ETL Conventions for use with PEDSnet CDM v3.6 OMOP V5.2

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 3.6 of the PEDSnet CDM reflects the ETL processes developed after several iterations of network development. As such, it proposes to align with version 5.2 of the PCORnet CDM.

This document provides the ETL processing assumptions and conventions developed by the PEDSnet data partners that should be used by a data partner for ensuring common ETL business rules. This document will be modified as new situations are identified, incorrect business rules are identified and replaced, as new analytic use cases impose new/different ETL rules, and as the PEDSnet CDM continues to evolve.

Comments on this specification and ETL rules are welcome. Please send email to 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 PEDSnet CDM v3.6 is based on the Observational Medical Outcomes Partnership (OMOP) data model, version 5.2.
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.
5. The data elements classified as PDEs and those included in the PCDM will be approved by the PEDSnet Executive Committee (comprised of each PEDSnet institution's site principal investigator).
6. Concept IDs are taken from OMOP 5 vocabularies for PEDSnet CDM v3.6, using the complete (restricted) version that includes licensed terminologies such as CPT and others.
7. PCORnet CDM v5.2 requires data elements that are not currently considered "standard concepts". Vocabulary version 5 has a new vocabulary (vocabulary*id=PCORNet) that was added by OMOP to capture all of the PCORnet concepts that are not in the standard terminologies. We use concept*ids from vocabulary*id=PCORNet where there are no existing standard concepts. We highlight where we are pulling concept*ids from vocabulary*id=PCORNet in the tables. While terms from vocabulary*id=PCORNet violates the OMOP rule to use only concept*ids from standard vocabularies vocabulary*id=PCORNet is a non-standard vocabulary), this convention enables a clean extraction from PEDSnet CDM to PCORnet CDM.
8. Some source fields may be considered sensitive by data sites. Potential examples include patient*source*value, provider*source*value, care*site*source*value. Many of these fields are used to generate an ID field, such as PERSON.patient*source*value PERSON.person*id, that is used as a primary key in PERSON and a foreign key in many other tables. Sites are free to obfuscate or not provide source values that are used to create ID variables. Sites must maintain a mapping from the ID variable back to the original site-specific value for local re-identification tasks.
    1. Source fields that contain clinical data, such as source condition occurrence, should be included
    2. The PEDSnet DCC will never release source values to external data partners.
    3. Source value obfuscation techniques may include replacing the real source value with a random number, an encrypted derivative value/string, or some other site-specific algorithm.
9. The PCORnet CDM has specific definitons for null values (as seen below). For the PEDSNet CDM, please use the following logic on which concept value to use for `source_concept_id` fields where there are null values in the source `*_source_value`.

Null Name	Definition of each field
NULL	A data field is not present in the source system. Note. This is not a 'NULL' string but the NULL value.
'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

Guidelines for populating `'*_concept_id'`, `'*_source_concept_id'` and `'*_source_value'` for flavors of null:

Null Name	'*_concept_id'	'*_source_concept_id'	'*_source_value'
'NI'	44814650	0	value as in source (leave as null)
'UN'	44814653	0	value as in source (denoting an unknown value)
'OT'	44814649	0	value as in source

10. **For populating `'*_source_concept_id'` (where there exists non-null values in the source) use the following Logic :**

Populate `'*_source_concept_id'` (i.e. non-zero) if the source_value is drawn from a standard vocabulary in OMOP.

Please use your local system knowledge to determine this or use the following criteria : All the values in the source*value field should be drawn from the concept*code in the concept table (for a given/relevant domain*id and a given vocabulary*id).

ELSE Use 0

(usually the case when the sites need to "manually" map the foo*source*value to foo*concept*id)

11. **For populating `*_source_value` please make a best effort to provide "human readable" values rather than a coded value where possible from the source.**

Example for `gender_source_value`, the source value at your site may be `1` for Female and `2` for Male. Please provide the label value of `Female` and `Male`.

ETL Recommendation: Due to PK/FK constraints, the most efficient order for ETL table is location, care*site, provider, person, visit*occurrence, condition*occurrence, observation, procedure*occurrence,measurement,measurement_organism,drug exposure

Table of Contents

Data Extraction Guide

Please use the table headings as a guide in extracting and submitting data. These specifications are indicative of DCC and Network Requirements. All fields must be submitted to the DCC even if you are not submitting data in a field. Here are examples of how the specification should be interpreted:

Field	NOT Null Constraint	Network Requirement	Data Type	Description	PEDSnet Conventions
Field Name	• Yes	• Yes	Data Type	Description	PEDSnet Conventions

- **The above example indicates the data in this field is required by both the DCC and Network. It absolutely must be provided in the data submission.**

Field	NOT Null Constraint	Network Requirement	Data Type	Description	PEDSnet Conventions
Field Name	• No	• Provide When Available	Data Type	Description	PEDSnet Conventions

- **The above example indicates the data in this field is required by Network if it is populated or available at your site. If it is available it must provided in the data submission.**

Field	NOT Null Constraint	Network Requirement	Data Type	Description	PEDSnet Conventions
Field Name	• No	• Site Preference	Data Type	Description	PEDSnet Conventions

- **The above example indicates the data in this field is not required by the DCC or Network. A site may choose to send this information if they desire to do so.**

Field	NOT Null Constraint	Network Requirement	Data Type	Description	PEDSnet Conventions
Field Name	• No	• Optional	Data Type	Description	PEDSnet Conventions

- **The above example indicates the data in this field is truly optional for submission. A site may choose to send this information if they desire to do so.**

1.1 PERSON

The person domain contains records that uniquely identify each patient in the source data who is time at-risk to have clinical observations recorded within the source systems. Each person record has associated demographic attributes, which are assumed to be constant 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.

