**OMOP Common Data Model**

**(CDM) Specifications**

**Release 2.0**

**September 28, 2009**

License

© 2009 Foundation for the National Institutes of Health (FNIH).

Licensed under the Apache License, Version 2.0 (the "License"); you may not use this document except in compliance with the License. You may obtain a copy of the License at <http://omop.fnih.org/publiclicense>.

Unless required by applicable law or agreed to in writing, documentation and software distributed under the License is distributed on an "AS IS" BASIS, WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied. Any redistributions of this work or any derivative work or modification based on this work should be accompanied by the following source attribution: "This work is based on work by the Observational Medical Outcomes Partnership (OMOP) and used under license from the FNIH at <http://omop.fnih.org/publiclicense>.

Any scientific publication that is based on this work should include a reference to <http://omop.fnih.org>.

Table of Contents

[1.0 Background 5](#_Toc242066353)

[1.1 Problem Description 5](#_Toc242066354)

[1.2 The Role of the Common Data Model 5](#_Toc242066355)

[1.3 Design Principles 6](#_Toc242066356)

[2.0 Glossary of Terms 8](#_Toc242066357)

[3.0 Conceptual Data Model 15](#_Toc242066358)

[3.1 CDM Core Module 15](#_Toc242066359)

[3.2 Dictionary 18](#_Toc242066360)

[3.2.1 SNOMED-CT and RxNorm 18](#_Toc242066361)

[4.0 CDM Logical Data Model 19](#_Toc242066362)

[4.1 Logical Entity-Relational Diagram 19](#_Toc242066363)

[4.2 Logical Entities and Attributes – CDM Core Module 21](#_Toc242066364)

[4.2.1 PERSON 21](#_Toc242066365)

[4.2.2 DRUG\_EXPOSURE 24](#_Toc242066366)

[4.2.3 DRUG\_ERA 27](#_Toc242066367)

[4.2.4 DRUG\_EXPOSURE\_REF 30](#_Toc242066368)

[4.2.5 CONDITION\_OCCURRENCE 31](#_Toc242066369)

[4.2.6 CONDITION\_ERA 34](#_Toc242066370)

[4.2.7 CONDITION\_OCCURRENCE\_REF 38](#_Toc242066371)

[4.2.8 VISIT\_OCCURRENCE 39](#_Toc242066372)

[4.2.9 PROCEDURE\_OCCURRENCE 41](#_Toc242066373)

[4.2.10 PROC\_OCCURRENCE\_REF 43](#_Toc242066374)

[4.2.11 OBSERVATION 43](#_Toc242066375)

[4.2.12 OBSERVATION\_TYPE\_REF 46](#_Toc242066376)

[4.2.13 OBSERVATION\_PERIOD 47](#_Toc242066377)

[4.3 Logical Entities and Attributes – Dictionary 50](#_Toc242066378)

[4.3.1 Dictionary Logical Data Model 50](#_Toc242066379)

[4.3.2 CONCEPT 51](#_Toc242066380)

[4.3.3 CONCEPT\_SYNONYM 53](#_Toc242066381)

[4.3.4 CONCEPT\_RELATIONSHIP 55](#_Toc242066382)

[4.3.5 CONCEPT\_ANCESTOR 57](#_Toc242066383)

[4.3.6 VOCABULARY\_REF 61](#_Toc242066384)

[4.3.7 SOURCE\_TO\_CONCEPT\_MAP 62](#_Toc242066385)

[4.3.8 RELATIONSHIP\_TYPE 63](#_Toc242066386)

[Appendix A: SNOMED-CT Conceptual Data Model 65](#_Toc242066387)

[Appendix B: RxNorm Conceptual Data Model 66](#_Toc242066388)

[Appendix C: Drug Exposure Type Codes 67](#_Toc242066389)

[Appendix D: Condition Occurrence Type Codes 68](#_Toc242066390)

[Appendix E: Procedure Occurrence Type Codes 71](#_Toc242066391)

[Appendix F: Relationship Types Codes 73](#_Toc242066392)

**Document Control**

**Authors and Contributors**

| Name | Organization | Title |
| --- | --- | --- |
| Patrick Ryan | OMOP | Co-Investigator |
| Don Griffin | CSC | Senior Data Architect |
| Christian Reich | OMOP | Program Manager, IT Implementation |
| Shesh Mudiyanur | CSC | Data Architect |
| Cortney Hayflinger | OMOP | Programmer |
| Aurelia Ford | CSC | Informaticist |
| Mark Khayter | Ephir | Documentation Specialist |

**Reviewers**

| Name | Role | Title | Date Reviewed |
| --- | --- | --- | --- |
| Peter Parsons | Technical Writer | Principal | 02-Aug-2009 |
| Aurelia Ford | CSC | Informaticist | 02-Aug-2009 |
| Vicki Fan | CSC | Business Analyst | 25-Aug-2009 |

**Document References**

| Document Title | Type of Reference | Document Location |
| --- | --- | --- |
| OMOP CDM ETL Mapping Specification | Business Rules | OMOP Basecamp |
| OMOP CDM ETL Technical Specification | Detail Technical Information | OMOP Basecamp |

**Change Record**

|  |  |  |  |
| --- | --- | --- | --- |
| Date | Author | Version | Change Reference |
| 02-Aug-2009 | Shesh Mudiyanur | 1.0 | New document, describes OMOP  Common Data Model Specification. |
| 23-Sep-2009 | Sanjay Dharmadhikari  Mark Khayter | 2.0 | Modified to reflect renamed Column names  Added License page  Modified data examples |

# Background

The Observational Medical Outcomes Partnership (OMOP) is a public-private partnership designed to protect human health by improving the monitoring of drugs for safety and effectiveness. The partnership, which began in the fourth quarter of 2008, is conducting a two-year research initiative to determine the contribution and utility of using existing healthcare databases to identify and evaluate safety issues associated with drugs that are already on the market.

OMOP is funded and managed through the Foundation for the National Institutes of Health, and draws on the expertise and resources of the pharmaceutical industry, academic institutions, non-profit organizations, the Food and Drug Administration (FDA), and other federal agencies. In addition to sponsoring specific research efforts, OMOP is creating a set of tools—such as data models, experimental protocols, and database evaluation tools—that will be placed in the public domain to encourage research by a broad community of scientific investigators. All project results will be made public in accordance with the public health mission of the partnership. These will include comprehensive reports on scientific and technical findings, lessons learned, and peer-reviewed articles on the experimental findings by OMOP’s sponsored investigators.

This document describes the design of—and the rationale behind—one of the aforementioned tools, the OMOP Common Data Model (CDM). The remainder of this chapter describes the CDM and its place in the larger OMOP toolset. Subsequent chapters of this document describe how the OMOP project team designed the CDM and how OMOP researchers will use the CDM to develop and evaluate new, data-driven research methods for drug safety surveillance.

This document consists of two primary sections: the Conceptual Data Model, and the Logical Data Model. The Conceptual Data Model section describes the overall developmental approach taken. The Logical Data Model contains two components; the Logical Entities and Attributes of the CDM Core Module, containing the actual data from the source systems, and the Logical Entities and Attributes of the Dictionary.

## Problem Description

One of OMOP’s goals is to define processes that can be used to assess the feasibility and utility of using observational data to identify and evaluate associations between drugs and health-related conditions. To facilitate its methodological research, the Partnership will evaluate the performance of various analytical methods for identifying drug-outcome associations across multiple disparate observational data sources (i.e., administrative claims data and Electronic Health Records). OMOP will partner with a number of different organizations with observational data to undertake this research, including licensing data that can be housed centralized in the OMOP Research Core and collaborating with data providers as a distributed network.

To facilitate this research, OMOP needs to develop a common structure and framework for organizing and standardizing observational data. Such is the role of the Common Data Model in the OMOP pilot infrastructure.

## The Role of the Common Data Model

The Common Data Model, combined with a method for standardizing its content (via the Dictionary, described below) will ensure that research methods can be systematically applied to produce meaningfully comparable results.

No single observational data source is likely to be sufficient to meet all expected drug safety analysis needs, so there is interest in assessing the feasibility and utility of analyzing multiple data sources concurrently. The CDM, however, is not intended to be an integration point for multiple source data sets; rather, OMOP researchers will create a separate CDM instance for each source data set. Analysis results from disparate sources can be brought together to facilitate comparisons and synthesis of the aggregated findings.

All analysis methods and code (e.g., SAS, SQL, or R programs) used to execute OMOP research protocols will be developed for the Common Data Model, with the express purpose of enabling a common set of procedures to be applied to (i.e., to be “portable” across) each participating data source. OMOP intends to test the feasibility of both distributed and centralized network architectures to enable analyses across disparate observational data sources. All participating data sources will be transformed into the Common Data Model structure and Dictionary standards, regardless of where the data reside either logically (e.g., in multiple databases) or physically (e.g., in multiple geographies).

## Design Principles

The OMOP Common Data Model intends to facilitate observational analyses of disparate healthcare databases, including, but not necessarily limited to, administrative claims and Electronic Health Records (EHRs). Observational research will be conducted to identify and evaluate associations between Drug Exposure and Condition Occurrence. Specific Health Outcomes of Interest (HOIs) may be defined by clinical events (e.g., diagnoses, observations, procedures, etc.) in predefined temporal relationships.

The CDM must include all observational data elements that are relevant to identifying Drug Exposures and defining Condition Occurrences. However, the model does not necessarily need to provide a mechanism for archiving all observational data elements. For example, cost information, which is a major component of administrative claims data but, which may not play a prominent role in identifying associations between Drug Exposures and conditions, may not have a place in the CDM.

The CDM design documented herein is guided by six design principles.

**Design Principle 1: The OMOP Common Data Model must accommodate all observational data elements that the partnership wishes to collect, including, but not necessarily limited to, those data elements relevant to identifying Drug Exposures, Condition Occurrences, and other clinical observations.**

**Design Principle 2: In designing the CDM, OMOP should not “reinvent the wheel.” The CDM design should leverage, where reasonable and appropriate, the learning inherent in industry-leading data modeling efforts, such as those associated with the HL7 RIM, the HIMSS EHR Definitional Model, the i2b2 Hive framework, the HMORN Virtual Data Warehouse, and others.**

**Design Principle 3: The CDM design must allow each datum, when applicable, to be related (i.e., mapped) to the appropriate corresponding standard healthcare Concept in the Dictionary.**

**Design Principle 4: The CDM design should anticipate the existence of an ideal Dictionary that maps each source datum to one and only one standard health care Concept. However, the CDM design should remain valid if the mapping of a source datum to multiple standard healthcare concepts should be required.**

CDM Design Principle 3 implies the existence of a Dictionary that pre-coordinates to-be-standardized values from source data sets to standard healthcare Concepts. Ideally, each unique CDM datum that must be standardized to the Dictionary will have its best match to exactly one of the Dictionary’s standard healthcare Concepts. Therefore, ideally, there will be a many-to-one relationship between the to-be-standardized data elements in the CDM and the standard healthcare Concepts in the Dictionary.

Consider the example in which source data set A indicates that person B, in the context of hospital visit C, had a diagnosis represented by ICD-9-CM diagnosis code 410.01, which means “Acute Myocardial Infarction, Anterolateral Wall, Initial Episode of Care.” Source data set A represents the diagnosis “Acute Myocardial Infarction, Anterolateral Wall, Initial Episode of Care” using ICD-9-CM diagnosis code 410.01. However, another source data set may represent this same diagnosis in a completely different way (e.g., using a different coding system, as a text description, etc.). Regardless, the Dictionary will contain a single standard Concept, having Concept code 321318, which means “Acute Myocardial Infarction, Anterolateral Wall, Initial Episode of Care”. Furthermore, the Dictionary will map to this Concept (i.e., standardize) all of the various source-specific representations of this diagnosis, including ICD-9-CM diagnosis code 410.01 from data source A. Queries against a CDM instance will use the standard Concept code for this diagnosis rather than its source-specific representation to ensure selection of all persons with this diagnosis regardless of how the data were originally represented in any source data set. That is, by standardizing the query to the Dictionary, it is ensured that the query will be portable to any CDM instance that has also been standardized to the Dictionary.

A Dictionary that is rich in semantic and ontological information is envisioned. For example, in addition to containing standard healthcare Concepts related to Drug Exposures and Condition Occurrences, the OMOP Dictionary will comprehensively address the relationships among those Concepts, such as parent-child (i.e., class-subclass) relationships, composite-component relationships, and so forth.

**Design Principle 5: The CDM design should discourage the use of Protected Health Information (PHI) except where necessary to conduct analyses to protect the public health.**

Observational analyses should be able to be supported by a CDM that minimizes the use of PHI. Such protections would ensure analysis results can inform public health interests without jeopardizing person privacy. For this reason, CDM tables that correspond to identifiable entities (e.g., Person) should not include columns for HIPAA-recognized identifiers, such as names, person identification numbers, addresses, telephone numbers, and dates of birth. Only those data elements required to facilitate analysis of drug safety issues should be captured in the CDM, including visit dates, prescription details, and enrollment information. Year of birth can be used as a minimally sufficient surrogate to measure age which, it is acknowledged, may limit the utility of the model for studying drug effects in infants.

**Design Principle 6: The CDM design, and the databases that instantiate it, must be usable. Of primary importance is the ability of the CDM design to provide a user with the data that he requires for his research. Of secondary importance is the ability of the CDM design to provide a user with data in the manner (i.e., format) that he prefers.**

The CDM design must ultimately be intuitive, not overly complex, and otherwise “researcher-friendly.” Researchers who find it difficult to understand the CDM design will find it difficult to formulate an accurate and efficient query against a CDM instance. And since CDM queries are the starting point for many data-driven research methods, an unwieldy and unintuitive Common Data Model design will effectively undermine the OMOP mission.

