Introducción al modelo OMOP-CDM

Alberto Labarga – Febrero 2026

Education is the key to unlock the golden door of freedom.

I am Alberto

alabarga

alabarga

/in/albertolabarga

alberto.labarga@gmail.com

Febrero
2026

OMOP-CDM
El curso

El curso tiene como objetivo proporcionar una comprensión profunda del modelo de datos **OMOP-CDM** de la OHDSI, incluida su **estructura**, tablas clave y **vocabularios**.

Abordaremos también como **transformar** datos al modelo y explorar y analizar la **calidad** de los datos transformados.

L	M	X	J	V	S	D
						1
2	3	4	5	6	7	8
9	10	11	12	13	14	15
16	17	18	19	20	21	22
23	24	25	26	27	28	

Objetivos del curso

01 Introducción al modelo OMOP-CDM, a sus tablas y sus vocabularios

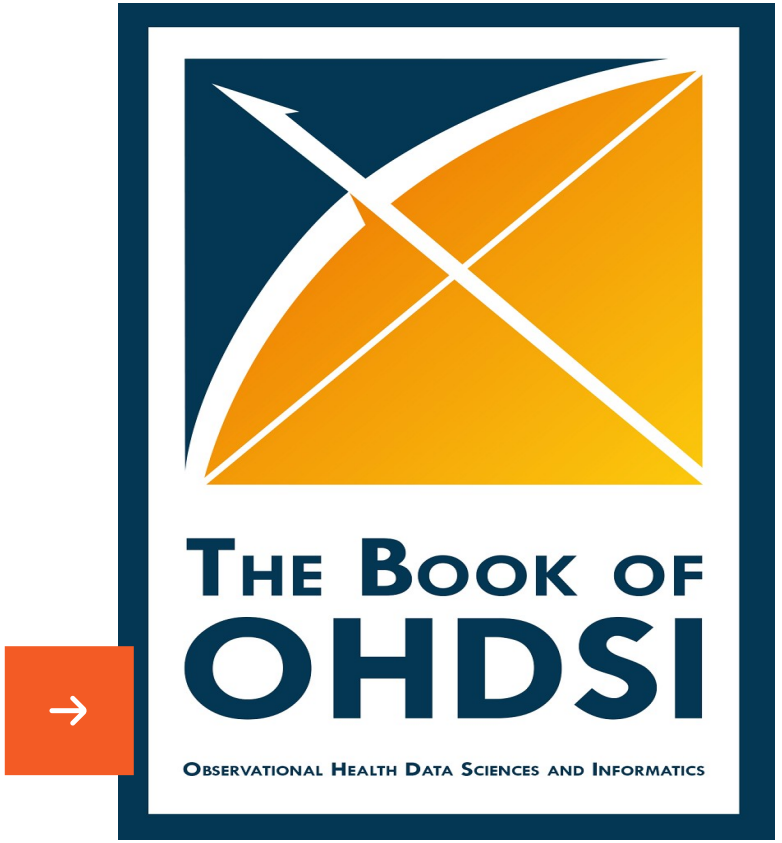
02 Transformación de datos al modelo OMOP-CDM utilizando la suite de herramientas de la OHDSI

03 Análisis de calidad y explotación de una base de datos OMOP-CDM

Recursos de aprendizaje

Sitio web: https://github.com/alabarga/omop-cdm-course

Libro https://ohdsi.github.io/TheBookOfOhdsi/



DATA*

IS THE NEW OIL

Supreme
PWS
Regular
93
89
87

HAZET
VANADIUM
19
16
20

(CLINICAL)DATA*

IS COMPLICATED

CLINICAL DATA
IS COMPLEX

El sector médico es uno de los sectores que genera un gran volumen de datos, de gran heterogeneidad, entre los que se incluyen las historias clínicas electrónicas (EMR), las imágenes médicas o los resultados de pruebas médicas, y muchos otros.

El análisis y la interpretación de estos datos pueden brindar información importante para la mejora de la calidad de la atención médica, la investigación clínica y la toma de decisiones.

A
B

CLINICAL DATA
IS SILOED

La falta de estandarización en los modelos de datos ha creado una serie de problemas. En primer lugar, dificulta la comparación y el intercambio de datos entre diferentes sistemas y organizaciones, lo que hace que sea difícil para los investigadores, clínicos y los responsables políticos obtener una imagen clara y completa de las enfermedades, los tratamientos y su eficacia.

CLOSED

OPEN

FAIR

FAIR
DATA

FINDABLE

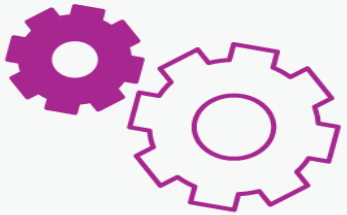
Identificadores únicos y metadatos que pueden usarse para encontrar los datos de manera sencilla

ACCESSIBLE

Los datos son abiertos, libres y universalmente disponibles para la investigación

INTER-OPERABLE

Los datos utilizan formatos libres y estándar utilizados por un amplio número de aplicaciones



REUSABLE

Todos los datos están descritos con formatos estándard y tienen licencias de uso claramente descritas

Armonización
DE DATOS

La **armonización** de datos se refiere al proceso de transformar y estandarizar los datos para que puedan ser comparables y analizados juntos. Al intentar combinar datos de diferentes fuentes, surge el siguiente problema: estos datos suelen tener distintas formas de recopilación, procesamiento e interpretación, lo que hace que los datos estén en diferentes formatos y no tengan las mismas estructuras y convenciones.

Para abordar este problema y alcanzar la armonización de los datos de salud, deben crearse **estándares y protocolos p**ara la recopilación y el análisis de estos, así como el desarrollo de herramientas y técnicas que hagan posible la combinación de los datos

ESTÁNDARES
DE DATOS

Los **estándares de datos** en medicina se refieren a un conjunto de reglas y formatos que se utilizan para definir la **estructura**, el **contenido** y el **intercambio** de datos médicos entre sistemas y aplicaciones de salud.

Estos estándares son necesarios para garantizar la **calidad y la interoperabilidad** de los datos médicos, y para permitir que el personal de atención médica y los pacientes compartan información de manera eficiente y segura.

¿QUÉ ESTÁNDARES CONOCEIS?

OHDSI

OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS

FHIR

openEHR

Estándares
de formato

Los estándares de formato establecen cómo se debe estructurar y presentar la información médica en un formato estandarizado. Ejemplos de este tipo de estándares en medicina son el **Health Level 7 (HL7)**, que se utiliza para **intercambiar información** entre sistemas de información médica, y **Digital Imaging and Communications in Medicine (DICOM)**, que se utiliza para almacenar y compartir imágenes médicas.

Estándares
de interoperabilidad

Los estándares de interoperabilidad son un conjunto de normas, directrices y especificaciones técnicas que establecen la forma en que diferentes sistemas de información y tecnologías deben comunicarse y transmitir datos entre sí. Este tipo de estándares es imprescindible para asegurar que los distintos sistemas u orígenes de datos pueden trabajar entre ellos. Ejemplos de este tipo de estándares son **Patient Summary**, así como el modelo de datos **OMOP-CDM**. Estñan orientados a la **explotación** de los datos.

Estándares
de modelado clínico

Los estándares de modelado clínico y semántico definen cómo debe representarse, estructurarse y dotarse de significado la información clínica de forma explícita y reutilizable, separando el conocimiento clínico de la implementación técnica. Un ejemplo representativo de este tipo de estándar es **openEHR**, que utiliza **modelos de referencia, arquetipos y plantillas** para garantizar la coherencia, comparabilidad y reutilización de los datos clínicos a lo largo del tiempo y entre diferentes sistemas. Están principalmente orientados a la **persistencia** de los datos.

Estándares
de codificación

La función de estos estándares es convertir información médica, como diagnósticos, procedimientos, medicamentos y resultados de laboratorio, en un formato particular para que pueda ser transmitida, almacenada o procesada de manera eficiente. Algunos de los estándares de codificación más comunes son el Sistema **Internacional de Clasificación Internacional de Enfermedades (CIE)** y el Código Universal para la **Identificación de Observaciones Clínicas y de Laboratorio (LOINC)**.

The Observational Medical Outcomes Partnership Common Data Model (**OMOP-CDM**) ha sido desarrollado por y para la comunidad científica, proporcionándo las herramientas para generar evidencia de alta calidad y reproducible a partir de datos de salud observacionales. Los datos pueden ser extraídos, transformados y cargados directamente desde los sistemas de información clínica.

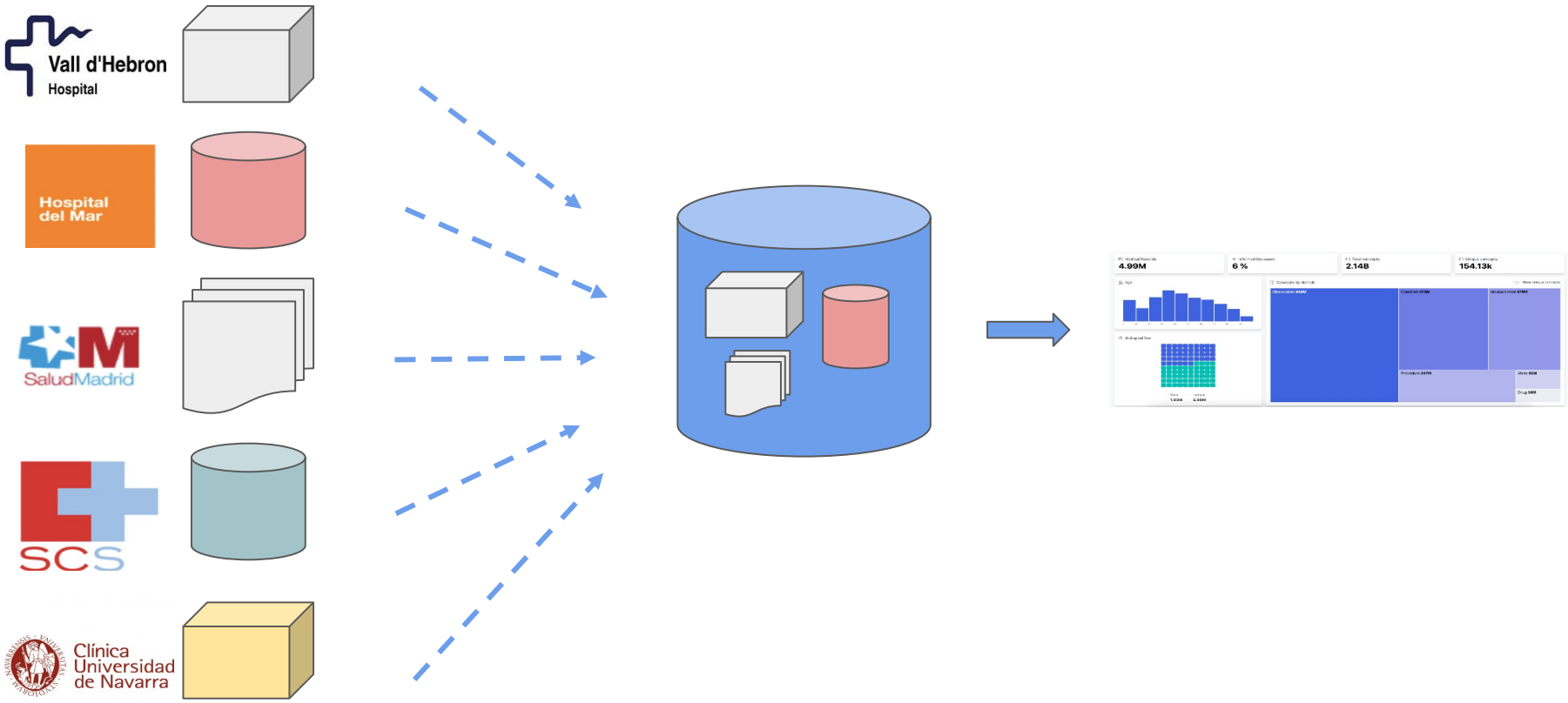
FHIR (Fast Healthcare Interoperability Resources) es un estándar universal para el intercambio de registros de salud de pacientes en sistemas de historia clínica electrónica. Está compuesto por lo que se denominan recursos, agrupados en perfiles que especifican cómo debe representarse un elemento de información particular.

openEHR es un estándar abierto que describe la administración y almacenamiento de información sanitaria basada en arquetipos, definiciones estandarizadas para estructurar y almacenar información médica. Estos arquetipos permiten que los sistemas de salud capturen, compartan e interpreten datos de manera coherente y estandarizada

Vall d'Hebron
Hospital
Hospital
del Mar
SaludMadrid
SCS
Clínica
Universidad
de Navarra

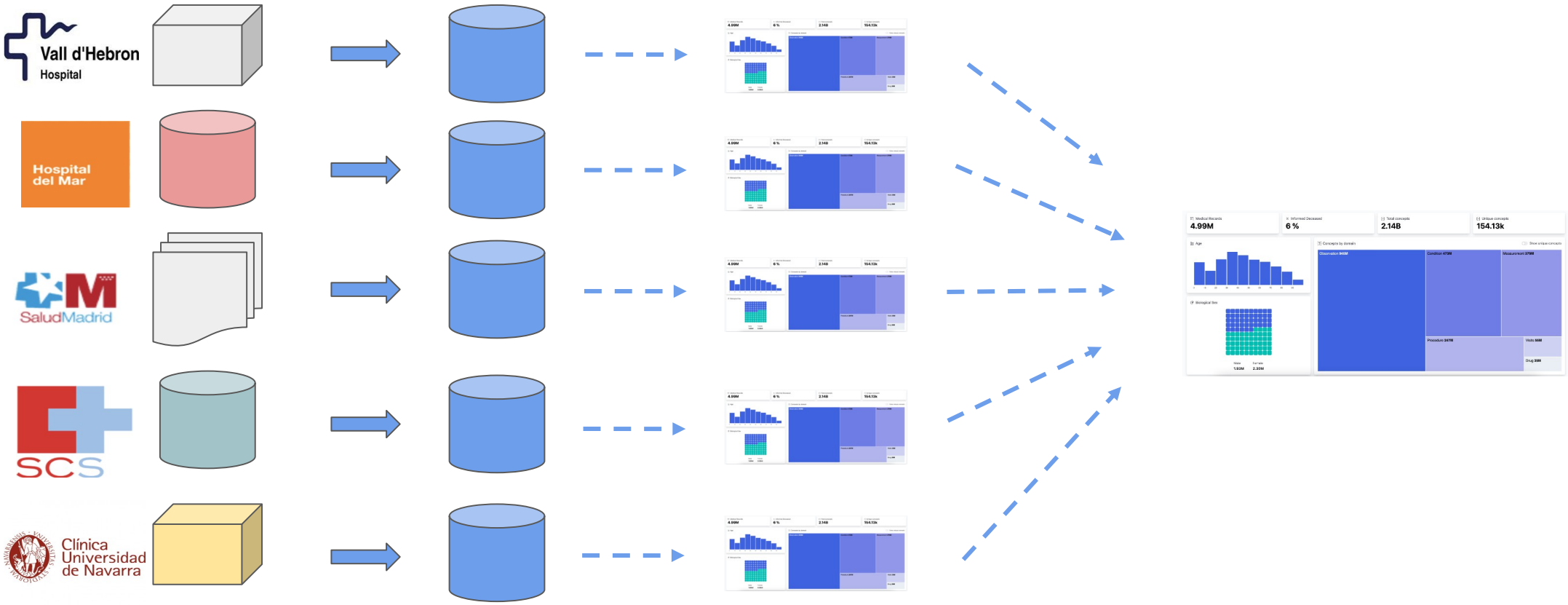
CRENTRALIZAR
LOS DATOS

Una solución podría ser la centralización de datos, que consiste en recolectar, almacenar y gestionar los datos en un único lugar. Sin embargo, este método no es posible para los datos clínicos, ya que la centralización de datos está pensada para datos de una misma organización.


Vall d'Hebron
Hospital
Hospital
del Mar
SaludMadrid
SCS
Clínica
Universidad
de Navarra
4.99M
6 %
2.14B
154.13k

FEDERAR
LOS DATOS

Otra solución son las **bases de datos federadas**, que permiten analizar los datos sin que sea necesario que estos se muevan del lugar donde se han recogido. No obstante, el tratamiento conjunto de los datos requiere que se tenga un modelo común para que las consultas puedan devolver datos sobre bases de datos armonizadas en este modelo de datos común.


Vall d'Hebron
Hospital
Hospital
del Mar
SaludMadrid
SCS
Clínica
Universidad
de Navarra
Medical Records
4.99M
Informed Deceased
6 %
Total concepts
2.14B
Unique concepts
154.13k
Age
Concepts by domain
Biological Sex
Male
1.92M
Female
2.20M
Observation 944M
Condition 473M
Measurement 379M
Procedure 247M
Visits 56M
Drug 20M

OHDSI*

EL ECOSISTEMA DE OMOP-CDM

OHDSI

Observational Health Data Sciences and Informatics

La Observational Health Data Sciences and Informatics (OHDSI) es una red internacional de **investigadores** y **bases de datos observacionales de salud**, la cual tiene como objetivo crear y aplicar soluciones analíticas de datos de código abierto a una gran red de bases de datos clínicos para mejorar la salud y el bienestar humanos.

Esta red de investigación surgió del **Observational Medical Outcomes Partnership (OMOP)**, que era una asociación público-privada establecida en los EE. UU. para informar el uso apropiado de las bases de datos observacionales de atención médica para **estudiar los efectos de los productos médicos**

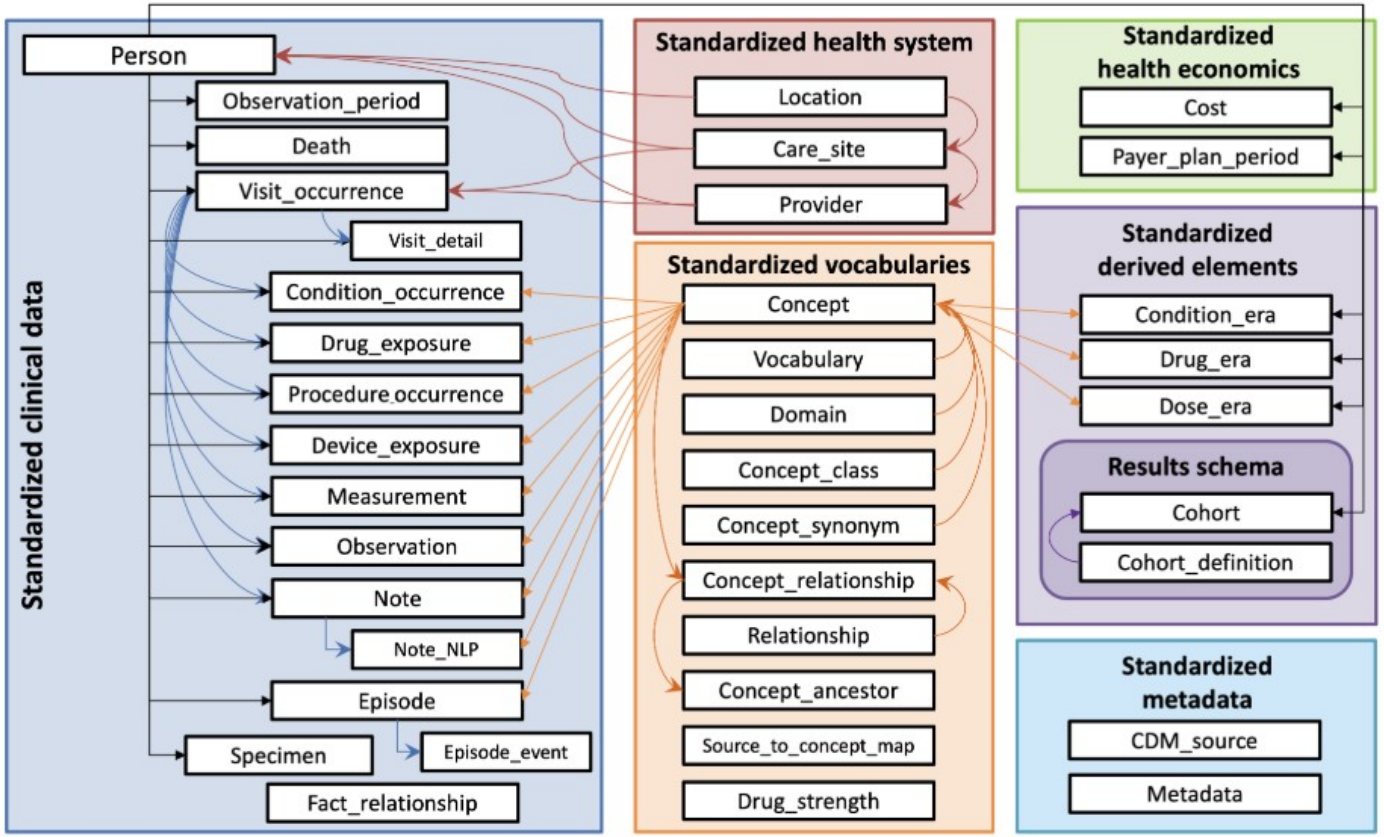
Para lograr la integración entre los datos de los múltiples participantes de la red, OHDSI utiliza el **Common Data Model (CDM)** desarrollado por OMOP

https://www.ohdsi-europe.org/

OMOP Common Data Model | Background | Conventions | CDM Versions | CDM Additions | How to | Support

The current CDM version is CDM v5.4, depicted below. This CDM version was developed over the course of a year by considering requests that were sent via our issues page. The list of proposed changes was then shared with the community in multiple ways: through discussions at the weekly OHDSI Community calls, discussions with the OHDSI Steering Committee, and discussions with all potentially affected workgroups. The final changes were then delivered to the Community through a new R package designed to dynamically generate the DDLs and documentation for all supported SQL dialects. Looking for an entity-relationshop diagram? Click here!