PEDSnet uses a specific definition of an active PEDSnet patient. Only patients who meet the PEDSnet definition of an active patient should be included in this table. The criteria for identifying an active patient are:

- Has a unique identifier AND
- At least 1 "in person" clinical encounter on or after January 1, 2009 AND
- At least 1 coded diagnoses recorded on or after January 1, 2009 AND
- Is not a test patient or a research-only patient

The definition of an "in person" clinical encounter remains heuristic -any encounter type that involves a meaningful **physical** interaction with a clinician that involved clinical content. An encounter for a telephone encounter or a lab blood draw does not meet this definition.

For reference `Visit_concept_ids` that correspond to an "in person" clinical encounter are:

Visit Type	Visit*concept*id
Inpatient Hospital Stay	9201
Ambulatory/Outpatient Visit (With a Physician)	9202
Outpatient Non Physician	2000000469
Emergency Department	9203
Long Term Care Visit	42898160
Non-Acute Institutional Stay	44814710
Emergency Department Admit to Inpatient Hospital Stay (If sites are unable to split the encounter)	2000000048
Observation Stay	2000000088

NOTE: While the 1/1/2009 date and "in person" clinical encounter restrictions apply to defining an active PEDSnet patient, once a patient has met this criteria, PEDSnet will extract *ALL* available clinical encounters/clinical data of any type across all available dates. That is, 1/1/2009 and 1 'in person' clinical encounter applies only to defining the active patient cohort. It does NOT apply to data extraction on active patients.

Field	NOT Null Constraint	Network Requirement	Data Type	Description	PEDSnet Conventions
person_id	Yes	Yes	Integer	A unique identifier for each person; this is created by each contributing site.	This is not a value found in the EHR. PERSON*ID must be unique for all patients within a single data set.* ***SITE RESPONSIBILITY: This field must remain a stable identifier across submissions to the DCC.*** *A mapping from the person*id to a real patient ID or MRN from the source EHR must be kept at the local site. This mapping is not shared with the data coordinating center. It is used only by the site for re-identification for study recruitment or for data quality review.
gender*concept*id	Yes	Yes	Integer	A foreign key that refers to a standard concept identifier in the Vocabulary for the gender of the person.	Please include valid concept ids (consistent with OMOP CDMv5.1). Predefined value set (valid concept*ids found in CONCEPT table select * from concept where ((domain*id='Gender' and concept*class*id='Gender')or (domain*id='Observation' and vocabulary*id='PCORNet' and concept*class*id in ('Gender','Undefined'))) and concept*code not in ('Sex-F','Sex-M') and invalid*reason is null: • Ambiguous: concept*id = 44814664* • *Female: concept*id = 8532 • Male: concept*id = 8507* • *No Information: concept*id = 44814650 (Vocabulary*id='PCORNet')* • *Unknown: concept*id = 44814653 • Other: concept_id = 44814649
gender*source*concept_id	Yes	Yes	Integer	A foreign key to the gender concept that refers to the code used in the source.	**If there is not a mapping for the source code in the standard vocabulary, use concept_id = 0**
year*of*birth	Yes	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. Please keep all accurate/real dates (No date shifting)
month*of*birth	No	Provide When Available	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. Please keep all accurate/real dates (No date shifting)
day*of*birth	No	Provide When Available	Integer	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. Please keep all accurate/real dates (No date shifting)
		Provide			

birth_date	No	When Available	Date	The birth date	Full date. Please keep all accurate/real dates (No date shifting).
birth_datetime	No	Provide When Available	Datetime	The birth date and time	Do not include timezone. Please keep all accurate/real dates (No date shifting). If there is no time associated with the date assert midnight.
race*concept*id	Yes	Yes	Integer	A foreign key that refers to a standard concept identifier in the Vocabulary for the race of the person.	Details of categorical definitions: • **-American Indian or Alaska Native**: A person having origins in any of the original peoples of North and South America (including Central America), and who maintains tribal affiliation or community attachment. • **-Asian**: A person having origins in any of the original peoples of the Far East, Southeast Asia, or the Indian subcontinent including, for example, Cambodia, China, India, Japan, Korea, Malaysia, Pakistan, the Philippine Islands, Thailand, and Vietnam. • **-Black or African American**: A person having origins in any of the black racial groups of Africa. • **-Native Hawaiian or Other Pacific Islander**: A person having origins in any of the original peoples of Hawaii, Guam, Samoa, or other Pacific Islands. • **-White**: A person having origins in any of the original peoples of Europe, the Middle East, or North Africa. For patients with multiple races (i.e. biracial), race is considered a single concept, meaning there is only one race slot. If there are multiple races in the source system, concatenate all races into one race*source*value (see below) and use concept*id code as 'Multiple Race.'* *Predefined values (valid concept*ids found in CONCEPT table where ((domain*id='Race' and vocabulary*id = 'Race') or (vocabulary*id='PCORNet' and concept*class*id='Undefined') or concept*id in (44814659,44814660)) and invalid*reason is null:* • *American Indian/Alaska Native: concept*id = 8657 • Asian: concept*id = 8515* • *Black or African American: concept*id = 8516 • Native Hawaiian or Other Pacific Islander: concept*id = 8557* • *White: concept*id = 8527 • Multiple Race: concept*id = 44814659 (vocabulary*id='PCORNet') • Refuse to answer: concept*id = 44814660 (vocabulary*id='PCORNet') • No Information: concept*id = 44814650 vocabulary*id='PCORNet') • Unknown: concept*id = 44814653* • *Other: concept*id = 44814649
race*source*concept_id	Yes	Yes	Integer	A foreign key to the race concept that refers to the code used in the source.	**If there is not a mapping for the source code in the standard vocabulary, use concept_id = 0**
ethnicity*concept*id	Yes	Yes	Integer	A foreign key that refers to the standard concept identifier in the Vocabulary for the ethnicity of the person.	For PEDSnet, a person with Hispanic ethnicity is defined as "A person of Cuban, Mexican, Puerto Rican, South or Central American, or other Spanish culture or origin, regardless of race." Please include valid concept ids (consistent with OMOP CDMv5). Predefined value set (valid concept*ids found in CONCEPT table where vocabulary*id ='Ethnicity' or (vocabulary*id=PCORNet and concept*class*id='Undefined) where noted):* • *Hispanic: concept*id = 38003563 • Not Hispanic: concept*id = 38003564* • *No Information: concept*id = 44814650 (vocabulary*id='PCORNet')* • *Unknown: concept*id = 44814653 (vocabulary*id='PCORNet')* • *Other: concept*id = 44814649 (vocabulary_id='PCORNet')
				A foreign key to the	