**Design Principle 7: Extensibility. Enable an unlimited number of Persons to be included in any CDM instance. Also, for each Person in the source data set, enable an unlimited number of types of Drug Exposures, Conditions, Observations, Procedures, and Visits to be included in the CDM instance. Accommodate all conceivable values of every conceivable attribute of those Drug Exposures, Conditions, Observations, and Procedures.**

**Design Principle 8: Flexibility. Because a CDM instance is populated based on source-to-target data mappings and other metadata rather than “hard-wired” between predefined source and target column pairs, the CDM is flexible on the allowed types and values of Drug Exposures, Conditions, Observations, and Procedures. For example, the Common Data Model described herein can evolve over time to accommodate different types of Observations without necessitating CDM design changes.**

**Design Principle 9: Reference Concept Codes. The CDM must be able to reference the standard Concepts in the Dictionary both for types and values of Drug Exposures, Conditions, Observations, and Procedures whenever available. It is anticipated that the Dictionary will contain standard Concept Codes for all drugs, medical conditions, clinical observations, and medical procedures in the source data sets, as well as standard Concept Codes for many of the values of the attributes of these entities.**

**Design Principle 10: Data Access. Each database that instantiates the CDM will be deployed as a relational database accessible from any SQL interface.**

**Design Principle 11: Column Order. Columns order as defined in the Logical Entities might not reflect order of the columns in the Physical Entities. All select statements must reference column names and not a wildcard (\*).**

# Glossary of Terms

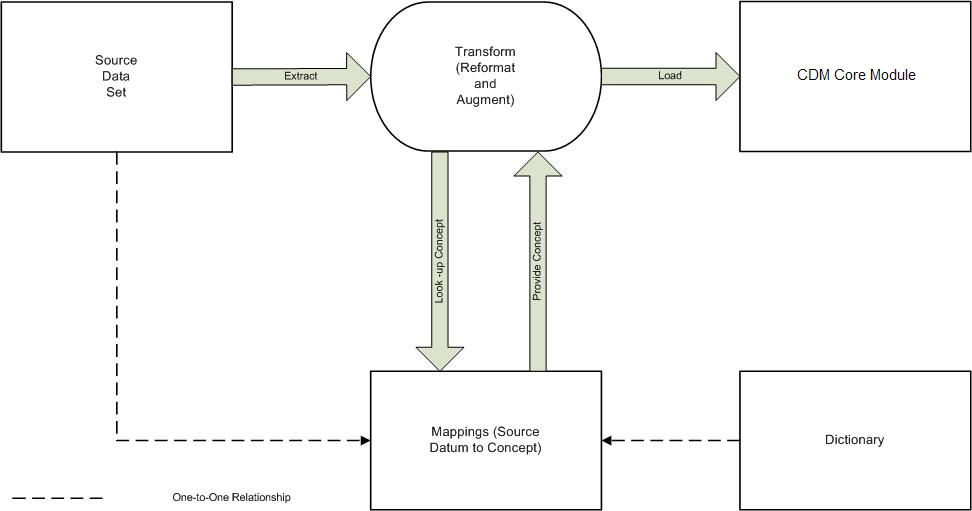
| **Term** | **Abbr.** | **Description** |
| --- | --- | --- |
| **Common Data Model** | **CDM** | The CDM intends to facilitate observational analyses of disparate healthcare databases. The CDM defines table structures for each of the data entities (e.g., Persons, Visit Occurrence, Drug Exposure, Condition Occurrence, Observation, Procedure-Occurrence, etc.). It includes all observational data elements that are relevant to identifying drug exposures and defining condition occurrence. The CDM includes both the dictionary of terms and the entity domain tables. |
| **Concept** |  | A concept is the basic unit of information. Concepts may be grouped into a given domain. A concept is a unique term that has a unique and static identifier/name, belongs to a Namespace, and may exist in relation to other concepts. The vertical relationships consist of "is a" statements that form a logical hierarchy. In general, concepts above a given concept are referred to as ancestors and those below as descendents. |
| **Conceptual Data Model** |  | A Conceptual Data Model is a map of concepts and their relationships. This describes the semantics of an organization and represents a series of assertions about its nature. Specifically, it describes the things of significance to an organization (entity classes), about which it is inclined to collect information, and characteristics of (attributes) and associations between pairs of those things of significance (relationships). |
| **Condition** |  | A condition is a disease, such as a heart condition, as in Medical condition**.** |
| **Condition Era (entity)** |  | A Condition Era entity consist of individual records of a Condition Occurrences that serve as indicators for the presence of a Person’s Condition, and are stored in the Condition\_Era table. Combining individual Condition Occurrences into a single Condition Era serves two purposes:   * Aggregation of chronic conditions that require continuous ongoing care that refers to the same underlying illness. * Aggregation of multiple, closely timed events whether either condition is chronic or acute |
| **Condition Occurrence** **(entity)** |  | Condition Occurrences record individual instances of a Person’s Conditions (i.e., diagnoses) extracted from source data. Conditions are recorded in various data sources in different forms with varying levels of standardization, and are stored in the Condition\_Occurrence table. |
| **Current Procedural Terminology, 4th edition** | **CPT-4** | A terminology that is maintained by the American Medical Association (AMA). It is used by hospitals for Medicare hospital outpatient and by physician for outpatient services. |
| **Data mapping** |  | It is the data element mappings between two distinct data models, terminologies, or concepts. Data mapping is the process of creating data element mappings between two distinct data models. Data mapping is used as a first step for a wide variety of data integration tasks. |
| **Demographics** |  | Demographics refer to selected population characteristics. Demographics may include data such as race, age, sex, date of birth, location, etc. |
| **Design Principle** |  | An organized arrangement of one or more elements or principles for a purpose. It identifies core principles and best practices to assist developers to produce software. Thoroughly understanding the goals of stakeholders and designing systems with those goals in mind are the best approaches to successfully deliver results. |
| **Dictionary** |  | A computerized list (as of items of data or words) used for reference (as for information retrieval or word processing). |
| **Domain** |  | A *data domain* refers to all the unique values which a data element may contain. For example, a database table that has information about people, with one record per person, might have a "gender" column. This gender column might be declared as a string data type, and allowed to have one of two known code values: "M" for male, "F" for female -- and NULL for records where gender is unknown or not applicable (or arguably "U" for unknown as a sentinel value). The data domain for the gender column is: "M", "F".  In database technology, domain refers to the description of an attribute's allowed values. The physical description is a set of values the attribute can have, and the semantic, or logical, description is the meaning of the attribute. |
| **Drug** |  | In pharmacology, a drug as "a chemical substance used in the treatment, cure, prevention, or diagnosis of disease or used to otherwise enhance physical or mental well-being." Drugs may be prescribed for a limited duration, or on a regular basis for chronic disorders. |
| **Drug Exposure** **(entity)** |  | A Drug Exposure entity contains individual records that suggest drug utilization by the person. Drug Exposure indicators store key information about each person medication and the timing thereof, including the drug (captured as standard Concept code in the CDM), quantity, beginning date of medication, number of days supply, period of exposure, and prescription refill data. Drug Exposures are stored in the Drug\_Exposure table. |
| **Drug Era** **(entity)** |  | A Drug Era entity consist of Drug Exposures that are recorded in successive periods (e.g. drug prescription and refill claims) and are combined to form one continuous period of exposure to a drug Concept. This is based on certain rules that consider Drug Exposure dates and persistence windows. Drug Eras are stored in the Drug\_Era table. |
| **Electronic Health Record** | **EHR** | EHR refers to an individual person's medical record in digital format. It may be made up of electronic medical records (EMRs) from many locations and/or sources. The Electronic Health Record (EHR) is a longitudinal electronic record of person health information generated by one or more encounters in any care delivery setting. Included in this information are person demographics, progress notes, problems, medications, vital signs, past medical history, immunizations, laboratory data and radiology reports. The EHR automates and streamlines the clinician's workflow. The EHR has the ability to generate a complete record of a clinical person encounter - as well as supporting other care-related activities directly or indirectly via interface - including evidence-based decision support, quality management, and outcomes reporting. |
| **Electronic Medical Record** | **EMR** | An electronic medical record (EMR) is usually a computerized legal medical record created in an organization that delivers care, such as a hospital and doctor's surgery. Electronic medical records tend to be a part of a local stand-alone health information system that allows storage, retrieval and manipulation of records. Electronic Medical Data: EMR is a subset of EHR. This document will reference EHR moving forward even if specific data source might internally use Electronic Medical Record (EMR) definition. |
| **Entity** |  | An entity is any distinguishable object or concept that is to be represented in the database. Entities are represented by a unique identifying code (the KEY) stored as a single record in a table, called the entity's master record. |
| **Extract Transform Load** | **ETL** | Process of getting data out of one data store (Extract), modifying it (Transform), and inserting it into a different data store (Load). |
| **Generic Product Information** | **GPI** | A proprietary unique identifier for a drug used by the commercial MediSpan formulary database. |
| **Healthcare Common Procedure Coding System** | **HCPCS** | HCPCS Level I codes are managed by the AMA (licensing fees apply). The HCPCS Level II codes are managed by CMS (Centers for Medicare & Medicaid Services). The Level II codes includes: alphanumeric HCPCS procedure and modifier codes, their long and short descriptions, and applicable Medicare administrative, coverage, and pricing data. These codes are used for Medicare outpatient services. |
| **Health Insurance claims** |  | An insurance claim is the actual application for benefits provided by an insurance company. Policyholders must first file an insurance claim before any money can be disbursed. Computerized health insurance claims databases are maintained largely for billing and administrative purposes. Unlike studies with primary data collection, claims data are not collected to meet specific research objectives. Nevertheless, these databases are useful for describing health care utilization, patterns of care, disease prevalence, drug and disease outcomes, and cost of care. |
| **Health Insurance Portability and Accountability Act** | **HIPAA** | A federal law that was designed to allow portability of health insurance between jobs. In addition, it required the creation of a federal law to protect personally identifiable health information; if that did not occur by a specific date (which it did not), HIPAA directed the Department of Health and Human Services (DHHS) to issue federal regulations with the same purpose. DHHS has issued HIPAA privacy regulations (the HIPAA Privacy Rule) as well as other regulations under HIPAA. |
| **Health Level Seven** | **HL7** | HL7 is an all-volunteer, not-for-profit organization involved in development of international healthcare standards. HL7 specifications primarily draw upon codes and vocabularies from a variety of sources. The standard contains some standardized values sets for data interoperability (e.g., gender). HIPSP (Health Information and Standards Panel) has recommended the use of HL7 demographic concepts for use in EHR interoperability. |
| **Health Outcomes of Interest** | **HOIs** | May be defined by clinical events (e.g., drugs, conditions, observations, procedures, etc.) in predefined temporal relationships. |
| **International Classification of Disease, 9th Revision, Clinical Modifications** | **ICD-9-CM** | The official system of assigning codes to diagnoses and procedures associated with hospital utilization in the United States. |
| **Logical Data Model** |  | Logical data models are graphical representation of the business requirements. They describe the things of importance to an organization and how they relate to one another, as well as business definitions and examples. The logical data model can be validated and approved by a business representative, and can be the basis of physical database design. |
| **Logical Observation Identifiers Names and Codes** | **LOINC** | Universal code names and identifiers to medical terminology related to the Electronic Health Record and assists in the electronic exchange and gathering of clinical results (such as laboratory tests, clinical observations, outcomes management and research). |
| **Medical Dictionary for Regulatory Activities** | **MedDRA** | MedDRA is a clinically-validated international medical terminology used by regulatory authorities and the regulated biopharmaceutical industry. The terminology is used through the entire regulatory process, from pre-marketing to post-marketing, and for data entry, retrieval, evaluation, and presentation. |
| **Metadata** |  | Metadata is "data about other data". For example, metadata would document data about data elements or attributes, (name, size, data type, etc.) and data about records or data structures (length, fields, columns, etc.) and data about data (where it is located, how it is associated, ownership, etc.). Metadata may include descriptive information about the context, quality and condition, or characteristics of the data. It may be recorded with high or low granularity. |
| **National Drug Codes** | **NDC** | Unique identifiers assigned to individual drugs. NDCs are used primarily as an inventory code and for prescriptions |
| **National Drug File Reference Terminology** | **NDF-RT** | A nonproprietary drug reference terminology that includes drug knowledge and classifies drugs, most notably by mechanism of action and physiologic effect. |
| **Observation** |  | An observation represents a conclusion reached after examination or investigation (i.e., something that has been found). It may be delivered as a statement or document containing an authoritative decision or conclusion. |
| **Observation (entity)** |  | The Observation entity contains all general observations that are tracked as attributes, including source Observation code, matching standard Concept Code, date of the Observation, type of Observation, type of result, number/text/Concept code, and reference range for numeric results. Observation entities are recorded in the Observation table. |
| **Observation Period (entity)** |  | An Observation Period records the periods of time in which a Person may have data recorded. In claims data, this is commonly represented by periods of enrollment in specific insurance plans. In Electronic Medical Record data, this may represent the span of time for which Observations have been recorded for a person. Observation Periods are stored in the Observation\_Period table. |
| **Observational Medical Outcomes Partnership** | **OMOP** | A public-private partnership designed to protect human health by improving the monitoring of drugs for safety and effectiveness. |
| **Person (entity)** |  | A Person entity is one of the basic dimensions of analysis. It presents the framework for active drug surveillance. The Person entity is Concept-driven, and its attribute values are stored as standard Concept codes rather than original (i.e., “raw”) source values and is stored in the logical Person table. |
| **Primary Care Physician** | **PCP** | A physician designated as responsible to provide specific care to a patient, including evaluation and treatment as well as referral to specialists. |
| **Procedure Occurrence (entity)** |  | A Procedure Occurrence records individual instances of medical procedures extracted from source data. Procedures are recorded in various data sources in different forms with varying levels of standardization such as CPT-4, ICD-9-CM, and HCPCS procedure codes. These are stored in the Procedure\_Occurrence table. |
| **Protected Health Information** | **PHI** | Protected health information (PHI) under HIPAA includes any *individually identifiable* health information. *Identifiable* refers not only to data that is explicitly linked to a particular individual (that's *identified* information). It also includes health information with data items which reasonably could be expected to allow individual identification. De-indentified information is that from which all potentially identifying information has been removed. (HIPAA also has a provision for a limited data set, from which most but not all potentially identifying information has been removed.) |
| **RxNorm** |  | A standardized nomenclature for clinical drugs and drug delivery devices is produced by the National Library of Medicine (NLM). In RxNorm, the name of a clinical drug combines its ingredients, strengths, and/or form.  RxNorm provides normalized names for clinical drugs and links its names to many of the drug vocabularies commonly used in pharmacy management and drug interaction software, including those of First DataBank, Micromedex, MediSpan, Gold Standard Alchemy, and Multum. By providing links between these vocabularies, RxNorm can mediate messages between systems not using the same software and vocabulary. |
| **Systematized Nomenclature of Medicine--Clinical Terms** | **SNOMED-CT** | A comprehensive clinical terminology, originally created by the College of American Pathologists (CAP) and, as of April 2007, owned, maintained, and distributed by the International Health Terminology Standards Development Organization (IHTSDO), a non-for-profit association in Denmark. The CAP continues to support SNOMED CT operations under contract to the IHTSDO and provides SNOMED-related products and services as a licensee of the terminology.  SNOMED CT is one of a suite of designated standards for use in U.S. Federal Government systems for the electronic exchange of clinical health information, and is also a required standard in interoperability specifications of the U.S. Healthcare Information Technology Standards Panel. SNOMED CT is also being implemented internationally as a standard within other IHTSDO Member countries. |
| **Visit Occurrence (entity)** |  | The Visit Occurrence entity contains the information available in the source data about person visits to healthcare providers, including inpatient, outpatient, and ER visits. Visits are recorded in various data sources in different forms with varying levels of standardization. The detail level of the classification and description of the visit differs by data source. Visit Occurrence entities are recorded in the Visit\_Occurrence table. |
| **Terminology** |  | Technical or special terms used in a business or special subject area. |

# Conceptual Data Model

The CDM defines table structures for each of the data entities (e.g., Persons, Visit Occurrence, Drug Exposure, Condition Occurrence, Observation, Procedure-Occurrence, etc.). Loading a CDM instance from a source data set standardizes the data, both in format and in representation, to ensure that research methods applied to the CDM instance will be portable to any other CDM instance. As the *Figure 1: CDM Conceptual Data Model* below depicts, the ETL process that loads the CDM instance from the source data set performs two actions:

1. Transforms (i.e., reformats) the source data set content to conform to the CDM table structures, and
2. Augments the source data with their corresponding standard Concept codes from the Dictionary.

*Figure 1:CDM Conceptual Data Model*

****

To achieve this augmentation, the ETL logic uses an individual source datum to perform a look-up on the Dictionary, finds the standard Concept that corresponds to the datum, and loads the Concept code into the appropriate table and column of the CDM instance. Predefined mappings between the distinct values of the source data set and the standard Concepts of the Dictionary make this look-up operation possible. For additional information on the ETL process, refer to the ETL Specification document.

