**PEDSnet Common Data Model Version 1.0**

Revision Date: October 22, 2014

The PEDSnet Common Data Model is an evolving specification, based in structure on the OMOP Common Data Model, but expanded to accommodate requirements of both the PCORnet Common Data Model and the primary research cohorts established in PEDSnet.

Version 1 of the PEDSnet CDM reflects the ETL processes developed during the first six months of network development. As such, it closely follows version 1 of the PCORnet CDM. We anticipate that the PEDSnet CDM will expand to include additional data domains such as medication usage, laboratory testing, and other types of clinical observations. However, in order to minimize discordance with the PCORnet CDM, specification of these domains has been deferred until the cognate portion of future PCORnet CDM specifications are released, or active research in one of the PEDSnet cohorts requires those data types sooner than PCORnet.

This document provides the data model specification, including names, data types, and description of key semantic points, for the PEDSnet CDM. For ETL specification documents, please consult the PEDSnet GitHub site. Included in other documents are ETL conventions for specific tables and fields to ensure common business rules and specific OMOP CONCEPT\_IDs that must be used to ensure consistent terminology coding across PEDSnet and to ensure accurate extraction of PCORnet CDM data elements from a PEDSnet CDM instance.

Comments on this specification, whether providing critique of the existing data elements or requesting additional data types, are welcome. Please send email to [pedsnetdcc@email.chop.edu](mailto:pedsnetdcc@email.chop.edu), or contact the PEDSnet project management office (details available via <http://www.pedsnet.info>).

**PEDSnet Data Standards and Interoperability Policies:**

1: The PEDSnet data network will store data using structures compatible with the PEDSnet Common Data Model (PCDM).

2: The PCDM will be based on the Observational Medical Outcomes Partnership (OMOP) data model, version 4. OMOP will be expanded to include the PCORnet and pediatric-specific data standards, as developed by PEDSnet.

The second release of PCDM will be based on OMOP CDM Version 5, which is still in draft form at the time of finalization of PCDM CDM V1.

3: A subset of data elements in the PCDM will be identified as principal data elements (PDEs). The PDEs will be used for population-level queries. Data elements which are NOT PDEs will be marked as “Optional” (ETL at site discretion) or “Non-PDE” (ETL required, but data need not be transmitted to DCC, and will not be used in queries without prior approval of site.

4: It is anticipated that PEDSnet institutions will make a good faith attempt to obtain as many of the data elements not marked as “Optional” as possible.

1. The data elements classified as PDEs and those included in the PCDM will be approved by the PEDSnet Leadership Council (comprised of each PEDSnet institution’s site principal investigator).
2. Concept IDs are taken from OMOP v4.5 vocabularies for PEDSnet CDM v1, using the complete (“restricted”) version that includes licensed terminologies such as CPT and others.
3. Regarding the nullability of all source value (string) fields only, in the presence of a missing or unmapped value in source data, one of the following values must be used:

|  |  |
| --- | --- |
| **Null Name** | **Definition of each field** |
| NULL | A data field is not present in the source system |
| NI = No Information | A data field is present in the source system, but the source value is null or blank |
| UN = Unknown | A data field is present in the source system, but the source value explicitly denotes an unknown value |
| OT = Other | A data field is present in the source system, but the source value cannot be mapped to the CDM |

**Please note: For all unique identifiers, please do not send actual cleartext identifiers from clinical source systems.**

The following tables are included in the CDM:

## PERSON

The person domain contains records that uniquely identify each patient in the source data who meets the network definition of a “PEDSnet Patient”. See ETL conventions document for the current definition of a “PEDSnet Patient.” Each person record has associated demographic attributes, which represent constant or summarized values for the patient throughout the course of their periods of observation. All other patient-related data domains have a foreign-key reference to the person domain.