ethnicity*source*concept_id	Yes	Yes	Integer	ethnicity concept that refers to the code used in the source.	**If there is not a mapping for the source code in the standard vocabulary, use concept_id = 0**
location_id	No	Provide When Available	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	Provide When Available	Integer	Foreign key to the primary care provider the person is seeing in the provider table.	For PEDSnet CDM v3.6.0: Sites will use site-specific logic to determine the best primary care provider and document how that decision was made (e.g., billing provider).
care*site*id	Yes	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	For patients who receive care at multiple care sites, use site-specific logic to select a care site that best represents where the patient obtains the majority of their recent care. If a specific site within the institution cannot be identified, use a care*site*id representing the institution as a whole.
pn*gestational*age	No	Provide When Available	Integer	The post-menstrual age in weeks of the person at birth, if known	Use granularity of age in weeks as is recorded in local EHR.
person*source*value	No	Provide When Available	Varchar	An encrypted key derived from the person identifier in the source data.	Insert a unique pseudo-identifier (random number, encrypted identifier) into the field. Do not insert the actual MRN or PAT*ID from your site. A mapping from the pseudo-identifier for person*source_value in this field to a real patient ID or MRN from the source EHR must be kept at the local site. This mapping is not shared with the data coordinating center. It is used only by the site for re-identification for study recruitment or for data quality review.
gender*source*value	Yes	Yes	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. See gender*concept*id
race*source*value	Yes	Yes	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. For patients with multiple races (i.e. biracial), race is considered a single concept, meaning there is only one race slot. If there are multiple races in the source system, concatenate all races into one source value, and use the concept_id for Multiple Race.
ethnicity*source*value	Yes	Yes	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.
language*concept*id	Yes	Yes	Integer	A foreign key that refers to the standard concept identifier in the Vocabulary for the language of the person.	For PEDSNet, please map your source codes to acceptable language values in appendix 2 **If there is not a mapping for the source code in the network language mapping, use concept_id = 44814649 (Other PCORNet Vocabulary)**
				A foreign key to the language concept	**If there is not a mapping for the source code in the standard**

language*source*concept_id	Yes	Yes	Integer	that refers to the code used in the source.	**vocabulary, use concept_id = 0**
language*source*value	Yes	Yes	Varchar	The source code for the language of the person as it appears in the source data	The person language is mapped to a standard language concept in the Vocabulary and the original code is stored here for reference.

If a field marked as "Provide when available" for the network requirement is not available at your site, please relay this information to the DCC

1.2 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	NOT Null Constraint	Network Requirement	Data Type	Description	PEDSnet Conventions
death*cause*id	Yes	Yes	Integer	A unique identifier for each death cause occurrence	This is not a value found in the EHR. Sites may choose to use a sequential value for this field
person_id	Yes	Yes	Integer	A foreign key identifier to the deceased person. The demographic details of that person are stored in the person table.	See PERSON.person_id (primary key)
death_date	Yes	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. When the date of death is not present in the source data, use the date the source record was created.
death_datetime	Yes	Yes	Datetime	The date the person was deceased.	**This field is custom to PEDSnet** 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. When the date of death is not present in the source data, use the date the source record was created. If there is no time associated with the date assert '23:59:59'.
death*type*concept_id	Yes	Yes	Integer	A foreign key referring to the predefined concept identifier in the Vocabulary reflecting how the death was represented in the source data.	Please include valid concept ids (consistent with OMOP CDMv5). Predefined value set (valid concept*ids found in CONCEPT table where domain*id ='Death Type') select * from concept where concept*class*id ='Death Type' yields 9 valid concept*ids. If none are correct, use concept*id = 0 Note: Most current ETLs are extracting data from EHR. The common concept*id to insert here is* • *38003569 ("EHR record patient status "Deceased")* *. Please assert* • *No information: concept*id = 44814650 where there is no information in the source **Note**: These terms only describe the source from which the death was reported. It does not describe our certainty/source of the date of death, which may have been created by one of the heuristics described in death_date.