## CDM Core Module

Conceptually, the Common Data Model core module has eight entities. These are:

1. Person
2. Observation Period (the time at which health care information may be available, which can be used to estimate event rates over time)
3. Drug Exposure (i.e., the association between Person and Drug for a specific time period)
4. Health Outcome of Interest, which may be based on a combination of:
5. The Person’s medical Condition(s)
6. One or more Clinical Observations about the Person (e.g., laboratory test results)
7. One or more Medical Procedures that the Person required

8. One or more Visits for health care services for the Person

The *Figure 2: The Conceptual View of the CDM Core Module* on the next page illustrates these conceptual entities, and their relationships with (i.e., mappings to) the standard healthcare Concepts stored in the Dictionary. Each conceptual entity may have many such mappings. For example, the Visit entity will map values for Visit Type (e.g., hospital inpatient, hospital outpatient, emergency room, ambulatory/office visit, etc.) to the standard Concepts that represent these values. Likewise, the Observation entity will map values for Observation Type (e.g., laboratory test result) and individual instances of specific Observation Types (e.g., blood glucose test, serum sodium test, etc.) to the standard Concepts that represent them.

Generally, any conceptual construct of the CDM that may be assigned a value should have a corresponding standard Concept in the Dictionary. For example, since a blood glucose test may be assigned a value (i.e., the test result expressed as a number and a unit of measure) there should be a standard Concept that represents the blood glucose test in the Dictionary. Standard Concepts that represent all of the possible units of measure that may qualify a blood glucose test result should also be in the Dictionary. Some other laboratory tests may actually have a small number of possible discrete values (e.g., positive or negative, present or absent, etc.) and as such the result values themselves should also have standard Concept codes in the Dictionary. For example, some laboratory tests have such values as “positive,” “negative”, etc., and the Dictionary should contain standard Concepts for this set of discrete result values.

*Figure 2: The Conceptual View of the CDM Core Module* 

## Standard Dictionary

The Standard Dictionary is a semantic network containing all of the Concepts, Concept-to-Concept relationships, and other metadata necessary to describe the meanings and structures of the data within the CDM. The Dictionary will accommodate Concepts for each of the entities of interest relative to drugs, conditions, procedures, visits, demographics, etc. The Conceptual data for the OMOP Dictionary is a standardized format designed to integrate and standardize terminologies for observational analysis.

The diagram below, *Figure 3: Dictionary Conceptual Data Model*, depicts its internal organization.

*Figure 3: Dictionary Conceptual Data Model*



For example, the Dictionary will contain a single standard Concept having Concept code 4249983, which means “Acute Myocardial Infarction, Anterolateral Wall, Initial Episode of Care.” Furthermore, the Dictionary will map to this Concept (i.e., standardize) all of the various source-specific representations of this diagnosis, including ICD-9-CM diagnosis code 410.01 from data source A. The Concept itself will be captured in the Concept section of the Dictionary. The mappings from the Concept to its various source-specific representations will be captured in the Source-to-Concept Mappings section of the Dictionary. The Concept Metadata section will capture additional information about each Concept, such as its properties (e.g., synonyms, preferred terms, etc.) and qualifiers (e.g., property sources, allowed values, ranges, etc.) The related nature of Concepts will be mapped in the Concept Relationship section, which captures inter Concepts relationships (e.g., parent-child, composite-component, etc.).

### Content of the Standard Dictionary

The logical design of the Dictionary is, by intention, designed to integrate any Vocabulary source that can be used for observational analysis. The concrete description of all source vocabularies, their implementations, the definitions of the relationships, the choice of hierarchical relationships to define ancestry between concepts as well as the mapping from non-standard vocabularies into the standard vocabularies is described in a separate specification document, the Terminology Specification.

# CDM Logical Data Model

## Logical Entity-Relational Diagram

The *Figure 4: Logical View of OMOP Common Data Model*, on the next page, depicts the entire CDM Logical Model that includes CDM Core Module and Dictionary.

*Figure 4: Logical View of OMOP Common Data Model*

# 

## Logical Entities and Attributes – CDM Core Module

### PERSON

The Person entity is one of the basic dimensions of analysis, and when combined with the Drug Exposure, Condition, Observation, and Procedure entities, presents the framework for active drug surveillance. The source data for the Person entity comes from person demographics data that will be de-identified to ensure HIPAA compliance. The extent of these data varies by data source. The Person entity attribute values are stored as standard Concept codes mapped to the original (i.e., “raw”) source values.

| **Field** | **Data Type** | **Required** | **Description and Notes** |
| --- | --- | --- | --- |
| PERSON\_ID | INTEGER | YES | A system-generated identifier to uniquely identify each PERSON. |
| YEAR\_OF\_BIRTH | INTEGER | NO | The year of birth of the Person. For data sources with date of birth, only 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. |
| GENDER\_CONCEPT\_ID | INTEGER | NO | The foreign key that refers to the standard Concept Code in the Dictionary for the Gender of the Person. |
| RACE\_CONCEPT\_ID | INTEGER | NO | The foreign key that refers to the standard Concept Code in the Dictionary for the Race of the Person. |
| LOCATION\_CONCEPT\_ID | INTEGER | NO | The foreign key that refers to the standard Concept Code in the Dictionary for the Location of the Person. |
| SOURCE\_PERSON\_KEY | VARCHAR(32) | NO | An Encrypted key derived from the Person identifier in the source data. This is necessary when a drug safety issue requires a link back to the Person data from the raw source data set.  The Source Person Key always needs to be an encrypted value and no identifier with any medical or demographic significance can be stored.  The OMOP Research environment stores the de-identified unique identifiers for that Person from the source data as the Source Person Key. |
| SOURCE\_GENDER\_ CODE | VARCHAR(20) | NO | The source code for the Gender of the Person.  The Person Gender is mapped to a standard Gender Concept in the Dictionary and the corresponding Concept Code is stored here as a reference.  The Concept Code references the Concept entity in the Dictionary. |
| SOURCE\_LOCATION\_CODE | VARCHAR(20) | NO | The source code for the Location of the Person.  The Person’s Location is mapped to a standard Geographic Location Concept in the Dictionary and the corresponding Concept Code is stored here as a reference.  The Concept Code references the Concept entity in the Dictionary. |
| SOURCE\_RACE\_CODE | VARCHAR(20) | NO | The source code for the Race of the Person.  The Person’s Race is mapped to a standard Race Concept in the Dictionary and the corresponding Concept Code is stored here as a reference.  The Concept Code references the CONCEPT entity in the Dictionary. |

#### 

#### Business Rules

* Person data will remain de-identified as much as possible to comply with Design Principle 5. Accordingly, the CDM will not store the precise date of birth, rather it will store only the year of birth, nor will it store any identifiers that could be used to re-identify the Person data.
* The granularity of the Person data from the source system will be maintained. There will be no consolidation or aggregation of individual Person records.
* Standard attributes will be stored as Concept codes. Original source values will be mapped to the corresponding standard Concept codes in the Dictionary.
* Person source data attributes that are in-scope are limited to race, gender, location, year of birth, and Person status. Other Person source attributes are out-of-scope.

#### 

#### Example of Loaded Table

Consider the following example of inbound source data.

| **Source person id** | **Source data set** | **Person year of birth** | **Person gender** | **Person location** | **Person race** |
| --- | --- | --- | --- | --- | --- |
| 121107 | GE-EHR | 1932 | Female | 850 | Hispanic |
| 127260 | GE-EHR | 1933 | Female | 850 | Caucasian |

Sample code representation of demographic data from the Dictionary follows.

| **Concept ID** | **Concept Code** | **Concept Name** | **Source Code Description** | **Mapping Type** |
| --- | --- | --- | --- | --- |
| 8527 | 2106-3 | WHITE | CAUCASIAN | RACE |
| 8558 | 2135-2 | HISPANIC | HISPANIC | RACE |
| 8532 | F | FEMALE | FEMALE | GENDER |
| 850 | 850 | ZIPCODE | 850 | ZIPCODE |

Unique system-generated identifiers are used for the PERSON\_ID: 121107 and 127260 for the two persons.

The one-way hash keys for the source person identifiers are determined in this example using MD5 hashing.

| **Source Person Identifier** | **Source Person Key** |
| --- | --- |
| 121107 | 57bcc40b9080b35f781bc87dd8dc77b7 |
| 127260 | cd759452e9577d52354cc327b86f0760 |

The above data are represented in the CDM PERSON table as follows:

| **Person ID** | **Year of Birth** | **Gender Concept ID** | **Race Concept ID** | **Location Concept ID** | **Source Person Key** | **Source Gender Code** | **Source Location Code** | **Source Race Code** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 121107 | 1932 | 8532 | 8527 | 850 | 57bcc40b9080b35f781bc87dd8dc77b7 | Female | 850 | Caucasian |
| 127260 | 1933 | 8532 | 8558 | 850 | cd759452e9577d52354cc327b86f0760 | Female | 850 | Hispanic |

## 

### DRUG\_EXPOSURE

Drug\_Exposure contains individual records that reflect drug utilization from within the observational source. Drug Exposure indicators include drug details (captured as standard Concept codes in the Dictionary), drug quantity, number of days supply, period of exposure, and prescription refill data. Drug Exposure is recorded in a variety of ways.

* The “Prescription” section of an EHR captures prescriptions written by physicians.
* Other drugs (both non-prescription products and medications prescribed by other providers) used by a Person are recorded in the “Medications” section of the EHR.
* Administrative claims systems capture prescriptions filled at dispensing providers.
* Drug Exposure information as a by-product of certain procedure codes (i.e., procedure codes that refer to professional services related to the administration of certain drugs).

Drug Exposures are indicated in the CDM by standard drug Concepts from the Dictionary. The standard Concept code for a drug is stored with the drug reference data; however, the Concept hierarchy and therapeutic class categorizations from the source data are not stored with the drug exposure data but rather in the Dictionary.

| **Field** | **Data Type** | **Required** | **Description and Notes** |
| --- | --- | --- | --- |
| DRUG\_EXPOSURE\_ID | INTEGER | YES | A system-generated identifier to uniquely identify each DRUG EXPOSURE. |
| DRUG\_EXPOSURE\_START\_DATE | DATE | YES | This is the Start Date for the current instance of drug utilization. Valid indicators include a start date of a prescription, the date a prescription was filled, or the date on which a drug administration procedure was recorded. |
| DRUG\_EXPOSURE\_END\_DATE | DATE | NO | This is the End Date for the current instance of drug utilization. It is not available from all sources. |
| PERSON\_ID | INTEGER | YES | System-generated foreign key identifier for the PERSON who is the subject of the Drug Exposure. The demographics for the Person are captured in the PERSON entity. |
| DRUG\_CONCEPT\_ID | INTEGER | NO | A foreign key that refers to the standard Concept Code in the Dictionary for the Drug Concept. It is used to map to standard drug information and the Concept hierarchy in the Dictionary. |
| DRUG\_EXPOSURE\_TYPE | VARCHAR(3) | YES | A foreign key to the predefined code in the DRUG\_EXPOSIRE\_REF entity reflecting the type of Drug Exposure recorded. It is the indicator by which the Drug Exposure, including medication history, filled prescriptions, etc., is identified, |
| STOP\_REASON | VARCHAR(20) | NO | The reason the medication was stopped, where available. Reasons include Regimen Completed, Changed, Removed, etc. |
| REFILLS | INTEGER | NO | The number of refills for the prescription. |
| DRUG\_QUANTITY | INTEGER | NO | The quantity of drug recorded in the corresponding Drug Exposure instance. |
| DAYS\_SUPPLY | INTEGER | NO | The number of days’ supply of the medication recorded in the corresponding Drug Exposure instance. |
| SOURCE\_DRUG\_CODE | VARCHAR(20) | YES | The drug identifier captured in the raw source data. The types of identifiers allowed include National Drug Codes (NDCs), Generic Product Identifier (GPI) codes, etc. |

#### 

#### Business Rules

* Source drug identifiers, including NDC Codes, Generic Product Identifiers, etc. are mapped to standard drug Concepts in the Dictionary (e.g., RxNorm RXCUI). When the Source Drug identifier cannot be translated into Standard Drug Concepts, a Drug exposure entry is stored with only the corresponding Source Drug Code.
* A Drug Exposure Type is assigned to each Drug Exposure, to track the indicator from which the data were drawn or inferred. The Drug Exposure Types are discussed in detail in the DRUG\_EXPOSURE\_REF section of this document.
* Drug Exposures extracted from some of the data sources do not include all desired attributes. The following attributes constitute the minimum set required for usable Drug Exposure data:
  + Person identifier
  + Source drug identifier/Concept
  + Date of Exposure
* Financial details related to the medications are out-of-scope.

#### Example of Loaded Table

Consider the following example of inbound source data on medications, from an EHR.

First from Person Medication List:

| **Person ID** | **Generic Product Identifier** | **Medication Name** | **Medication Start Date** | **MedicationEndDate** | **Stop Reason** |
| --- | --- | --- | --- | --- | --- |
| 121798 | 83200030200320 | Warfarin 7.5 MG Oral Tablet | 29-NOV-08 | 29-NOV-08 | Other |
| 121798 | 83200030200320 | Warfarin 7.5 MG Oral Tablet | 19-JAN-09 |  |  |
| 121798 | 83200030200320 | Warfarin 7.5 MG Oral Tablet | 25-JUN-08 |  | Removed |

Next, from list of prescriptions for the person:

| **Person ID** | **Generic Product Identifier** | **Medication Name** | **Prescription Date** | **Refills** | **Drug Quantity** | **Days Supply** |
| --- | --- | --- | --- | --- | --- | --- |
| 121801 | 83200030200310 | Warfarin 2.5 MG Oral Tablet | 14-AUG-07 | 0 | 30 |  |
| 121798 | 83200030200320 | Warfarin 7.5 MG Oral Tablet | 16-JUN-08 | 11 | 25 |  |

Sample Concept code representation of drug data from the Dictionary follows.

| **Concept ID** | **Concept Code** | **Concept Description** | **Concept Vocabulary Code** |
| --- | --- | --- | --- |
| 1310213 | 313734 | Warfarin 2.5 MG Oral Tablet | 08 |
| 1310217 | 313739 | Warfarin 7.5 MG Oral Tablet | 08 |

Since the data were drawn from Person Medication lists in the EHR, Drug Exposure type is set to “MED HISTORY,” which is described in the DRUG\_EXPOSURE\_REF section of this chapter as “Medication History from Electronic Health Records.”

