INTRODUCCIÓN AL MODELO DE DATOS OMOP-CDM

Alberto Labarga

2023-09-14

Telecommunications Engineer
Head of Biomedical Data Hub @BSC
More than 20 years teaching

open data - open source – open science

I am Alberto

alabarga
alabarga
/in/albertolabarga

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Descripción del curso

El curso tiene como objetivo proporcionar una comprensión profunda del modelo de datos **OMOP-CDM de la OHDSI**, sus vocabularios y cómo usar la **aplicación web Atlas** para crear cohortes y extraer datos para **estudios observacionales**. El curso cubrirá los conceptos y principios básicos del modelo de datos OMOP-CDM, incluida su estructura, tablas clave y vocabularios. Además, el curso cubrirá los aspectos prácticos de trabajar con **Atlas**, incluido cómo crear y ejecutar análisis de **cohortes**.

Introducción al modelo de datos OMOP-CDM

- Descripción general del modelo de datos OMOP-CDM
- Conceptos y principios clave del modelo de datos OMOP-CDM
- Comprensión de la estructura y tablas clave del modelo de datos OMOP-CDM
- Introducción a los vocabularios OMOP-CDM
- Resultados del aprendizaje: los participantes comprenderán los **conceptos básicos del modelo de datos OMOP-CDM**, su estructura y los vocabularios utilizados.

Materiales del curso

alabarga / **omop-cdm-course** Public

Code | Issues | Pull requests | Actions | Projects | Security | Insights

main | 1 branch | 0 tags | Go to file | Code

alabarga Update README.md af00948 8 hours ago **18** commits

docs	Add course outline	6 months ago
queries	Add queries	6 months ago
slides	Update README.md	8 hours ago
.gitignore	Initial commit	6 months ago
INSTALL.md	Update INSTALL.md	10 hours ago
LICENSE	Initial commit	6 months ago
README.md	Update README.md	8 hours ago
docker-compose.yml	Add files via upload	10 hours ago

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

(Clinical) Data is complicated

(Clinical) Data is complicated

- 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 complicated

- 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

Protección de Datos

- El **Reglamento General de Protección de Datos (RGPD)** es una ley de la Unión Europea que entró en vigor el 25 de mayo de 2018, y que establece las reglas para la protección de datos personales de los ciudadanos europeos.
- El objetivo principal del RGPD es proteger los **derechos de privacidad** de los ciudadanos europeos, y proporcionar un **marco legal claro y coherente** para el **tratamiento de datos personales**. Para ello, establece un **conjunto de normas y principios** para garantizar que las empresas y organizaciones que manejan datos personales, incluyendo empresas en línea, procesen la información de forma justa y transparente.

Protección de Datos

- Entre los principales objetivos del RGPD se incluyen:
  - La necesidad de obtener el **consentimiento expreso** del titular de los datos antes de recopilar, utilizar o compartir sus datos personales.
  - La obligación de las empresas y organizaciones de **informar** clara y concisamente a los titulares de datos sobre cómo se utilizarán sus datos personales.
  - El derecho de los titulares de datos a **solicitar el acceso** a sus datos personales, su rectificación o supresión, y a obtener una copia de sus datos personales en un formato legible por máquina.
  - La obligación de las empresas y organizaciones de tomar **medidas de seguridad adecuadas** para proteger los datos personales, y notificar a los titulares de datos y a las autoridades competentes en caso de violación de seguridad de los datos.

OPEN

FAIR

Datos FAIR

FINDABLE

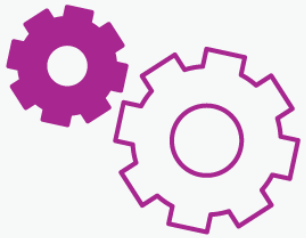
Identificadores únicos y metadatos que pueden usarse para encontrar los datos de manera sencilla

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

(Clinical) Data is complicated

- 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** para la recopilación y el análisis de estos, así como el desarrollo de **herramientas** y técnicas que hagan posible la combinación de los datos

(Clinical) Data is complicated

- 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.
- 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.

(Clinical) Data is complicated

- Los **estándares de datos** en medicina se refieren a un conjunto de **reglas y formato**s 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.

(Clinical) Data is complicated

- **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.

(Clinical) Data is complicated

- **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).**

(Clinical) Data is complicated

- **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**.