- Link to DDLs for CDM v5.4
- Link to ReadMe for instructions on how to use the R package



CommonDataModel / inst / ddl / 5.4 /

clairblacketer Removing codespell and older ddl folder for CRAN submission

Name
..
bigquery
duckdb
hive
impala
netezza
oracle
pdw
postgresql
redshift
snowflake
spark
sql_server
sqlite
sqlite_extended
synapse

https://ohdsi.github.io/CommonDataModel/

ATHENA

SEARCH | DOWNLOAD | Alberto Labarga

Show all | SHOW HISTORY | DOWNLOAD VOCABULARIES

	ID (CDM V4.5)	CODE (CDM V5)	NAME	REQUIRED	LATEST UPDATE
☑	1	SNOMED	Systematic Nomenclature of Medicine - Clinical Terms (IHTSDO)		01-Mar-2025
☑	2	ICD9CM	International Classification of Diseases, Ninth Revision, Clinical Modification, Volume 1 and 2 (NCHS)		01-Oct-2014
☑	3	ICD9Proc	International Classification of Diseases, Ninth Revision, Clinical Modification, Volume 3 (NCHS)		01-Oct-2014
☑	4	CPT4	Current Procedural Terminology version 4 (AMA)	EULA required	05-May-2025
☑	5	HCPCS	Healthcare Common Procedure Coding System (CMS)		01-Jul-2025
☑	6	LOINC	Logical Observation Identifiers Names and Codes (Regenstrief Institute)		26-Feb-2025
☐	7	NDFRT	National Drug File - Reference Terminology (VA)		06-Aug-2018
☑	8	RxNorm	RxNorm (NLM)		02-Jun-2025
☑	9	NDC	National Drug Code (FDA and manufacturers)		17-Aug-2025

https://athena.ohdsi.org

HADES

Health Analytics Data-to-Evidence Suite

Cohort construction and evaluation

Capr
Develop and manipulate complex cohort definitions in R
Learn more…

CirceR
An R wrapper for Circe, a library for creating cohort definitions, expressing them as JSON, SQL, or Markdown.
Learn more…

CohortGenerator
Instantiating cohorts in a database based on a set of cohort definitions.
Learn more…

PhenotypeLibrary
The OHDSI Phenotype Library: a collection of community-maintained pre-definined cohorts.
Learn more…

CohortDiagnostics
Generate a wide set of diagnostics to evaluate cohort definitions against databases in the CDM.
Learn more…

PheValuator
Semi-automated evaluation of cohorts, producing metrics suchs as sensitivity, specificity, and positive and negative preditive value.
Learn more…

CohortExplorer
Visually explore all individual-level data of patients in a cohort
Learn more…

Keeper
Knowledge-Enhanced Electronic Profile Review
Learn more…

Characterization

Characterization
Various types of characterizations of a target and outcome cohort.
Learn more…

CohortIncidence
Compute incidence rates and proportions
Learn more…

TreatmentPatterns
Analyze treatment patterns of a study population of interestÊ
Learn more…

Patient-level prediction

PatientLevelPrediction
Build and evaluate predictive models for user-specified outcomes, using a wide array of machine learning algorithms.
Learn more…

DeepPatientLevelPrediction
Performing patient level prediction using deep learning
Learn more…

EnsemblePatientLevelPrediction
Building and validating ensemble patient-level predictive models.
Learn more…

Population-level estimation

CohortMethod
New-user cohort studies using large-scale regression for propensity and outcome models.
Learn more…

SelfControlledCaseSeries
Self-Controlled Case Series analysis using few or many predictors, includes splines for age and seasonality.
Learn more…

SelfControlledCohort
A self-controlled cohort design, where time preceding exposure is used as control.
Learn more…

EvidenceSynthesis
Routines for combining causal effect estimates and study diagnostics across multiple data sites in a distributed study.
Learn more…

Evidence Quality

Achilles
Generate descriptive statistics on an entire OMOP CDM database
Learn more…

DataQualityDashboard
Expose and evaluate observational data quality.
Learn more…

EmpiricalCalibration
Use negative control exposure-outcome pairs to profile and calibrate a particular analysis design.
Learn more…

MethodEvaluation
Use real data and established reference sets as well as simulations injected in real data to evaluate the performance of methods.
Learn more…

https://ohdsi.github.io/Hades/

OHDSI Studies

OHDSI is a global, open-science community that is committed to generating real-world evidence to both support clinical decision-making and advance the methodology within this field. We have collaborated on many network studies across our community, many of which (both past and ongoing) are listed in this table. Please click on any listing that interests you to learn more about the study, and how you can potentially collaborate to generate reliable, reproducible evidence.

Show 15 entries

Search:

Title	Use cases	Study type	Tags	Status	Lead(s)	Start date	Last change
All	All	All	All	All	All	All	All
Evidence Synthesis with Negative Controls	Population-Level Estimation	Methods Research		Started	Martijn Schuemie	2025-10-09	2026-02-06
GDE2025 Diabetic Retinopathy Screening				Started			2026-02-02
QRISKvalidation	Patient-Level Prediction	Methods Research	Prediction Models	Repo Created	Alexander Saelmans		2026-01-30
Small-Sample Comparative-Effect Estimation Evaluat...	Population-Level Estimation	Methods Research		Started	Martijn Schuemie	2020-12-03	2026-01-27
Pandemic Prediction	Patient-Level Prediction	Methods Research	Covid19, Prediction models	Repo Created	Egill Fridgeirsson		2026-01-22
MTXprediction	Patient-Level Prediction	Methods Research	Prediction models	Repo Created	Alexander Saelmans		2026-01-15
Exploring Causal Estimands	Population-Level Estimation	Methods Research		Started	Martijn Schuemie	2025-04-01	2026-01-13
Large-scale Evidence Generation and Evaluation for...	Population-Level Estimation	Clinical Application		Results Available	Hsin Yi Chen, George Hripcsak		2026-01-07
Improving Evidence Synthesis	Population-Level Estimation	Methods Research		Results Available	Martijn Schuemie	2025-04-01	2025-12-16
RCRIValidationGP	Patient-Level Prediction	Methods Research	Prediction models	Repo Created	Alexander Saelmans		2025-12-10
Use of GLP-1 receptor agonists and subsequent risk...	Population-level Estimation	Clinical Application		Started	Evelyn Goh	2024-10-01	2025-11-20
Large-scale Evidence Generation and Evaluation acr...	Population-Level Estimation	Clinical Application		Results Available	Marc A. Suchard		2025-11-14
RCRIvalidation	Patient-Level Prediction	Methods Research	Prediction models	Repo Created	Alexander Saelmans		2025-11-11
Lung Cancer Prognostic	Patient-Level Prediction	Clinical Application	PatientLevelPrediction	Results Available	Urmila Chandran, Jenna Reps		2025-10-21
Comparative Effectiveness of Different Dosing Sche...	`Population-Level Estimation`	`Clinical Application`		Started	Fan Bu and Kayoko Shioda	2025-02-01	2025-10-16

Showing 1 to 15 of 137 entries

Previous 1 2 3 4 5 ... 10 Next

Select a study to see details

Last updated: 2026-02-06 11:06:30 (Updated every 24 hours)

https://data.ohdsi.org/OhdsiStudies/

Posters & Abstracts

In this section, you will find all posters and abstracts presented during the OHDSI Europe Symposium 2025 Collaborator Showcase, ranked by poster number. *Links will be added as they become available and when authors have kindly agreed to share their work online.*

#	Title	Author(s)	Link to poster	Link to abstract
1	**OMOP model evaluation for representing GENIE oncology research project data**	Vojtech Huser, Sebastian van Sandijk	01_poster.pdf	01_abstract.pdf
2	**Cardiovascular Effects of SGLT2 Inhibitors in Advanced CKD: A Real-World Cohort Study Using TMUCRD OMOP-CDM and OHDSI Tools— A Foundation for Future Multinational OHDSI Network Collaboration**	Christianus Heru Setiawan, Septi Melisa, Nguyen Thi Kim Hien, Phan Thanh-Phuc, Muhammad Solihuddin Muhtar, Nguyen Phung-Anh, Jason C. Hsu	02_poster.pdf	02_abstract.pdf
3	**Accelerating OMOP Common Data Model Adoption for Sustainable Cross-National Atopic Dermatitis Research: Implementation Across Five European Registers**	Axel Hertzschuch, MD Nazrul Islam, Niels Steen Krogh, Man Fung Tsoi, Bolaji Coker, Merel C. Postema, Ahmet Akkoc, David Prieto Merino, Jochen Schmitt, Carsten Flohr, Elizaveta Gribaleva, Thomas Birkner	03_poster.pdf	03_abstract.pdf

https://www.ohdsi-europe.org/index.php/symposium/51-archive-symposium-2025

Software Demos

501	dqdbt: Continuous Data Quality Testing for OMOP ETL with dbt	Katy Sadowski, Lawrence Adams, Thomas Wylie
502	Summarizing FHIR® to OMOP Transformation Exceptions using Generative AI	Ron Sweeney, Hannah Kimura, Qi Li
503	Usagi-on-the-Web: A Cloud-Based Collaborative Platform for Vocabulary Mapping	Natthawut Adulyanukosol
504	Advancing Electronic Clinical Quality Measure (eCQM) Interoperability: Model Context Protocol (MCP)-Orchestrated CQL-to-OMOP Translation	Star Liu, Robert B Barrett, Kyle Zollo-Venecek, Benjamin Riesser, Benjamin Martin
505	OmniMapper: Hybrid Cloud-based Generative AI for OMOP Vocabulary Mapping of Cancer Registries	Maisie Ng, Cindy Ho, Estella Ye, Ismail Mohd, Erwin Tantoso, Leon Tjandra, Mengling Feng, Sebastian Maurer-Stroh, Johan G Eriksson, Mukkesh Kumar
506	Enhancing OMOP Concept Mapping in Data2Evidence: A Comparative Study of Full-Text and Semantic Search	Zhi Min, Peter Hoffmann
507	The OMOP Annotator: A Database Agnostic Tool for Reviewing and Augmenting the Patient Record	Amy Yates, Erik Benton, Izabelle Humes, Matthew Lawhead, Heath Harrelson, Imogen Bentley, Rumel Mahmood, William Hersh, Steven Bedrick
508	Automated OMOP Concept Mapping Using Multi-Agent Large Language Models and Graph-Enhanced Semantic Retrieval ** Best Community Contribution winner*	Adil Ahmed, Selvin Soby, Boudewijn Aasman, Parsa Mirhaji
509	EHR Browser: A Web Tool to Explore OMOP-CDM Health Records by Concept Hierarchy, Mappings, and Temporal Trends	Veronica Lorenzini, Javier Gracia-Tabuenca, Nicola Cerioli, FinnGen, Mary Pat Reeve

https://www.ohdsi.org/2025-global-collaborator-showcase/

EHDEN

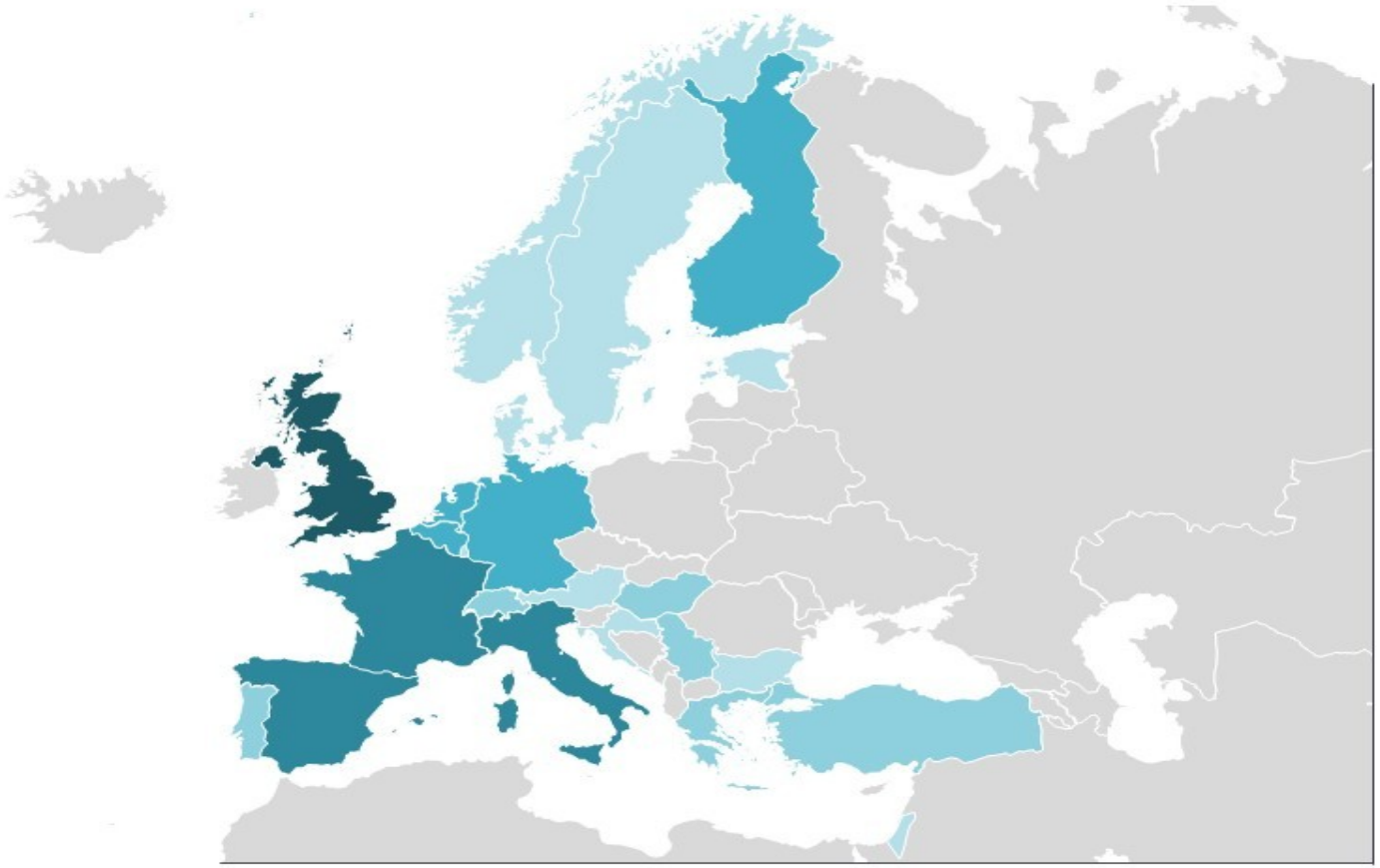
Red Europea de Datos y Evidencia en Salud

La **Red Europea de Datos y Evidencia en Salud (EHDEN)** ha desarrollando un ecosistema federado y equitativo de instituciones que generan datos clínicos, junto con investigadores del ámbito académico y de la industria, con el apoyo de pymes certificadas y cualificadas, armonizando datos clínicos y creando una tecnología de red para la investigación con datos del mundo real.

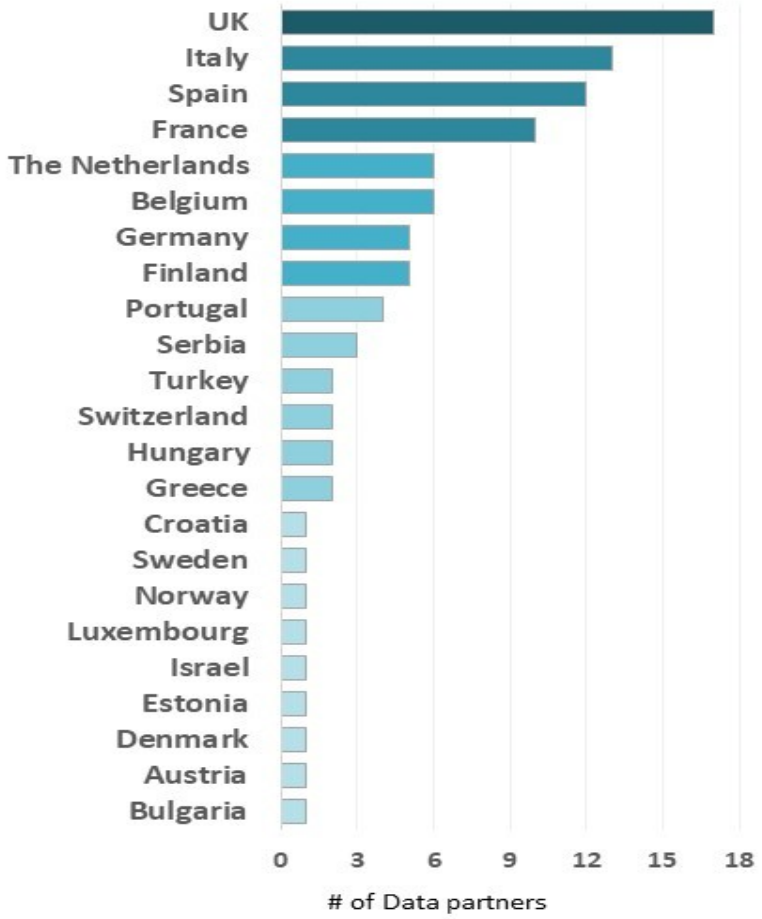
Los centros que se incorporan a la red EHDEN pueden ser hospitales universitarios o no académicos, redes de atención primaria, registros existentes, etc.

La ejecución de los estudios se realiza de manera federada; es decir, el código analítico se envía a la fuente de datos, se ejecuta localmente en su propio entorno seguro, y únicamente se comparten los resultados agregados.

https://www.ehden.eu/



Geographic spread of data partners. The shade of blue indicates the # of data partners in that country (darker = more)



Alberto Labarga

- HOME
- CATALOGUE
- DASHBOARD
- ACADEMY
- EHDEN
- PUBLICATIONS
- STATUS

PORTAL

- ABOUT
- GET STARTED
- PROFILE
- SIGN OUT

INCLIVA | VLC Instituto de Investigación Sanitaria	ABUCASIS	ABUCASIS Instituto de Investigación Sanitaria INCLIVA	3882419 Spain
S/	AQUAS - CatSalut CMBD	AQUAS - CatSalut CMBD AQUAS - CatSalut	19779530 Spain
Base de Análisis Resultados DE NAvarra	BARDENA	Base de Analisis Resultados de Navarra Navarre Health Service - SNS-O	600000 Spain
bifap	BIFAP	BIFAP - Base de Datos para la Investigación Farmacoepidemiológica en el Ámbito Público (Pharmacoepidemiological Research Database for Public Health Systems) Agencia Española de Medicamentos y Productos Sanitarios (AEMPS)	30000000 Spain
BiGÁN	BIGAN	BIGAN - Big Data in Healthcare, Aragon (Spain) Aragon Institute for Health Sciences - Instituto Aragonés de Ciencias de la Salud	2240000 Spain

« 1 2 3 4 5 »

https://portal.ehden.eu/

100% complete

Tool

Extract, Transform and Load

Map raw observational data to the OMOP CDM.

100% complete

Methods

Health Technology Assessment

Basic understanding of HTA agencies and their importance

100% complete

Tool

Infrastructure

Install and configure the OHDSI infrastructure.

0% complete

Skill

Introduction to Data Quality

Introduction to data quality and data quality dashboards....

100% complete

Tool

OHDSI-in-a-Box

Deploy a single instance implementation of OHDSI tools...

100% complete

Tool

OMOP CDM and Standardised Vocabularies

The structure of the common data model and its vocabularies.

https://academy.ehden.eu/

159 results found

Sort by Newest

March 25, 2025 (v1) Publication Open

Acceptability of Using Real-World Data to Estimate Relative Treatment Effects in Health Technology Assessments: Barriers and Future Steps

Gomes, Manuel; Turner, Alex J.; Sammon, Cormac; and 4 others

Objectives Evidence about the comparative effects of new treatments is typically collected in randomized controlled trials (RCTs). In some instances, RCTs are not possible, or their value is limited by an inability to capture treatment effects over the longer term or in all relevant population subgroups. In these cases, nonrandomized studies (NRS) using real-world data (RWD) are…

Part of European Health Data and Evidence Network

Uploaded on February 26, 2025 | Published in: Value in Health, 2025.

115 124

January 24, 2024 (v1) Publication Open

Advancing the use of real world evidence in health technology assessment: insights from a multi-stakeholder workshop

Claire, Ravinder; Elvidge, Jamie; Hanif, Shahid; and 5 others

Introduction: Real-world evidence (RWE) in health technology assessment (HTA) holds significant potential for informing healthcare decision-making. A multistakeholder workshop was organised by the European Health Data and Evidence Network (EHDEN) and the GetReal Institute to explore the status, challenges, and opportunities in incorporating RWE into HTA, with a…

Part of EU Open Research Repository, European Health Data and Evidence Network

Uploaded on February 26, 2025 | Published in: Frontier Pharmacology, 2024.

69 96

April 19, 2024 (v1) Publication Open

Markov modeling for cost-effectiveness using federated health data network

Haug, Markus; Oja, Marek; Pajusalu, Maarja; and 9 others

Objective: To introduce 2 R-packages that facilitate conducting health economics research on OMOP-based data networks, aiming to standardize and improve the reproducibility, transparency, and transferability of health economic models. Materials and methods: We developed the software tools and demonstrated their utility by replicating a UK-based heart failure data analysis…

Part of European Health Data and Evidence Network

Uploaded on February 26, 2025 | Published in: JAMIA, 2024.