Drug Exposure types are determined, based on the source from which the exposure is recorded, as follows:

| **Drug Exposure Type** | **Drug Exposure Type Description** |
| --- | --- |
| 2 | Prescriptions Written |
| 3 | Medication List |

The above data are represented in the CDM DRUG\_EXPOSURE table as follows:

| **Drug Exposure ID** | **Drug Exposure Start Date** | **Drug Exposure End Date** | **Person ID** | **Drug Concept ID** | **Drug Exposure Type** | **Stop Reason** | **Refills** | **Drug Quantity** | **Days Supply** | **Source Drug Code** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 150807907 | 14-AUG-07 |  | 121801 | 1310213 | 2 |  | 0 | 30 |  | 83200030200310 |
| 32830 | 29-NOV-08 | 29-NOV-08 | 121798 | 1310217 | 3 | Other |  |  |  | 83200030200320 |
| 32832 | 19-JAN-09 |  | 121798 | 1310217 | 3 |  |  |  |  | 83200030200320 |
| 131428827 | 16-JUN-08 |  | 121798 | 1310217 | 2 |  | 11 | 25 |  | 83200030200320 |
| 32829 | 25-JUN-08 |  | 121798 | 1310217 | 3 | Removed |  |  |  | 83200030200320 |

### 

### DRUG\_ERA

Drug Era is defined as a span of time when the Person is assumed to be using a particular drug of a particular strength. A Drug Era is not the same as a Drug Exposure; successive periods of Drug Exposure may, under certain rules, be combined to produce one continuous Drug Era. Each drug ingredient and strength combination is mapped to a separate Concept, and exposure to each is treated as a separate Drug Era.

| **Field** | **Data Type** | **Required** | **Description and Notes** |
| --- | --- | --- | --- |
| DRUG\_ERA\_ID | INTEGER | YES | A system-generated identifier to uniquely identify each Drug Era that is constructed from Drug Exposure data. |
| DRUG\_ERA\_START\_DATE | DATE | YES | The start date for the Drug Era constructed from the individual instances of Drug Exposure. It is the start date of the very first chronologically recorded instance of utilization of a drug. |
| DRUG\_ERA\_END\_DATE | DATE | YES | The end date for the Drug Era constructed from the individual instance of Drug Exposure. It is the end date of the final continuously recorded instance of utilization of a drug. |
| PERSON\_ID | INTEGER | YES | A system-generated foreign key identifier for the PERSON who is the subject of the Drug Exposure. The demographics for the Person are captured in the PERSON entity. |
| DRUG\_EXPOSURE\_TYPE | VARCHAR(3) | YES | A foreign key to the predefined code in the DRUG\_EXPOSURE\_REF entity reflecting the type of Drug Exposure recorded. It is the indicator by which the Drug Exposure, including medication history, filled prescriptions, etc., is identified. |
| DRUG\_CONCEPT\_ID | INTEGER | YES | A foreign key that refers to the standard Concept Code in the Dictionary for the Drug Concept. It is used to map to standard drug information and the Concept hierarchy in the Dictionary. |
| DRUG\_EXPOSURE\_COUNT | INTEGER | YES | The number of Drug Exposure occurrences used to construct the Drug Era. |

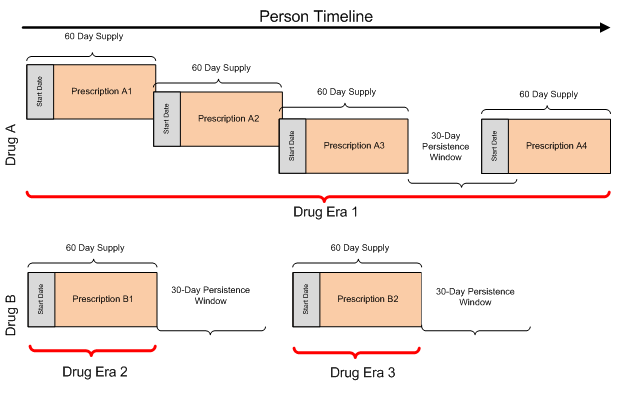
#### 

#### Business Rules

* For claims related to pharmacy prescriptions, the dispensed date and number of days supply are used to extrapolate the end date for the period of Drug Exposure. When a Person receives recurring prescriptions for the same product and strength, the multiple prescriptions may need to be treated as a single Drug Era. To determine whether this is indeed the case, the drug’s “persistence window”, which is the number of days after the Person stops taking a drug and during which the Person is deemed to still be affected by the drug, must be taken into account. If the number of days between the end date of the prior Drug Exposure and the start date of the subsequent Drug Exposure falls within the persistence window, then the two exposures are considered to belong to the same Drug Era.
* For EHRs, the medications data include the start and end dates for the medication. The prescription data track only the date on which the medication was prescribed and the date on which the record was created. While the prescription data often include a pointer to identify the corresponding medication record, this is not always the case.
* For a Drug Exposure indicated by procedure codes, usually only a single date is available (i.e., the administration date). This may occasionally pose a challenge in determining Drug Eras but generally should be able to be determined from the source.

For example, consider a Person who is taking two drugs: Drug A and Drug B. The Person has had four prescriptions for Drug A (A1, A2, A3, A4), each with a sixty-day supply. The Person has also had two prescriptions for Drug B (B1, B2). *Figure 5: Drug Era Examples* below illustrates the above scenario.

*Figure 5: Drug Era Examples*



To define the Drug Era for Drug A, the timing, duration, overlap, and persistence of the Person’s prescriptions for Drug A must be considered. A2 was filled before the expected completion of A1. Similarly, A3 was filled before the expected completion of A2. A4 was filled after A3 was completed, but within the persistence window for Drug A. Therefore, the four prescriptions for Drug A will be consolidated into a single Drug Era (DrugEra1), with the start for prescription A1 recorded as the start date for the consolidated record and the end date for prescription A4 recorded as the end date.

As the persistence window was exceeded between filling the two prescriptions for Drug B, they are defined as two distinct Drug Eras. The start and end dates for DrugEra2 and DrugEra3 are the start and end dates for prescriptions B1 and B2, respectively.

Persistence windows of zero days and thirty days will be used.

#### Example of Loaded Table

Consider the following example excerpt from the CDM DRUG\_EXPOSURE table.

| **Drug Exposure ID** | **Drug Exposure Start Date** | **Drug Exposure End Date** | **Person ID** | **Drug Exposure Type** | **Drug Concept Id** | **Stop Reason** | **Refills** | **Drug Quantity** | **Days Supply** | **Source Drug Code** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1001 | 09-MAY-2003 | 09-MAY-2003 | 121107 | MED HISTORY | 375374009 | Regimen Completed |  |  |  | 375374009 |
| 1002 | 30-APR-/2003 |  | 127260 | MED HISTORY | 375383004 |  |  |  |  | 375383004 |
| 1003 | 27-JUL-2003 | 27-JUL-2003 | 127260 | MED HISTORY | 375383004 |  |  |  |  | 375383004 |
| 1004 | 22-AUG-2003 | 22-AUG-2003 | 127260 | MED HISTORY | 375383004 |  |  |  |  | 375383004 |
| 1005 | 07-SEP-2003 | 07-SEP-2003 | 127260 | MED HISTORY | 375378007 |  | 1 | 30 | 30 | 375378007 |
| 1006 | 02-OCT-2003 | 02-OCT-2003 | 127260 | MED HISTORY | 375378007 | Regimen Completed | 1 | 90 | 90 | 375378007 |

The above example uses the following hypothetical drug Concept codes from the Dictionary.

|  |  |  |  |
| --- | --- | --- | --- |
| **Concept ID** | **Concept Code** | **Concept Description** | **Concept Level** |
| 1310216 | 313738 | Warfarin 6 MG Oral Tablet | 1 |
| 1310213 | 313734 | Warfarin 2.5 MG Oral Tablet | 1 |
| 1310217 | 313739 | Warfarin 7.5 MG Oral Tablet | 1 |

The drug hierarchy in the Dictionary indicates that all of the above drug Concepts are children of the following high-level drug class Concept.

| **Concept ID** | **Concept Code** | **Concept Description** | **Concept Level** |
| --- | --- | --- | --- |
| 1310149 | 11289 | Warfarin | 2 |

Drug Exposure types are determined for Drug Eras with a 30-day persistence window as follows:

| **Drug Exposure Type** | **Drug Exposure Type Description** |
| --- | --- |
| 7 | Drug Era, 30-day window |

The Drug Eras constructed from the above data, based on the higher-level drug class Concept and using a 30-day persistence window would be reflected in the CDM DRUG\_ERA table as follows:

| **Drug Era ID** | **Drug Era Start Date** | **Drug Era End Date** | **Person ID** | **Drug Exposure Type** | **Drug Concept ID** | **Drug Exposure Count** |
| --- | --- | --- | --- | --- | --- | --- |
| 20001 | 09-MAY-2003 | 09-MAY-2003 | 121107 | 7 | 1310149 | 1 |
| 20002 | 30-APR-2003 | 30-APR-2003 | 127260 | 7 | 1310149 | 1 |
| 20003 | 27-JUL-2003 | 02-OCT-2003 | 127260 | 7 | 1310149 | 4 |

### 

### DRUG\_EXPOSURE\_REF

This is a reference listing of various types of Drug Exposures recorded for analysis. The Drug Exposure Type conveys the indicator(s) from which the Drug Exposure was captured, and defines the characteristic of the exposure and the level of aggregation.

The Drug Exposure Types follow:

* Prescription Written (from Electronic Health Records)
* Medication History (from Electronic Health Records)
* Filled Prescription (from Pharmacy Claims)
* Drug from Procedure Code (from Medical Claims and Electronic Health Records)
* Drug Era using a persistence window

| **Field** | **Data Type** | **Required** | **Description and Notes** |
| --- | --- | --- | --- |
| DRUG\_EXPOSURE\_TYPE | VARCHAR(3) | YES | A predefined code for the type of Drug Exposure recorded. The Drug Exposure Type is used to define the data source and the type of representation of the Drug utilization recorded.  A detailed listing of condition Drug Exposure Types is recorded in [Appendix C: Drug Exposure Type Codes](#_Appendix_C:_Drug) |
| DRUG\_EXPOSURE\_TYPE\_DESC | VARCHAR(120) | YES | A detailed description for the type of Drug Exposure recorded. |
| PERSISTENCE\_WINDOW | INTEGER | NO | The Persistence Window used to build the Drug Era based on Drug Exposure data.  A Persistence Window defines the longest period of time between two instances of Drug Exposure for them to be considered as part of the same continuous exposure.  This is applicable only when constructing Drug Eras.  Persistence windows of 0 days and 30 days are used. |

#### 

Drug Exposure References have been populated based on Thomson and GE References. Different databases might need additional list of data to reflect information that is not currently captured.

#### Example of Loaded Table

Reference data for Drug Exposure Types are stored in the DRUG\_EXPOSURE\_REF table, which includes the type codes identified and their detailed descriptions. The allowed Drug Exposure Type Codes and their descriptions are listed in [Appendix C](#_Appendix_C:_Drug).

### 

### CONDITION\_OCCURRENCE

Condition Occurrences record individual instances of Person conditions extracted from source data. Conditions are recorded in various data sources in different forms with varying 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 capture Person 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.
* Death when observed in Person status codes (such as discharge status).

Condition Occurrences are analyzed based on standard condition Concepts in the Dictionary, which also maintains the standard condition hierarchy for use in class-based CDM queries.

| **Field** | **Data Type** | **Required** | **Description and Notes** |
| --- | --- | --- | --- |
| CONDITION\_OCCURRENCE\_ID | INTEGER | YES | A system-generated identifier to uniquely identify each Person condition.  Every Person with a Condition Occurrence must have corresponding demographics data in the PERSON entity. |
| CONDITION\_START\_DATE | DATE | YES | The date when the instance of the Condition is first recorded. |
| PERSON\_ID | INTEGER | YES | A system-generated foreign key identifier for the PERSON who is the subject of the Drug Exposure. The demographics for the Person are captured in the PERSON entity. |
| CONDITION\_END\_DATE | DATE | NO | The date when the instance of the Condition is last recorded. |
| CONDITION\_OCCURRENCE\_TYPE | VARCHAR(3) | YES | A foreign key to the predefined Code in the CONDITION\_OCCURRENCE\_REF entity reflecting the type of Condition Occurrence recorded.  The Condition Occurrence Type is used to define the source data from which the Condition was recorded, the level of standardization, and the type of occurrence. Conditions are defined as Primary/Secondary Diagnoses, Problem List, and Person Status (including Mortality when available in the source data set).  A detailed listing of Condition Occurrence Types is recorded in [Appendix D: Condition Occurrence Type Codes](#Appendix_D). |
| CONDITION\_CONCEPT\_ID | INTEGER | NO | A foreign key to the standard Condition Concept Code in the Dictionary  The source code and/or the description is mapped to a standard Condition Concept and the corresponding Concept ID is stored here as a reference.  It is used to map to standard conditions Vocabulary and Concept hierarchy in the Dictionary. |
| STOP\_REASON | VARCHAR(20) | NO | The reason, if available, that the condition was no longer recorded, as indicated in the source data. Valid values include Discharged, Resolved, etc. |
| DX\_QUALIFIER | VARCHAR(20) | NO | An indicator for the category of Diagnosis as recorded in the source data. Valid values include qualifiers such as Major Diagnosis, Family History of, History of, Hospitalization, Recurrence, Risk of, Rule-Out, etc. |
| SOURCE\_CONDITION\_CODE | VARCHAR(20) | YES | The Condition Code as captured from the source data. Values primarily include ICD-9-CM diagnosis codes from medical claims and EHRs and specific Discharge Status/Disposition codes from medical claims. |

#### 

#### Business Rules

The approach to extraction of Condition Occurrence data is based on the individual data source, but the following guidelines are common to all data sources.

* Source attributes mapped to conditions are checked for standardization. If the source attributes are available as standard diagnosis codes (e.g., ICD-9-CM Diagnosis Codes) or specific discharge status codes, then they are mapped to standard Concepts in the Dictionary.
* If the source data are not coded to a national or international standard, then a finite listing of attribute values is created and mapped to standard condition Concepts in the Dictionary.
* A Condition Occurrence Type is assigned based on the data source and type of condition attribute, including:
  + ICD-9-CM Primary Diagnosis from Medical Claims and EHRs
  + ICD-9-CM Secondary Diagnoses from Medical Claims and EHRs
  + Person Status from Medical Claims from EHRs
  + Problem Concept from EHRs

More details regarding the Condition Occurrence Types appear in the CONDITION\_OCCURRENCE\_REF section of this chapter.

* Each Condition for every Person, along with its matching standard Concept code from the Dictionary, is extracted from the source data along with the Person identifier, start/onset date of the condition, end date for condition, and diagnosis qualifier (DX\_QUALIFIER) attributes, where available.
* Special handling is necessary for source data in which Person condition entries are updated by expiration of the current entry and addition of an updated entry. In such cases, only the final version of the record is extracted for inclusion in the CDM.

#### Example of Loaded Table

Consider the following example of inbound source data on medications, from GE EHRs.

| **Person ID** | **ICD-9–CM Diagnosis Code** | **Problem Description** | **Problem Start Date** | **Problem End Date** | **Dx Qualifier** | **Stop Reason** |
| --- | --- | --- | --- | --- | --- | --- |
| 127260 | 787.02 | Nausea | 03-MAY-2003 | 03-MAY-2003 | Diagnosis Of | Unknown |
| 127260 | 787.02 | Nausea | 29-JUL-2003 | 29-JUL-2003 | Diagnosis Of | Resolved |
| 127260 | 531.01 | Acute gastric ulcer without hemorrhage or perforation without obstruction | 23-AUG-2003 | 23-AUG-2003 | Diagnosis Of | Resolved |

The following Concept codes correspond to the meanings of the problems captured in the source data:

| **Concept ID** | **Concept Code** | **Concept Description** |
| --- | --- | --- |
| 4197598 | 313425006 | Gastric Ulcers |
| 31967 | 422587007 | Nausea |

Condition Occurrence types are determined for Conditions extracted from EHR Problem list as follows:

| **Condition Occurrence Type** | **Condition Occurrence Type Description** |
| --- | --- |
| 63 | EHR - Problem List |

The CDM CONDITION\_OCCURRENCE table, loaded with the above data, would appear as follows:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Condition Occurrence ID** | **Condition Start Date** | **Person ID** | **Condition End Date** | **Condition Occur Type** | **Condition Concept ID** | **Stop Reason** | **Dx Qualifier** | **Source Condition Code** |
| 3003 | 03-MAY-2003 | 127260 | 03-MAY-2003 | 63 | 4244606 | Unknown | Diagnosis Of | 787.02 |
| 3004 | 29-JUL-2003 | 127260 | 29-JUL-2003 | 63 | 4244606 | Resolved | Diagnosis Of | 787.02 |
| 3005 | 23-AUG-2003 | 127260 | 23-AUG-2003 | 63 | 4197598 | Resolved | Diagnosis Of | 787.02 |

### 

### CONDITION\_ERA

Similar to Drug Eras, Condition Eras are chronological periods of Condition Occurrence. Combining individual Condition Occurrences into a single Condition Era serves at least two purposes:

* It allows aggregation of chronic conditions that require frequent ongoing care, instead of treating each Condition Occurrence as an independent event.
* It allows aggregation of multiple, closely-timed doctor visits for the same condition to avoid double-counting the Condition Occurrences.