| **Field** | **Required** | **Type** | **Description** |
| --- | --- | --- | --- |
| person\_id | Yes | integer | A unique identifier for each person; this is created by each contributing site. Note: This is not a value found in the EHR. However, a map to patient ID or MRN from the EHR must be kept at the site and not shared with the data coordinating center for re-identification in the future.. |
| gender\_concept\_id | Yes | integer | A foreign key that refers to a standard concept identifier in the Vocabulary for the gender of the person. |
| year\_of\_birth | Yes | integer | The year of birth of the person. For data sources with date of birth, the year is extracted. For data sources where the year of birth is not available, the approximate year of birth is derived based on any age group categorization available. |
| month\_of\_birth | No | integer | The month of birth of the person. For data sources that provide the precise date of birth, the month is extracted and stored in this field. |
| day\_of\_birth | No | Date | The day of the month of birth of the person. For data sources that provide the precise date of birth, the day is extracted and stored in this field. |
| pn\_time\_of\_birth | No | Time | The time of birth at the birthday (UTC). NOTE:. |
| race\_concept\_id | No | integer | A foreign key that refers to a standard concept identifier in the Vocabulary for the race of the person. |
| ethnicity\_concept\_id | No | integer | A foreign key that refers to the standard concept identifier in the Vocabulary for the ethnicity of the person. |
| location\_id | No | integer | A foreign key to the place of residency (ZIP code) for the person in the location table, where the detailed address information is stored. |
| provider\_id | No | Integer | Foreign key to the primary care provider – the person is seeing in the provider table. |
| care\_site\_id | Yes | integer | A foreign key to the site of primary care in the care\_site table, where the details of the care site are stored. |
| pn\_gestational\_age | No | Number | The post-menstrual age in weeks of the person at birth, if known. |
| person\_source\_value | No | varchar | An encrypted key derived from the person identifier in the source data. For site-specific data extracts, this may be identical to the person\_id.. |
| gender\_source\_value | No | varchar | The source code for the gender of the person as it appears in the source data. The person’s gender is mapped to a standard gender concept in the Vocabulary; the original value is stored here for reference. |
| race\_source\_value | No | varchar | The source code for the race of the person as it appears in the source data. The person race is mapped to a standard race concept in the Vocabulary and the original value is stored here for reference. |
| ethnicity\_source\_value | No | varchar | The source code for the ethnicity of the person as it appears in the source data. The person ethnicity is mapped to a standard ethnicity concept in the Vocabulary and the original code is, stored here for reference. |

### Conventions

1. For site-speicific data extracts, the person\_id is unique across all patients within that data partner. For PEDSnet-wide databases and resultsets containing individual records, the person\_id is unique across PEDSnet.
2. An individual may receive care at multiple contributing institutions, and therefore be represented more than once in the data.
3. The provider\_id in the Person table is intended to represent a primary or principal provider associated with a person. Logic for this is site-specific in which case the ETL documentation for a site must specify how a single provider id is selected or leave the provider\_id field null.
4. The care\_site\_id in the Person table is intended to represent a principal site of care for a patient. Sites may choose to designate a particular clinic or department during their ETL, in which case the ETL documentation must specify how a single care\_site id is selected, or to set the care\_site\_id for all persons to a single care\_site\_id representing the institution as a whole.

## DEATH

The death domain contains the clinical event for how and when a person dies. Living patients should not contain any information in the death table.

|  |  |  |  |
| --- | --- | --- | --- |
| Field | Required | Type | Description |
| person\_id | Yes | integer | A foreign key identifier to the deceased person. The demographic details of that person are stored in the person table. |
| death\_date | Yes | date | The date the person was deceased. If the precise date including day or month is not known or not allowed, December is used as the default month, and the last day of the month the default day. If no date available, use date recorded as deceased. |
| death\_type\_concept\_id | Yes | integer | A foreign key referring to the predefined concept identifier in the Vocabulary reflecting how the death was represented in the source data. |
| cause\_of\_death\_concept\_id | No | integer | A foreign key referring to a standard concept identifier in the Vocabulary for conditions. |
| cause\_of\_death\_source\_value | No | varchar | The source code for the cause of death as it appears in the source data. This code is mapped to a standard concept in the Vocabulary and the original code is, stored here for reference. |

### Conventions

* Each Person may have more than one record of death in the source data. It is the task of the ETL to pick the most plausible or most accurate records to be stored to this table.
* If the Death Date cannot be precisely determined from the data, the best approximation should be used.