58 100

https://zenodo.org/communities/ehden/

DARWIN

Data Analysis and Real World Interrogation Network

La iniciativa DARWIN (Data Analysis and Real World Interrogation Network) de la **Agencia Europea del Medicamento** representa el siguiente paso natural tras el proyecto EHDEN, consolidando y ampliando la capacidad europea para generar evidencia del mundo real de forma rápida, fiable y a gran escala.

Basándose en la red, los estándares y la experiencia metodológica desarrollados en EHDEN, DARWIN establece una infraestructura federada y sostenible que permite responder a preguntas regulatorias prioritarias mediante el uso de datos de salud del mundo real, apoyando así la toma de decisiones regulatorias a lo largo de todo el ciclo de vida de los medicamentos en Europa.

https://www.darwin-eu.org/

DARWIN EU

Status ⇅	Type ⇅	EU PAS Register number ⇅	Official Title ⇅
Archived	C1: Off-the-shelf	EUPAS1000000088	DARWIN EU® - Comparing direct and indirect methods to estimate prevalence of chronic diseases using real-world data
Archived	C3: Complex	EUPAS103936	DARWIN EU® - Background rates of serious adverse events to contextualise safety assessments in clinical trials and non-interventional studies in adolescent and adult patients with severe asthma
Archived	C1: Off-the-shelf	EUPAS105033	DARWIN EU® - Multiple myeloma: patient characterisation, treatments and survival in the period 2012-2022
Archived	C1: Off-the-shelf	EUPAS105641	DARWIN EU® - Drug utilisation study of prescription opioids

SJD Sant Joan de Déu Barcelona · Hospital
INGEMM
idis
IMIBIC
IBSAL

OMOP-CDM*

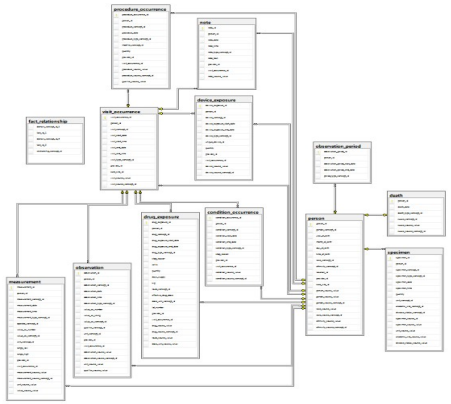
A COMMON DATA MODEL

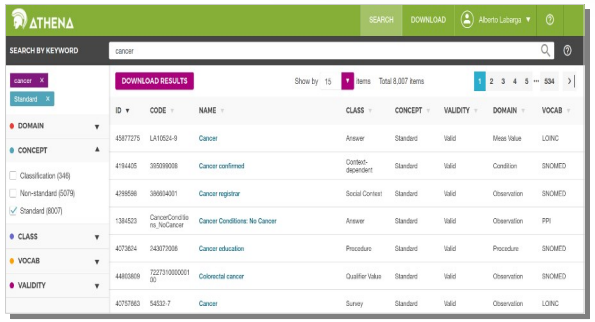
OMOP-CDM

Common Data Model

OMOP-CDM pretende servir como modelo de datos común para poder comprender los datos clínicos sin importar su origen. Este es un estándar de **licencia abierta** cuyo objetivo es estandarizar los datos y estructurarlos para permitir análisis y observaciones eficientes cuyos resultados sean fiables.

Propone **interoperabilidad sintáctica** (modelo de datos) e **interoperabilidad semántica** (vocabularios)





https://www.darwin-eu.org/

OMOP-CDM

Common Data Model

OMOP está optimizado para **identificar** conjuntos de la población con determinadas intervenciones clínicas y sus resultados; **caracterizar** estos conjuntos de población para distintas características como pueden ser información demográfica, historial de enfermedades, tratamientos, etc; **predecir** los resultados de la exposición a un medicamento o la aplicación de una determinada intervención médica para un paciente a nivel individual; **estimar** el efecto que puede tener una determinada intervención para un conjunto de la población.

Características del modelo

01 Idoneidad para el propósito: este CDM se centra en su capacidad para ofrecer información útil para el análisis, más que en satisfacer las necesidades operativas de los proveedores de atención médica o los pagadores.

02 un modelo centrado en la persona, lo cual se traduce en que todas las tablas de eventos se encuentran enlazadas a la tabla PERSON (persona), exceptuando las tablas de datos estandarizados del sistema de salud

03 Protección de datos: los datos personales, como pueden ser el nombre, los apellidos, o el lugar exacto de domicilio, están limitados, exceptuando aquellas investigaciones que requieran de datos exactos.

04 Diseño de dominios: los dominios se representan como tablas relacionales en las que se captura, como mínimo, la identidad de la persona y una fecha para cada registro.

Características del modelo

05 Se trata de un modelo de datos relacional en el que los datos se representan en forma de tablas que están interconectadas mediante claves primarias (primary key, PK) y foráneas (foreign key, FK).

06 Neutralidad tecnológica: OMOP no requiere una tecnología específica, sino que se puede utilizar en distintas bases de datos relaciones, como pueden ser Oracle o PostgreSQL, así como diferente software para el análisis de datos.

07 Los premisos del usuario final sobre las tablas de vocabulario y las tablas de eventos son de solo lectura. Aquellas tablas que necesiten ser manipuladas por el usuario final son almacenadas en el esquema “Results”.

08 Compatibilidad con versiones anteriores

Características del modelo

09 Vocabularios estandarizados: se utilizan vocabularios estandarizados para estandarizar los registros, con el fin de que estos contengan todos los conceptos correspondientes.

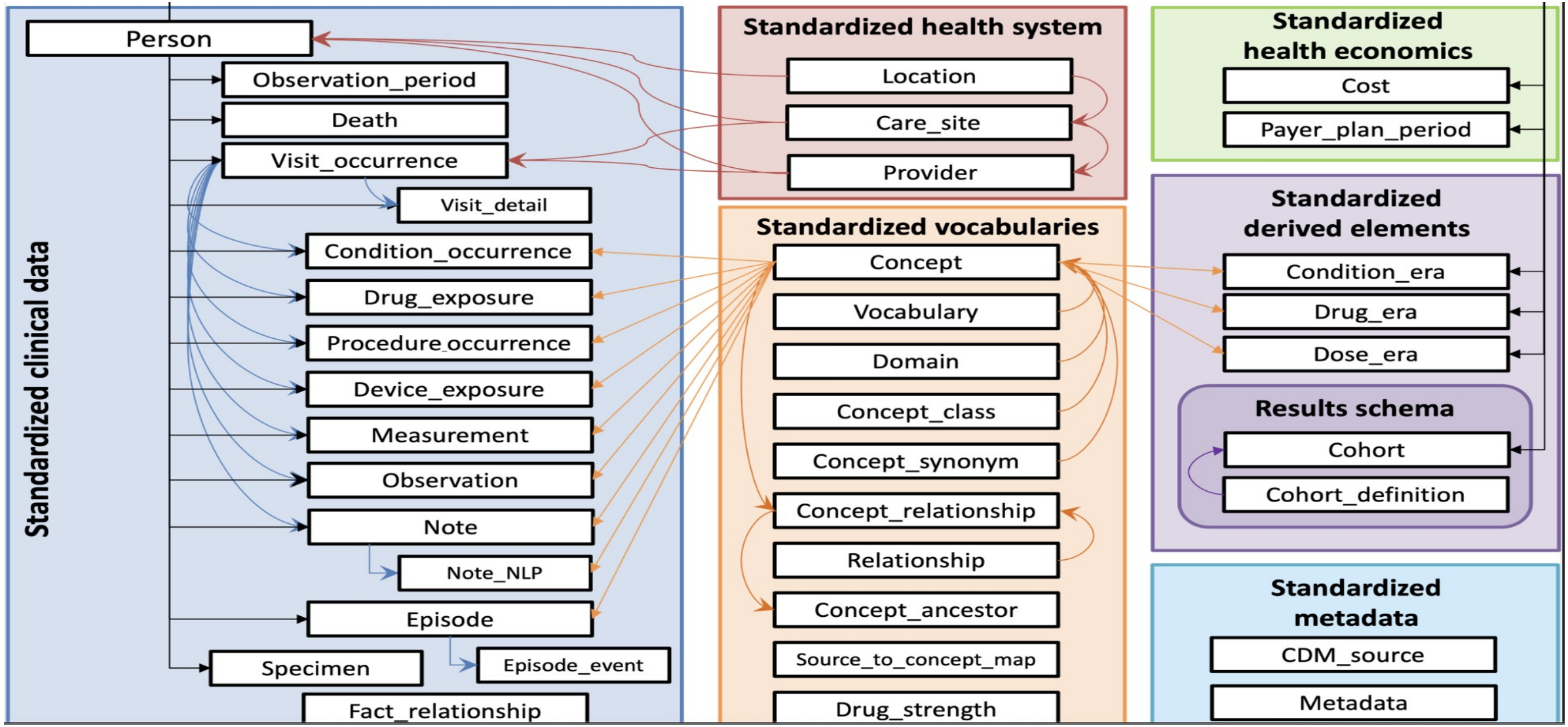
10 Reutilización de vocabularios existentes: siempre que sea posible, el origen de los conceptos serán estándares industriales o nacionales, así como definiciones provenientes de organizaciones médicas.

11 Conservación de códigos de origen: todos los códigos están conectados a los vocabularios estandarizados, pero aún así OMOP almacena el código de origen con el fin de no perder información.

12 Cada concepto estándar tiene una asignación única de dominio que define en qué tabla se registran. Aunque la asignación de dominio correcta puede ser objeto de debate, la correspondencia estricta entre dominio-tabla-campo asegura una ubicación inequívoca para cualquier código o concepto.

OMOP-CDM*

MODELO DE DATOS


Standardized clinical data
Person
Observation_period
Death
Visit_occurrence
Visit_detail
Condition_occurrence
Drug_exposure
Procedure_occurrence
Device_exposure
Measurement
Observation
Note
Note_NLP
Episode
Episode_event
Specimen
Fact_relationship
Standardized health system
Location
Care_site
Provider
Standardized vocabularies
Concept
Vocabulary
Domain
Concept_class
Concept_synonym
Concept_relationship
Relationship
Concept_ancestor
Source_to_concept_map
Drug_strength
Standardized health economics
Cost
Payer_plan_period
Standardized derived elements
Condition_era
Drug_era
Dose_era
Results schema
Cohort
Cohort_definition
Standardized metadata
CDM_source
Metadata

person		
person_id	bigint	NN
location_id	bigint	
provider_id	bigint	
care_site_id	bigint	
birth_datetime	timestamp	
year_of_birth	integer	NN
gender_source_concept_id	integer	NN
gender_concept_id	integer	NN
ethnicity_concept_id	integer	NN
ethnicity_source_concept_id	integer	NN
race_concept_id	integer	NN
race_source_concept_id	integer	NN
month_of_birth	integer	
day_of_birth	integer	
person_source_value	text	
gender_source_value	text	
race_source_value	text	
ethnicity_source_value	text	

La tabla PERSON se utiliza para representar a los individuos incluidos en la base de datos clínica.

Incluye un identificador único del paciente que permite relacionarlo con el resto de las tablas del modelo OMOP-CDM.

Contiene información demográfica básica, como el año de nacimiento, el sexo biológico, el género, la raza y la etnia, codificados mediante vocabularios estandarizados.

Esta tabla actúa como eje central del modelo, ya que todas las observaciones clínicas, visitas, diagnósticos, procedimientos y tratamientos se vinculan a una persona concreta a través de su identificador.

CONCEPTOS

ESTANDARIZACIÓN SEMÁNTICA

En las tablas de datos del CDM, cada registro está representado por un conjunto de conceptos normalizados, que se almacenan en tablas de eventos con sus valores CONCEPT_ID, los cuales son claves foráneas a la tabla CONCEPT, que sirve como tabla de referencia general.

Todas las instancias de CDM utilizan la misma tabla CONCEPT como referencia para los conceptos, lo que junto con el modelo de datos común es un mecanismo clave de interoperabilidad y la base de la red de investigación OHDSI.

La tabla CONCEPT contiene información detallada sobre cada concepto, como su nombre, dominio y clase, entre otros.

CONCEPTOS

ESTANDARIZACIÓN SEMÁNTICA

En las tablas de datos del CDM, cada registro está representado por un conjunto de conceptos normalizados, que se almacenan en tablas de eventos con sus valores CONCEPT_ID, los cuales son claves foráneas a la tabla CONCEPT, que sirve como tabla de referencia general.

Todas las instancias de CDM utilizan la misma tabla CONCEPT como referencia para los conceptos, lo que junto con el modelo de datos común es un mecanismo clave de interoperabilidad y la base de la red de investigación OHDSI.

La tabla CONCEPT contiene información detallada sobre cada concepto, como su nombre, dominio y clase, entre otros.

concept		
concept_id	integer	NN
valid_start_date	date	NN
valid_end_date	date	NN
concept_name	text	NN
domain_id	text	NN
vocabulary_id	text	NN
concept_class_id	text	NN
concept_code	text	NN
standard_concept	text	
invalid_reason	text	

DETAILS

Domain ID	Condition
Concept Class ID	Clinical Finding
Vocabulary ID	SNOMED
Concept ID	320128
Concept code	59621000
Validity	Valid
Concept	Standard
Synonyms	Essential hypertension (disorder) Idiopathic hypertension Primary hypertension Systemic primary arterial hypertension
Valid start	31-Jan-2002
Valid end	31-Dec-2099

concept		
concept_id	integer	NN
valid_start_date	date	NN
valid_end_date	date	NN
concept_name	text	NN
domain_id	text	NN
vocabulary_id	text	NN
concept_class_id	text	NN
concept_code	text	NN
standard_concept	text	
invalid_reason	text	

- Proporciona una visión estandarizada de los conceptos clínicos en distintos dominios
- Reutiliza vocabularios existentes
- Se actualiza constantemente
- NO es cerrado (podemos añadir conceptos usando concept_id > 200000000)
- En caso de que no exista un concepto estándar o no se puede identificar, el valor del CONCEPT_ID se establece en 0, lo que representa un concepto inexistente, desconocido o no mapeable.

Looking to send us a pull request for a bug fix? Please see the readme on the main github page.

Clinical Data Tables
- PERSON
- OBSERVATION_PERIOD
- VISIT_OCCURRENCE
- VISIT_DETAIL
- CONDITION_OCCURRENCE
- DRUG_EXPOSURE
- PROCEDURE_OCCURRENCE
- DEVICE_EXPOSURE
- MEASUREMENT
- OBSERVATION
- DEATH
- NOTE
- NOTE_NLP
- SPECIMEN
- FACT_RELATIONSHIP

Health System Data Tables

Health Economics Data Tables

Clinical Data Tables

PERSON

Table Description

This table serves as the central identity management for all Persons in the database. It contains records that uniquely identify each person or patient, and some demographic information.

User Guide

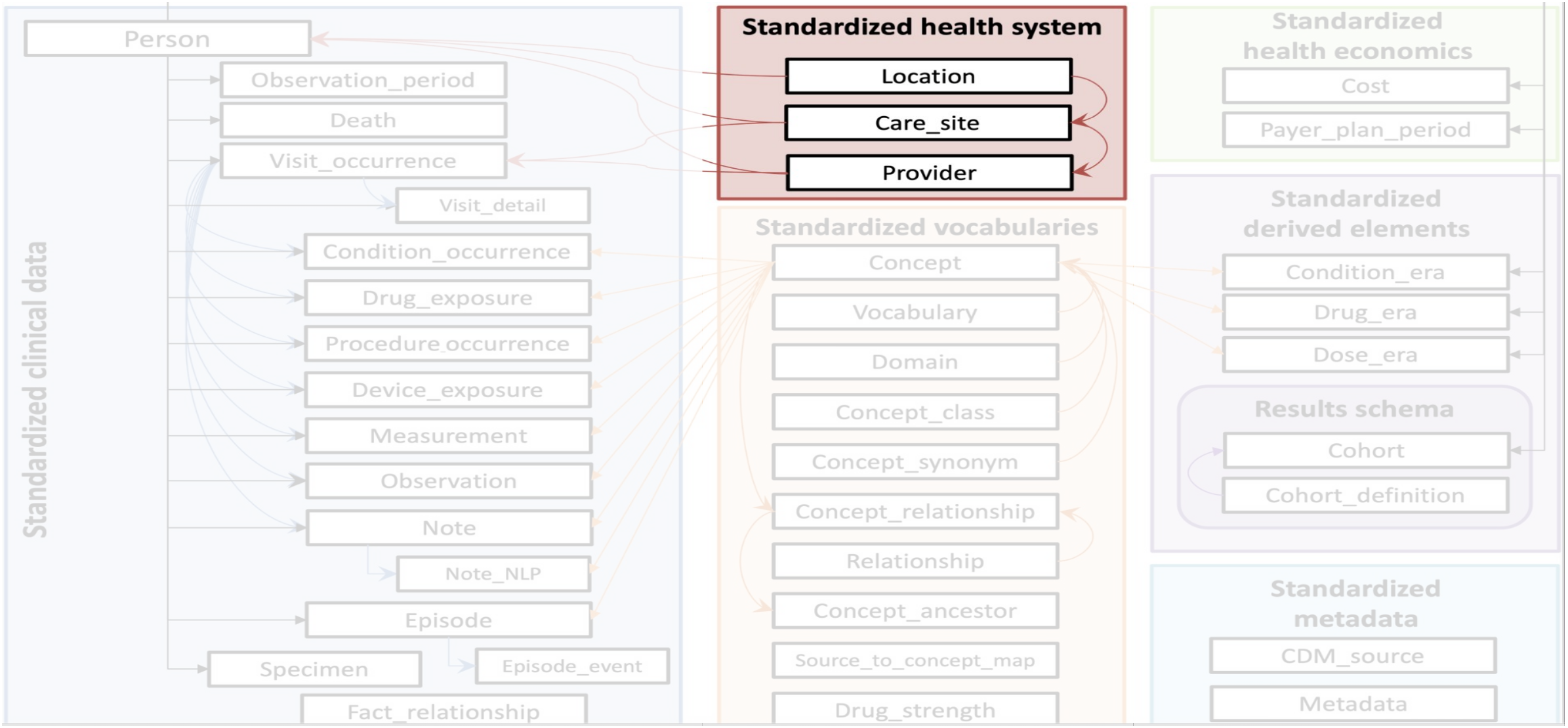
All records in this table are independent Persons.