For example, consider a Person who visits his Primary Care Physician (PCP), who diagnoses the Person with a specific condition and refers the Person to a Specialist. One week later, the Person visits the Specialist, who confirms the PCP’s diagnosis and provides the appropriate treatment to resolve the condition with no further care required. These two independent doctor visits should be aggregated into one Condition Era. Just as with Drug Eras, the persistence windows to be used in determining Condition Eras are also zero days and thirty days.

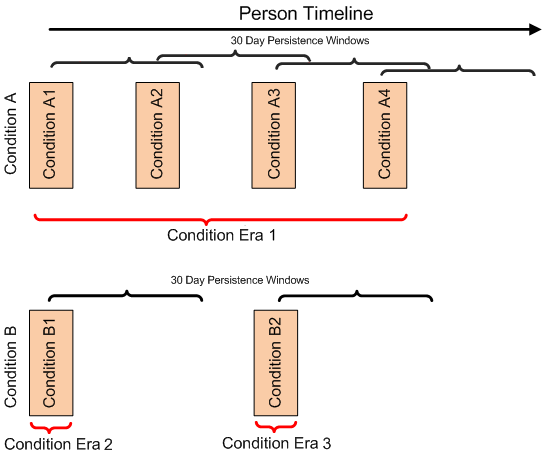
This model generally fits well for acute conditions, but may be less robust for chronic conditions. For example, chronic conditions that do not require regular follow-up may be recorded as multiple Condition Eras because the absence of data in the periods between visits does not justify the aggregation of the eras. Because the persistence window is small, it is likely that multiple visits will be captured in rapid succession for the same condition; however, it is unlikely that infrequent visits for chronic conditions (e.g. a Person with Rheumatoid Arthritis who visits his rheumatologist every three months) will be captured. However, the small window also reduces the likelihood that independent events will be falsely classified as the same Condition Era.

| **Field** | **Data Type** | **Required** | **Description and Notes** |
| --- | --- | --- | --- |
| CONDITION\_ERA\_ID | INTEGER | YES | A system-generated identifier used to uniquely identify each Condition Era. |
| CONDITION\_ERA\_START\_DATE | DATE | YES | The start date for the Condition Era constructed from the individual instances of Condition Exposures. It is the start date of the very first chronologically recorded instance of the condition. |
| PERSON\_ID | INTEGER | YES | A system-generated foreign key identifier for the PERSON who is the subject of the Drug Exposure. The demographics for the Persons are captured in the PERSON entity. |
| CONFIDENCE | INTEGER | NO | Degree of confidence based on the source data for Condition and the type of Condition recorded. A method for determining this has not yet been determined. |
| CONDITION\_ ERA\_END\_DATE | DATE | YES | The end date for the Condition Era constructed from the individual instance of Condition Occurrence. It is the end date of the final continuously recorded instance of the condition. |
| CONDITION\_CONCEPT\_ID | INTEGER | YES | A foreign key to the standard Condition Concept Code in the Dictionary.  The source code and/or the description is mapped to a standard Condition Concept and the corresponding Concept id is stored here as a reference.  It is used to map to standard conditions Vocabulary and Concept hierarchy in the Dictionary. |
| CONDITION\_OCCURRENCE\_TYPE | VARCHAR(3) | YES | A foreign key to the predefined Code in the CONDITION\_OCCURRENCE\_REF entity reflecting the type of Condition Occurrence recorded.  The Condition Occurrence Type is used to define the source data from which the Condition was recorded, the level of standardization, and the type of occurrence. Conditions are defined as Primary/Secondary Diagnoses, Problem List, and Person Status (including Mortality when available in the source data set).  A detailed listing of Condition Occurrence Types is recorded in [Appendix D: Condition Occurrence Type Codes](#Appendix_D). |
| CONDITION\_OCCURRENCE\_COUNT | INTEGER | NO | The number of Condition Occurrences used to construct the Condition Era. |

#### Business Rules

A Condition Era represents the span of time for which a Person can be considered to have a given condition. An example is illustrated graphically in *Figure 6: Condition Era Examples* on the next page. Imagine a Person who has been diagnosed with two conditions during his insurance coverage period: Condition A and Condition B. The Person has been diagnosed with Condition A four times (A1, A2, A3, A4), and has been diagnosed with Condition B twice (B1, B2).

*Figure 6: Condition Era Examples*



To define condition persistence for Condition A, the timing of successive diagnoses is considered. Here, A2 is within the persistence window of A1. Similarly, A3 is within the persistence window of A2, and A4 is within the persistence window of A3. Thus, the four diagnoses of Condition A should be consolidated into Condition Era1, with the start date equal to the diagnosis date for A1, and the end date equal to the diagnosis date for A4.

With Condition B, significant time has elapsed between diagnoses B1 and B2. Therefore, it cannot be assumed that there is dependence between the diagnoses as the time exceeded the persistence window for B1. Therefore two distinct Condition Eras are defined, one each that corresponds to B1 and B2.

#### Example of Loaded Table

Consider the following example excerpt from the CDM CONDITON\_OCCURRENCE table.

| **Condition Occurrence ID** | **Condition Start Date** | **Person ID** | **Condition End Date** | **Condition Occur Type** | **Condition Concept ID** | **Stop Reason** | **Dx Qualifier** | **Source Condition Code** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 3003 | 03-MAY-2003 | 127260 | 03-MAY-2003 | DIAGNOSIS FROM PROBLEM LIST | 42446006 | Unknown | Diagnosis Of | 787.02 |
| 3004 | 29-JUL-2003 | 127260 | 29-JUL-2003 | DIAGNOSIS… | 42446006 | Resolved | Diagnosis Of | 787.02 |
| 3005 | 23-AUG-2003 | 127260 | 23-AUG-2003 | DIAGNOSIS… | 42446006 | Resolved | Diagnosis Of | 787.02 |

The above example uses the following hypothetical condition Concept codes from the Dictionary.

| **Concept ID** | **Concept Code** | **Concept Description** |
| --- | --- | --- |
| 4197598 | 313425006 | Gastric Ulcers |
| 31967 | 422587007 | Nausea |

The ontology within the Dictionary indicates that the above two Concepts are children of the following higher-level condition class Concept.

| **Concept ID** | **Concept Code** | **Concept Description** |
| --- | --- | --- |
| 4201745 | 53619000 | Digestive system disorders |

Condition Occurrence types are determined for Condition Eras, based on a 30-day persistence window, as follows:

| **Condition Occurrence Type** | **Condition Occurrence Type Description** |
| --- | --- |
| 65 | Condition Era, 30 day persistence window |

The sample CONDITION\_ERA table is constructed from the condition data based on a 30-day persistence window, indicated by a CONDITION\_OCCURRENCE\_TYPE of “Era, 30-Day Persistence.” The sample representation of the above data in the CONDITION\_ERA table follows.

| **Condition Era ID** | **Condition Start Date** | **Person ID** | **Confidence** | **Condition End Date** | **Condition Concept ID** | **Condition Occurrence Type** | **Condition Occurrence Count** |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 4197598 | 03-MAY-2003 | 121107 |  | 03-MAY-2003 | 4197598 | 63 | 1 |
| 31967 | 29-JUL-2003 | 127260 |  | 23-AUG-2003 | 4197598 | 63 | 2 |

### 

### CONDITION\_OCCURRENCE\_REF

The Condition Occurrence Reference reflects the indicator(s) from which the Condition Occurrence was drawn or inferred, and indicates whether a condition (diagnosis) was primary or secondary and the relative positioning within a Person’s condition record.

A detailed listing of the Condition Occurrence Types is included in [Appendix D](#Appendix_D).

| **Field** | **Data Type** | **Required** | **Description and Notes** |
| --- | --- | --- | --- |
| CONDITION\_OCCURRENCE\_TYPE | VARCHAR(3) | YES | The code indicating the type of Condition Occurrence.  Condition Occurrence Type is used to define the source data indicators from which the Condition was identified, the level of standardization, and the type of occurrence. Conditions Types are defined as Primary/Secondary Diagnosis, Problem Report, and Person Status (including Mortality when available in the source data).  A detailed listing of Condition Occurrence Types is recorded in [Appendix D: Condition Occurrence Type Codes](#Appendix_D) |
| CONDITION\_OCCURRENCE\_TYPE\_DESC | VARCHAR(120) | YES | A detailed description of the Condition Occurrence Type. Valid values include Diagnosis from Inpatient Claims, Diagnosis from Outpatient visits, Diagnosis from ER visits, etc. |
| PERSISTENCE\_WINDOW | INTEGER | NO | The Persistence Window used to build the Condition Era based on Condition Occurrence data.  A Persistence Window defines the longest period of time between two instances of the Condition for them to be considered as part of the same continuous event.  This is applicable only when constructing Condition Eras. Persistence windows of 0 days and 30 days are used. |
| CONDITION\_OCCURRENCE\_POSITION | VARCHAR(20) | YES | The sequential position of the occurrence when multiple conditions are observed. |

### 

Condition Occurrence References have been populated based on Thomson and GE References. Different databases might need additional list of data to reflect information that is not currently captured.

### VISIT\_OCCURRENCE

The Visit Occurrence entity contains all Person visits to health care providers, including inpatient, outpatient, and ER visits. Visits are recorded in various data sources in different forms with varying levels of standardization. For example:

* Medical Claims include Inpatient Admissions, Outpatient Services, and Emergency Room visits.
* Electronic Health Records may capture Person visits as part of the activities recorded.

| **Field** | **Data Type** | **Required** | **Description and Notes** |
| --- | --- | --- | --- |
| VISIT\_OCCURRENCE\_ID | INTEGER | YES | A system-generated identifier to uniquely identify each occurrence of a Person’s Visit to a healthcare provider. |
| VISIT\_START\_DATE | DATE | YES | The date on which the Visit started. |
| VISIT\_END\_DATE | DATE | NO | The date on which the Visit ended. |
| PERSON\_ID | INTEGER | YES | A system-generated foreign key identifier for the PERSON who is the subject of the Drug Exposure. The demographics for the Person are captured in the PERSON entity. |
| VISIT\_CONCEPT\_ID | INTEGER | YES | A foreign key to the standard Visit Concept Code in the Dictionary.  The source code and/or the description is mapped to a standard Visit Concept and the corresponding Concept ID is stored here as a reference.  It is used to map to standard Visit Vocabulary and Concept hierarchy in the Dictionary. |
| SOURCE\_VISIT\_CODE | VARCHAR(20) | YES | The Source Visit Code is used to reflect the type/source of the visit data. Valid entries include office visits, hospital admissions, etc.  For the OMOP research environment, type-of-service codes and activity type codes may be used as source visit codes. |

#### 

#### Business Rules

A Visit Occurrence is recorded for each visit to a healthcare facility. Each visit is standardized by assigning a corresponding Concept code based on the type of facility visited and the type of services rendered.

#### Example of Loaded Table

Consider the following example visit data extracted from Medical Claims.

| **Person ID** | **Place of Service Code** | **Type of Visit** | **Visit Start Date** | **Visit End Date** |
| --- | --- | --- | --- | --- |
| 127260 | 21 | Hospital Admission | 03-MAY-2003 | 04-MAY-2003 |
| 127260 | 22 | Outpatient Dialysis | 29-JUL-2003 | 29-JUL-2003 |

The following Concept codes correspond to the meanings of the types of visits that were indicated in the source data.

| **Concept ID** | **Concept Code** | **Concept Description** |
| --- | --- | --- |
| 8715 | HOSPITAL ADMSISSION | Hospital Admission |
| 8614 | OFFICE VISIT | Outpatient Visit |

The above data, represented in the CDM VISIT\_OCCURRENCE table, follows.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Visit Occurrence ID** | **Visit Start Date** | **Visit End Date** | **Person ID** | **Visit Concept ID** | **Source Visit Code** |
| 5003 | 03-MAY-2003 | 04-MAY-2003 | 127260 | 8715 | HOSPITAL ADMISSION |
| 5004 | 29-JUL-2003 | 29-JUL-2003 | 127260 | 8614 | OFFICE VISIT |

### 

### PROCEDURE\_OCCURRENCE

Procedure occurrences record individual instances of Person procedures extracted from source data. Procedures are recorded in various data sources in different forms with varying levels of standardization. For example:

* Medical Claims include CPT-4, ICD-9-CM (Procedures), and HCPCS procedure codes that are submitted as part of a claim for health services rendered, including procedures performed.
* Electronic Health Records that capture CPT-4, ICD-9-CM (Procedures), and HCPCS procedures as orders.

| **Field** | **Data Type** | **Required** | **Description and Notes** |
| --- | --- | --- | --- |
| PROCEDURE\_OCCURRENCE\_ID | INTEGER | YES | A system-generated identifier to uniquely identify each Procedure Occurrence. |
| PROCEDURE\_DATE | DATE | YES | The date on which the Procedure was performed. |
| PERSON\_ID | INTEGER | YES | A system-generated foreign key identifier for the Person who is the subject of the Drug Exposure. The demographics for the Person are captured in the PERSON entity. |
| PROCEDURE\_CONCEPT\_ID | INTEGER | YES | A foreign key to the standard Procedure Concept Code in the Dictionary.  The Procedure Code and/or description from the source data is mapped to a standard Procedure Concept in the Dictionary and is used to map to standard Procedure Vocabulary and Concept hierarchy. The corresponding Concept code is stored here as a reference. |
| SOURCE\_PROCEDURE\_CODE | VARCHAR(6) | YES | The Procedure Code as captured from the source data. Values include CPT-4, ICD-9-CM (Procedure), HCPCS, and other procedure codes. |
| PROCEDURE\_OCCURRENCE\_TYPE | VARCHAR(3) | YES | A foreign key to the standard Procedure Occurrence Type in the PROC\_OCCURRENCE\_REF table.  The source procedure is used to define the Procedure Occurrence Type that was recorded, as well as the type of occurrence. [Appendix E](#Appendix_E) reflects all Procedure Occurrence Types that have been recorded. |

#### 

#### Business Rules

Procedure Occurrences are recorded for each procedure performed on a Person. Each procedure is standardized by assigning a Concept code corresponding to the definition of the procedure code and code type used.

#### Example of Loaded Table

Consider the following example procedure data extracted from Electronic Health Records.

| **Person ID** | **CPT-4 Code** | **Problem Description** | **Procedure Date** |
| --- | --- | --- | --- |
| 127260 | 71020 | Chest X-Ray | 03-MAY-2003 |
| 127260 | 93925 | Lower Extremity Arterial Duplex, Bilateral | 29-JUL-2003 |
| 127260 | 72110 | X-ray exam of lower spine | 23-AUG-2003 |

The following Concept codes correspond to the definition of the procedures that were captured in the source data:

| **Concept ID** | **Concept Code** | **Concept Description** |
| --- | --- | --- |
| 2211361 | 71020 | Chest X-Ray |
| 2313985 | 93925 | Lower Extremity Arterial Duplex Bilateral |
| 2211398 | 72110 | X-ray exam of lower spine |

Procedure Occurrence types are determined for Procedures extracted from the EHR Order list as follows:

|  |  |
| --- | --- |
| **Procedure Occurrence Type** | **Procedure Occurrence Type Description** |
| 27 | EHR - Order List |

The above data, represented in the CDM PROCEDURE\_OCCURRENCE table, follows:

| **Procedure Occurrence ID** | **Procedure Date** | **Person ID** | **Procedure Concept ID** | **Source Procedure Code** | **Procedure Occurrence Type** |
| --- | --- | --- | --- | --- | --- |
| 5003 | 03-MAY-2003 | 127260 | 2211361 | 71020 | 27 |
| 5004 | 29-JUL-2003 | 127260 | 2313985 | 93925 | 27 |
| 5005 | 23-AUG-2003 | 127260 | 2211398 | 72110 | 27 |

### PROC\_OCCURRENCE\_REF

The Procedure Occurrence Reference defines the indicators from which the Procedure Occurrence is drawn or inferred, and indicates whether a Procedure was primary or secondary and their relative positioning within a Person Procedure record.