## LOCATION

The Location table represents a generic way to capture physical location or address information. Locations are used to define the addresses for Persons, Providers, and Care Sites.

|  |  |  |  |
| --- | --- | --- | --- |
| **Field** | **Required** | **Type** | **Description** |
| location\_id | Yes | integer | A unique identifier for each geographic location. |
| state | No | varchar | The state field as it appears in the source data. |
| zip | No | varchar | The zip code. For US addresses, valid zip codes can be 3, 5 or 9 digits long, depending on the source data. |
| location\_source\_value | No | varchar | The verbatim information that is used to uniquely identify the location as it appears in the source data. |
| address\_1 | No | varchar | Optional - Do not transmit to DCC |
| address\_2 | No | varchar | Optional - Do not transmit to DCC |
| city | No | varchar | Optional - Do not transmit to DCC |
| county | No | varchar | Optional - Do not transmit to DCC |

### Conventions

* Each address or Location is unique and is present only once in the table

## CARE\_SITE

The Care Site table contains a list of uniquely identified physical or organizational units where healthcare delivery is practiced (offices, wards, hospitals, clinics, etc.).

|  |  |  |  |
| --- | --- | --- | --- |
| Field | Required | Type | Description |
| care\_site\_id | Yes | integer | A unique identifier for each defined location of care within an organization. Here, an organization is defined as a collection of one or more care sites that share a single EHR database. |
| place\_of\_service\_concept\_id | No | integer | A foreign key that refers to a place of service concept identifier in the Vocabulary |
| location\_id | No | integer | A foreign key to the geographic location of the administrative offices of the organization in the location table, where the detailed address information is stored. |
| care\_site\_source\_value | No | varchar | The identifier for the organization in the source data, stored here for reference. |
| place\_of\_service\_source\_value | No | varchar | The source code for the place of service as it appears in the source data, stored here for reference. |
| organization\_id | Yes | Integer | A foreign key to the organization record. |

### Conventions

* Care sites are primarily identified based on the specialty or type of care provided, and secondarily on physical location, if available (e.g. North Satellite Endocrinology Clinic)
* The Care Site Source Value typically contains the name of the Care Site.
* The Place of Service Concepts are based on a catalog maintained by the CMS (see vocabulary for values)

## ORGANIZATION

|  |  |  |  |
| --- | --- | --- | --- |
| Field | Required | Type | Description |
| organization\_id | Yes | Integer | A unique identifier for each defined location of care within an organization. Here, an organization is defined as a collection of one or more care sites that share a single EHR database. |
| place\_of\_service\_concept\_id | No | Integer | A foreign key that refers to a place of service concept identifier in the Vocabulary |
| location\_id | No | Integer | A foreign key to the geographic location of the administrative offices of the organization in the location table, where the detailed address information is stored. |
| place\_of\_service\_source\_value | No | varchar | The source code for the place of service as it appears in the source data, stored here for reference. |
| organization\_source\_value | Yes | Integer | The identifier for the organization in the source data, stored here for reference |

## PROVIDER

The Provider table contains a list of uniquely identified health care providers. These are typically physicians, nurses, etc.

|  |  |  |  |
| --- | --- | --- | --- |
| **Field** | **Required** | **Type** | **Description** |
| provider\_id | Yes | Integer | A unique identifier for each provider. Each site must maintain a map from this value to the identifier used for the provider in the source data. |
| specialty\_concept\_id | No | Integer | A foreign key to a standard provider's specialty concept identifier in the Vocabulary. This is an optional field. PEDSnet has not yet defined an ETL convention for selecting one specialty code for providers who have multiple specialties. |
| care\_site\_id | Yes | Integer | A foreign key to the main care site where the provider is practicing.. This field is required in PEDSnet, which is a deviation from OMOP CDM V4. |
| NPI | No | varchar | The National Provider Identifier (NPI) of the provider. Optional - Do not transmit to DCC |
| DEA | No | varchar | The Drug Enforcement Administration (DEA) number of the provider. Optional - Do not transmit to DCC |
| provider\_source\_value | No | varchar | The identifier used for the provider in the source data, stored here for reference. Sites should create a random ID, but keep the mapping. |
| specialty\_source\_value | No | varchar | The source code for the provider specialty as it appears in the source data, stored here for reference. |

### Conventions

* Providers are not duplicated in the table.
* Valid Specialty Concepts are based on the CDC specialty classification contained in the OMOP Vocabluary.