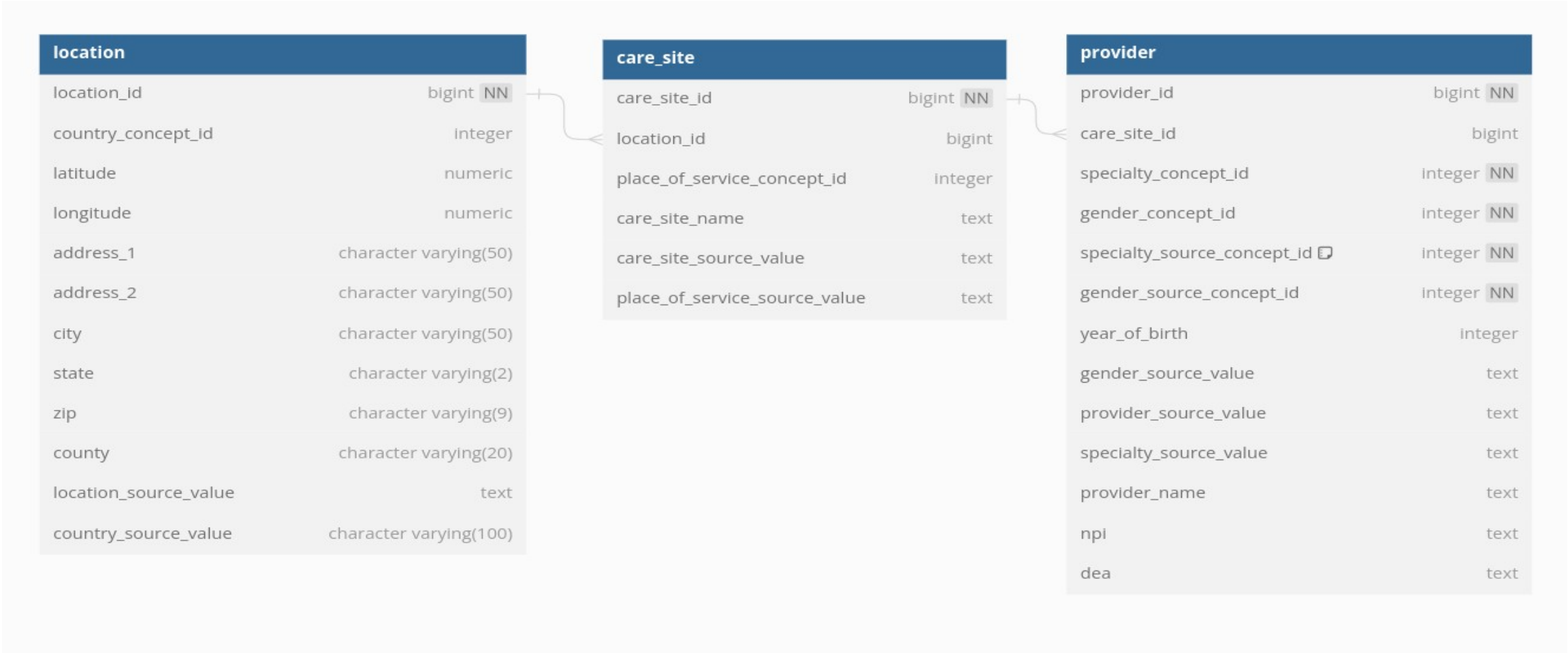
ETL Conventions

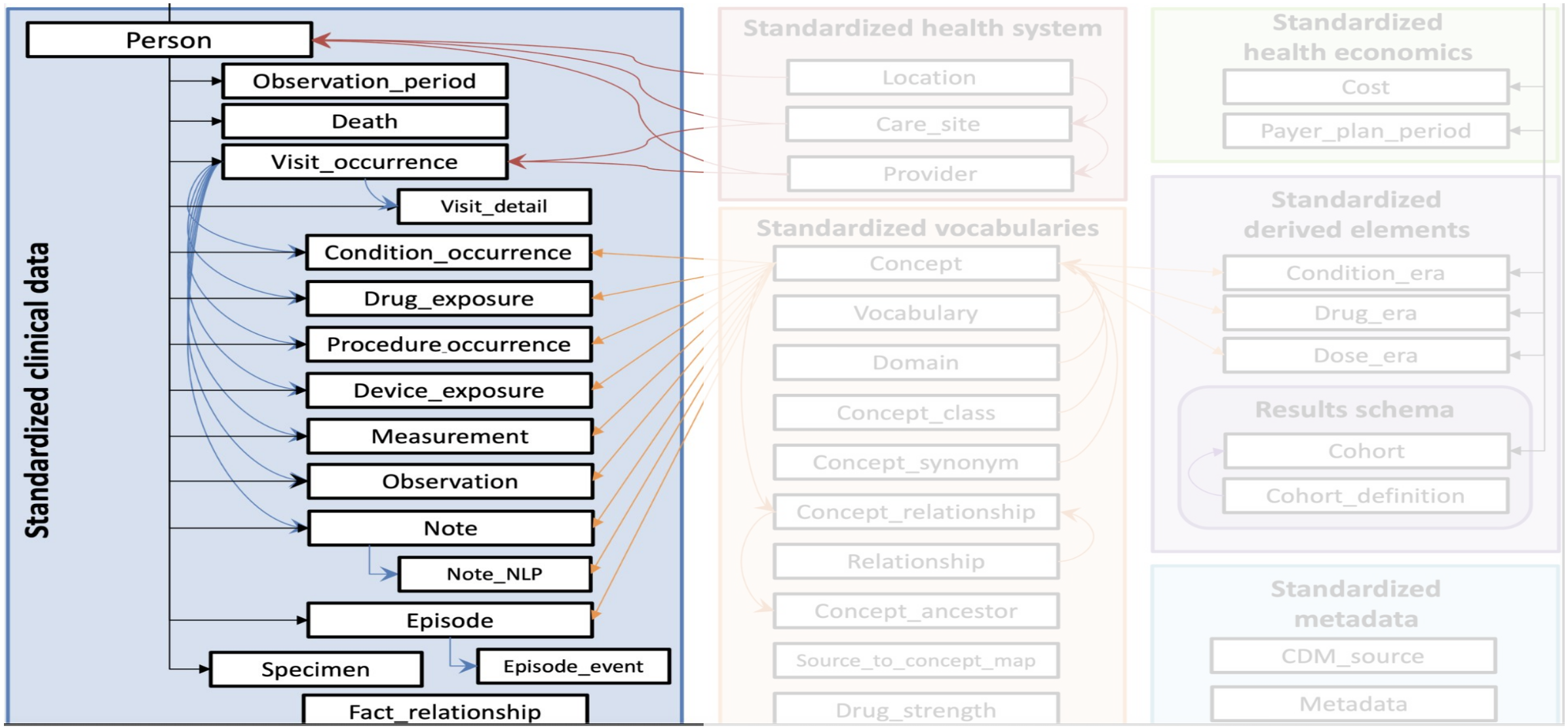
All Persons in a database needs one record in this table, unless they fail data quality requirements specified in the ETL. Persons with no Events should have a record nonetheless. If more than one data source contributes Events to the database, Persons must be reconciled, if possible, across the sources to create one single record per Person. The content of the BIRTH_DATETIME must be equivalent to the content of BIRTH_DAY, BIRTH_MONTH and BIRTH_YEAR.

CDM Field	User Guide	ETL Conventions	Datatype	Required	Primary Key	Foreign Key	FK Table	FK Domain
person_id	It is assumed that every person with a different unique identifier is in fact a different person and should be treated independently	Any person linkage that needs to occur to uniquely identify Persons ought to be done prior to writing this table. This identifier can be the original id from the source data provided if it is an integer, otherwise it can	integer	Yes	Yes	No		

https://ohdsi.github.io/CommonDataModel/cdm54.html


Standardized clinical data
Person
Observation_period
Death
Visit_occurrence
Visit_detail
Condition_occurrence
Drug_exposure
Procedure_occurrence
Device_exposure
Measurement
Observation
Note
Note_NLP
Episode
Specimen
Episode_event
Fact_relationship
Standardized health system
Location
Care_site
Provider
Standardized vocabularies
Concept
Vocabulary
Domain
Concept_class
Concept_synonym
Concept_relationship
Relationship
Concept_ancestor
Source_to_concept_map
Drug_strength
Standardized health economics
Cost
Payer_plan_period
Standardized derived elements
Condition_era
Drug_era
Dose_era
Results schema
Cohort
Cohort_definition
Standardized metadata
CDM_source
Metadata


location
location_id bigint NN
country_concept_id integer
latitude numeric
longitude numeric
address_1 character varying(50)
address_2 character varying(50)
city character varying(50)
state character varying(2)
zip character varying(9)
county character varying(20)
location_source_value text
country_source_value character varying(100)
care_site
care_site_id bigint NN
location_id bigint
place_of_service_concept_id integer
care_site_name text
care_site_source_value text
place_of_service_source value text
provider
provider_id bigint NN
care_site_id bigint
specialty_concept_id integer NN
gender_concept_id integer NN
specialty_source_concept_id integer NN
gender_source_concept_id integer NN
year_of_birth integer
gender_source_value text
provider_source_value text
specialty_source_value text
provider_name text
npi text
dea text


Standardized clinical data
Person
Observation_period
Death
Visit_occurrence
Visit_detail
Condition_occurrence
Drug_exposure
Procedure_occurrence
Device_exposure
Measurement
Observation
Note
Note_NLP
Episode
Episode_event
Specimen
Fact_relationship
Standardized health system
Location
Care_site
Provider
Standardized vocabularies
Concept
Vocabulary
Domain
Concept_class
Concept_synonym
Concept_relationship
Relationship
Concept_ancestor
Source_to_concept_map
Drug_strength
Standardized health economics
Cost
Payer_plan_period
Standardized derived elements
Condition_era
Drug_era
Dose_era
Results schema
Cohort
Cohort_definition
Standardized metadata
CDM_source
Metadata

person	
person_id	bigint NN
location_id	bigint
provider_id	bigint
care_site_id	bigint
birth_datetime	timestamp
year_of_birth	integer NN
gender_source_concept_id	integer NN
gender_concept_id	integer NN
ethnicity_concept_id	integer NN
ethnicity_source_concept_id	integer NN
race_concept_id	integer NN
race_source_concept_id	integer NN
month_of_birth	integer
day_of_birth	integer
person_source_value	text
gender_source_value	text
race_source_value	text
ethnicity_source_value	text

death	
person_id	bigint NN
death_datetime	timestamp
death_date	date NN
death_type_concept_id	integer NN
cause_concept_id	integer
cause_source_concept_id	integer
cause_source_value	text

La tabla DEATH se utiliza para registrar la información relacionada con el fallecimiento de una persona incluida en la base de datos. Incluye el identificador del paciente, que permite vincular este evento con la tabla PERSON y el resto de la información clínica asociada.
Contiene la fecha de la muerte y, cuando está disponible, lla causa del fallecimiento, codificada mediante vocabularios estandarizados.

visit_occurrence		
visit_occurrence_id	bigint	NN
person_id	bigint	NN
visit_start_datetime	timestamp	NN
visit_end_datetime	timestamp	NN
provider_id	bigint	
care_site_id	bigint	
preceding_visit_occurrence_id	bigint	
visit_concept_id	integer	NN
visit_type_concept_id	integer	NN
visit_source_concept_id	integer	NN
admitted_from_concept_id	integer	NN
discharged_to_concept_id	integer	NN
visit_start_date	date	
visit_end_date	date	
visit_source_value	text	
admitted_from_source_value	text	
discharged_to_source_value	text	

La tabla VISIT_OCCURRENCE se utiliza para representar las visitas de un paciente a un establecimiento de atención médica u otro tipo de encuentros con un proveedor de salud

Incluye el identificador del paciente y una clave primaria para relacionarla con otras tablas.

Contiene información sobre la fecha de inicio y fin de la visita.

Almacena el tipo de visita, que puede ser una visita ambulatoria, una hospitalización, una visita de emergencia, entre otros.

VISIT_OCCURRENCE

visit_concept_id

Inpatient Visit: Person visiting hospital, at a Care Site, in bed, for duration of more than one day, with physicians and other Providers permanently available to deliver service around the clock
Emergency Room Visit: Person visiting dedicated healthcare institution for treating emergencies, at a Care Site, within one day, with physicians and Providers permanently available to deliver service around the clock
Emergency Room and Inpatient Visit: Person visiting ER followed by a subsequent Inpatient Visit, where Emergency department is part of hospital, and transition from the ER to other hospital departments is undefined
Outpatient Visit: Person visiting dedicated ambulatory healthcare institution, at a Care Site, within one day, without bed, with physicians or medical Providers delivering service during Visit
Home Visit: Provider visiting Person, without a Care Site, within one day, delivering service

_TYPE_CONCEP_ID

- EHR: An Electronic Health (or Medical) Record. These are usually systems supporting hospitals or ambulatory healthcare providers
- EHR administration record: A record in the EHR indicating the administration of a service, drug or treatment.
- EHR discharge summary: A record in the EHR usually created by the discharging Provider upon discharge of a Person from the Provider's care (when a Person is discharged from inpatient hospitalization, discharged from a physical therapy series, discharged from counseling, etc.)
- EHR note: Unspecified note recorded in the EHR
- EHR Pathology report: Pathology Laboratory Report based on findings derived from examination of a specimen
- EHR radiology report: Radiology Department Report based on findings derived from the analysis of imaging modalities
- NLP: Information derived by Natural Language Processing
- Survey: A survey in general

VISIT_OCCURRENCE

visit_type_concept_id

EHR: An Electronic Health (or Medical) Record. These are usually systems supporting hospitals or ambulatory healthcare providers
EHR administration record: A record in the EHR indicating the administration of a service, drug or treatment.
EHR discharge summary: A record in the EHR usually created by the discharging Provider upon discharge of a Person from the Provider's care (when a Person is discharged from inpatient hospitalization, discharged from a physical therapy series, discharged from counseling, etc.)
EHR note: Unspecified note recorded in the EHR
EHR Pathology report: Pathology Laboratory Report based on findings derived from examination of a specimen
EHR radiology report: Radiology Department Report based on findings derived from the analysis of imaging modalities
NLP: Information derived by Natural Language Processing
Survey: A survey in general

ATHENA

SEARCH BY KEYWORD

aspirin

Type Concept ×

Standard ×

DOMAIN

CONCEPT

CLASS

VOCAB

VALIDITY

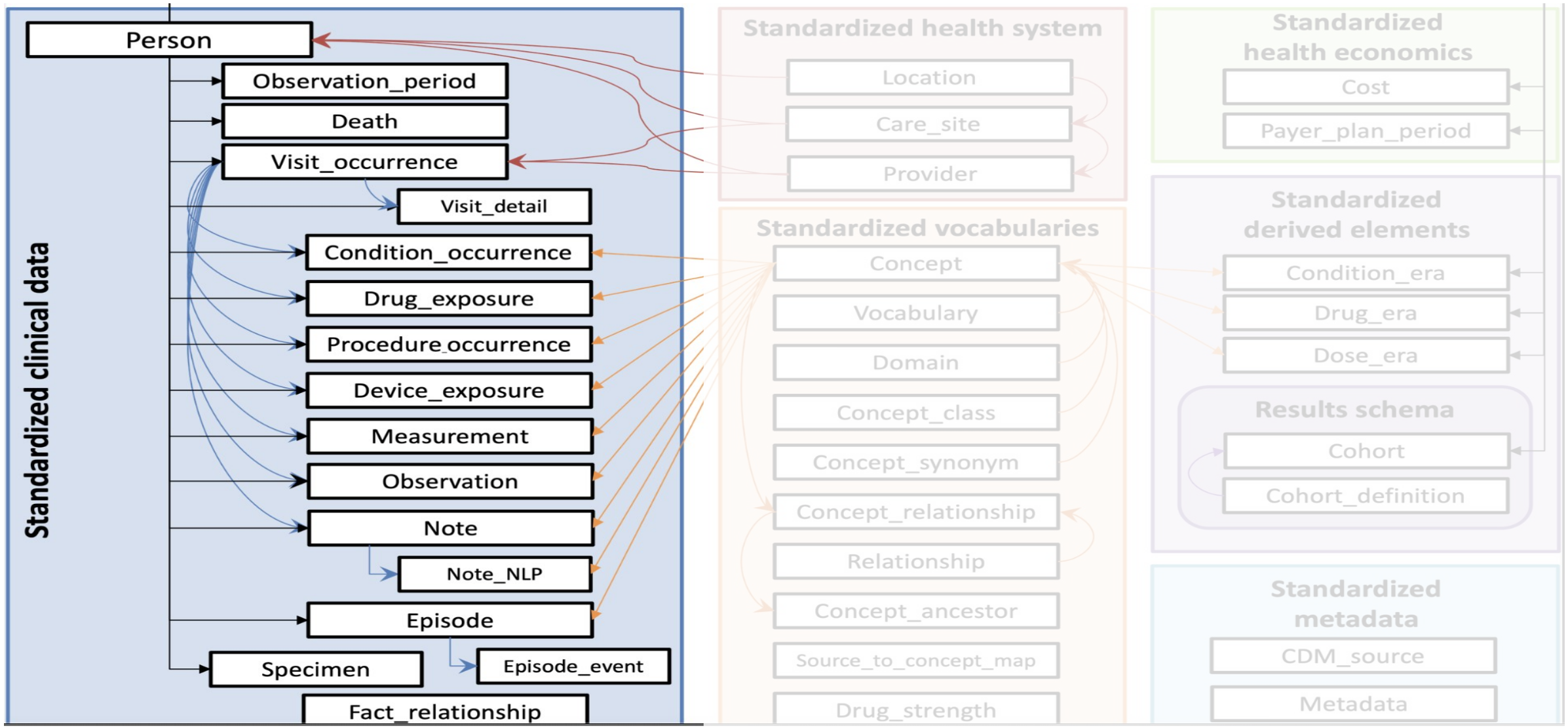
DOWNLOAD RESULTS

ID	CODE	NAME	CLASS
32809	OMOP4976882	Case Report Form	Type Concept
32810	OMOP4976883	Claim	Type Concept
32811	OMOP4976884	Claim authorization	Type Concept
32812	OMOP4976885	Claim discharge record	Type Concept
32813	OMOP4976886	Claim enrolment record	Type Concept
32814	OMOP4976887	Cost record	Type Concept
32815	OMOP4976888	Death Certificate	Type Concept
32816	OMOP4976889	Dental claim	Type Concept

https://athena.ohdsi.org/search-terms/terms?domain=Type+Concept&standardConcept=Standard

visit_detail	
visit_detail_id	bigint NN
visit_detail_start_datetime	timestamp
visit_detail_end_datetime	timestamp
person_id	bigint NN
visit_occurrence_id	bigint NN
provider_id	bigint
care_site_id	bigint
preceding_visit_detail_id	bigint
parent_visit_detail_id	bigint
visit_detail_concept_id	integer NN
visit_detail_type_concept_id	integer NN
discharged_to_concept_id	integer
admitted_from_concept_id	integer
visit_detail_source_concept_id	integer
visit_detail_start_date	date
visit_detail_end_date	date
admitted_from_source_value	text
visit_detail_source_value	text
discharged_to_source_value	text

La tabla VISIT_DETAIL se utiliza para proporcionar detalles adicionales sobre las visitas del paciente que no están cubiertos en la tabla VISIT_OCCURRENCE, en general, cuando la visita tiene múltiples entradas para PROVIDER o CARE_SITE


Standardized clinical data
Person
Observation_period
Death
Visit_occurrence
Visit_detail
Condition_occurrence
Drug_exposure
Procedure_occurrence
Device_exposure
Measurement
Observation
Note
Note_NLP
Episode
Episode_event
Specimen
Fact_relationship
Standardized health system
Location
Care_site
Provider
Standardized vocabularies
Concept
Vocabulary
Domain
Concept_class
Concept_synonym
Concept_relationship
Relationship
Concept_ancestor
Source_to_concept_map
Drug_strength
Standardized health economics
Cost
Payer_plan_period
Standardized derived elements
Condition_era
Drug_era
Dose_era
Results schema
Cohort
Cohort_definition
Standardized metadata
CDM_source
Metadata

condition_occurrence	
condition_occurrence_id	bigint NN
person_id	bigint
provider_id	bigint
visit_occurrence_id	bigint
visit_detail_id	bigint
condition_start_datetime	timestamp
condition_end_datetime	timestamp
condition_concept_id	integer
condition_type_concept_id	integer
condition_status_concept_id	integer
condition_source_concept_id	integer
condition_start_date	date
condition_end_date	date
stop_reason	text
condition_source_value	text
condition_status_source_value	text

La tabla **CONDITION_OCCURRENCE** tiene como objetivo principal registrar las ocurrencias individuales de condiciones médicas,y diagnósticos, que han sido observadas o documentadas en pacientes en un contexto de atención médica.

El diagnóstico se recoge en el campo **condition_concept_id** usando SNOMED.

El campo **condition_status_concept_id** se utiliza para registrar el contexto o la circunstancia en la que se dio el diagnóstico de una condición médica en una visita médica específica. Esta información detalla las condiciones bajo las cuales se hizo el diagnóstico, como si se tratara de un diagnóstico de ingreso, un diagnóstico final, si se basó en hallazgos de laboratorio

procedure_occurrence		
procedure_occurrence_id	bigint	NN
person_id	bigint	NN
procedure_datetime	timestamp	NN
provider_id	bigint	
visit_occurrence_id	bigint	
visit_detail_id	bigint	
procedure_end_datetime	timestamp	
procedure_concept_id	integer	NN
procedure_source_concept_id	integer	NN
procedure_type_concept_id	integer	NN
modifier_concept_id	integer	NN
quantity	integer	
procedure_date	date	
procedure_end_date	date	
procedure_source_value	text	
modifier_source_value	text	

La tabla **PROCEDURE_OCCURRENCE** se utiliza para registrar eventos individuales de procedimientos médicos realizados en pacientes durante su atención médica.

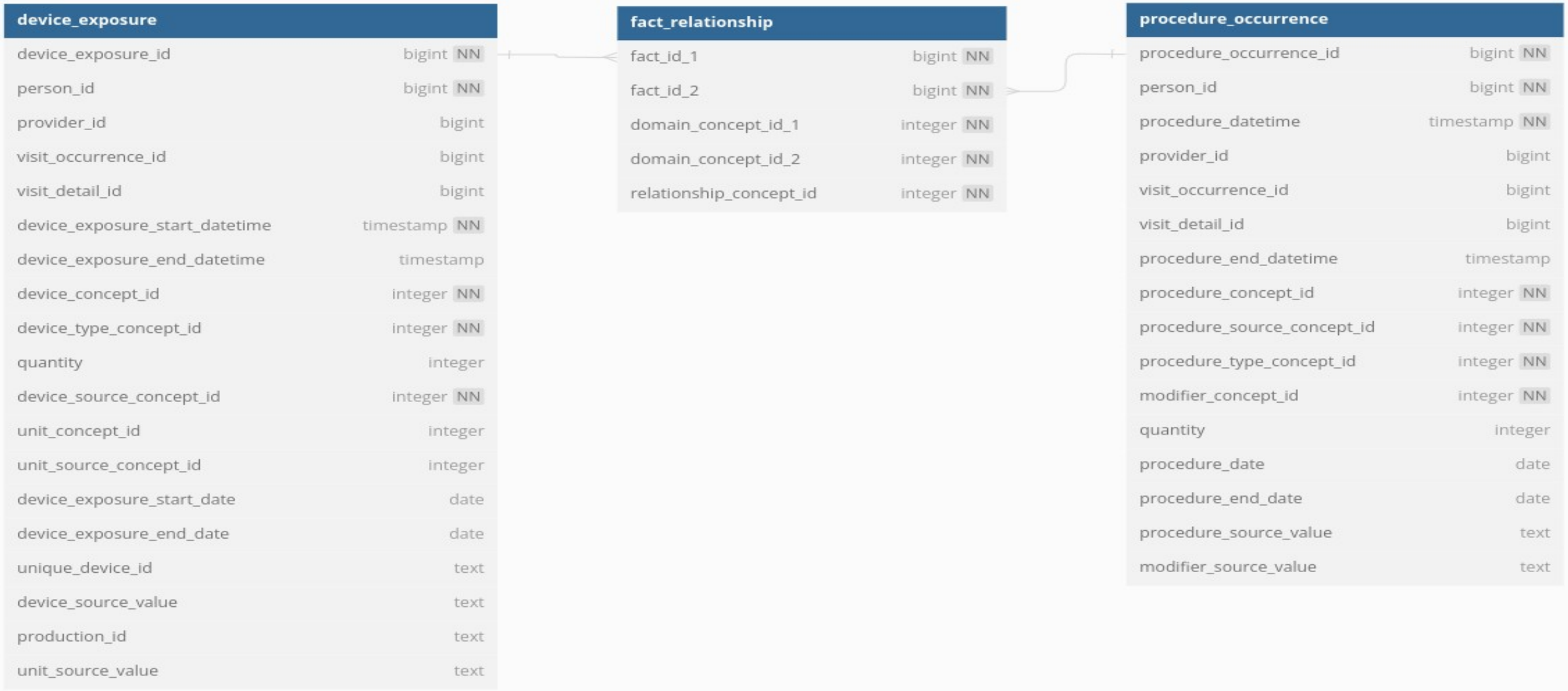
procedure_concept_id: Representa el concepto del procedimiento médico utilizando un identificador de concepto estándar, que está vinculado a un vocabulario médico controlado, como SNOMED-CT u otro.