A detailed listing of the Procedure Occurrence Type codes and their associated descriptions are listed in [Appendix E](#Appendix_E).

| **Field** | **Data Type** | **Required** | **Description and Notes** |
| --- | --- | --- | --- |
| PROCEDURE\_OCCURRENCE\_TYPE | VARCHAR(3) | YES | The code indicating the type of Procedure Occurrence.  The Procedure Occurrence Type is used to define the source data indicators for which the Procedure was identified, the level of standardization, and the type of procedure.  A detailed listing of the Procedure Occurrence Types is recorded in [Appendix E: Procedure Occurrence Type Codes](#Appendix_E). |
| PROCEDURE\_OCCURRENCE\_TYP\_DESC | VARCHAR(120) | YES | A detailed description of Procedure Occurrence Types. Valid values include Procedures from Inpatient Claims, Procedures from Outpatient visits, etc. |
| PROCEDURE\_OCCURRENCE\_POSITION | VARCHAR(20) | YES | The sequential position of the occurrence when multiple positions are observed. |

Procedure Occurrence References have been populated based on Thomson and GE References. Different databases might need additional list of data to reflect information that is not currently captured.

### OBSERVATION

The Observation entity contains all general observations from the following categories:

* Lab observations (i.e., test results) from Medical Claims
* Lab and other observations from Electronic Health Records
* A Person chief complaint as captured in Electronic Health Records
* Other observations from various data sources that cannot be otherwise categorized within the entities provided (Drug, Condition, Procedure, Visit)

| **Field** | **Data Type** | **Required** | **Description and Notes** |
| --- | --- | --- | --- |
| OBS\_OCCURRENCE\_ID | INTEGER | YES | A system-generated identifier to uniquely identify each Observation. |
| PERSON\_ID | INTEGER | YES | A system-generated foreign key identifier for the PERSON who is the subject of the Drug Exposure. The demographics for the Person are captured in the PERSON entity. |
| SOURCE\_OBS\_CODE | VARCHAR(20) | YES | The Observation Code as it appears in the source data. |
| OBS\_CONCEPT\_ID | INTEGER | YES | A foreign key to the standard Observation Concept Code in the Dictionary  The source Observation is mapped to a standard Observation Concept and the corresponding Concept id is stored here as a reference.  It is used to map to standard conditions Vocabulary and Concept hierarchy in the Dictionary. |
| OBS\_VALUE\_AS\_NUMBER | NUMBER(14,3) | NO | The Observation result stored as a numeric value. This is applicable to Observations where the result is expressed as a numeric value. |
| OBS\_DATE | DATE | YES | The date of the Observation. |
| OBS\_RANGE\_LOW | NUMBER(14,3) | NO | The lower limit of the numeric range of the Observation Value. 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. |
| OBS\_RANGE\_HIGH | NUMBER(14,3) | NO | The upper limit of the numeric range of the Observation Value. 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. |
| OBS\_TYPE | VARCHAR(3) | YES | A foreign key to an Observation Type result stored as Concept Code in the Dictionary. |
| OBS\_VALUE\_AS\_STRING | VARCHAR(60) | NO | The Observation result stored as character string. It is applicable to Observations where the result is expressed as a character string. |
| OBS\_VALUE\_AS\_CONCEPT\_ID | INTEGER | NO | A foreign key to an Observation result stored as Concept Code in the Dictionary.  Applicable to Observations where the result can be expressed as a standard Concept from the Dictionary (e.g., positive/negative, present/absent, low/high, etc.). |
| OBS\_UNITS\_CONCEPT\_ID | INTEGER | NO | Unit of measure used for Observation result when measured as a numeric value. A foreign key referencing the units that are stored as a Concept code in the Dictionary. |

### 

#### Business Rules

The approach to extraction and representation of Observation data are based on the individual data source, but the following guidelines are common to all data sources.

* Source attribute values mapped to Observations are checked for standardization. If the source attribute values are available as national or international standard codes (e.g. LOINC codes) then they are mapped to standard Concepts in the Dictionary.
* If the source data are not coded to a national or international standard then a finite listing of attribute values is created and mapped to standard Observation Concepts in the Dictionary.
* The type of result recorded for the Observation is important for further processing of the Observation data. Knowledge of whether an Observation result is captured as a numeric value (with the range of values considered normal), standard Concept code or non-standard text will determine the handling of the Observation data.
* An Observation Type is assigned based on the type of source data from which the Observation is extracted and type of result expected. More details regarding the Observation types appear in the OBSERVATION\_TYPE\_REF section of this chapter.
* Each Observation for every Person, along with its matching standard Concept code from the Dictionary, is extracted from the source data along with the Person identifier. Related attributes including date of the Observation, type of observation, type of result, result as a number/text/Concept code, and reference range for numeric results are also extracted.

Special handling is necessary for source data in which Person condition entries are updated by expiration of the current entry and addition of an updated entry. In such cases, only the final version of the record is extracted for transformation and loading into the CDM.

#### Example of Loaded Table

Consider the following example Observation data extracted from Lab Claims and Observations.

| **Person ID** | **Observation Code** | **Observation Description** | **Observation Date** |
| --- | --- | --- | --- |
| 127260 | 13457-7 | LDL CHOLESTEROL | 03-MAY-2003 |
| 127260 | 6690-2 | White Blood Count | 29-JUL-2003 |
| 127260 | 4766 | Smoking Status | 23-AUG-2003 |

The following Concept codes correspond to the observations that were captured in the source data.

| **Concept ID** | **Concept Code** | **Concept Description** |
| --- | --- | --- |
| 3028288 | 13457-7 | Lipid Panel – LDL Check |
| 3000905 | 6690-2 | White Blood Count Check |
| 4275495 | 4766 | Smoking Status |

The following Concept codes correspond to the meanings of the units of measure associated with the lab observations in the source data:

| **Concept ID** | **Concept Code** | **Concept Description** |
| --- | --- | --- |
| 8840 | mg/dL | mg/DL |
| 8784 | {cells}/uL | Cells per microliter |

The above data, represented in the CDM OBSERVATION table, follows.

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Obs Occurrence ID** | **Person ID** | **Source Obs Code** | **Obs Concept ID** | **Obs Value as Number** | **Obs Date** | **Obs Range High** | **Obs Range Low** | **Obs Type** | **Obs Value as String** | **Obs Value as Concept ID** | **Obs Units Concept ID** |
| 5003 | 127260 | 13457-7 | 3028288 | 124 | 03-MAY-2003 | 130 | 0 | LON |  |  | 8840 |
| 5004 | 127260 | 6690-2 | 3000905 | 6000 | 29-JUL-2003 | 10000 | 4500 | LON |  |  | 8784 |
| 5005 | 127260 | 4766 | 4275495 |  | 23-AUG-2003 |  |  | LOT | PASSIVE SMOKER |  |  |

### 

### OBSERVATION\_TYPE\_REF

Assignment of an Observation type is essential to determine the type of source data, level of standardization, and coding, as well as the type of result recorded for the Observation. The Observation Types include the following.

* Lab Observation Numeric Result
* Lab Observation Text
* Lab Observation Concept Code Result
* Numeric Observations from EHRs (e.g., blood pressure). These are tracked separately and not rolled into other Lab Observation categories
* EHR observations with text results (e.g., reason for visit)
* Chief Complaint

| **Field** | **Data Type** | **Required** | **Description and Notes** |
| --- | --- | --- | --- |
| OBSERVATION\_TYPE | VARCHAR(3) | YES | A code representing the type of Observation that was recorded. |
| OBSERVATION\_TYPE\_DESC | VARCHAR(255) | YES | A detailed description of the type of Observation recorded. |

#### 

Observation Type References have been populated based on Thomson and GE References. Different databases might need additional list of data to reflect information that is not currently captured.

#### Data in the OBSERVATION\_TYPE\_REF Table

The Observation Types identified so far follow.

| **Observation Type** | **Observation Type Description** |
| --- | --- |
| LON | Lab Observation Numeric Result |
| LOT | Lab Observation Text |
| LOC | Lab Observation Concept Code Result |
| EHR | Observation recorded from Electronic Health Records |
| TEM | Observation recorded from Electronic Health Records with text results |
| CHC | Chief Complaint |

### OBSERVATION\_PERIOD

The Observation Period entity is designed to track Person status during a period of study. Person Status is mapped to the corresponding Concept from the OMOP Vocabulary and the corresponding Concept code is stored for each status entry during the period of study. Observation Period entity is also used to track Prescription/Medication coverage availability for the Person during the period of study.

Knowing the Person Status during a period of time under study will help refine active drug surveillance by accounting for changes in the Person’s medical coverage data availability. However, not all Person Status details are available from data sources.

| **Field** | **Data Type** | **Required** | **Description and Notes** |
| --- | --- | --- | --- |
| OBSERVATION\_PERIOD\_ID | INTEGER | YES | A system-generated identifier to uniquely identify each Observation Period. |
| OBSERVATION\_PERIOD\_START\_DATE | DATE | YES | The start date of the Observation Period for which Person history data was available from the data provider. |
| OBSERVATION\_PERIOD\_END\_DATE | DATE | YES | The end date of the Observation Period for which Person history data was available from the data provider. |
| PERSON\_ID | INTEGER | YES | A system-generated foreign key identifier for the PERSON who is the subject of the Drug Exposure. The demographics for the Person are captured in the PERSON entity. |
| PERSON\_STATUS\_CONCEPT\_ID | INTEGER | NO | A foreign key to a Person Status stored as Concept Code in the Dictionary.  The represents the clinical status of the Person with valid values that include Active, Deceased, and Unknown. |
| RX\_DATA\_AVAILABILITY | VARCHAR(1) | NO | A flag to indicate whether medication/prescription coverage was available for the person during the time period covered by the study.  The valid values, not available from all sources, are:  Y: Drug Coverage available  N: No Drug Coverage  U: Unknown |

#### 

#### Business Rules

Tracking Person Status during an Observation Period requires unique handling for each raw data source from which Person data are extracted.

* The status of a Person is determined to ensure that the Person was active during the Observation Period. Determination from this step will enable tracking of changes in the Person Status related to non-eligibility for medical coverage or mortality.
* For data sources in which the status of a Person for each calendar month or year is recorded as a separate entry, even if there are no changes, a single consolidated Person Status entry is recorded in the Common Data Model.
* Medication/prescription coverage is tracked for the period of study for data sources in which it is applicable.
* Understanding of the availability of Person data during the Observation Period will help isolate time periods when reliable data for a Person is not available in the data source being used for the analysis.

#### Example of Loaded Table

Consider the following example data extracted from Medical Claims.

| **Person ID** | **Person Status** | **Rx Coverage Indicator** | **Start Date** | **End Date** |
| --- | --- | --- | --- | --- |
| 127260 | Active | Available | 01-JAN-2003 | 30-SEP-2003 |
| 127260 | Active | Available | 01-OCT-2003 | 31-JAN-2003 |
| 127260 | Deceased | N/A | 01-JAN-2004 |  |

The following Concept codes correspond to the meanings of the Person Status values that were present in the source data.

| **Concept ID** | **Person Status Concept Code** | **Person Status Concept Code Description** |
| --- | --- | --- |
| 9181 | 55561003 | Person Active |
| 9176 | 18632008 | Person Deceased |

The above data are consolidated and represented in the OBSERVATION\_PERIOD table as follows:

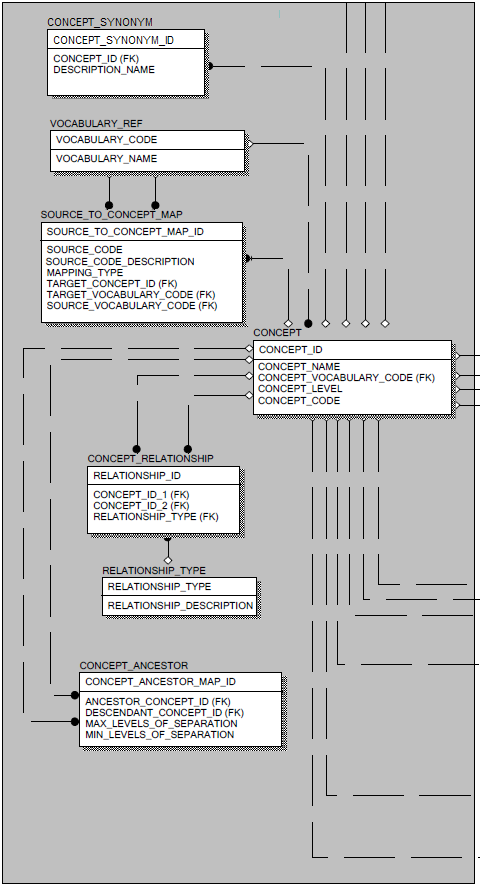
| **Observation Period ID** | **Observation Start Date** | **Observation End Date** | **Person ID** | **Person Status Concept ID** | **Rx Data Availability** |
| --- | --- | --- | --- | --- | --- |
| 80001 | 01-JAN-2003 | 31-JAN-2003 | 127260 | 9181 | Y |
| 80003 | 01-JAN-2004 |  | 127260 | 9176 |  |

## Logical Entities and Attributes – Standard Dictionary

### Dictionary Logical Data Model

The logical data model for the Standard Dictionary is shown in *Figure 7: Dictionary Logical Data Model*:

*Figure 7: Dictionary Logical Data Model*



**Assumptions**:

* There is one design which will accommodate all different source Terminologies and Classifications
* Only information relevant for the OMOP research is loaded into design
* Source Vocabulary data needs to be transformed prior to loading CDM data.
* Source Concepts are preserved, but the Source relationships might be adapted somewhat.
* Concept hierarchy contains both individual relationships as well as Ancestor and Child relationships to simplify building of Drug and Condition Eras.

**Pros:**

* Preservation of Source relationships but not Source data structure.
* Simpler design, unified access, and optimization of performance for OMOP analysis.
* Navigation does not require any knowledge of Source Vocabulary.
* Consistent approach that can be adopted into any future Vocabulary integration.

**Cons:**

* Requires extensive transformation of Source data to conform to consistent design, as well as greater initial effort to load Source Vocabulary.
* Not every Source data structure and Original Source hierarchy is retained.

### CONCEPT

In the context of the Standard Dictionary, a Concept is a basic unit of medical information that is identified by a unique static identifier. Concepts can represent broad categories (like “Cardiovascular disease”), detailed clinical elements (”Myocardial infarction of the anterolateral wall”) or characteristics and relationships that define Concepts at various levels of detail (severity of a disease, associated morphology, etc.).