## VISIT\_OCCURRENCE

The visit domain contains the spans of time a person continuously receives medical services from one or more providers at a care site in a given setting within the health care system. Visits are classified into 4 settings: outpatient care, inpatient confinement, emergency room, and long-term care

| **Field** | **Required** | **Type** | **Description** |
| --- | --- | --- | --- |
| visit\_occurrence\_id | Yes | integer | A unique identifier for each person’s visits or encounter at a healthcare provider. Sites can provide whatever integers (DCC will replace the value). |
| person\_id | Yes | integer | A foreign key identifier to the person for whom the visit is recorded. The demographic details of that person are stored in the person table. |
| visit\_start\_date | Yes | date | The start date of the visit. |
| visit\_end\_date | No | date | The end date of the visit. If this is a one-day visit the end date should match the start date. If the encounter is on-going at the time of ETL, this should be null. |
| provider\_id | No | integer | A foreign key to the provider in the provider table who was associated with the visit.  **NOTE: this is NOT in OMOP CDM v4, but appears in OMOP CDMv5. PEDSnet is including the field at this time due to an existing use case (Obesity cohort).** |
| care\_site\_id | Yes | integer | A foreign key to the care site in the care site table that was visited. |
| place\_of\_service\_concept\_id | Yes | integer | A foreign key that refers to a place of service concept identifier in the vocabulary. |
| place\_of\_service\_source\_value | No | varchar | The source code used to reflect the type or source of the visit in the source data. Valid entries include office visits, hospital admissions, etc. These source codes can also be type-of service codes and activity type codes. |

### Conventions

* A Visit Occurrence is recorded for each visit to a healthcare facility.
* Each Visit is standardized by assigning a corresponding Concept Identifier in place\_of\_service\_concept\_id based on the type of facility visited and the type of services rendered.
* At any one day, there could be more than one visit.
* One visit may involve multiple providers, in which case the ETL must specify how a single provider id is selected or leave the provider\_id field null.
* One visit may involve multiple care sites, in which case the ETL must specify how a single care\_site id is selected or leave the care\_site\_id field null.

## CONDITION\_OCCURRENCE

The condition occurrence domain captures records of a disease or a medical condition based on diagnoses, signs and/or symptoms observed by a provider or reported by a patient.

Conditions are recorded in different sources and levels of standardization. For example:

* Medical claims data include ICD-9-CM diagnosis codes that are submitted as part of a claim for health services and procedures.
* EHRs may capture a person’s conditions in the form of diagnosis codes and symptoms as ICD-9-CM codes, but may not have a way to capture out-of-system conditions.

|  |  |  |  |
| --- | --- | --- | --- |
| Field | Required | Type | Description |
| condition\_occurrence\_id | Yes | integer | A unique identifier for each condition occurrence event. |
| person\_id | Yes | integer | A foreign key identifier to the person who is experiencing the condition. The demographic details of that person are stored in the person table. |
| condition\_concept\_id | Yes | integer | A foreign key that refers to a standard condition concept identifier in the Vocabulary. |
| condition\_start\_date | Yes | date | The date when the instance of the condition is recorded. |
| condition\_end\_date | No | date | The date when the instance of the condition is considered to have ended. If this information is not available, set to NULL. |
| condition\_type\_concept\_id | Yes | integer | A foreign key to the predefined concept identifier in the Vocabulary reflecting the source data from which the condition was recorded, the level of standardization, and the type of occurrence. For example, conditions may be defined as primary or secondary diagnoses, problem lists and person statuses. |
| stop\_reason | No | varchar | The reason, if available, that the condition was no longer recorded, as indicated in the source data. Valid values include discharged, resolved, etc. Note that a stop\_reason does not necessarily imply that the condition is no longer occurring. |
| associated\_provider\_id | No | integer | A foreign key to the provider in the provider table who was responsible for determining (diagnosing) the condition. |
| visit\_occurrence\_id | No | integer | A foreign key to the visit in the visit table during which the condition was determined (diagnosed). |
| condition\_source\_value | No | varchar | The source code for the condition as it appears in the source data. This code is mapped to a standard condition concept in the Vocabulary and the original code is, stored here for reference. Condition source codes are typically ICD-9-CM diagnosis codes from medical claims or discharge status/visit diagnosis codes from EHRs. |