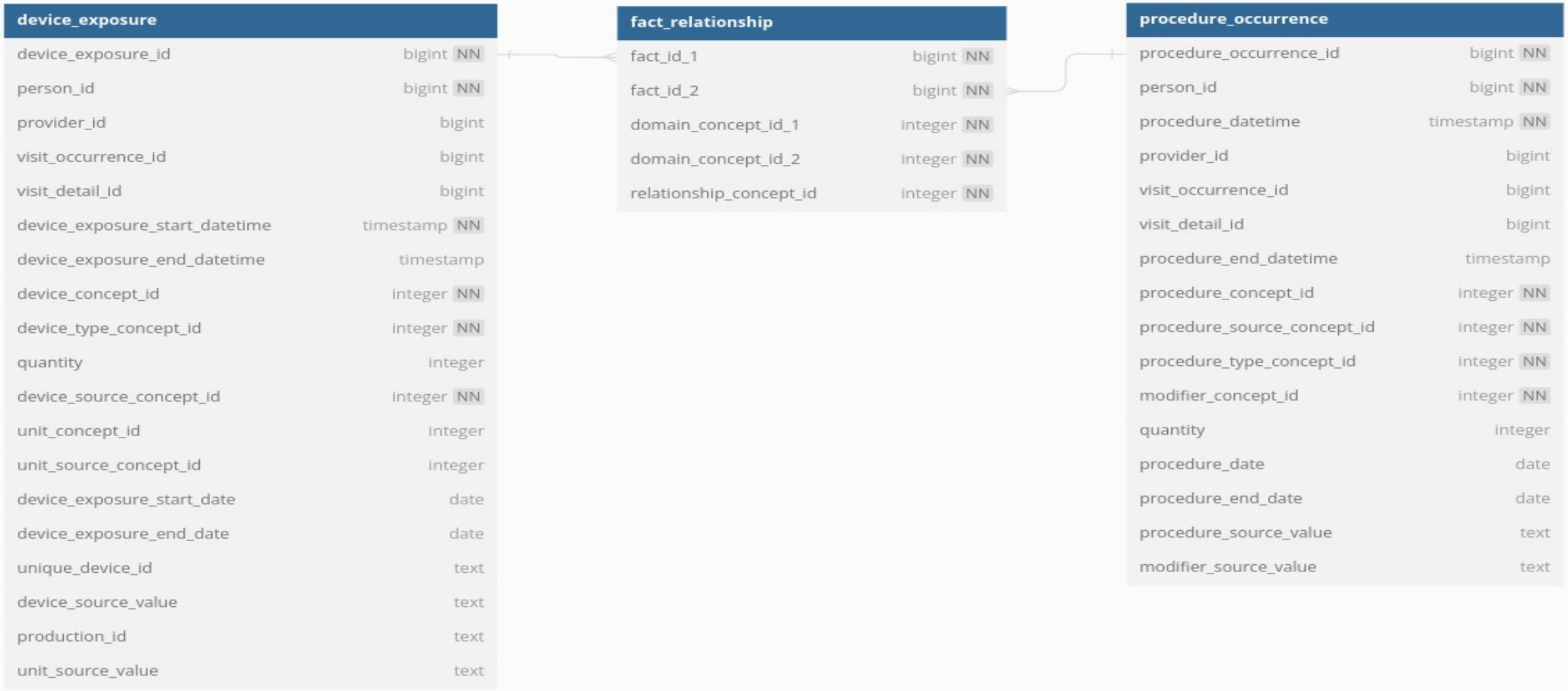
modifier_concept_id: Opcionalmente, se puede registrar un modificador que proporciona información adicional sobre el procedimiento, como el enfoque o la técnica utilizada.

device_exposure	
device_exposure_id	bigint NN
person_id	bigint NN
provider_id	bigint
visit_occurrence_id	bigint
visit_detail_id	bigint
device_exposure_start_datetime	timestamp NN
device_exposure_end_datetime	timestamp
device_concept_id	integer NN
device_type_concept_id	integer NN
quantity	integer
device_source_concept_id	integer NN
unit_concept_id	integer
unit_source_concept_id	integer
device_exposure_start_date	date
device_exposure_end_date	date
unique_device_id	text
device_source_value	text
production_id	text
unit_source_value	text

La tabla **DEVICE_EXPOSURE** tiene como objetivo principal registrar eventos de exposición a dispositivos médicos que incluyen objetos implantables (por ejemplo, marcapasos, *stents*, articulaciones artificiales), equipos y suministros médicos (por ejemplo, vendajes, muletas, jeringas), otros instrumentos utilizados en procedimientos médicos (por ejemplo, suturas, desfibriladores) y material utilizado en la atención clínica (por ejemplo, adhesivos, material corporal, material dental, material quirúrgico).

El campo **device_concept_id** representa el concepto del dispositivo médico utilizando un identificador de concepto estándar, que está vinculado a un vocabulario médico controlado, como SNOMED-CT u otro.


device_exposure
device_exposure_id bigint NN
person_id bigint NN
provider_id bigint
visit_occurrence_id bigint
visit_detail_id bigint
device_exposure_start_datetime timestamp NN
device_exposure_end_datetime timestamp
device_concept_id integer NN
device_type_concept_id integer NN
quantity integer
device_source_concept_id integer NN
unit_concept_id integer
unit_source_concept_id integer
device_exposure_start_date date
device_exposure_end_date date
unique_device_id text
device_source_value text
production_id text
unit_source_value text
fact_relationship
fact_id_1 bigint NN
fact_id_2 bigint NN
domain_concept_id_1 integer NN
domain_concept_id_2 integer NN
relationship_concept_id integer NN
procedure_occurrence
procedure_occurrence_id bigint NN
person_id bigint NN
procedure_datetime timestamp NN
provider_id bigint
visit_occurrence_id bigint
visit_detail_id bigint
procedure_end_datetime timestamp
procedure_concept_id integer NN
procedure_source_concept_id integer NN
procedure_type_concept_id integer NN
modifier_concept_id integer NN
quantity integer
procedure_date date
procedure_end_date date
procedure_source_value text
modifier_source_value text



44818892	OMOP generated	Device used by (SNOMED)	Relationship	Non-standard	Valid	Metadata	Relationship

drug_exposure	
drug_exposure_id	bigint NN
person_id	bigint NN
provider_id	bigint
visit_occurrence_id	bigint
visit_detail_id	bigint
drug_exposure_start_datetime	timestamp
drug_exposure_end_datetime	timestamp
drug_concept_id	integer NN
drug_type_concept_id	integer NN
drug_source_concept_id	integer NN
route_concept_id	integer NN
refills	integer
days_supply	integer
drug_exposure_start_date	date
drug_exposure_end_date	date
verbatim_end_date	date
quantity	numeric
sig	text
lot_number	text
drug_source_value	text
route_source_value	text
dose_unit_source_value	text
stop_reason	text

La tabla **DRUG_EXPOSURE** tiene como objetivo principal registrar eventos individuales de exposición a medicamentos o sustancias farmacológicas por parte de los pacientes durante su atención médica. drug_concept_id representa el concepto del medicamento o la sustancia utilizando un identificador de concepto estándar vinculado generalmente al **vocabulario RxNorm.** En este contexto, un medicamento se define como un **principio activo**.

Los registros en esta tabla representan recetas médicas escritas, medicamentos dispensados y medicamentos administrados por un proveedor, entre otros. El **DRUG_TYPE_CONCEPT_ID** se puede utilizar para buscar y filtrar estos tipos.

drug_exposure	
drug_exposure_id	bigint NN
person_id	bigint NN
provider_id	bigint
visit_occurrence_id	bigint
visit_detail_id	bigint
drug_exposure_start_datetime	timestamp
drug_exposure_end_datetime	timestamp
drug_concept_id	integer NN
drug_type_concept_id	integer NN
drug_source_concept_id	integer NN
route_concept_id	integer NN
refills	integer
days_supply	integer
drug_exposure_start_date	date
drug_exposure_end_date	date
verbatim_end_date	date
quantity	numeric
sig	text
lot_number	text
drug_source_value	text
route_source_value	text
dose_unit_source_value	text
stop_reason	text

El **dominio DRUG**, forma una jerarquía compleja por lo que un DRUG_SOURCE_CONCEPT_ID puede mapear a múltiples identificadores de conceptos estándar si se trata de un producto combinado.

El contenido de información más detallado es preferido durante el proceso de mapeo, indicado por el **CONCEPT_CLASS_ID** del Concepto y se registran en el siguiente orden de precedencia: ‘Branded Pack’, ‘Clinical Pack’, ‘Branded Drug’, ‘Clinical Drug’, ‘Branded Drug Component’, ‘Clinical Drug Component’, ‘Branded Drug Form’, ‘Clinical Drug Form’ and ‘Ingredient’.

measurement		
measurement_id	bigint	NN
person_id	bigint	NN
provider_id	bigint	
visit_occurrence_id	bigint	
visit_detail_id	bigint	
measurement_event_id	bigint	
measurement_datetime	timestamp	NN
measurement_concept_id	integer	NN
measurement_type_concept_id	integer	NN
measurement_source_concept_id	integer	NN
measurement_date	date	
operator_concept_id	integer	
value_as_concept_id	integer	
unit_concept_id	integer	
unit_source_concept_id	integer	
meas_event_field_concept_id	integer	
value_as_number	numeric	
range_low	numeric	
range_high	numeric	
measurement_time	text	
measurement_source_value	text	
unit_source_value	text	
value_source_value	text	

La tabla **MEASUREMENT** contiene registros de mediciones, es decir, **valores estructurados** (numéricos o categóricos) obtenidos a través de un examen o **prueba sistemática y estandarizada** de una persona o de una muestra de la persona.

La tabla MEASUREMENT contiene tanto las solicitudes de pruebas como los resultados, como pruebas de laboratorio, signos vitales, hallazgos cuantitativos de informes de patología, etc.

Las mediciones se almacenan como pares atributo-valor, donde el atributo es el concepto medido y el valor representa el resultado.

measurement_concept_id se codifica generalmente usando el vocabulario LOINC

observation		
observation_id	bigint	NN
person_id	bigint	NN
observation_datetime	timestamp	NN
provider_id	bigint	
visit_occurrence_id	bigint	
visit_detail_id	bigint	
observation_event_id	bigint	
observation_date	date	
observation_concept_id	integer	NN
observation_type_concept_id	integer	NN
qualifier_concept_id	integer	NN
observation_source_concept_id	integer	NN
obs_event_field_concept_id	integer	
value_as_concept_id	integer	
unit_concept_id	integer	
value_as_number	numeric	
value_as_string	text	
observation_source_value	text	
unit_source_value	text	
qualifier_source_value	text	
value_source_value	text	

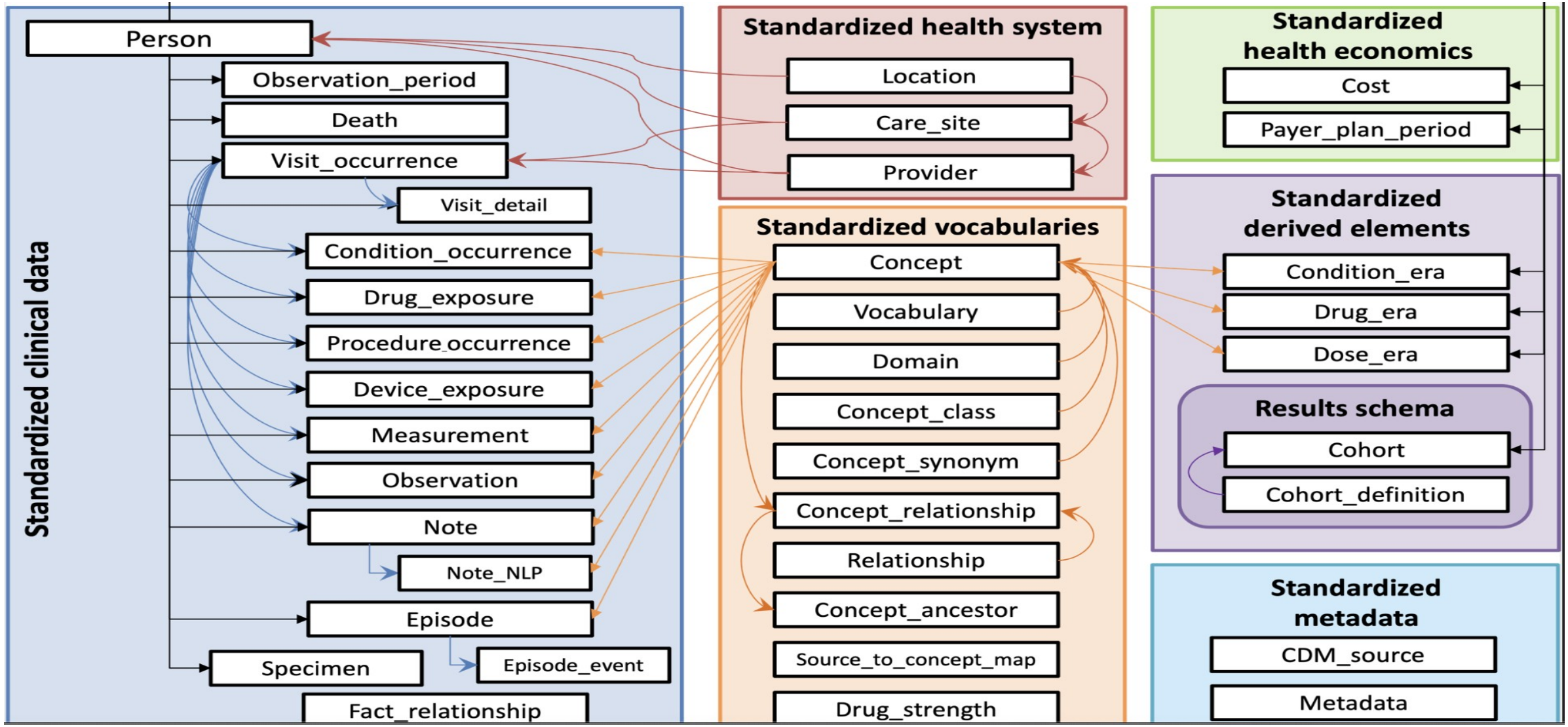
La tabla **OBSERVATION** captura información clínica sobre una persona obtenida en el contexto de un examen, una entrevista o un procedimiento. Cualquier dato que no pueda ser representado por otros dominios, como datos sociales y de estilo de vida, historial médico, historial familiar, etc., se registra aquí.

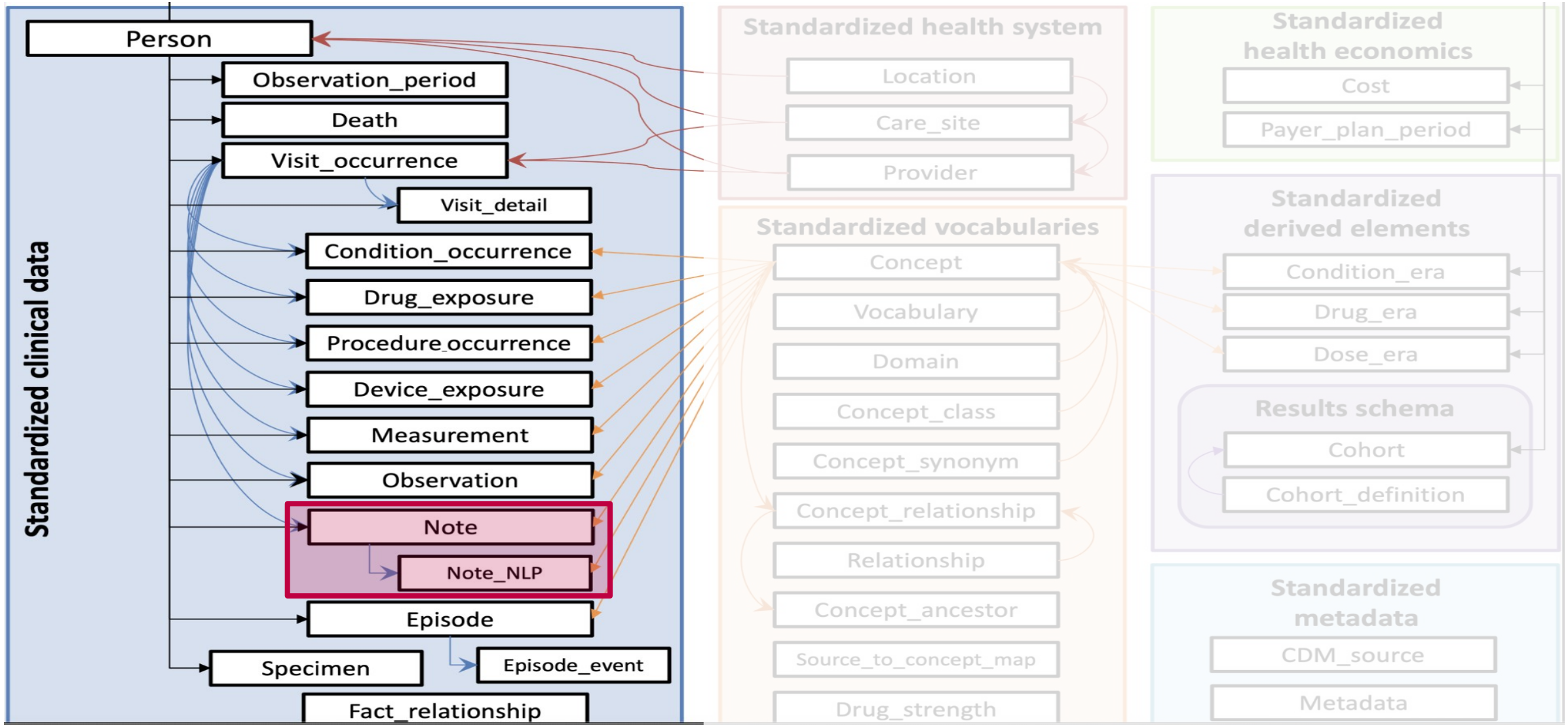
Las observaciones difieren de las mediciones en el sentido de que no requieren una prueba estandarizada u otra actividad para generar información clínica. Las observaciones típicas incluyen historial médico, historial familiar, la necesidad declarada de cierto tratamiento, circunstancias sociales, elecciones de estilo de vida, patrones de utilización de atención médica, etc.

Si la generación de información clínica requiere una prueba estandarizada, como análisis de laboratorio o imágenes, y da como resultado un resultado estandarizado, el dato se registra en la tabla MEASUREMENT.

specimen		
specimen_id	bigint	NN
person_id	bigint	NN
specimen_datetime	timestamp	NN
specimen_concept_id	integer	NN
specimen_type_concept_id	integer	NN
anatomic_site_concept_id	integer	NN
disease_status_concept_id	integer	NN
unit_concept_id	integer	
specimen_date	date	
quantity	numeric	
specimen_source_id	text	
specimen_source_value	text	
unit_source_value	text	
anatomic_site_source_value	text	
disease_status_source_value	text	

La tabla **SPECIMEN** almacena información sobre muestras biológicas


Standardized clinical data
Person
Observation_period
Death
Visit_occurrence
Visit_detail
Condition_occurrence
Drug_exposure
Procedure_occurrence
Device_exposure
Measurement
Observation
Note
Note_NLP
Episode
Episode_event
Specimen
Fact_relationship
Standardized health system
Location
Care_site
Provider
Standardized vocabularies
Concept
Vocabulary
Domain
Concept_class
Concept_synonym
Concept_relationship
Relationship
Concept_ancestor
Source_to_concept_map
Drug_strength
Standardized health economics
Cost
Payer_plan_period
Standardized derived elements
Condition_era
Drug_era
Dose_era
Results schema
Cohort
Cohort_definition
Standardized metadata
CDM_source
Metadata


Standardized clinical data
Person
Observation_period
Death
Visit_occurrence
Visit_detail
Condition_occurrence
Drug_exposure
Procedure_occurrence
Device_exposure
Measurement
Observation
Note
Note_NLP
Episode
Specimen
Episode_event
Fact_relationship
Standardized health system
Location
Care_site
Provider
Standardized vocabularies
Concept
Vocabulary
Domain
Concept_class
Concept_synonym
Concept_relationship
Relationship
Concept_ancestor
Source_to_concept_map
Drug_strength
Standardized health economics
Cost
Payer_plan_period
Standardized derived elements
Condition_era
Drug_era
Dose_era
Results schema
Cohort
Cohort_definition
Standardized metadata
CDM_source
Metadata

note	
note_id	bigint NN
person_id	bigint NN
note_event_id	bigint
provider_id	bigint
visit_occurrence_id	bigint
visit_detail_id	bigint
note_datetime	timestamp NN
note_event_field_concept_id	integer NN
note_type_concept_id	integer NN
note_class_concept_id	integer NN
encoding_concept_id	integer NN
language_concept_id	integer NN
note_date	date
note_title	text
note_text	text
note_source_value	text

La tabla **NOTE** se utiliza para almacenar notas o comentarios relacionados con eventos clínicos y de atención médica en los datos del paciente.