Records in the Concept tables are derived from standard national or international vocabularies such as SNOMED-CT, NDF-RT, and MedDRA, or custom Concepts defined to cover various aspects of observational data analysis. The detailed description of all source vocabularies, their implementation, the definitions of the relationships, the choice of hierarchical relationships to define ancestry between concepts as well as the mapping from non-standard vocabularies into the standard vocabularies is described in a separate specification document, the Terminology Specification.

NOTE: OMOP will reserve the concept identifiers within the range of 10-12 million for assignment by data partners for their source concepts.

| **Field** | **Data Type** | **Required** | **Description and Notes** |
| --- | --- | --- | --- |
| CONCEPT\_ID | INTEGER | YES | A unique system generated identifier for the Concept. |
| CONCEPT\_NAME | VARCHAR(256) | YES | A unique and unambiguous name for the Clinical Concept. The name is descriptive of the Clinical Concept. |
| CONCEPT\_VOCABULARY\_CODE | VARCHAR(3) | YES | A code indicating the source Vocabulary from which the Concept has been adapted.  The abbreviated code can be a maximum of 3 characters. |
| CONCEPT\_LEVEL | INTEGER | No | The Level of hierarchy associated with the Concept for Concepts with a clearly defined hierarchy. The Concept Level is used for Drug and Condition Concepts with a clearly defined hierarchy, where lower level Concepts need to be aggregated into Ancestor Concepts for observational analysis. |
| CONCEPT\_CODE | VARCHAR(20) | YES | The Concept code for the Clinical Concept as represented in the Vocabulary source from which it has been adapted.  Examples include SNOMED-CT Concept ID, RxNorm RXCUI etc. |
| CONCEPT\_CLASS | VARCHAR(60) | NO | The type of concept both along the hierarchical tree as well as between different domains within the Standard Dictionary. Examples are “Clinical Drug or Pack”, “Ingredient”, “Clinical Finding” etc. |

#### Business Rules

Concepts in the Common Data Model are derived from any of the standard terminologies such as SNOMED-CT and MedDRA, or custom generated to standardize aspects of observational data. Both standard and custom Concepts are integrated based on the following rules:

* For all Concepts, whether they are custom generated or adopted from published terminologies, a unique numeric identifier is assigned and used as the key to link all observational data to the corresponding Concept reference data.
* For Concepts adopted from published terminology (such as SNOMED-CT), only active and current Concepts are retained. Concept status, history, and mapping of current to obsolete Concepts are outside the scope of observational analysis.
* A descriptive name for each Concept is stored as the Concept name as part of the Concept entity. Additional names and descriptions for the Concept are stored as Synonyms in the Concept Synonym entity.
* For standard Concepts inherited from published terminologies, the source Concept code is retained as part of the Concept reference data and used to reference the source terminology.
* All logical data elements associated with a clinical Concept, including defining characteristics, qualifying attributes, etc., are also stored as Concepts. These elements are linked to the entities to which they are applicable using Concept relationships.

#### Example of Loaded Table

Following examples illustrate the representation of some of the Concepts from standard terminologies in the Common Data Model.

1. For SNOMED-CT Clinical finding “Chronic Atrial Fibrillation”, Concept information is as follows:

Vocabulary Source: SNOMED-CT

Source Concept identifier: 426749004

Concept Name: Chronic atrial fibrillation (disorder)

A unique numeric identifier is generated and used as the CONCEPT\_ID for the Concept.

Concept ID: 4141360

Vocabulary code related to SNOMED-CT is determined from the Vocabulary reference table.

Concept Vocabulary Code: 01

For Concepts that share ontology with a well defined hierarchy, the Concept Level attribute is used to store the location of the Concept within the Concept hierarchy.

Concept Level: 2

The concept class defines for each concept the class it is in. Concept Classes depend on the Vocabulary Domain (Drug, Condition etc.) and the hierarchical position:

Concept Class: Clinical Finding

The Concept data is represented in the Concept entity as follows:

| **Concept ID** | **Concept Code** | **Concept Name** | **Concept Vocabulary Code** | **Concept Level** | **Concept Class** |
| --- | --- | --- | --- | --- | --- |
| 4141360 | 426749004 | Chronic atrial fibrillation (disorder) | 01 | 2 | Clinical Finding |

1. For RxNorm drug Concept “Celecoxib”, Concept data is represented in the CDM as follows:

Source Vocabulary: (RxNorm)

Source Concept identifier: 140587

Concept Name: celecoxib

A unique numeric identifier is generated and used as the CONCEPT\_ID for the Concept.

Concept ID: 1118084

Vocabulary code related to SNOMED-CT is determined from the Vocabulary reference table.

Concept Vocabulary Code: 08

For Concepts that share ontology with a well defined hierarchy, the Concept Level attribute is used to store the location of the Concept within the Concept hierarchy.

Concept Level: 2

The concept class defines for each concept the class it is in. Concept Classes depend on the Vocabulary Domain (Drug, Condition etc.) and the hierarchical position.

Concept Class: Ingredient

| **Concept ID** | **Concept Code** | **Concept Name** | **Concept Vocabulary Code** | **Concept Level** | **Concept Class** |
| --- | --- | --- | --- | --- | --- |
| 1118084 | 140587 | Celecoxib | 08 | 2 | Ingredient |

### CONCEPT\_SYNONYM

The Synonym entity is used to store all alternate names and descriptions for a Concept. Each Synonym is assigned its own unique identifier and contains the text of a description and the identifier of the Concept that it represents.

Each Concept may include zero, one, or more Synonyms in the Synonym table. As an example, for a SNOMED-CT Concept, if the fully specified name is stored as the Concept name in the Concept entity, then the Preferred Term and Synonyms associated with the Concept are stored in the Synonym entity. Synonyms are used to express Descriptions that may denote the same basic Concept but are expressed in different terms. Only synonyms that are active and current are stored in the Synonym entity. Tracking synonym/description history and mapping of obsolete synonyms to current Concepts/Synonyms is out of scope for observational analysis. Synonyms entities are stored in the Concept\_Synonym table.

| **Field** | **Data Type** | **Required** | **Description and Notes** |
| --- | --- | --- | --- |
| CONCEPT\_SYNONYM\_ID | INTEGER | YES | A unique identifier for the Clinical Description. |
| CONCEPT\_ID | INTEGER | YES | A foreign key to a Concept stored in the Concept table of the Dictionary. |
| DESCRIPTION\_NAME | VARCHAR(1000) | YES | The text of the term used to describe the Concept associated with the Description. |

### 

#### Example of Loaded Table

The following example illustrates the representation of Descriptions associated with a SNOMED-CT Concept “Chronic Atrial Fibrillation’. The Concept can be described in the following ways:

Concept Code: 426749004

Concept Fully Specified Name: Chronic atrial fibrillation (disorder)

Preferred Term for the Concept: Chronic atrial fibrillation

Source Vocabulary: SNOMED\_CT

Description Status: Current for both Preferred Term and Fully Specified Name

If the Fully Specified Name for the Concept is used as the Concept Name in the Concept entity, then the preferred term is stored as a Synonym in the Synonym entity. The particulars for the Synonym being added are as follows:

Concept ID: 3409069

Concept Synonym ID: 3351850

Description Name: Chronic atrial fibrillation

| **Concept Synonym ID** | **Concept ID** | **Description Name** |
| --- | --- | --- |
| 3409069 | 4141360 | Chronic atrial fibrillation |

### CONCEPT\_RELATIONSHIP

The Concept Relationship entity stores the relationship between two Concepts. The relationships described in the CDM are directional and are intended to include a Source Concept and a Target Concept with an explicit relationship from Source to Target.

Concept Relationship includes many different types of relationships between Concepts. The type of relationships described between Concepts includes the following:

* Hierarchical or ”ISA” Relationships (Subsumes relationships), which are used to define a direct hierarchies between Concepts. Only the hierarchy relationships between two closest Concepts are captured in the Relationship table. All the Ancestor and descendent relationships that are not immediate parent-child relationships are recorded in the Concept Ancestor table. The hierarchical relationships can be between Concepts that are adopted from the same Vocabulary source or between Concepts adopted from difference Vocabulary sources. The Relationship Type attribute is used to identify various types of relationships that are captured in the Concept Relationships table.
* Object-Attribute-Value (OAV) relationships, that are used to define various defining and qualifying attributes/characteristics associated with a Concept. The OAV relationships include the Concept Code (Object), Relationship Type (Attribute), and Attribute Value (another Concept).

Hierarchical relationships where the order in which the subtype/child Concepts are displayed is important are not handled in the Relationships table.

The parent-child Concepts hierarchy in the Standard Dictionary is of special importance to OMOP, as it allows researchers to query a CDM instance for classes of Concepts without needing to know the underlying subclasses. For example, a researcher will be able to query a CDM instance for all drugs within a specific therapeutic class without needing to know the specific Concepts codes of each drug within the class, and will be able to query a CDM instance for a particular medical condition without the necessity to know the individual indicators (i.e., diagnoses) of that condition. The researcher would search or browse the Dictionary to find the class of Concepts on which to query or analyze, then transfer the appropriate Concept Code of that class from the Dictionary to the query and analysis tool.

| **Field** | **Data Type** | **Required** | **Description and Notes** | |
| --- | --- | --- | --- | --- |
| RELATIONSHIP\_ID | INTEGER | YES | A unique identifier for the Concept Relationship. | |
| CONCEPT\_ID\_1 | INTEGER | YES | A foreign key to the Reference Concept associated with the Relationship, stored in the Concept table of the Dictionary.  As the relationships are directional, the source and target Concept designation is clearly recorded.  In case of relationship, this would represent the Parent Concept.  For OAV relationships, this would represent the parent Concept. | |
| CONCEPT\_ID\_2 | INTEGER | YES | A foreign key to the target Concept associated with the relationship, stored in the Concept table of the Dictionary.  In case of (subtype) relationships, this would represent the child Concept.  In terms of OAV (Attribute) relationships, this is the Attribute Concept. . |
| RELATIONSHIP\_TYPE | VARCHAR(3) | YES | The type of relationship captured by the Relationship record.  In terms of OAV relationships, the Relationship Type holds the Concept that defines the Attribute. |

#### Example of Loaded Table

The following examples illustrate the representation of Concept relationships in the Common Data Model.

Concept “Chronic Atrial Fibrillation” includes a hierarchic subtype relationship with the “Atrial Fibrillation” Concept and an OAV (attribute) relationship with a defining characteristic “Severity” Concept. The relationship is represented as follows:

Concept ID1: 4141360

Source Concept Code 1: 426749004

Concept Description: Chronic atrial fibrillation (disorder)

Concept ID2: 313217

Concept Code 2: 49436004

Concept Description 2: Atrial Fibrillation

Relationship Type: 010 (“Is a” or “Subsumes” or “subtype” relationship representing hierarchy)

Source Vocabulary Code: SNOMED-CT

Concept ID1: 4141360

Concept Code 1: 426749004

Concept Description: Chronic atrial fibrillation (disorder)

Concept ID2: 4153899

Concept Code 2: 272141005

Concept Description 2: Severities

Relationship Type: 034 (Severity)

Source Vocabulary Code: SNOMED-CT

Unique identifiers are generated for both Relationships and the data is represented in the CDM as a Relationship as follows:

| **Relationship ID** | **Concept ID 1** | **Concept ID 2** | **Relationship Type** |
| --- | --- | --- | --- |
| 1 | 4141360 | 313217 | 010 |
| 2 | 4141360 | 4153899 | 034 |

### 

### CONCEPT\_ANCESTOR

Concept Ancestor entity is a custom entity designed to simplify observational analysis by consolidating the hierarchical relationship between various Concepts. Parent-child relationships between Concepts are stored in the Concept Relationship table. However, it is stored in a form that is harder to navigate due to the interlocking nature of the relationships and the multiplicity of parent-child relationships for many Concepts.

The Ancestor – Descendant relationship captures hierarchical relationships between Ancestor and any Descendant Concepts, along with indicators for the shortest and longest navigation path between them.

The Ancestor relationship is primarily targeted at observational analysis that would involve:

* Rollup of lower level Concepts into higher level aggregation Concepts.
* Collection of all lower level Concepts in the hierarchy that follow from a high level Concept.

To help simplify the navigation of the hierarchy, the Concept Ancestor captures Ancestor-Descendant relationships between all Concepts linked through a hierarchy, not just immediate parent-child relationships. Two attributes, namely Maximum levels of separation and Minimum levels of separation, are also captured as indicators of the range of the levels of hierarchy that separate them. The two attributes indicate the number of traversals necessary to navigate the hierarchy from Ancestor to Descendant and also offer an insight into existence of alternate navigation paths of varying distance.

| **Field** | **Data Type** | **Required** | **Description and Notes** |
| --- | --- | --- | --- |
| CONCEPT\_ANCESTOR\_MAP\_ID | INTEGER | YES | A unique identifier for the Concept Ancestor relationship. |
| ANCESTOR\_CONCEPT\_ID | INTEGER | YES | A foreign key to the Concept Code for the higher level Concept that forms the Ancestor in the relationship. |
| DESCENDANT\_CONCEPT\_ID | INTEGER | YES | A foreign key to the Concept Code for the lower level Concept that forms the Descendant in the relationship. |
| MAX\_LEVELS\_OF\_SEPARATION | INTEGER | NO | The maximum separation in number of levels of hierarchy between Ancestor and Descendant Concepts.  This is an optional attribute that is used to simplify hierarchic analysis. |
| MIN\_LEVELS\_OF\_SEPARATION | INTEGER | NO | The minimum separation in number of levels of hierarchy between Ancestor and Descendant Concepts.  This is an optional attribute that is used to simplify hierarchic analysis. |

#### 

#### Example of Loaded Table

*Figure 8: Ancestor* illustrates the representation of the Ancestor – Descendant relationships between Concepts in the Common Data Model.

The Non-steroidal anti-inflammatory drugs (NSAID) hierarchy is represented in the following diagram based on hypothetical drug ontology.

*Figure 8: Ancestor*



Based on the ontology described in *Figure 8: Ancestor*, two of the Ancestor – Descendant relationships could be captured as follows:

Relationship 1:

Ancestor Concept: “Non Steroidal Anti-inflammatory drugs”

Ancestor Concept ID: 16403005

Descendant Concept: “Naproxen”

Descendant Concept ID: 4186860

Maximum Levels of separation: 1

Minimum Levels of separation: 1

Relationship 2:

Ancestor Concept: “Non Steroidal Anti-inflammatory drugs”

Ancestor Concept ID: 16403005

Descendant Concept: “Celecoxib”

Descendant Concept ID: 4021058

Ancestor Level to Root: 2

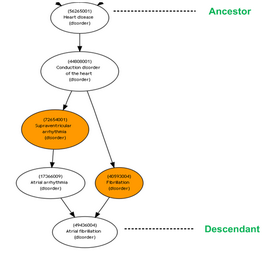
Ancestor Maximum Levels to Leaf: 2

Consolidated, the data represented in the Common Data Model as follows:

| **Concept Ancestor Map ID** | **Ancestor Concept ID** | **Descendant Concept ID** | **Max Levels of Separation** | **Min Levels of Separation** |
| --- | --- | --- | --- | --- |
| 1 | 16403005 | 4186860 | 1 | 1 |
| 2 | 16403005 | 4021058 | 2 | 2 |

*Figure 9: Ancestor to Descendant* displays an example of a partial hierarchy, where navigating the hierarchy from the Ancestor to the Descendant involves multiple paths.