### Conventions

* Condition records are inferred from diagnostic codes recorded in the source data by a clinician or abstractionist for a specific visit. In the current version of the CDM, problem list entries are not used, nor are diagnoses extracted from unstructured data, such as notes.
* Source code system, like ICD-9-CM, ICD-10-CM, etc., provide coverage of conditions. However, if the code does not define a condition, but rather an observation or a procedure, then such information is not stored in the CONDITION\_OCCURRENCE table, but in the respective observation or procedure tables instead (see ETL conventions document).
* Source Condition identifiers are mapped to Standard Concepts for Conditions in the Vocabulary. When the source code cannot be translated into a Standard Concept, a CONDITION\_OCCURRENCE entry is stored with only the corresponding source\_value and a condition\_concept\_id of 0.
* Codes written in the process of establishing the diagnosis, such as "question of" of and "rule out", are not represented here.

## PROCEDURE\_OCCURRENCE

The procedure domain contains records of significant activities or processes ordered by and/or carried out by a healthcare provider on the patient to have a diagnostic and/or therapeutic purpose that are not fully captured in another table (e.g. drug\_exposure).

Procedures records are extracted from structured data in Electronic Health Records that capture CPT-4, ICD-9-CM (Procedures), HCPCS or OPCS-4 procedures as orders.

| **Field** | **Required** | **Type** | **Description** |
| --- | --- | --- | --- |
| procedure\_occurrence\_id | Yes | integer | A system-generated unique identifier for each procedure occurrence. Sites can use any integer- DCC will do a substitution. |
| person\_id | Yes | integer | A foreign key identifier to the person who is subjected to the procedure. The demographic details of that person are stored in the person table. |
| procedure\_concept\_id | Yes | integer | A foreign key that refers to a standard procedure concept identifier in the Vocabulary. |
| procedure\_date | Yes | date | The date on which the procedure was performed. |
| procedure\_type\_concept\_id | Yes | integer | A foreign key to the predefined concept identifier in the Vocabulary reflecting the type of source data from which the procedure record is derived. |
| associated\_provider\_id | No | integer | A foreign key to the provider in the provider table who was responsible for carrying out the procedure. |
| visit\_occurrence\_id | No | integer | A foreign key to the visit in the visit table during which the procedure was carried out. |
| relevant\_condition\_concept\_id | No | integer | A foreign key to the predefined concept identifier in the vocabulary reflecting the condition that was the cause for initiation of the procedure. Note that this is not a direct reference to a specific condition record in the condition table, but rather a condition concept in the vocabulary. |
| procedure\_source\_value | No | varchar | The source code for the procedure as it appears in the source data. This code is mapped to a standard procedure concept in the Vocabulary and the original code is, stored here for reference. Procedure source codes are typically ICD-9-Proc, ICD-10-Proc, CPT-4, HCPCS or OPCS-4 codes. |

### Conventions

* Procedure source values may be based on a variety of vocabularies: SNOMED-CT, ICD-9-Proc, CPT-4, HCPCS and OPCS-4.
* Procedures could reflect the administration of a drug, in which case the procedure is recorded in the procedure table and simultaneously the administered drug in the drug table.
* The Visit during which the procedure was performed is recorded through a reference to the VISIT\_OCCURRENCE table. This information is not always available.
* The Provider carrying out the procedure is recorded through a reference to the PROVIDER table. This information is not always available.