cause*concept*id	No	Provide When Available	Integer	A foreign referring to a standard concept identifier in the Vocabulary for conditions.	
cause*source*value	No	Provide When Available	Varchar	The source code for the cause of death as it appears in the source. This code is mapped to a standard concept in the Vocabulary and the original code is stored here for reference.	
cause*source*concept_id	No	Provide When Available	Integer	A foreign key to the vocabulary concept that refers to the code used in the source.	This links to the concept id of the vocabulary of the cause of death concept id as stored in the source. For example, if the cause of death is "Acute myeloid leukemia, without mention of having achieved remission" which has an icd9 code of 205.00 the cause source concept id is 44826430 which is the icd9 code concept that corresponds to the diagnosis 205.00. **If there is not a mapping for the source code in the standard vocabulary, use concept_id = 0**
death*impute*concept_id	Yes	Yes	Varchar	A foreign key referring to a standard concept identifier in the vocabulary for death imputation.	p>Please include valid concept ids (consistent with OMOP CDMv5). Predefined value set (valid concept*ids found in CONCEPT table where concept*class*id ='Death Imput Type')* *select * from concept where (concept*class*id ='Death Imput Type' or (vocabulary*id='PCORNet' and concept*class*id='Undefined')) and invalid*reason is null yields 8 valid concept*ids. If none are correct, use concept*id = 0* • *Both month and day imputed: 2000000034* • *Day imputed: 2000000035* • *Month imputed: 2000000036* • *Full Date imputed: 2000000038* • *Not imputed:2000000037* • *No Information: concept*id = 44814650 (Vocabulary*id='PCORNet')* • *Unknown: concept*id = 44814653 • Other: concept_id = 44814649

If a field marked as "Provide when available" for the network requirement is not available at your site, please relay this information to the DCC

1.2.1 Additional Notes

- Each Person may have more than one record of death in the source data. It is OK to insert multiple death records for an individual.
- If the Death Date cannot be precisely determined from the data, the best approximation should be used.

1.3 LOCATION

The Location domain represents a generic way to capture physical location or address information. Locations are used to define the addresses for Persons and Care Sites. The most important field is ZIP for location-based queries.

Field	NOT Null Constraint	Network Requirement	Data Type	Description	PEDSnet Conventions
location_id	Yes	Yes	Integer	A unique identifier for each geographic location.	This is not a value found in the EHR. Sites may choose to use a sequential value for this field
city	Yes	Provide When Available	Varchar	The city field as it appears in the source data.	
state	No	Provide When Available	Varchar	The state field as it appears in the source data.	
zip	No	Provide When Available	Varchar	The zip code. For US addresses, valid zip codes can be 3, 5 or 9 digits long, depending on the source data.	While optional, this is the most important field in this table to support location-based queries.
location*source*value	No	Provide When Available	Varchar	The verbatim information that is used to uniquely identify the location as it appears in the source data.	If location source values are deemed sensitive by your organization, insert a pseudo-identifier (random number, encrypted identifier) into the field. Sites electing to obfuscate location*source*values will keep the mapping between the value in this field and the original clear text location source value. This value is only used for site-level re-identification for study recruitment and for data quality review. Sites may consider using the location*id field value in this table as the pseudo-identifier as long as a local mapping from location*id to the real site identifier is maintained.
address_1	No	NO	Varchar		Do not transmit to DCC
address_2	No	NO	Varchar		Do not transmit to DCC
county	No	NO	Varchar		Do not transmit to DCC

If a field marked as "Provide when available" for the network requirement is not available at your site, please relay this information to the DCC

1.3 Additional Notes

- Each address or Location is unique and is present only once in the table
- Locations in this table are restricted to locations that are applicable to persons and care_sites in the Pedsnet cohort at each site. When external data is implemented, valid(data containing) locations may be expanded beyond locations of those only present in clinical tables.

1.4 CARE_SITE

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

Field	NOT Null Constraint	Network Requirement	Data Type	Description	PEDSnet Conventions
care*site*id	Yes	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.	This is not a value found in the EHR. Sites may choose to use a sequential value for this field
care*site*name	No	Provide When Available	Varchar	The description of the care site	
					Please include valid concept ids (consistent with OMOP CDMv5.1).