Figure 9: Ancestor to Descendant



Details of the examples are as follows:

Ancestor: SNOMED-CT Concept 56265001 “Heart disease (disorder)”

Ancestor Concept ID: 321588

Descendant: SNOMED-CT Concept 49436004 “Atrial Fibrillation (disorder)”

Descendant Concept ID: 4344544

Max Levels of Separation: 3

Min. Levels of Separation: 2

| **Concept Ancestor Map ID** | **Ancestor Concept ID** | **Descendant Concept ID** | **Max Levels of Separation** | **Min Levels of Separation** |
| --- | --- | --- | --- | --- |
| 1 | 321588 | 4344544 | 3 | 2 |

Variance in minimum and maximum levels is an indicator of the complex traversal path that offers alternate navigation paths.

### VOCABULARY\_REF

The Vocabulary Reference entity includes a list of all standard terminologies from which Concepts have been extracted for observational analysis using the Common Data Model. The reference table is populated with a single record for each Vocabulary source and includes a descriptive name for the Vocabulary source.

| **Field** | **Data Type** | **Required** | **Description and Notes** |
| --- | --- | --- | --- |
| VOCABULARY\_CODE | VARCHAR(3) | YES | A unique identifier for each of the Vocabulary sources used in the observational analysis. |
| VOCABULARY\_NAME | VARCHAR(256) | YES | An elaborative name for each of the Vocabulary sources |

#### 

#### Example of Loaded Table

For a complete listing of this reference table see separate Standard Terminology Specifications Document.

### SOURCE\_TO\_CONCEPT\_MAP

Observational data elements that need to be standardized into Concepts require mapping tables to translate source identifiers into standard Concept codes. The mapping entity serves as a lookup table that stores a cross-reference between source identifiers and standard Concept Codes. Concept mappings need to be created for both:

* Mapping of national standard concept identifiers to other national standard Concept identifiers, such as mapping of ICD-9-CM diagnosis codes to SNOMED-CT Clinical finding Concepts or NDC codes to RxNorm Drug Concepts.
* Mapping of source specific or original codes to standard Concepts wherever possible, such as mapping of source specific Problem strings to SNOMED-CT clinical findings or source specific Gender codes to HL7 Sex Concepts.

The mapping table serves the critical function of inferring standard Concepts from source data created with various objectives in mind. The ETL logic that loads the CDM instance from the source data set looks-up Concepts from the Dictionary as it transforms (i.e., reformats) the source and Concepts data to conform to the CDM table structures. The look-up process entails matching the inbound source data (vocabulary code and source coce) against the Source-to-Concept Mappings in the Dictionary to determine which standard Concept Code values are mapped to which Target Concept ID.

| **Field** | **Data Type** | **Required** | **Description and Notes** |
| --- | --- | --- | --- |
| SOURCE\_TO\_CONCEPT\_MAP\_ID | INTEGER | YES | A unique identifier for each Source identifier to Concept mapping record. |
| SOURCE\_CODE | VARCHAR(20) | YES | The source Identifier being translated into a standard Concept. |
| SOURCE\_CODE\_DESCRIPTION | VARCHAR(256) | NO | An optional description for the source identifier. This is included as a convenience so that the description for the source identifier is readily available. |
| MAPPING\_TYPE | VARCHAR(20) | YES | A string identifying the observational data element being translated. Examples include ‘Drug’, ‘Condition’, ‘Procedure’, ‘Procedure Drug’, etc.  It is important to pick the appropriate mapping record when the same source identifier is being mapped to different Concepts for different contexts in observational analysis.  As an example a Procedure code for Drug Administration can be mapped to a Procedure Concept or a Drug Concept depending on the context. |
| TARGET\_CONCEPT\_ID | INTEGER | YES | A foreign key to the Concept to which the source code is being mapped. |
| TARGET\_VOCABULARY\_CODE | VARCHAR(3) | YES | A foreign key to the Vocabulary\_Ref table defining the vocabulary from which the target Concept has been adapted.. |
| SOURCE\_VOCABULARY\_CODE | VARCHAR(3) | YES | A foreign key to the Vocabulary\_Ref table defining the vocabulary that is being mapped to the Standard Dictionary. |

#### 

#### Example of Loaded Table

The following are several examples of sample records in the Source to Concept mapping table.

1. Mapping of national standard identifiers to standard Concepts. In this case the mapping of ICD-9-CM diagnosis code 140. 0 for “Malignant Neoplasm of Upper lip, vermilion border” is mapped to SNOMED-CT Concept 187602007 “Malignant neoplasm of upper lip, vermilion border NOS (disorder)” (Concept ID 4093013).

The Source to Concept mapping can be represented as follows:

| **Source to Concept ID** | **Mapping Type** | **Source Vocabulary Code** | **Source Code** | **Target Concept ID** | **Target Vocabulary Code** |
| --- | --- | --- | --- | --- | --- |
| 1 | CONDITION | 02 | 140.0 | 4093013 | 01 |

1. Mapping of source specific identifiers to standard Concepts. In this case a source specific GE – EHR Observation key 3000963 with description of “hemoglobin, blood” is being mapped to LOINC Concept 718-7 with description of “Hemoglobin [Mass/volume] in Blood” (Concept ID: 3000963).

| **Source to Concept ID** | **Mapping Type** | **Source Vocabulary Code** | **Source Code** | **Target Concept ID** | **Target Vocabulary Code** |
| --- | --- | --- | --- | --- | --- |
| 2 | OBSERVATION | 51 | 718-7 | 3000963 | 06 |

### 

### RELATIONSHIP\_TYPE

A Concept Relationship is standardized via the Relationship Type entity. The Relationship Type codes are adopted from the various source vocabularies the Standard Dictionary is derived from. Some of the relationships are hierarchical and define ancestry (see above), and others are preserved for the convenience of the researcher from their original source.

| **Field** | **Data Type** | **Required** | **Description and Notes** |
| --- | --- | --- | --- |
| RELATIONSHIP\_TYPE | VARCHAR(3) | YES | The type of relationship captured by the Relationship record. |
| RELATIONSHIP\_DESCRIPTION | VARCHAR(256) | NO | The text that describes relationship type. |

#### Example of Loaded Table

For a complete listing of this reference table and the discussion of its content see separate Standard Terminology Specifications Document.



# Appendix C: Drug Exposure Type Codes

Drug Exposure Types are used to define the indicators from which exposures have been extracted. They also define the characteristics of the exposure and the level of aggregation. The following Drug Exposure Types are allowed.

| **Drug Exposure Type** | **Drug Exposure Type Description** | **Persistence Window (In Days)** |
| --- | --- | --- |
| 1 | Prescription Dispensed |  |
| 2 | Prescription Written |  |
| 3 | Medication List |  |
| 4 | Physician Administered Drug (Identified as Procedure) |  |
| 5 | Inpatient Administration |  |
| 6 | Drug Era – 0 day window | 0 |
| 7 | Drug Era – 30 days window | 30 |

# Appendix D: Condition Occurrence Type Codes

The Condition Occurrence Type reflects the indicator(s) from which the Condition Occurrence is drawn or inferred, whether a condition (diagnosis) was primary or secondary, and their relative positioning within a person condition record.

The following Condition Occurrence Types are allowed.

| **Condition Occurrence Type** | **Condition Occurrence Type Description** | **Condition Occurrence Position** | **Persistence Window (In Days)** |
| --- | --- | --- | --- |
| 1 | Inpatient Detail | Primary |  |
| 2 | Inpatient Detail | 1 |  |
| 3 | Inpatient Detail | 2 |  |
| 4 | Inpatient Detail | 3 |  |
| 5 | Inpatient Detail | 4 |  |
| 6 | Inpatient Detail | 5 |  |
| 7 | Inpatient Detail | 6 |  |
| 8 | Inpatient Detail | 7 |  |
| 9 | Inpatient Detail | 8 |  |
| 10 | Inpatient Detail | 9 |  |
| 11 | Inpatient Detail | 10 |  |
| 12 | Inpatient Detail | 11 |  |
| 13 | Inpatient Detail | 12 |  |
| 14 | Inpatient Detail | 13 |  |
| 15 | Inpatient Detail | 14 |  |
| 16 | Inpatient Detail | 15 |  |
| 17 | Inpatient Header | Primary |  |
| 18 | Inpatient Header | 1 |  |
| 19 | Inpatient Header | 2 |  |
| 20 | Inpatient Header | 3 |  |
| 21 | Inpatient Header | 4 |  |
| 22 | Inpatient Header | 5 |  |
| 23 | Inpatient Header | 6 |  |
| 24 | Inpatient Header | 7 |  |
| 25 | Inpatient Header | 8 |  |
| 26 | Inpatient Header | 9 |  |
| 27 | Inpatient Header | 10 |  |
| 28 | Inpatient Header | 11 |  |
| 29 | Inpatient Header | 12 |  |
| 30 | Inpatient Header | 13 |  |
| 31 | Inpatient Header | 14 |  |
| 32 | Inpatient Header | 15 |  |
| 33 | Outpatient Detail | 1 |  |
| 34 | Outpatient Detail | 2 |  |
| 35 | Outpatient Detail | 3 |  |
| 36 | Outpatient Detail | 4 |  |
| 37 | Outpatient Detail | 5 |  |
| 38 | Outpatient Detail | 6 |  |
| 39 | Outpatient Detail | 7 |  |
| 40 | Outpatient Detail | 8 |  |
| 41 | Outpatient Detail | 9 |  |
| 42 | Outpatient Detail | 10 |  |
| 43 | Outpatient Detail | 11 |  |
| 44 | Outpatient Detail | 12 |  |
| 45 | Outpatient Detail | 13 |  |
| 46 | Outpatient Detail | 14 |  |
| 47 | Outpatient Detail | 15 |  |
| 48 | Outpatient Header | 1 |  |
| 49 | Outpatient Header | 2 |  |
| 50 | Outpatient Header | 3 |  |
| 51 | Outpatient Header | 4 |  |
| 52 | Outpatient Header | 5 |  |
| 53 | Outpatient Header | 6 |  |
| 54 | Outpatient Header | 7 |  |
| 55 | Outpatient Header | 8 |  |
| 56 | Outpatient Header | 9 |  |
| 57 | Outpatient Header | 10 |  |
| 58 | Outpatient Header | 11 |  |
| 59 | Outpatient Header | 12 |  |
| 60 | Outpatient Header | 13 |  |
| 61 | Outpatient Header | 14 |  |
| 62 | Outpatient Header | 15 |  |
| 63 | Problem List |  |  |
| 64 | Condition Era |  | 0 |
| 65 | Condition Era |  | 30 |
| 66 | Death at Discharge |  |  |

# Appendix E: Procedure Occurrence Type Codes

The Procedure Occurrence Type defines the indicators from which the Procedure Occurrence is drawn or inferred, whether a Procedure was primary or secondary, and their relative positioning within a Person Procedure record.

The following Procedure Occurrence Types are allowed.

| **Procedure Occurrence Type** | **Procedure Occurrence Type Description** | **Procedure Occurrence Position** |
| --- | --- | --- |
| 1 | Inpatient Detail | Primary |
| 2 | Inpatient Detail | 1 |
| 3 | Inpatient Header | Primary |
| 4 | Inpatient Header | 1 |
| 5 | Inpatient Header | 2 |
| 6 | Inpatient Header | 3 |
| 7 | Inpatient Header | 4 |
| 8 | Inpatient Header | 5 |
| 9 | Inpatient Header | 6 |
| 10 | Inpatient Header | 7 |
| 11 | Inpatient Header | 8 |
| 12 | Inpatient Header | 9 |
| 13 | Inpatient Header | 10 |
| 14 | Inpatient Header | 11 |
| 15 | Inpatient Header | 12 |
| 16 | Inpatient Header | 13 |
| 17 | Inpatient Header | 14 |
| 18 | Inpatient Header | 15 |
| 19 | Outpatient Detail | Primary |
| 20 | Outpatient Detail | 1 |
| 21 | Outpatient Header | Primary |
| 22 | Outpatient Header | 1 |
| 23 | Outpatient Header | 2 |
| 24 | Outpatient Header | 3 |
| 25 | Outpatient Header | 4 |
| 26 | Outpatient Header | 5 |
| 27 | Outpatient Header | 6 |
| 28 | EHR Order |  |

# 

# Appendix F: Relationship Types Codes

The following Dictionary Relationship Types are allowed:

| **Relationship Type** | **Relationship Description** |
| --- | --- |
| 001 | LOINC Map To |
| 002 | RXNORM Has precise ingredient |
| 003 | RXNORM Has tradename |
| 004 | RXNORM Has dose form |
| 005 | RXNORM Has form |
| 006 | RXNORM Has ingredient |
| 007 | RXNORM Constitutes |
| 008 | RXNORM Contains |
| 009 | RXNORM Reformulation of |
| 010 | Subsumes |
| 011 | NDFRT Has DoseForm |
| 012 | NDFRT Induces |
| 013 | NDFRT May Diagnose |
| 014 | NDFRT Has PE |
| 015 | NDFRT CI PE |
| 016 | NDFRT Has Ingredient |
| 017 | NDFRT CI ChemClass |
| 018 | NDFRT Has MoA |
| 019 | NDFRT CI MoA |
| 020 | NDFRT Has PK |
| 021 | NDFRT May Treat |
| 022 | NDFRT CI With |
| 023 | NDFRT May Prevent |
| 024 | NDFRT Has Active Metabolites |
| 025 | NDFRT Site of Metabolism |
| 026 | NDFRT Effect May Be Inhibited By |
| 027 | NDFRT Has Chemical Structure |
| 028 | NDFRT RXN RELA |
| 101 | OMOP Intermediate Condition Concept To SNOMED |
| 102 | OMOP Intermediate Drug Concept To RxNorm |
| 041 | Indirect morphology |
| 072 | Procedure site - Direct |
| 060 | Scale type |
| 093 | CPT - SNOMED |
| 074 | Procedure device |
| 032 | Pathological process |
| 045 | Has intent |
| 050 | Episodicity |
| 037 | Occurrence |
| 056 | Associated morphology |
| 042 | Indirect device |
| 030 | Procedure site |
| 091 | HLI ICD9CM Procedure to SNOMED Category |
| 083 | Using substance |
| 080 | Surgical approach |
| 043 | Has specimen |
| 089 | HLI ICD-9-CM to SNOMED Category |
| 049 | Finding site |
| 054 | Component |
| 040 | Interprets |
| 039 | Laterality |
| 038 | Method |
| 068 | Subject relationship context |
| 062 | Specimen procedure |
| 036 | Access |
| 065 | Specimen source topography |
| 085 | Clinical course |
| 087 | Finding method |
| 064 | Specimen source morphology |
| 048 | Has active ingredient |
| 058 | Measurement Method |
| 044 | Has interpretation |
| 070 | After |
| 076 | Finding context |
| 053 | Direct device |
| 071 | Associated procedure |
| 073 | Procedure site - Indirect |
| 063 | Specimen source identity |
| 092 | HLI ICD9CM Procedure to SNOMED Specific |
| 035 | Revision status |
| 079 | Associated with |
| 046 | Has focus |
| 031 | Priority |
| 033 | Part of |
| 075 | Procedure morphology |
| 029 | Recipient category |
| 061 | Time aspect |
| 082 | Using energy |
| 057 | Associated finding |
| 069 | Has dose form |
| 067 | Due to |
| 047 | Has definitional manifestation |
| 088 | Finding informer |
| 094 | CPT EQUAL SNOMED |
| 059 | Property |
| 055 | Causative agent |
| 077 | Procedure context |
| 051 | Direct substance |
| 090 | Hli ICD-9-CM to SNOMED Specific |
| 078 | Temporal context |
| 034 | Severity |
| 084 | Using access device |
| 086 | Route of administration |
| 052 | Direct morphology |
| 081 | Using device |
| 066 | Specimen substance |