El campo **note_class_concept_id** se completa con LOINC Document Ontology, que es un estándar para nombrar documentos de manera coherente con el fin de respaldar una variedad de casos de uso: recuperación, organización, visualización e intercambio.

Content	Description
193240	Corresponds to LOINC 34127-1, Dentistry Hygienist Outpatient Progress note
55443322132	Corresponds to LOINC LP173418-7, Kind of Document = Note
55443322175	Corresponds to LOINC LP173213-2, Type of Service = Progress
55443322166	Corresponds to LOINC LP173051-6, Setting = Outpatient
55443322107	Corresponds to LOINC LP172934-4, Subject Matter Domain = Dentistry
55443322146	Corresponds to LOINC LP173071-4, Role = Hygienist

LOINC Document Ontology (DO)

La Document ontology anota cada documento con 5 dimensiones:

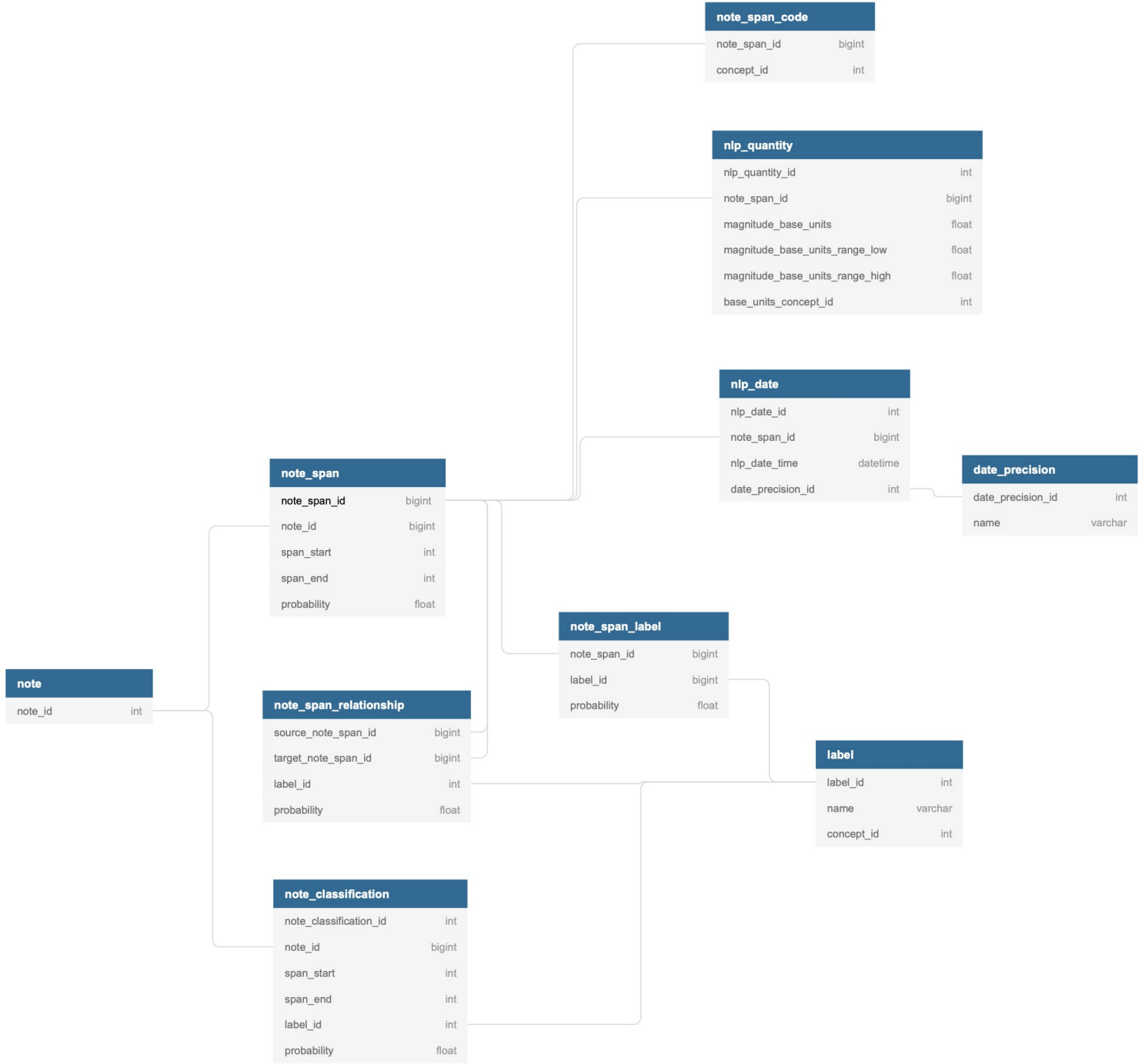
- Tipo de Documento: Caracteriza la estructura general del documento a nivel macro (por ejemplo, Consentimiento de Anestesia).
- Tipo de Servicio: Caracteriza el tipo de servicio o actividad (por ejemplo, evaluaciones, consultas y resúmenes). La noción de secuencia temporal, por ejemplo, al comienzo (admisión) o al final (alta), está subsumida en este eje. Ejemplo: Enseñanza al Alta.
Configuración: La configuración es una extensión de las definiciones de CMS (por ejemplo, Hospitalización, Ambulatorio).
Dominio del Tema: Caracteriza el dominio del tema de una nota (por ejemplo, Anestesiología).
Rol: Caracteriza el nivel de formación o profesionalismo del autor del documento, pero no se desglosa en especialidad o subespecialidad (por ejemplo, Médico). Cada combinación de estas 5 dimensiones se traduce en un código LOINC único.

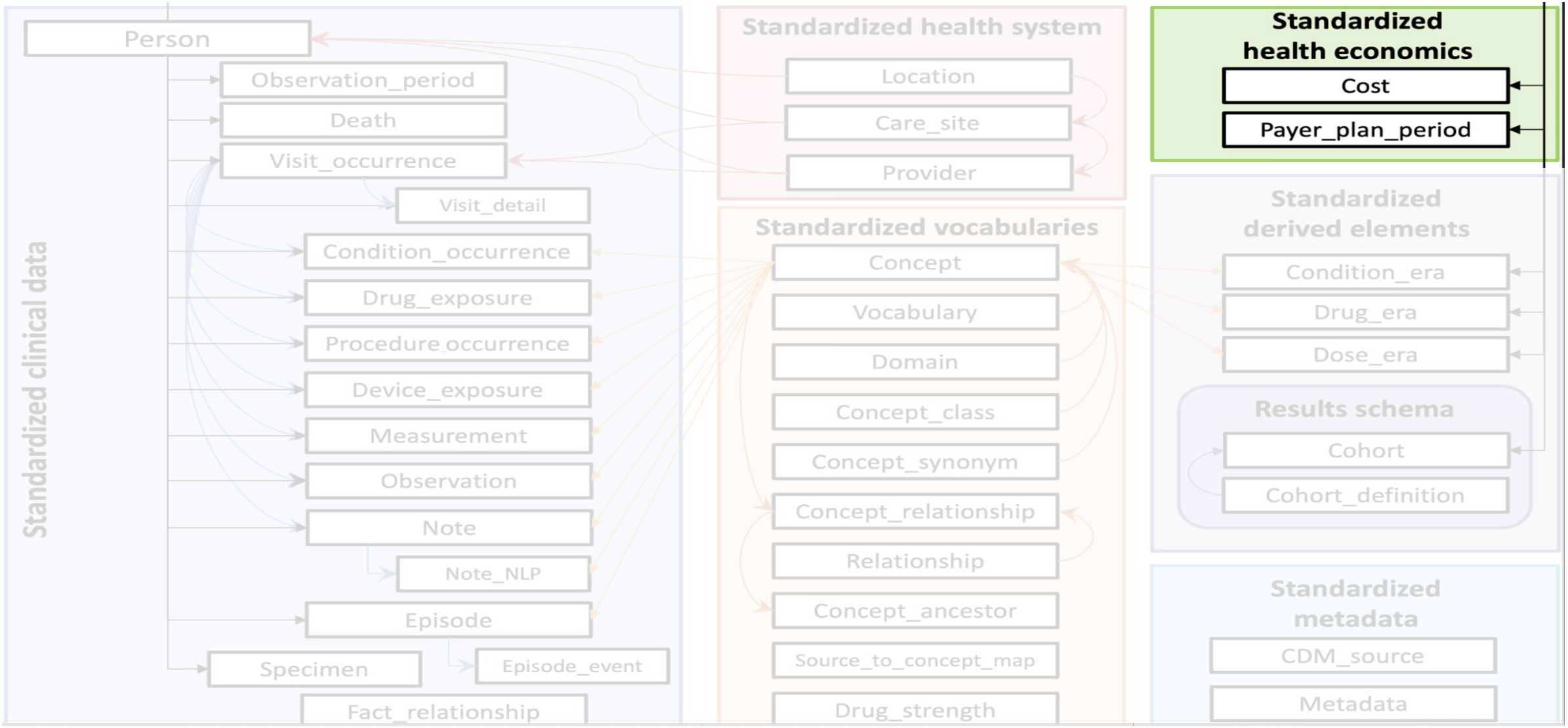
note_nlp		
note_nlp_id	bigint	NN
note_id	bigint	NN
section_concept_id	integer	NN
snippet	text	
offset	text	
lexical_variant	text	NN
note_nlp_concept_id	integer	NN
nlp_system	text	
nlp_date	date	NN
nlp_datetime	timestamp	
term_exists	text	
term_temporal	text	
term_modifiers	text	
note_nlp_source_concept_id	integer	NN

La tabla **NOTE_NLP** tiene como objetivo principal almacenar los resultados del procesamiento de lenguaje natural de notas clínicas y comentarios, lo que incluye la extracción de información estructurada a partir de texto no estructurado, como diagnósticos, medicamentos, síntomas, y otra información relevante contenida en las notas.

note_nlp_concept_id: Representa el concepto detectado o extraído del texto utilizando NLP. Puede ser un diagnóstico, un medicamento, un síntoma u otra entidad relevante.

El campo **term_modifiers** se utiliza para registrar información adicional o modificadores relacionados con el término extraído o detectado a través del procesamiento de lenguaje natural (NLP).
Negation = false
Subject = patient
Conditional = false
Rule_out = false
Uncertain = true or high or moderate or even low
History = true
Concept_date = 2018-01-23


note_span_code
note_span_id bigint
concept_id int
nlp_quantity
nlp_quantity_id int
note_span_id bigint
magnitude_base_units float
magnitude_base_units_range_low float
magnitude_base_units_range_high float
base_units_concept_id int
nlp_date
nlp_date_id int
note_span_id bigint
nlp_date_time datetime
date_precision_id int
date_precision
date_precision_id int
name varchar
note_span
note_span_id bigint
note_id bigint
span_start int
span_end int
probability float
note_span_label
note_span_id bigint
label_id bigint
probability float
note
note_id int
note_span_relationship
source_note_span_id bigint
target_note_span_id bigint
label_id int
probability float
label
label_id int
name varchar
concept_id int
note_classification
note_classification_id int
note_id bigint
span_start int
span_end int
label_id int
probability float


Standardized clinical data
Person
Observation_period
Death
Visit_occurrence
Visit_detail
Condition_occurrence
Drug_exposure
Procedure_occurrence
Device_exposure
Measurement
Observation
Note
Note_NLP
Episode
Specimen
Episode_event
Fact_relationship
Standardized health system
Location
Care_site
Provider
Standardized vocabularies
Concept
Vocabulary
Domain
Concept_class
Concept_synonym
Concept_relationship
Relationship
Concept_ancestor
Source_to_concept_map
Drug_strength
Standardized health economics
Cost
Payer_plan_period
Standardized derived elements
Condition_era
Drug_era
Dose_era
Results schema
Cohort
Cohort_definition
Standardized metadata
CDM_source
Metadata

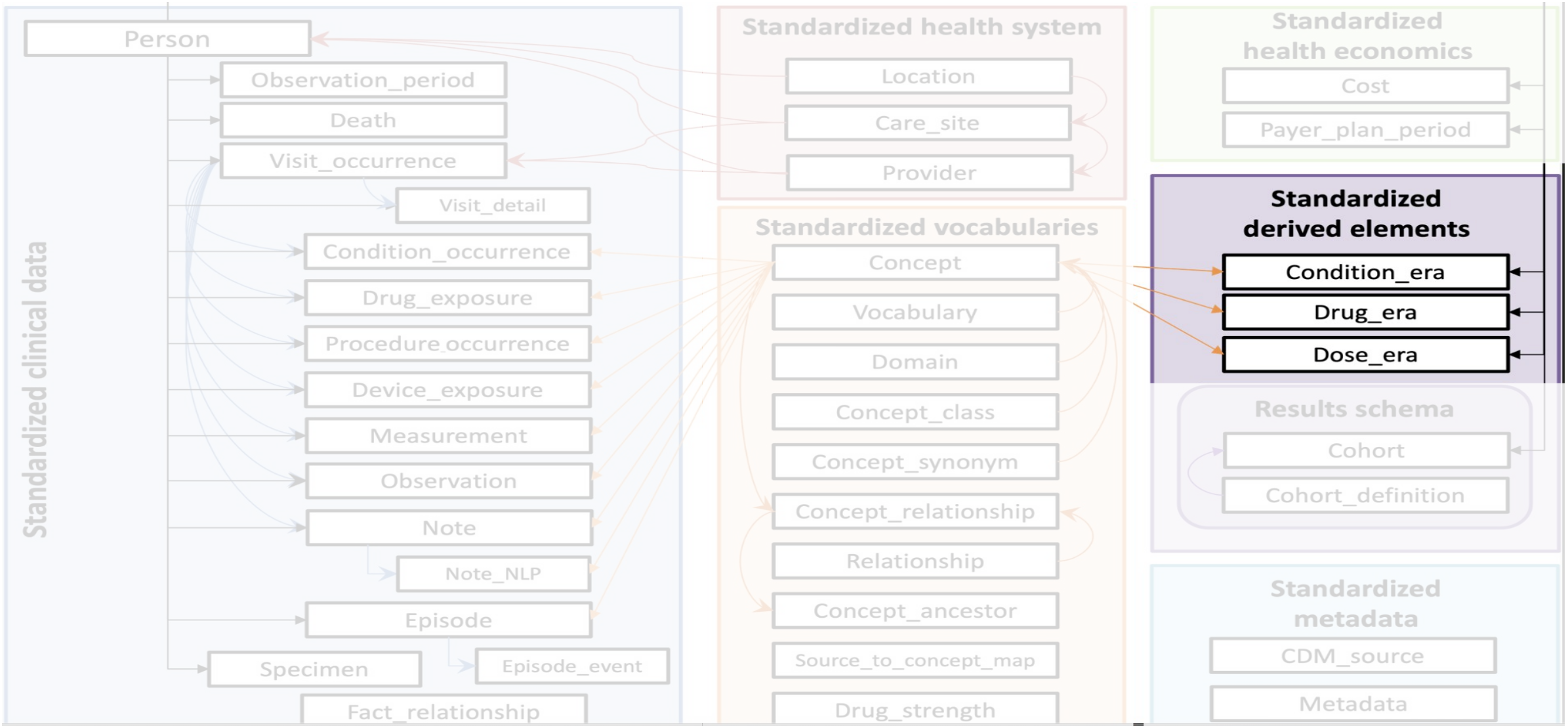
cost		
cost_id	bigint	NN
cost_event_id	bigint	NN
cost_domain_id	text	NN
payer_plan_period_id	bigint	
cost_type_concept_id	integer	NN
currency_concept_id	integer	
revenue_code_concept_id	integer	
drg_concept_id	integer	
total_charge	double	
total_cost	double	
total_paid	double	
paid_by_payer	double	
paid_by_patient	double	
paid_patient_copay	double	
paid_patient_coinsurance	double	
paid_patient_deductible	double	
paid_by_primary	double	
paid_ingredient_cost	double	
paid_dispensing_fee	double	
amount_allowed	double	
revenue_code_source_value	text	
drg_source_value	text	

La tabla **COST** se utiliza para registrar información relacionada con los costes asociados a eventos clínicos y de atención médica en los datos de los pacientes.

Esta tabla es fundamental para el análisis económico y de costes en el contexto de la atención médica y la **Evaluación de tecnologías sanitarias (HTA)**

payer_plan_period		
payer_plan_period_id	bigint	NN
person_id	bigint	NN
contract_person_id	bigint	
payer_plan_period_start_date	date	NN
payer_plan_period_end_date	date	NN
payer_concept_id	integer	NN
plan_concept_id	integer	NN
contract_concept_id	integer	
sponsor_concept_id	integer	
stop_reason_concept_id	integer	NN
payer_source_concept_id	integer	NN
plan_source_concept_id	integer	NN
contract_source_concept_id	integer	
sponsor_source_concept_id	integer	NN
stop_reason_source_concept_id	integer	NN
payer_source_value	text	
plan_source_value	text	
contract_source_value	text	
sponsor_source_value	text	
family_source_value	text	
stop_reason_source_value	text	

La tabla **PAYER_PLAN_PERIOD** se utiliza para registrar información relacionada con los tipos y períodos de cobertura de planes de pago o seguros médicos para los pacientes y se referencia desde la tabla **COST**


Standardized clinical data
Person
Observation_period
Death
Visit_occurrence
Visit_detail
Condition_occurrence
Drug_exposure
Procedure_occurrence
Device_exposure
Measurement
Observation
Note
Note_NLP
Episode
Specimen
Episode_event
Fact_relationship
Standardized health system
Location
Care_site
Provider
Standardized vocabularies
Concept
Vocabulary
Domain
Concept_class
Concept_synonym
Concept_relationship
Relationship
Concept_ancestor
Source_to_concept_map
Drug_strength
Standardized health economics
Cost
Payer_plan_period
Standardized derived elements
Condition_era
Drug_era
Dose_era
Results schema
Cohort
Cohort_definition
Standardized metadata
CDM_source
Metadata

condition_era	
condition_era_id	INTEGER
person_id	INTEGER
condition_concept_id	INTEGER
condition_era_start_date	DATE
condition_era_end_date	DATE
condition_occurrence_count	INTEGER

La tabla **CONDITION_OCCURRENCE** almacena eventos individuales de condiciones médicas que se han registrado para un paciente.

En la tabla **CONDITION_ERA**, se agrupan las condiciones médicas que ocurren en secuencia continua sin interrupciones para un paciente específico, es decir, aquellas que se superponen o se encuentran en secuencia temporal sin un período de tiempo significativo entre ellas (normalmente 30 días).

drug_era	
drug_era_id	bigint NN
person_id	bigint NN
drug_concept_id	integer NN
drug_exposure_count	integer
gap_days	integer
drug_era_start_date	date
drug_era_end_date	date

Para crear **DRUG_ERA**, se agrupan las exposiciones a medicamentos que ocurren en secuencia continua sin interrupciones significativas para un paciente específico.

Las exposiciones a medicamentos que cumplen con este criterio se agrupan juntas y se registran en DRUG_ERA como una única "era" de uso de medicamento.

La columna drug_era_start_date de la era de uso de medicamento se establece en la fecha de inicio de la primera exposición a medicamento en el grupo, y la columna drug_era_end_date se establece en la fecha de finalización de la última exposición a medicamento en el grupo.

Se contabiliza la cantidad de exposiciones a medicamentos en esa era y se registra en la columna drug_exposure_count.

dose_era		
dose_era_id	bigint	NN
person_id	bigint	NN
dose_era_start_datetime	timestamp	NN
dose_era_end_datetime	timestamp	NN
drug_concept_id	integer	NN
unit_concept_id	integer	NN
dose_value	numeric	NN

Del mismo modo **DOSE_ERA** se genera agrupando las exposiciones a dosis de medicamentos que ocurren en secuencia continua sin interrupciones significativas para un paciente específico.

Las exposiciones a dosis que cumplen con este criterio se agrupan y se registran en DOSE_ERA como una única "era" de exposición a dosis de medicamento.

La columna dose_era_start_date de la era de exposición a dosis se establece en la fecha de inicio de la primera exposición a dosis en el grupo, y la columna dose_era_end_date se establece en la fecha de finalización de la última exposición a dosis en el grupo.