place*of*service*concept*id	No	Provide When Available	Integer	A foreign key that refers to a place of service concept identifier in the Vocabulary	Predefined value set (valid concept*ids found in CONCEPT table where concept*class*id = 'Place of Service' and invalid*reason is null) select * from concept where concept*class*id = 'Place of Service' and invalid*reason is null yields 49 valid concept*ids. Please use the following value set for PEDSnet v3.6: • **Urgent Care Facility = 8782** • Rural Health Clinic = 8761 • Outpatient (Examples: Hospital Dialysis, HOD, Day Hospital, Day Medicine) = 8756 • Office =8940 • Inpatient Psychiatric Facility =8971 • Inpatient Hospital =8717 • Independent Clinic =8716 • Emergency Room - Hospital = 8870 • Other Place of Service =8844 • Other Inpatient Care =8892 • Unknown: concept*id = 44814653* • *Other: concept*id = 44814649 • No information: concept_id = 44814650
location_id	No	Provide When Available	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	Yes	Yes	Varchar	The identifier for the organization in the source data, stored here for reference.	If care site source values are deemed sensitive by your organization, insert a pseudo-identifier (random number, encrypted identifier) into the field. Sites electing to obfuscate care site*source*values will keep the mapping between the value in this field and the original clear text location source value. This value is only used for site-level re-identification for study recruitment and for data quality review. For EPIC EHRs, map care*site*id to Clarity Department. Sites may consider using the care*site*id field value in this table as the pseudo-identifier as long as a local mapping from care*site*id to the real site identifier is maintained.
place*of*service*source*value	No	Provide When Available	Varchar	The source code for the place of service as it appears in the source data, stored here for reference.	
specialty*concept*id	No	Provide When	Integer	The specialty of the department linked to a standard specialty concept as	Care sites could have one or more specialties or a Care site could have no specialty information. **Valid specialty concept ids for PEDSnet are found in the appendix** **Please use the following rules:** • If care site specialty information is unavailable, please follow the convention on reporting values that are unknown,null or unavailable. • If a care site has a single specialty associated with it, sites should link the specialty to the **valid specialty concepts as assigned in the appendix**. If the specialty does not correspond to a value in this listing, please use the NUCC Listing (vocabulary*id='NUCC') provided in the vocabulary as a reference.* • *If there are multiple specialties associated with a particular care site and sites are not able to assign a specialty value on the visit*

		Available		it appears in the Vocabulary	*occurrence level, sites should use the specialty concept id=38004477 "Pediatric Medicine".* • *If there are multiple specialties associated with a particular care site and this information is attainable, sites should document the strategy used to obtain this information and the strategy used to link the correct care site/specialty pair for each visit occurrence. Sites should also link the specialty to the* ***valid specialty concepts as assigned in the appendix*** *If the specialty does not correspond to a value in this listing, please use the NUCC Listing (vocabulary*id='NUCC') provided in the vocabulary as a reference. • If the speciality does not correspond to a value in the NUCC Listing and no value in the ABMS Listing, please use the Specialty listing (vocabulary_id=' Medicare Specialty') as a reference
specialty*source*value	No	Provide When Available	Varchar	The source code for the specialty as it appears in the source data, stored here for reference.	

If a field marked as "Provide when available" for the network requirement is not available at your site, please relay this information to the DCC

1.4.1 Additional Notes

- 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 Place of Service Concepts are based on a catalog maintained by the CMS (see vocabulary for values)

1.5 PROVIDER

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

Field	NOT Null Constraint	Network Requirement	Data Type	Description	PEDSnet Conventions
provider_id	Yes	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.	This is not a value found in the EHR. **SITE RESPONSIBILITY: This field must remain a stable identifier across submissions to the DCC.** A mapping from the provider_id to a real provider from the source EHR must be kept at the local site. This mapping is not shared with the data coordinating center. It is used only by the site for re-identification for study recruitment or for data quality review. Sites should document who they have included as a provider.
provider_name	No	NO	Varchar	A description of the provider	DO NOT TRANSMIT TO DCC
gender*concept*id	No	Provide When Available	Integer	The gender of the provider	A foreign key to the concept that refers to the code used in the source.
specialty*concept*id	No	Provide When Available	Integer	A foreign key to a standard provider's specialty concept	Please map the source data to the mapped provider specialty concept associated with the American Medical Board of Specialties as seen in Appendix A1. Predefined value set (valid concept*ids found in CONCEPT table where vocabulary*id in ('Medicare Specialty', 'ABMS','NUCC','PEDSnet')) select * from concept where vocabulary*id in (' Medicare Specialty', 'ABMS','NUCC','PEDSnet') and invalid*reason is null yields 2200 valid concept*ids.* *If none are correct, use concept*id = 0 For providers with more than one specialty, use site-specific logic to select one

				identifier in the Vocabulary.	specialty and document the logic used. For example, sites may decide to always assert the **first** specialty listed in their data source. As a first guide please use the ABMS and PEDsnet vocabulary specialty listing listing to map your specialtity values. If the specialty does not correspond to a value in these listings, please use the NUCC Listing (vocabulary*id='NUCC') provided in the vocabulary as a reference and the Specialty (vocabulary*id='Medicare Specialty') if no correspond value exists in the NUCC Listing.
care*site*id	Yes	Yes	Integer	A foreign key to the main care site where the provider is practicing.	See CARE*SITE.care*site_id (primary key)
year*of*birth	No	Provide When Available	Integer	The year of birth of the provider	
NPI	No	Site Preference	Varchar	The National Provider Identifier (NPI) of the provider.	
DEA	No	Site Preference	Varchar	The Drug Enforcement Administration (DEA) number of the provider.	
provider*source*value	Yes	Yes	Varchar	The identifier used for the provider in the source data, stored here for reference.	Insert a pseudo-identifier (random number, encrypted identifier) into the field. Do not insert the actual PROVIDER*ID from your site. A mapping from the pseudo-identifier for provider*source*value in this field to a real provider ID from the source EHR must be kept at the local site. This mapping is not shared with the data coordinating center. It is used only by the site for re-identification for study recruitment or for data quality review. *Sites may consider using the provider*id field value in this table as the pseudo-identifier as long as a local mapping from provider_id to the real site identifier is maintained.
specialty*source*value	No	Provide When Available	Varchar	The source code for the provider specialty as it appears in the source data, stored here for reference.	Optional. May be obfuscated if deemed sensitive by local site.
specialty*source*concept_id	No	Provide When Available	Integer	A foreign key to a concept that refers to the code used in the source.	If providing this information, sites should document how they determine the specialty associated with the provider. **Valid specialty concept ids for PEDSnet are found in the appendix** If the specialty does not correspond to a value in this listing, please use the NUCC Listing (vocabulary*id='NUCC') provided in the vocabulary as a reference. ***If there is not a mapping for the source code in the standard vocabulary, use concept*id = 0**
gender*source*value	No	Provide When Available	Varchar	The source value for the provider gender.	
gender*source*concept_id	No	Provide When Available	Integer	The gender of the provider as represented in the source that maps to a	**If there is not a mapping for the source code in the standard vocabulary, use concept_id = 0**

				concept in the vocabulary	

If a field marked as "Provide when available" for the network requirement is not available at your site, please relay this information to the DCC