## OBSERVATION

The observation domain captures clinical facts about a patient obtained in the context of examination, questioning or a procedure. For the PEDSnet CDM version 1, the following observations are extracted from source data:

* Height/length in cm
* Height/length type
* Weight in kg
* Body mass index in kg/m2
* Systolic blood pressure in mmHg
* Diastolic blood pressure in mmHg

In addition, the following observations are derived via the DCC:

* Body mass index in kg/m2 if not directly extracted
* Height/length z score for age/sex using NHANES 2000 norms for measurements at which the person was <240 months of age. In the absence of a height/length type for the measurement, recumbent length is assumed for ages <24 months, and standing height thereafter.
* Weight z score for age/sex using NHANES 2000 norms for measurements at which the person was <240 months of age.
* BMI z score for age/sex using NHANES 2000 norms for visits at which the person was between 20 and 240 months of age.
* Systolic BP z score for age/sex/height using NHBPEP task force fourth report norms.
* Diastolic BP z score for age/sex/height using NHBPEP task force fourth report norms.

|  |  |  |  |
| --- | --- | --- | --- |
| Field | Required | Type | Description |
| observation\_id | Yes | integer | A unique identifier for each observation. |
| person\_id | Yes | integer | A foreign key identifier to the person about whom the observation was recorded. The demographic details of that person are stored in the person table. |
| observation\_concept\_id | Yes | integer | A foreign key to the standard observation concept identifier in the Vocabulary. |
| observation\_date | Yes | date | The date of the observation (UTC). |
| observation\_time | No | time | The time of the observation (UTC). |
| observation\_type\_concept\_id | Yes | integer | A foreign key to the predefined concept identifier in the Vocabulary reflecting the type of the observation. |
| value\_as\_number | No | float | The observation result stored as a number. This is applicable to observations where the result is expressed as a numeric value. |
| value\_as\_string | No | varchar | The observation result stored as a string. This is applicable to observations where the result is expressed as verbatim text. |
| value\_as\_concept\_id | No | integer | A foreign key to an observation result stored as a concept identifier. This is applicable to observations where the result can be expressed as a standard concept from the Vocabulary (e.g., positive/negative, present/absent, low/high, etc.). |
| unit\_concept\_id | No | integer | A foreign key to a standard concept identifier of measurement units in the Vocabulary. |
| associated\_provider\_id | No | integer | A foreign key to the provider in the provider table who was responsible for making the observation. |
| visit\_occurrence\_id | No | integer | A foreign key to the visit in the visit table during which the observation was recorded. |
| relevant\_condition\_concept\_id | No | Integer | A foreign key to the condition concept related to this observation, if this relationship exists in the source data (*e.g.* indication for a diagnostic test). |
| observation\_source\_value | No | varchar | The observation code as it appears in the source data. This code is mapped to a standard concept in the Vocabulary and the original code is, stored here for reference. |
| unit\_source\_value | No | varchar | The source code for the unit as it appears in the source data. This code is mapped to a standard unit concept in the Vocabulary and the original code is, stored here for reference. |
| range\_low | No | float | The lower limit of the normal range of the observation. It is not applicable if the observation results are non-numeric or categorical, and must be in the same units of measure as the observation value. Optional - Do not transmit to DCC |
| range\_high | No | float | The upper limit of the normal range of the observation. It is not applicable if the observation results are non-numeric or categorical, and must be in the same units of measure as the observation value. Optional - Do not transmit to DCC |

### Conventions

* Observations must have a value represented by one of a concept ID and a string value or a concept ID and a numeric value.
* The Visit during which the observation was made is recorded through a reference to the VISIT\_OCCURRENCE table. This information is not always available.
* The Provider making the observation is recorded through a reference to the PROVIDER table. This information is not always available.

## OBSERVATION PERIOD

The observation period table is designed to capture the time intervals in which data are being recorded for the person. An observation period is the span of time when a person is expected to have the potential of drug and condition information recorded. This table is used to generate the PCORnet CDM enrollment table.

While analytic methods can be used to calculate gaps in observation periods that will generate multiple records (observation periods) per person, for PEDSnet, the logic has been simplified to generate a single observation period row for each patient.

|  |  |  |  |
| --- | --- | --- | --- |
| Field | Required | Data Type | Description |
| Observation\_period\_id | Yes | Integer | A system-generate unique identifier for each observation period |
| person\_id | Yes | Integer | A foreign key identifier to the person who is experiencing the condition. The demographic details of that person are stored in the person table. |
| Observation\_period\_start\_date | Yes | Date | The start date of the observation period for which data are available from the data source |
| Observation\_period\_stop\_date | No | Date | The end date of the observation period for which data are available from the source. |

### Additional Notes