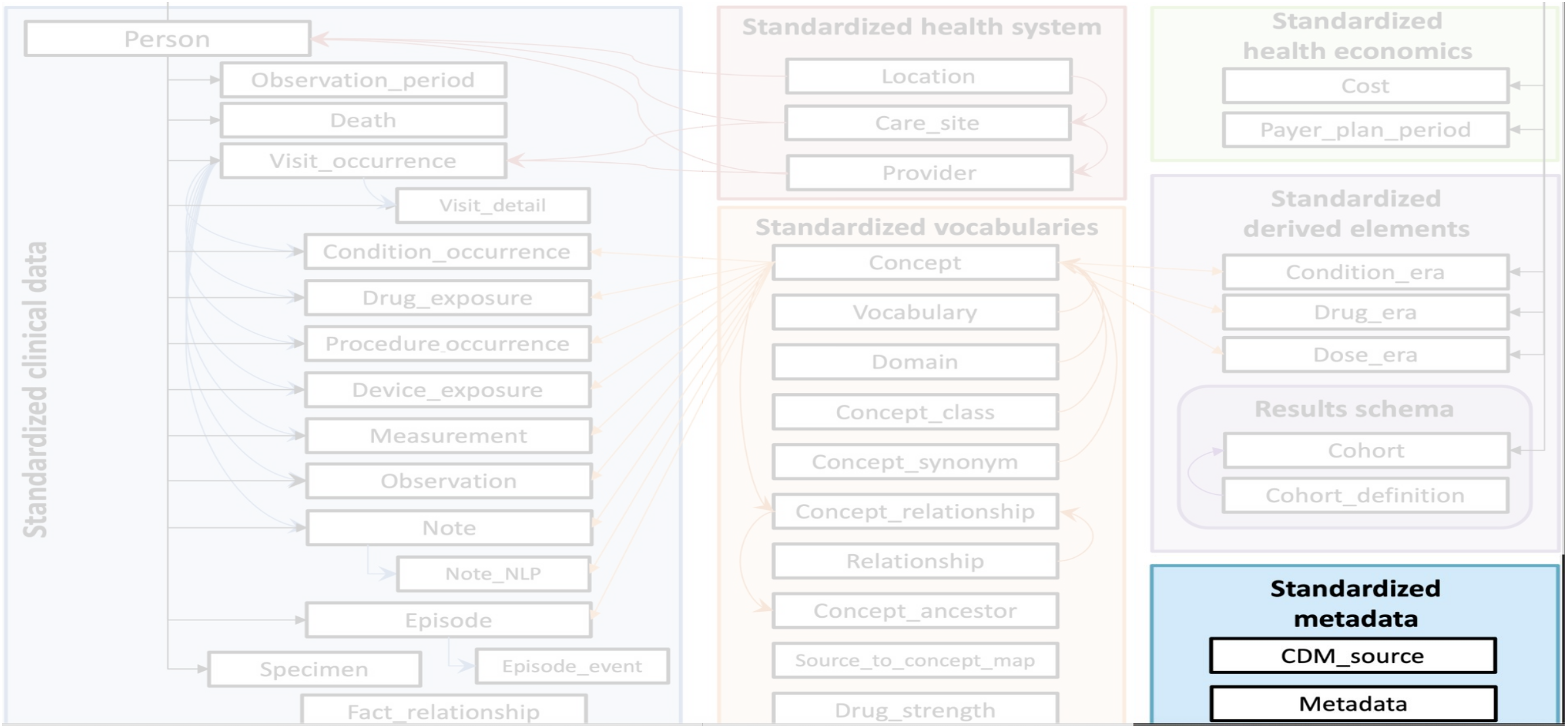
observation_period		
observation_period_id	**INTEGER**	**NN**
person_id	INTEGER	NN
observation_period_start_date	DATE	NN
observation_period_end_date	DATE	NN
period_type_concept_id	INTEGER	

La tabla **OBSERVATION_PERIOD** se utiliza para registrar los períodos de tiempo durante los cuales un paciente es observado o seguido en el sistema de atención médica.

Contiene información sobre la fecha de inicio (primer registro) y fin del período de observación para cada paciente (último registro + 60 días, fecha de muerte + 60 días, fecha de extracción).

Puede haber **múltiples períodos** de observación para un paciente a lo largo del tiempo (gap de 18 meses).

Ayuda a determinar el tiempo durante el cual se recopilaron datos clínicos para un paciente en particular.


Standardized clinical data
Person
Observation_period
Death
Visit_occurrence
Visit_detail
Condition_occurrence
Drug_exposure
Procedure_occurrence
Device_exposure
Measurement
Observation
Note
Note_NLP
Episode
Specimen
Episode_event
Fact_relationship
Standardized health system
Location
Care_site
Provider
Standardized vocabularies
Concept
Vocabulary
Domain
Concept_class
Concept_synonym
Concept_relationship
Relationship
Concept_ancestor
Source_to_concept_map
Drug_strength
Standardized health economics
Cost
Payer_plan_period
Standardized derived elements
Condition_era
Drug_era
Dose_era
Results schema
Cohort
Cohort_definition
Standardized metadata
CDM_source
Metadata

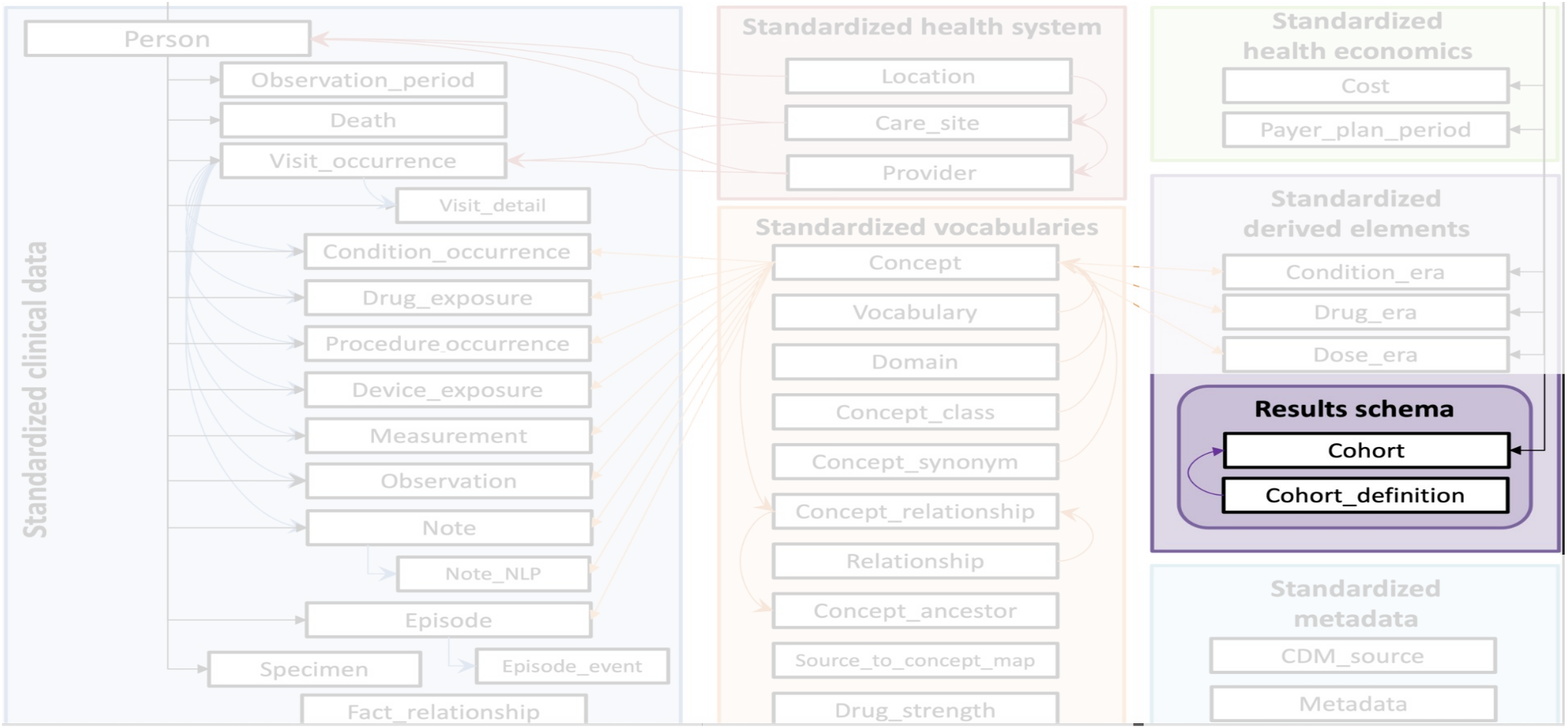
cdm_source	
cdm_version_concept_id	integer
source_release_date	date NN
cdm_release_date	date
cdm_source_name	text NN
cdm_source_abbreviation	text NN
cdm_holder	text NN
source_description	text
source_documentation_reference	text
cdm_etl_reference	text
cdm_version	text
vocabulary_version	text

Información sobre esta
instalación de OMOP-CDM

metadata	
metadata_datetime	timestamp
metadata_concept_id	integer NN
metadata_type_concept_id	integer NN
value_as_concept_id	integer
metadata_date	date
name	text NN
value_as_string	text

Posibles metadatos:

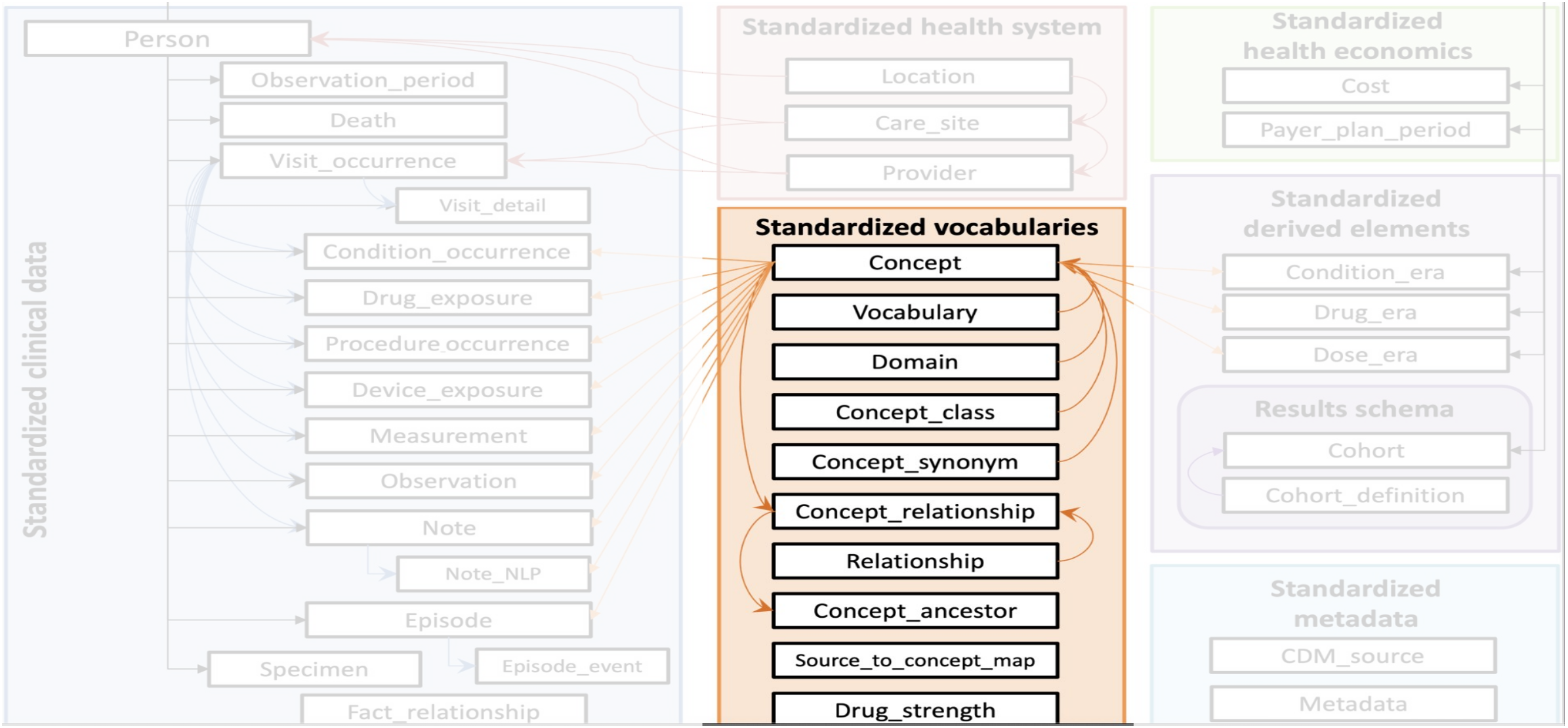
Data Source
Data Extraction Date
ETL Version
Last ETL Refresh Date
Total Patients Loaded
Domains covered
Mapping Conventions
Data Quality Checks
Data License Information

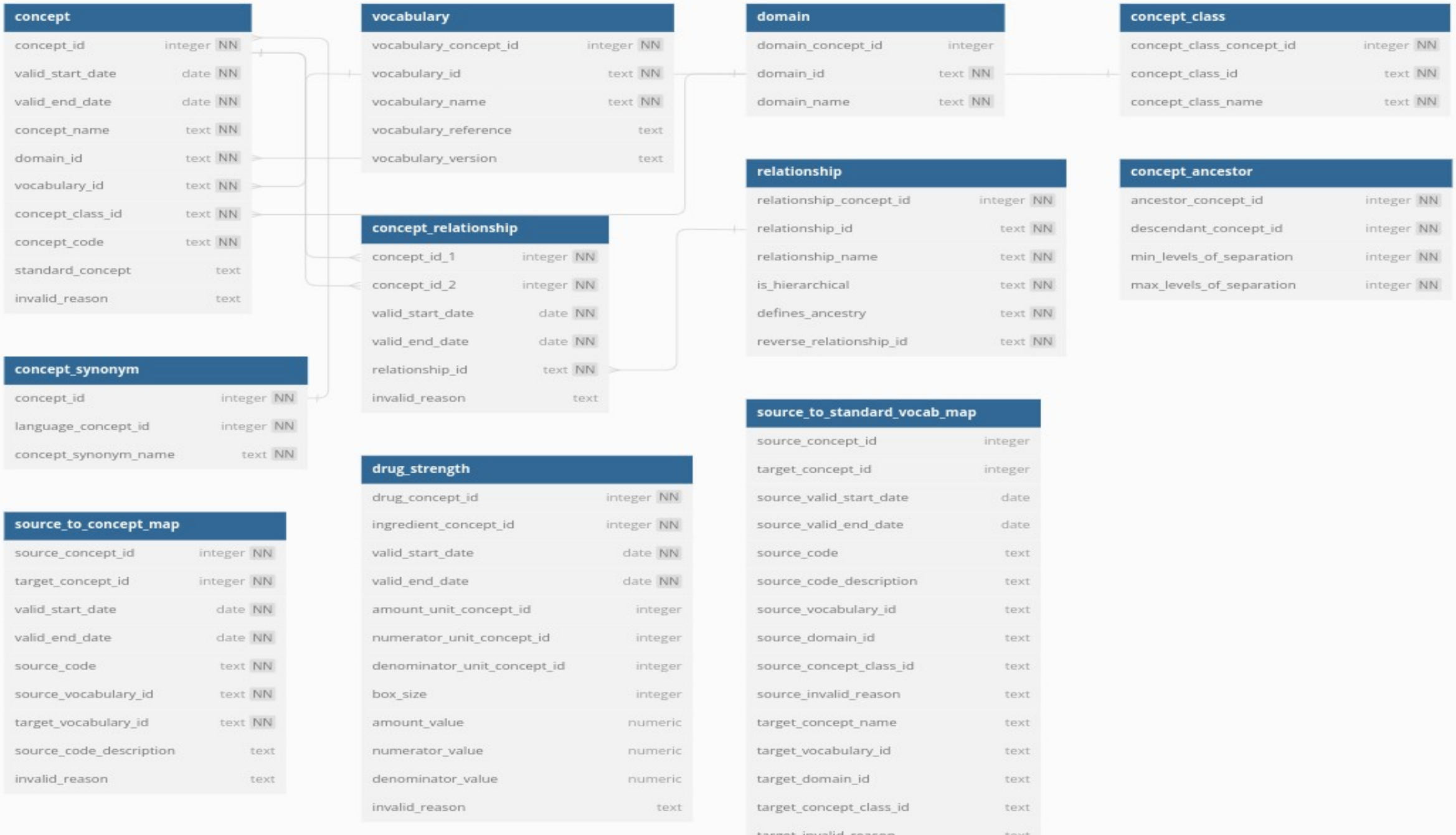

Standardized clinical data
Person
Observation_period
Death
Visit_occurrence
Visit_detail
Condition_occurrence
Drug_exposure
Procedure_occurrence
Device_exposure
Measurement
Observation
Note
Note_NLP
Episode
Specimen
Episode_event
Fact_relationship
Standardized health system
Location
Care_site
Provider
Standardized vocabularies
Concept
Vocabulary
Domain
Concept_class
Concept_synonym
Concept_relationship
Relationship
Concept_ancestor
Source_to_concept_map
Drug_strength
Standardized health economics
Cost
Payer_plan_period
Standardized derived elements
Condition_era
Drug_era
Dose_era
Results schema
Cohort
Cohort_definition
Standardized metadata
CDM_source
Metadata

COHORTES

cohort_definition and cohort

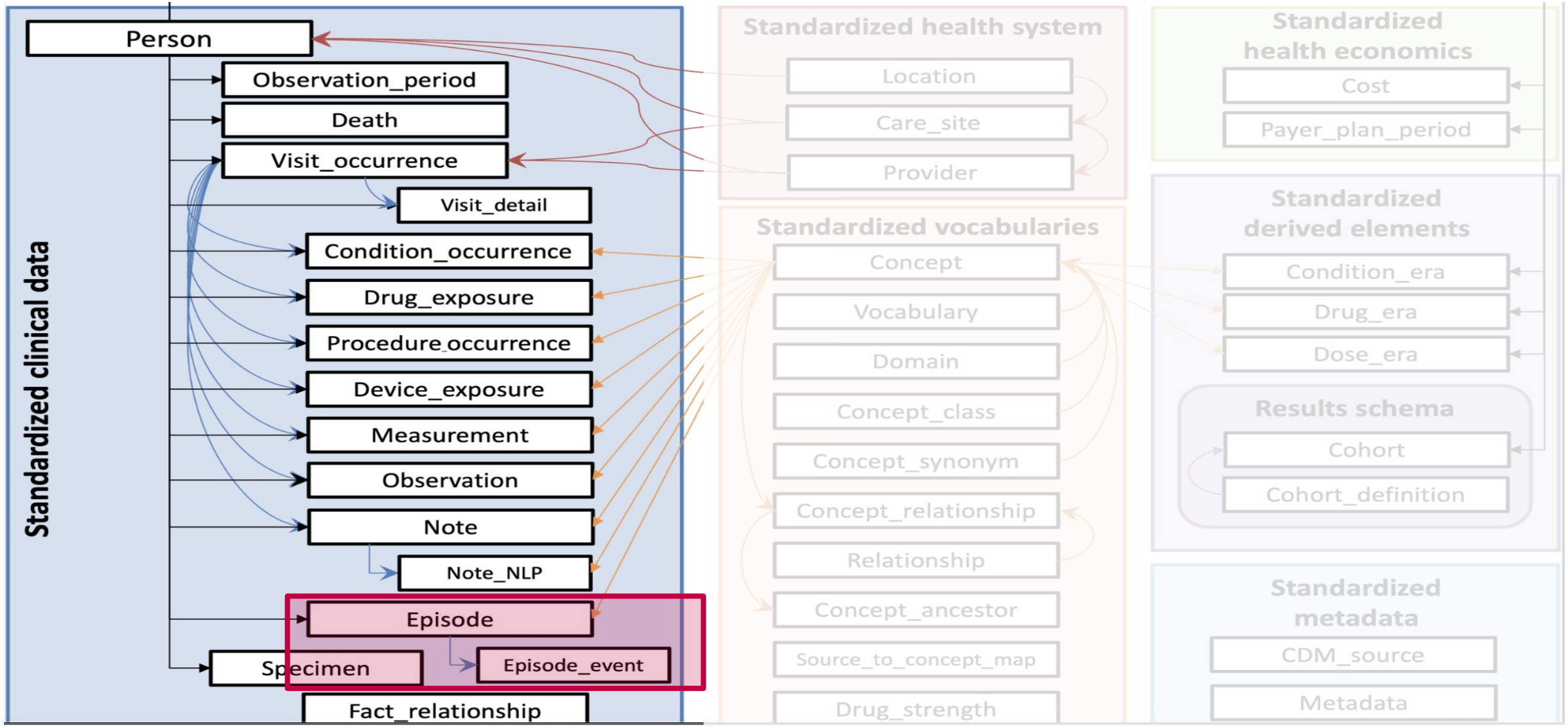
Otro de los aspectos importantes del modelo común de datos OMOP es el concepto de cohorte. Una cohorte es un conjunto de personas que satisfacen uno o más criterios de inclusión al conjunto durante un período determinado de tiempo. Estos conjuntos se utilizan en todas las herramientas analíticas y estudios de OHDSI como los elementos fundamentales para llevar a cabo una investigación. Las cohortes se definen de forma independiente a los demás conjuntos en una investigación