1.5.1 Additional Notes

- For PEDSnet, a provider is any individual (MD, DO, NP, PA, RN, etc) who is authorized to document care.
- Providers are not duplicated in the table.

1.6 VISIT_OCCURRENCE

The visit occurrence 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.

Exclusions:

1. Future Vists
2. Cancelled Visits (where the patient was not seen)

Note 1: Please use the following logic to assign visit concept ids:

Visit Concept Id	Concept Name	Visit Type Inclusion	In Person	Examples/Logic (includes but is not limited to)
9201	Inpatient Visit (IP)	Visits that resulted in a patient admission	Yes	Hospital Admissions
9202	Ambulatory Visit (AV)/Outpatient	In person Outpatient Visits visits where the patient was seen by a physician	Yes	Office Visits or Appointments
2000000469	Outpatient Non Physician (OP-Non Physician)	In person Outpatient Visits visits where the patient was **NOT** seen by a physician	Yes	Lab Visits, Radiology
9203	Emergency Department Visit (ED)	Emergency Department Visits and Urgent Care	Yes	Emergency Room Visits and Urgent Care
44814711	Other ambulatory Visit (OA)	Outpatient visits where the patient was not seen in person.	No	Telemedicine,Telephone, Emails, Refills and Orders Only Encounters
42898160	Long Term Care Visit	Formal or Informal long term care for chronic illness management	Yes	Site discretion
44814710	Non-Acute Institutional	Non-Acute long term management of care	Yes	Site discretion
2000000048	Emergency Department Admit to Inpatient Hospital Stay	Combination of 9203 and 9201 visits	Yes	Use only if unable to split the ED and inpatient visit.
2000000088	Observation Visit	Please discern what defines an observation visit at your site	Yes	Only map to the observation visit type if the patient leaves the hospital or is discharged from what has been determined to be an observation visit. For sites splitting visits, ED->Observation visits are only to be mapped as Observation Stay Visits. The split in this case is not required.
2000000104	Administrative Visit	Other visits that are in the source system for administrative purposes.	No	Professional Billing or Hospital Abstractions

Field	NOT Null Constraint	Network Requirement	Data Type	Description	PEDSnet Conventions

visit*occurrence*id	Yes	Yes	Integer	A unique identifier for each person's visits or encounter at a healthcare provider.	This is not a value found in the EHR. VISIT*OCCURRENCE*ID must be unique for all patients within a single data set. **SITE RESPONSIBILITY: This field must remain a stable identifier across submissions to the DCC.** A mapping from the visit occurrence id to a real patient encounter from the source EHR must be kept at the local site. This mapping is not shared with the data coordinating center. It is used only by the site for re-identification for study recruitment or for data quality review. Do not use institutional encounter ID.
person_id	Yes	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	Yes	Date	The start date of the visit.	No date shifting. Full date.
visit*end*date	No	Provide When Available	Date	The end date of the visit.	No date shifting. Full date. 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.
visit*start*datetime	Yes	Yes	Datetime	The start date of the visit.	No date shifting. Full date and time. **If there is no time associated with the date assert midnight for the start time**
visit*end*datetime	No	Provide When Available	Datetime	The end date of the visit.	No date shifting. 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. Full date and time. **If there is no time associated with the date assert 11:59:59 pm for the end time**
provider_id	No	Provide When Available	Integer	A foreign key to the provider in the provider table who was associated with the visit.	Use attending or billing provider for this field if available, even if multiple providers were involved in the visit. Otherwise, make site-specific decision on which provider to associate with visits and document. **NOTE: this is NOT required in OMOP CDM v4, but appears in OMOP CDMv5.**
care*site*id	No	Provide When Available	Integer	A foreign key to the care site in the care site table that was visited.	See CARE*SITE.care*site_id (primary key)
visit*concept*id	Yes	Yes	Integer	A foreign key that refers to a place of service concept identifier in the vocabulary.	**In PEDSnet CDM v1, this field was previously called place*of*service*concept*id** Please include valid concept ids (consistent with OMOP CDMv5). Predefined value set (valid concept*ids found in CONCEPT table where (domain*id='Visit' and ~ 'Enc
visit*type*concept_id	Yes	Yes	Integer	A foreign key to the predefined concept identifier in the standard vocabulary reflecting the type of source data from which the visit record is derived.	select * from concept where concept*class*id='Visit Type' yields 3 valid concept*ids.* *If none are correct, user concept*id=0. The majority of visits should be type 'Visit derived from EHR record' which is concept_id=44818518

