly as a code (e.g. ICD9/10) by a coder	Corresponding SNOMED concept id	Corresponding concept for site diagnosis code (must correspond to ICD9/ICD10 concept mapping)	Diagnosis Name "|" IMO Code "|" Diagnosis Code

source value: diagnosis name | IMO code | ICD code

source concept id: ICD9/ICD10 standardized code

PEDSnet
A Pediatric Learning Health System

PEDSnet Specific Standards: Source Values

Note 6: Please use the following table as a guide to determine how to populate the `drug_source_value`, `drug_source_concept_id` and `drug_concept_id` for Drug Exposure Values

You have in your source system	Drug_source_value	Drug_source_conept_id	Drug_concept_id
Drug code is GPI/Multum/Other code	• GPI/Multum/Other Code • Local name | GPI/Multum/Other (any above are OK)	OMOP's concept_id for GPI/Multum/Other code	RxNorm code that corresponds to a mapping from `concept_relationship`
Drug code is RxNorm	• RxNorm Code • Local name or • Local name | RxNorm code (any above are OK)	Corresponding RxNorm concept_id mapping	Corresponding RxNorm concept_id mapping

source_value: drug_code | local_name

source_concept_id: standardized concept for local code

PEDSnet
A Pediatric Learning Health System

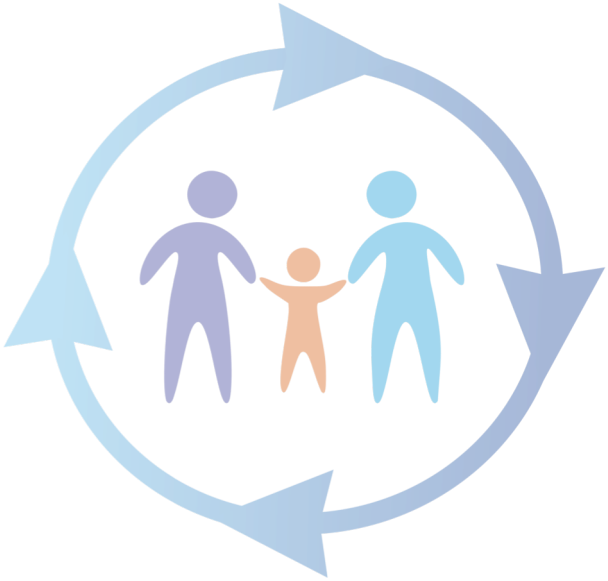
PEDSnet Specific Standards: Source Values

your source system	Listing Lab	Measurement_source_value	Measurement_source_concept_id
Lab code is institutional-specific code (not CPT/not LOINC)	**Yes**	• Local code or • Local name or • Local name | Local code/li> (any above are OK)	0 (zero)
Lab code is CPT code	**Yes**	• CPT Code • Local name or • Local name |CPT code (any above are OK)	OMOP's concept_id for CPT code
Lab code is LOINC code that is same as PEDSnet's LOINC code	**Yes**	• LOINC Code • Local name or • Local name | LOINC code (any above are OK)	PEDSnet's LOINC code's concept_id (provided by DCC)
Lab code is LOINC code that is different than PEDSnet	**Yes**	Same as above	OMOP's concept_id for your LOINC code

PEDSnet Specific Standards: Source Values

- PEDSnet list of standardized labs
  - available on PEDSnet conventions
  - ~ >700 standardized labs (most common)
- ***source_value*** is local code and local name
  - source values should represent what is in source
  - all attempts should be made to map to PEDSnet standard lab from source values for the **concept_id**
- ***source_concept_id*** is standardized concept from local code (if available)
  - PEDSnet list of labs are priority over other OMOP codes
  - if source is CPT, then field should be standardized CPT code

Remaining Tables



Data Model: Tables

- condition_occurrence
- procedure_occurrence
- drug_exposure
- measurement
- person
- visit_occurrence
- measurement_organism

- measurement_organism
- adt_occurrence
- immunization
- observation
- visit_payer
- era tables
- care_site
- provider

PEDSnet
A Pediatric Learning Health System

PEDSnet Data Model Additions

TABLE	DESCRIPTION
measurement_organism	• contains organism related to laboratory culture results in the measurement table • foreign key, *measurement_id*, back to measurement table • metadata about time, results, and source values
adt_occurrence	• contains transfers within inpatient hospitalization – currently limited to ICU • captures transfers in, out, and census • foreign key, *visit_occurrence_id*, back to visit_occurrence table • metadata about types of transfer, time,
immunization	• provides immunization information at the person-level • foreign key, *person_id,* back to person table • optional foreign keys, *visit_occurrence_id* and *procedure_occurrence_id*, back to visit_occurrence and procedure_occurrence tables, respectively

PEDSnet
A Pediatric Learning Health System

`observation` Table

- `observation` table an EAV structure
- Allows flexibility in adding variables to database that do not fit anywhere else
- Currently PEDSnet stores:
  - discharge status
  - tobacco use
  - smoking
  - birth information
  - DRG's

observation Table

Field name
observation_id (PK)
person_id (FK)
provider_id (FK)
visit_occurrence_id (FK)
observation_concept_id
observation_type_concept_id
qualifier_concept_id
observation_date (and datetime)
value_as_number
value_as_string
value_as_concept_id
unit_concept_id

primary and foreign keys

provides information about the observation of interest; can be any data point for which to make an observation

result from observation; allow for qualitative and quantiatative results

PEDSnet
A Pediatric Learning Health System

`provider` and `care_site` Tables

`provider` Table

- provider metadata, including **specialty**
- provider_id is primary key
- foreign key into almost all patient fact tables: person, visit_occurrence, procedure_occurrence, condition_occurrence, observation
  - usually indicates the provider responsible for instance of fact
  - provider_id in person table is either primary care provider or provider seen most often

`care_site` Table

- care site metadata, including **specialty**
- care_site_id is primary key
- foreign key into visit_occurrence and provider tables

- combination of **provider and care site specialty** often used to identify specialty a patient saw for particular visit
- some sites very good at provider specialty; others good at care site specialty
- specialty concepts in conventions

PEDSnet
A Pediatric Learning Health System

PEDSnet Remaining Tables

TABLE	DESCRIPTION
`visit_payer`	• visit level payer information • plan_type (e.g., HMO, PPO, POS, FFS, Other) • plan_class (e.g., Private, Medicaid, Medicare, Self-pay, etc)
`condition_era`, `drug_era`, `observation_period`	• purpose of table to provide summary information about patient facts • observation period provides total patient follow up • condition era / drug era: length of time a patient is continuously assigned a particular condition or drug • drug era rolled up to ingredient level → moving towards SCDG level

PEDSnet
A Pediatric Learning Health System