Standardized clinical data
Person
Observation_period
Death
Visit_occurrence
Visit_detail
Condition_occurrence
Drug_exposure
Procedure_occurrence
Device_exposure
Measurement
Observation
Note
Note_NLP
Episode
Specimen
Episode_event
Fact_relationship
Standardized health system
Location
Care_site
Provider
Standardized vocabularies
Concept
Vocabulary
Domain
Concept_class
Concept_synonym
Concept_relationship
Relationship
Concept_ancestor
Source_to_concept_map
Drug_strength
Standardized health economics
Cost
Payer_plan_period
Standardized derived elements
Condition_era
Drug_era
Dose_era
Results schema
Cohort
Cohort_definition
Standardized metadata
CDM_source
Metadata


concept
concept_id integer NN
valid_start_date date NN
valid_end_date date NN
concept_name text NN
domain_id text NN
vocabulary_id text NN
concept_class_id text NN
concept_code text NN
standard_concept text
invalid_reason text
vocabulary
vocabulary_concept_id integer NN
vocabulary_id text NN
vocabulary_name text NN
vocabulary_reference text
vocabulary_version text
domain
domain_concept_id integer
domain_id text NN
domain_name text NN
concept_class
concept_class_concept_id integer NN
concept_class_id text NN
concept_class_name text NN
relationship
relationship_concept_id integer NN
relationship_id text NN
relationship_name text NN
is_hierarchical text NN
defines_ancestry text NN
reverse_relationship_id text NN
concept_ancestor
ancestor_concept_id integer NN
descendant_concept_id integer NN
min_levels_of_separation integer NN
max_levels_of_separation integer NN
concept_relationship
concept_id_1 integer NN
concept_id_2 integer NN
valid_start_date date NN
valid_end_date date NN
relationship_id text NN
invalid_reason text
concept_synonym
concept_id integer NN
language_concept_id integer NN
concept_synonym_name text NN
source_to_standard_vocab_map
source_concept_id integer
target_concept_id integer
source_valid_start_date date
source_valid_end_date date
source_code text
source_code_description text
source_vocabulary_id text
source_domain_id text
source_concept_class_id text
source_invalid_reason text
target_concept_name text
target_vocabulary_id text
target_domain_id text
target_concept_class_id text
drug_strength
drug_concept_id integer NN
ingredient_concept_id integer NN
valid_start_date date NN
valid_end_date date NN
amount_unit_concept_id integer
numerator_unit_concept_id integer
denominator_unit_concept_id integer
box_size integer
amount_value numeric
numerator_value numeric
denominator_value numeric
invalid_reason text
source_to_concept_map
source_concept_id integer NN
target_concept_id integer NN
valid_start_date date NN
valid_end_date date NN
source_code text NN
source_vocabulary_id text NN
target_vocabulary_id text NN
source_code_description text
invalid_reason text

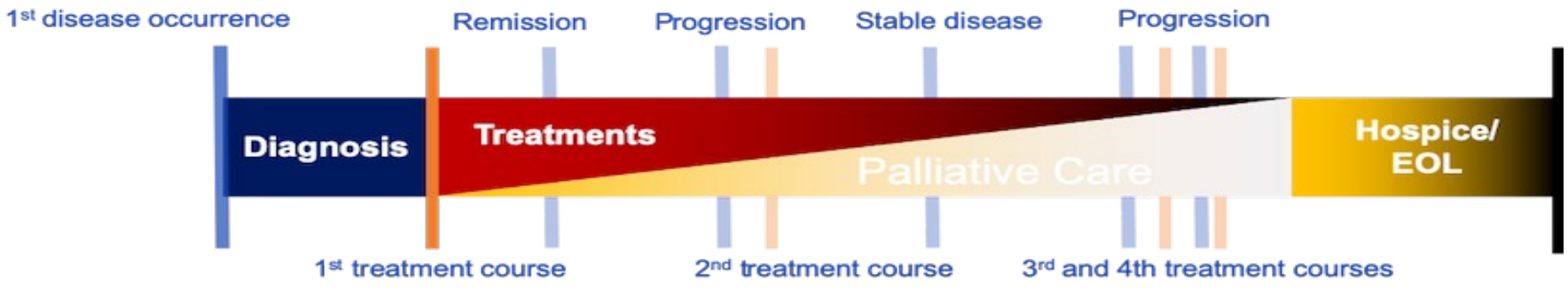
OMOP-CDM*

EXTENSIONES AL MODELO


Standardized clinical data
Person
Observation_period
Death
Visit_occurrence
Visit_detail
Condition_occurrence
Drug_exposure
Procedure_occurrence
Device_exposure
Measurement
Observation
Note
Note_NLP
Episode
Specimen
Episode_event
Fact_relationship
Standardized health system
Location
Care_site
Provider
Standardized vocabularies
Concept
Vocabulary
Domain
Concept_class
Concept_synonym
Concept_relationship
Relationship
Concept_ancestor
Source_to_concept_map
Drug_strength
Standardized health economics
Cost
Payer_plan_period
Standardized derived elements
Condition_era
Drug_era
Dose_era
Results schema
Cohort
Cohort_definition
Standardized metadata
CDM_source
Metadata

ONCOLOGY EXTENSION

EPISODE and EPISODE_EVENT



Example of disease and treatment episodes in the Episode table

'First occurrence'-of-'Carcinoma of breast'

episode_id	12345
person_id	1
episode_concept_id	32528
episode_start_datetime	June 9, 2019
episode_object_concept_id	4116071
episode_type_concept_id	32535

OMOP concept 'First disease occurrence'

SNOMED concept 'Carcinoma of breast'

OMOP concept 'Tumor registry'

'Treatment regimen'-of-' Paclitaxel and Bevacizumab'

episode_id	12346
person_id	1
episode_concept_id	32531
episode_start_datetime	July 9, 2019
episode_parent_id	12345
episode_object_concept_id	35804255
episode_type_concept_id	32545

OMOP concept 'Treatment Regimen'

Foreign key to the disease Episode record

HemOnc concept Paclitaxel and Bevacizumab'

OMOP concept 'Episode algorithmically derived from EHR'

Oncology Extension

episode	
episode_id	bigint NN
person_id	bigint NN
episode_concept_id	integer NN
episode_start_datetime	timestamp NN
episode_end_datetime	timestamp
episode_parent_id	bigint
episode_number	integer
episode_object_concept_id	integer NN
episode_type_concept_id	integer NN
episode_source_value	text
episode_source_concept_id	integer

episode_event	
episode_id	bigint NN
event_id	bigint NN
episode_event_field_concept_id	integer NN

measurement	
measurement_id	bigint NN
person_id	bigint NN
measurement_concept_id	integer NN
measurement_date	date
measurement_datetime	timestamp NN
measurement_time	text
measurement_type_concept_id	integer NN
operator_concept_id	integer
value_as_number	numeric
value_as_concept_id	integer
unit_concept_id	integer
range_low	numeric
range_high	numeric
provider_id	bigint
visit_occurrence_id	bigint
visit_detail_id	bigint
measurement_source_value	text
measurement_source_concept_id	integer NN
unit_source_value	text
value_source_value	text
modifier_of_event_id	bigint
modifier_of_field_concept_id	integer

modifier_of_event_id	**No**	**bigint**	**A foreign key identifier to the event (e.g. condition, procedure, episode) record for which the modifier is recorded.**
modifier_of_field_concept_id	**No**	**integer**	**The concept representing the table field concept that contains the value of the event id for which the modifier is recorded (e.g. CONDITION_OCCURRENCE.condition_occurrence_id).**

Oncology Extension

Example of cancer diagnosis in the OMOP CDM

Histology+Site diagnosis in **Condition_Occurrence**

condition_occurrence_id	123456789	
person_id	1	
condition_concept_id	4116071	SNOMED concept 'Carcinoma of breast'
condition_start_datetime	June 9, 2019	
condition_type_concept_id	32535	
condition_source_value	8010/3-C50.9	Precoordinated concept of ICD-O Histology & Site
condition_source_concept_id	44505310	

measurement_id	567890	
person_id	1	
measurement_datetime	June 9, 2019	
measurement_concept_id	35918640	NAACCR concept 'Grade Pathological'
measurement_date	June 9, 2019	
value_as_concept_id	35922509	NAACCR concept 'G3: High combined histologic grade (unfavorable); SBR score of 8-9 points'
measurement_type_concept_id	32534	OMOP concept 'Tumor registry'
measurement_source_value	3844	NAACCR code for 'Grade Pathological'
measurement_source_concept_id	35918640	NAACCR concept 'Grade Pathological'
value_source_value	breast@3844@3	NAACCR code for 'G3: High combined histologic grade (unfavorable); SBR score of 8-9 points'
modifier_of_event_id	123456789	Value of the respective condition record condition_occurrence_id
modifier_field_concept_id	1147127	Concept for 'condition_occurrence.condition_occurrence_id'

Oncology Extension

Cancer Modifier ×
Staging/Grading ×

- DOMAIN
- CONCEPT
- CLASS

filter

- [ ] Dimension (29)
- [ ] Extension/Invasion (545)
- [ ] Histopattern (444)
- [ ] Margin (433)
- [ ] Metastasis (579)
- [ ] Morph Abnormality (15)
- [ ] Nodes (192)
- [ ] Qualifier Value (7)
- [x] Staging/Grading (3281)
- [ ] Topography (518)

DOWNLOAD RESULTS

Show by 15 items Total 3,281 items

1 2 3 4 5 ··· 219 >

ID	CODE	NAME	CLASS	CONCEPT	VALIDITY	DOMAIN	VOCAB
1634568	advanced	Advanced Stage	Staging/Grading	Standard	Valid	Measurement	Cancer Modifier
1635479	AJCC-TRG-0	AJCC Tumor Regression Grading Grade 0	Staging/Grading	Standard	Valid	Measurement	Cancer Modifier
1635394	AJCC-TRG-1	AJCC Tumor Regression Grading Grade 1	Staging/Grading	Standard	Valid	Measurement	Cancer Modifier
1634517	AJCC-TRG-2	AJCC Tumor Regression Grading Grade 2	Staging/Grading	Standard	Valid	Measurement	Cancer Modifier
1634617	AJCC-TRG-3	AJCC Tumor Regression Grading Grade 3	Staging/Grading	Standard	Valid	Measurement	Cancer Modifier
1634194	c-6th_AJCC/UICC-M0	AJCC/UICC 6th clinical M0 Category	Staging/Grading	Standard	Valid	Measurement	Cancer Modifier
1634042	c-6th_AJCC/UICC-M0(i+)	AJCC/UICC 6th clinical M0(i+) Category	Staging/Grading	Standard	Valid	Measurement	Cancer Modifier
1634829	c-6th_AJCC/UICC-M1	AJCC/UICC 6th clinical M1 Category	Staging/Grading	Standard	Valid	Measurement	Cancer Modifier
1633709	c-6th_AJCC/UICC-M1a	AJCC/UICC 6th clinical M1a Category	Staging/Grading	Standard	Valid	Measurement	Cancer Modifier
1634442	c-6th_AJCC/UICC-M1b	AJCC/UICC 6th clinical M1b Category	Staging/Grading	Standard	Valid	Measurement	Cancer Modifier

NAACCR ×
NAACCR Variable ×

- DOMAIN
- CONCEPT
- CLASS

filter

- [ ] NAACCR Proc Schema (30)
- [ ] NAACCR Procedure (707)
- [ ] NAACCR Schema (195)
- [ ] NAACCR Value (31179)
- [x] NAACCR Variable (1881)
- [ ] Permissible Range (481)

DOWNLOAD RESULTS

Show by 15 items Total 1,881 items

1 2 3 4 5 ··· 126 >

ID	CODE	NAME	CLASS	CONCEPT	VALIDITY	DOMAIN	VOCAB
35917581	rectum@2864	18q Loss of Heterozygosity (LOH)	NAACCR Variable	Standard	Valid	Measurement	NAACCR
35917706	appendix@2864	18q Loss of Heterozygosity (LOH)	NAACCR Variable	Standard	Valid	Measurement	NAACCR
35917755	colon@2864	18q Loss of Heterozygosity (LOH)	NAACCR Variable	Standard	Valid	Measurement	NAACCR
35918610	3803	Adenoid Cystic Basaloid Pattern	NAACCR Variable	Standard	Valid	Measurement	NAACCR
35917191	lacrimal_gland@2930	Adenoid Cystic Carcinoma - Presence of Basaloid Pattern	NAACCR Variable	Standard	Valid	Measurement	NAACCR
35918528	3804	Adenopathy	NAACCR Variable	Standard	Valid	Measurement	NAACCR
35918321	3805	AFP Post-Orchiectomy Lab Value	NAACCR Variable	Standard	Valid	Measurement	NAACCR

Oncology Extension

Cómo puedo encontrar pacientes con tumores de menos de 2cm en mi modelo?

concept_id	concept_name	domain_id	vocabulary_id	concept_class_id	standard_concept	concept_code
45883535	Mild	Meas Value	LOINC	Answer	S	LA6752-5
35920053	Described as "less than 1 centimeter (cm)"	Meas Value	NAACCR	Cancer Identification	S	melanoma_skin@2800@991
35919956	Described as "less than 2 cm," or "greater than 1 cm," or "between 1 cm and 2 cm"	Meas Value	NAACCR	Cancer Identification	S	melanoma_mouth_other@2800@992

Oncology Extension

Cómo puedo encontrar pacientes con tumores de menos de 2cm en mi modelo?

```
SELECT *
FROM measurement
WHERE measurement_concept_id = 4139794  -- Tumor size
AND ( (value_as_number <= 2 and unit_concept_id = 8582) -- Centimeter
OR (value_as_number <= 20 and unit_concept_id = 8588)) --Millimeter
AND COALESCE(operator_concept_id, 4171754) = 4171754 -- Less than or equal
```

Oncology Extension

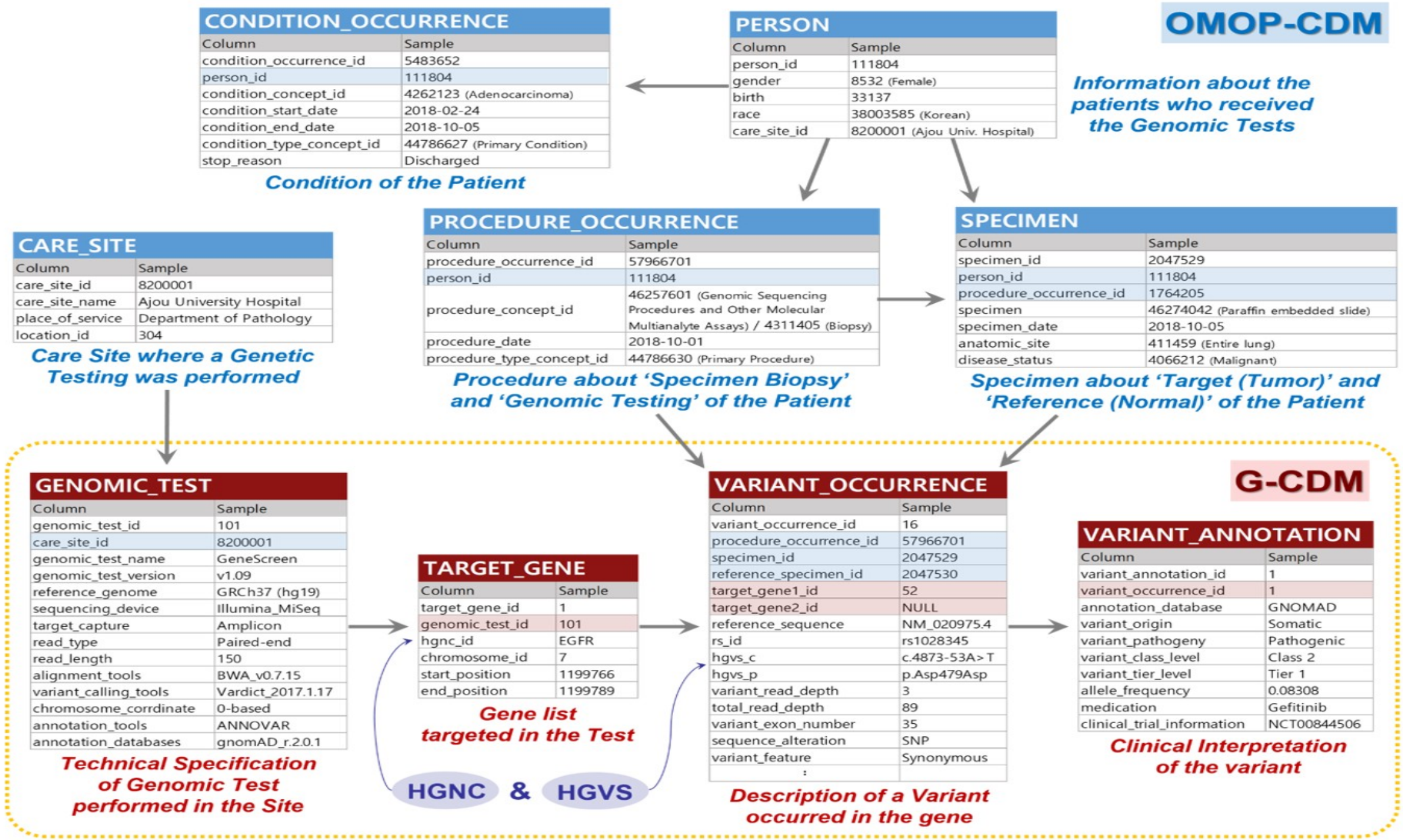
concept_numeric	
concept_id	integer
value_as_number	float
unit_concept_id	integer
operator_concept_id	float

Cómo puedo encontrar pacientes con tumores de menos de 2cm en mi modelo?

CONCEPT_NUMERIC

concept_id	value_as_number	unit_concept_id	operator_concept_id	Description
45883535	1			1 is numeric value of 'Mild'
35920053	1	8582	4171756	'Less than 1 cm'
35919956	1	8582	4171755	Lower limit of 'between 1 cm and 2 cm'
35919956	2	8582	4171754	Upper limit of 'between 1 cm and 2 cm'

Genomics Extension



Oncology Extension

Genomic Vocabularies

- HGNC (HUGO Gene Nomenclature Committee)
- ClinVar
- CiVIC (Clinical Interpretation of Variants in Cancer)
- NCIt (National Cancer Institute Thesaurus)
- CGI (Cancer Genome Interpreter)
- CAP (College of American Pathologists Cancer Checklists)
- JAX CKB (The Clinical Knowledgebase by The Jackson Laboratory)

Medical Imaging

DICOM

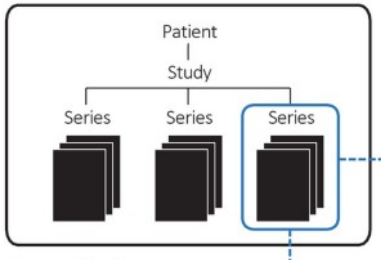
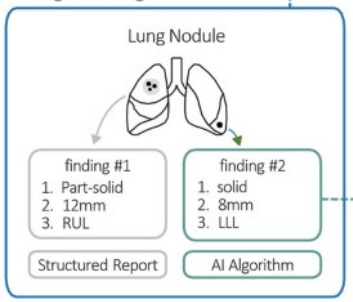


Image Findings



Image_occurrence Table

Column names	Row values
image_occurrence_id	37899
person_id	1234
procedure_occurrence_id	34445
visit_occurrence_id	30
anatomic_site_concept_id	4118108
wadors_uri	https://server.com/studies/1.2.3. ... 5630178
local_path	\\johnshopkins\micdm\lung_example
image_occurrence_date	2021-10-14
image_study_UID	1.2.3.1.4.1.14519.5.2.1.6279. ... 5630178
image_series_UID	1.2.3.1.4.1.14519.5.2.1.6279. ... 2063192
modality_concept_id	4300757

Procedure_occurrence Table

Column names	Row values
procedure_occurrence_id	34445
person_id	1234
procedure_concept_id	2211378
procedure_date	2021-10-14
procedure_end_date	2021-10-14
procedure_type_concept_id	32817
quantity	1
provider_id	78965412
procedure_source_value	71250
procedure_source_concept_id	2211378

Image_feature Table

Column names	Row values 1	Row values 2	Row values 3
image_feature_id	9825	9826	9827
person_id	1234	1234	1234
image_occurrence_id	37899	37899	37899
image_feature_event_field_concept_id	1147330	1147330	1147330
image_feature_event_id	9	10	11
image_feature_concept_id*	2000180050	20000180060	2100018077
image_feature_type_concept_id*	2000580000	20000580000	2000500000
image_finding_concept_id*	NULL	NULL	2037206719
image_finding_id*	NULL	NULL	2100046813
anatomic_site_concept_id	4118108	4118108	4213162
alg_system	NULL	NULL	http://jhu.ai_linear_learner/version/2/...
alg_datetime	NULL	NULL	2022-01-03 00:00

Measurement Table

Column names	Row values 1	Row values 2	Row values 3
measurement_id	9	10	11
person_id	1234	1234	1234
measurement_concept_id*	2000180050	20000180060	2100018077
measurement_source_value*	00180050	00180060	0018077
measurement_date	2021-10-14	2021-10-14	2021-10-14
measurement_type_concept_id	32817	32817	32817
value_as_number	1	150	8
value_as_concept_id	NULL	NULL	NULL
unit_concept_id	8588	200001234*	8588
unit_source_value	millimeter	kVp	millimeter
value_source_value	1	150	8

Person Table and
Other OMOP Clinical Data Tables

Image_occurrence Legends

Columns	Descriptions
anatomic_site_concept_id	4118108 refers to entire thorax
wadors_uri	DICOMweb URI
local_path	User's local path of the series
image_occurrence_date	DICOM tag (0008,0020) Study Date
image_study_UID	DICOM tag (0020,000D) Study Instance UID
image_series_UID	DICOM tag(0020,000E) Series Instance UID
modality_concept_id	OMOP concept id for the DICOM tag (0008,0060) Modality

Image_feature Legends

Columns	Descriptions
image_feature_event_field_concept_id	1147330 refers to the MEASUREMENT table.
image_feature_concept_id*	custom IDs generated based on DICOM tag, e.g., 2000180050 for (0018,0050) slice thickness
image_feature_type_concept_id*	2000580000 refers to acquisition parameters; 2000500000 refers to algorithms
image_finding_concept_id*	2037206719 refers to a nodule
image_finding_id*	2100046813 refers to the nodule that the feature is describing.
anatomic_site_concept_id	4213162 refers to lung structure

Preguntas?

GRACIAS POR VUESTRA ATENCIÓN

You are free to:

Share — copy and redistribute the material in any medium or format

Under the following terms:

Attribution - You must give appropriate credit , provide a link to the license, and indicate if changes were made . You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.

NonCommercial - You may not use the material for commercial purposes .

NoDerivatives - If you remix, transform, or build upon the material, you may not distribute the modified material.