visit*source*value	No	Provide When Available	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.	
visit*source*concept_id	No	Provide When Available	Integer	A foreign key to a concept that refers to the code used in the source.	If a site is using HCPS or CPT for their visit source value, the standard concept id that maps to the particular vocabulary can be used here. **If there is not a mapping for the source code in the standard vocabulary, use concept_id = 0**
preceding*visit*occurrence_id	No	NO	Integer	A foreign key to the VISIT_OCCURRENCE table record of the visit immediately preceding this visit.	Do not transmit to DCC
admitted*from*concept_id	No	Provide When Available	Integer	A foreign key to the predefined concept in the Place of Service Vocabulary reflecting the admitting source for a visit.	Please use the following valid concept id set for Admitting source: • Adult Foster Home=44814670 • Assisted Living Facility=44814671 • Ambulatory Visit=44814672 • Emergency Department=8870= • Home Health=44814674 • Home / Self Care=44814675 • Hospice=8546 • Other Acute Inpatient Hospital=38004279 • Nursing Home (Includes ICF)=44814678 • Rehabilitation Facility=44814679 • Residential Facility=44814680 • Skilled Nursing Facility=8863 • No information=44814650 • Unknown=44814653 • Other=44814649 This should be populated for inpatient encounters in the source but may vary for emergency department (ED) visits and outpatient encounters (AV,OA).
discharge*to*concept_id	No	Provide When Available	Integer	A foreign key to the predefined concept in the Place of Service Vocabulary reflecting the discharge disposition (destination) for a visit.	Please use the following valid concept id set for Discharge Destination: • Adult Foster Home=38004205 • Assisted Living Facility=38004301 • Against Medical Advice=4021968 • Absent without leave=44814693 • Expired=4216643 • Home Health=38004195 • Home / Self Care=8536 • Hospice=8546 • Other Acute Inpatient Hospital=38004279 • Nursing Home (Includes ICF)=8676 • Rehabilitation Facility=8920 • Residential Facility=44814701 • Still In Hospital=8717 • Skilled Nursing Facility=8863 • No information=44814650 • Unknown=44814653 • Other=44814649 This should be populated for inpatient encounters in the source but may vary for emergency department (ED) visits and outpatient encounters (AV,OA).
		Provide		The source code for the admitting	This should be populated for inpatient encounters in the

admitted*from*source_value	No	When Available	Varchar	source as it appears in the source data.	source but may vary for emergency department (ED) visits and outpatient encounters (AV,OA).
discharge*to*source_value	No	Provide When Available	Varchar	The source code for the discharge disposition as it appears in the source data.	This should be populated for inpatient encounters in the source but may vary for emergency department (ED) visits and outpatient encounters (AV,OA).

**If a field marked as "Provide when available" for the network requirement is not available at your site, please relay this information to the DCC

1.6.1 Additional Notes

- The 1/1/2009 date limitation that is used to define a PEDSnet active patient is **NOT** applied to visit_occurrence. All visits, of all types (physical and virtual) are included for an active patient.
- A Visit Occurrence is recorded for each visit to a healthcare facility.
- **If a visit includes moving between different visit*concepts (ED -> inpatient) sites may opt to split the record into separate visit*occurrence records**.

To show the relationship of the split (ED -> inpatient) encounter, use the FACT_RELATIONSHIP table.

An example of this is below:

VISIT_OCCURRENCE

visit*occurrence*id	person_id	visit*start*date	visit*end*date	provider_id	care*site*id	place*of*service*concept*id	place*of*service*source*value
35022489	209846	2011-11-14 17:36:00-05	2011-11-14 22:25:00-05	2238	322	9203	Emergency
35022490	209846	2011-11-14 22:25:00-05	2011-11-15 16:33:00-05	2238	43	9201	Emergency

FACT_RELATIONSHIP

Domain*concept*id_1	fact*id*1	Domain*concept*id_2	fact*id*2	relationship*concept*id
Visit	35022489	Visit	35022490	Occurs before
Visit	35022490	Visit	35022489	Occurs after

Because the domain*concept*id and relationship*concept*id are actually numeric values the following is an example of how the table is stored:

Domain*concept*id_1	fact*id*1	Domain*concept*id_2	fact*id*2	relationship*concept*id
8	35022489	8	35022490	44818881
8	35022490	8	35022489	44818783

- Operating and Anesthesia encounters that occur as apart of the Inpatient stay should be rolled up into one Inpatient encounter.
- Each Visit is standardized by assigning a corresponding Concept Identifier 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 attending or billing providers (e.g. billing, attending, etc), 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.

1.7 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 or ICD-10-CM codes, but may not have a way to capture out-of-system conditions.
- EHRs may also capture External Injury codes in different place in the source system. These types of codes are also to be included.

For the PEDSNet network, please provide **clinical physician based diagnosis** as opposed to billing or claim based diangosis data.

Note 1: For the PEDSNet network, we are coding all diagnosis codes to the SNOMED-CT Vocabulary. Research has showed that the IMO to SNOMED native mapping and IMO to ICD

to SNOMED OMOP mapping produces highly variable results. For a particular IMO Code, when comparing the two mapping options, the same SNOMED concept id is only produced 25% of the time. See below examples of the mapping differences (IMO-SNOMED, ICD10 and ICD9):

IMO Description	Direct SNOMED	Via ICD
Numbness of Toes	Numbness of toe	Altered Sensation of Skin
Cerebellar ataxia/dyskinesia	Cerebellar Disorder	Cerebellar Ataxia
Choking episode	Choking sensation	Finding of head and neck region
Intestional malrotation	Congenital malrotation of intestine	Congenital anomaly of fixation of intestine
Genetic disease carrier status testing	Genetic finding	Genetic disorder carrier
Duchenne muscular dystrophy	Duchenne muscular dystrophy	Hereditary progressive muscular dystrophy

For diagnosis codes, please provide the IMO to SNOMED mapping where it exists in the source system.

If the IMO to SNOMED mapping is not available in the system, utilize the IMO to ICD to SNOMED OMOP mapping in the vocabulary.

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

Note 2: For the PEDSNet network, please provide clinical physician based diagnosis as opposed to billing or claim based diangosis data. The clinical physician based diagnosis corresponds to the "Order origin" concept ids for `condition_type_concept_id` . If you are providing billing or claim diagnosis data, please use the "Billing" or "Claim" concept_ids for `condition_type_concept_id` .

Use the following logic to determine the correct `condition_type_concept_id` as it pertains to the visit the diagnosis stems from:

Visit*concept*id	Condition*type*concept_id
9201 (Inpatient)	Inpatient header
9202 (Outpatient)	Outpatient header
9203 (Emergency)	Emergency header
2000000048 (ED to Inpatient)	Inpatient header
2000000088 (Observation)	Inpatient header

Note 3: We have been made aware that there are a significant amount of conditions that route to a domain of Procedure, Measurement etc. Please **DO NOT** route these conditions to those domains or tables (i.e. Procedure*Occurrence, Measurement). Instead, include all records coming out of our source tables for diagnosis data in the Condition*Occurrence table.

Field	NOT Null Constraint	Network Requirement	Data Type	Description	PEDSnet Conventions
condition*occurrence*id	Yes	Yes	Integer	A unique identifier for each condition occurrence event.	This is not a value found in the EHR. Sites may choose to use a sequential value for this field
person_id	Yes	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	Yes	Integer	A foreign key that refers to a standard condition concept identifier in the Vocabulary.	Please include valid concept ids (consistent with OMOP CDMv5). Predefined value set (valid concept*ids found in CONCEPT table where vocabulary*id ='SNOMED') select * from concept where vocabulary*id ='SNOMED' yields ~440,000 valid concept*ids.

					If none are correct, use concept_id = 0
condition*start*date	Yes	Yes	Date	The date when the instance of the condition is recorded.	No date shifting.
condition*end*date	No	Provide When Available	Date	The date when the instance of the condition is considered to have ended	No date shifting. If this information is not available, set to NULL.
condition*start*datetime	Yes	Yes	Datetime	The date and time when the instance of the condition is recorded.	No date shifting. Full date and time. **If there is no time associated with the date assert midnight for the start time**
condition*end*datetime	No	Provide When Available	Datetime	The date and time when the instance of the condition is considered to have ended	No date shifting. If this information is not available, set to NULL. Full date and time. **If there is no time associated with the date assert 11:59:59 pm for the end time**
condition*type*concept_id	Yes	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.	Please include valid concept ids (consistent with OMOP CDMv5). Predefined value set (valid concept*ids found in CONCEPT table where concept*class*id ='Condition Type' and vocabulary*id='PEDSnet') select * from concept where concept*class*id ='Condition Type' and vocabulary*id='PEDSnet' yields 21 valid concept*ids. If none are correct, use concept_id = 0 **For the primary diagnosis for the inpatient, outpatient or emergency setting (may be identified as Dx#1 in a source system), Please use concepts the following concepts:** • Outpatient header - 1st position - Order Origin=2000000095 • Outpatient header - 1st position - Billing Origin=2000000096 • Outpatient header - 1st position - Claim Origin=2000000097 • Inpatient header - primary - Order Origin=2000000092 • Inpatient header - primary - Billing Origin =2000000093 • Inpatient header - primary - Claim Origin= 2000000094 • Emergency Header - 1st Position - Order Origin=2000001280 • Emergency Header - 1st Position - Claim Origin=2000001281 • Emergency Header - 1st Position - Billing Origin=2000001282 **All other diagnosis that is not the primary (or Dx#1) in the inpatient, outpatient or emergency setting should correspond to the following concept ids:** • Inpatient header - 2nd position - Order Origin=2000000098 • Inpatient header - 2nd position - Billing Origin = 2000000099 • Inpatient header - 2nd position - Claim Origin = 2000000100 • Outpatient header - 2nd position - Order Origin=2000000101 • Outpatient header - 2nd position - Billing Origin

					as a reference
specialty*source*value	No	Provide When Available	Varchar	The source code for the specialty as it appears in the source data, stored here for reference.	

If a field marked as "Provide when available" for the network requirement is not available at your site, please relay this information to the DCC

1.4.1 Additional Notes

- 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 Place of Service Concepts are based on a catalog maintained by the CMS (see vocabulary for values)

1.5 PROVIDER

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

Field	NOT Null Constraint	Network Requirement	Data Type	Description	PEDSnet Conventions
provider_id	Yes	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.	This is not a value found in the EHR. **SITE RESPONSIBILITY: This field must remain a stable identifier across submissions to the DCC.** A mapping from the provider_id to a real provider from the source EHR must be kept at the local site. This mapping is not shared with the data coordinating center. It is used only by the site for re-identification for study recruitment or for data quality review. Sites should document who they have included as a provider.
provider_name	No	NO	Varchar	A description of the provider	DO NOT TRANSMIT TO DCC
gender*concept*id	No	Provide When Available	Integer	The gender of the provider	A foreign key to the concept that refers to the code used in the source.
					Please map the source data to the mapped provider specialty concept associated with the American Medical Board of