OMOP-CDM

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

Introducción al modelo de datos OMOP-CDM

OHDSI

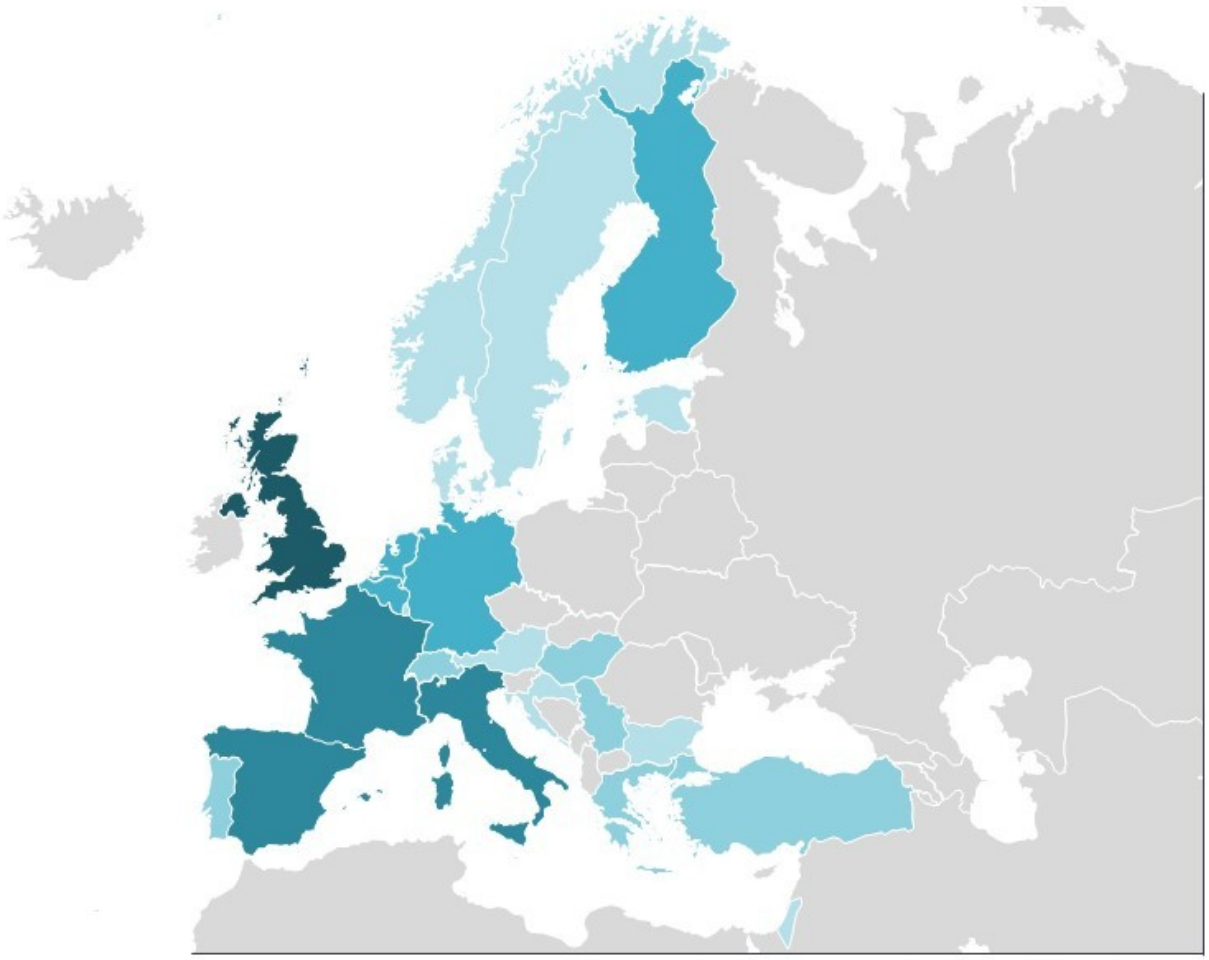
- La **Observational Health Data Sciences and Informatics (OHDSI)** es una red internacional de investigadores y bases de datos observacionales de salud, la cual tiene como objetivo crear y aplicar soluciones analíticas de datos de código abierto a una gran red de bases de datos clínicos para mejorar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

OHDSI
OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS

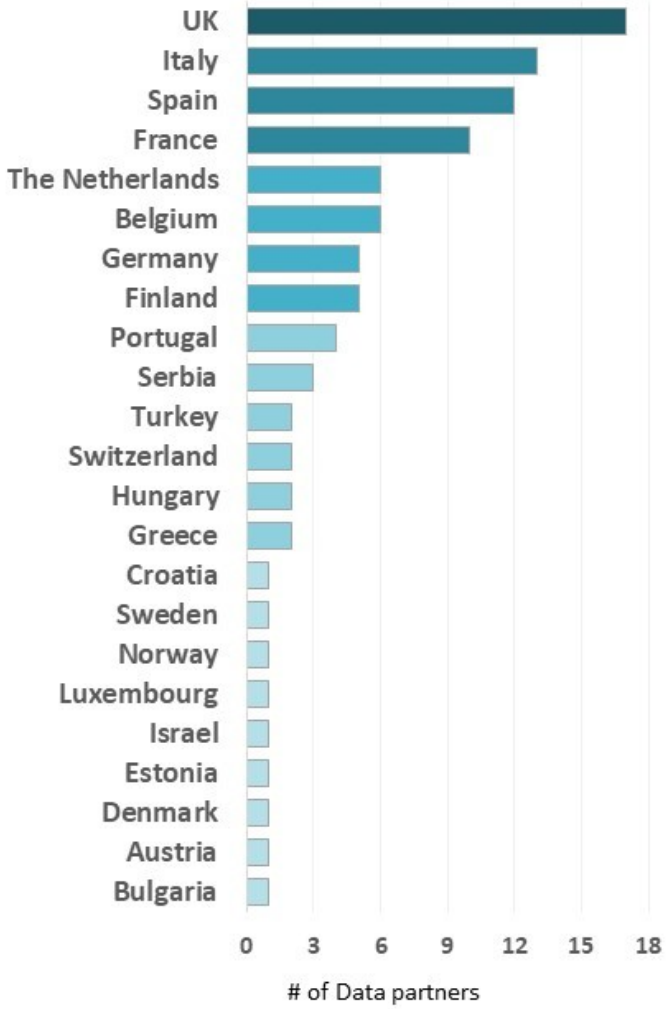
EHDEN

- The **European Health Data and Evidence Network (EHDEN)** is developing a federated and equitable ecosystem of institutions generating clinical data, with researchers across academia and industry, supported by certified and qualified SMEs, **harmonising clinical data** and creating a network technology for real-world data research.
- Centres joining the EHDEN network could be university or non-academic hospitals, primary care networks, existing registries, etc.
- Study execution is performed in a **federated** manner, i.e. analytical code is sent to the data source, run locally in their own safe haven, and only the aggregated results are shared.
- https://www.ehden.eu/

EHDEN
EUROPEAN HEALTH DATA & EVIDENCE NETWORK

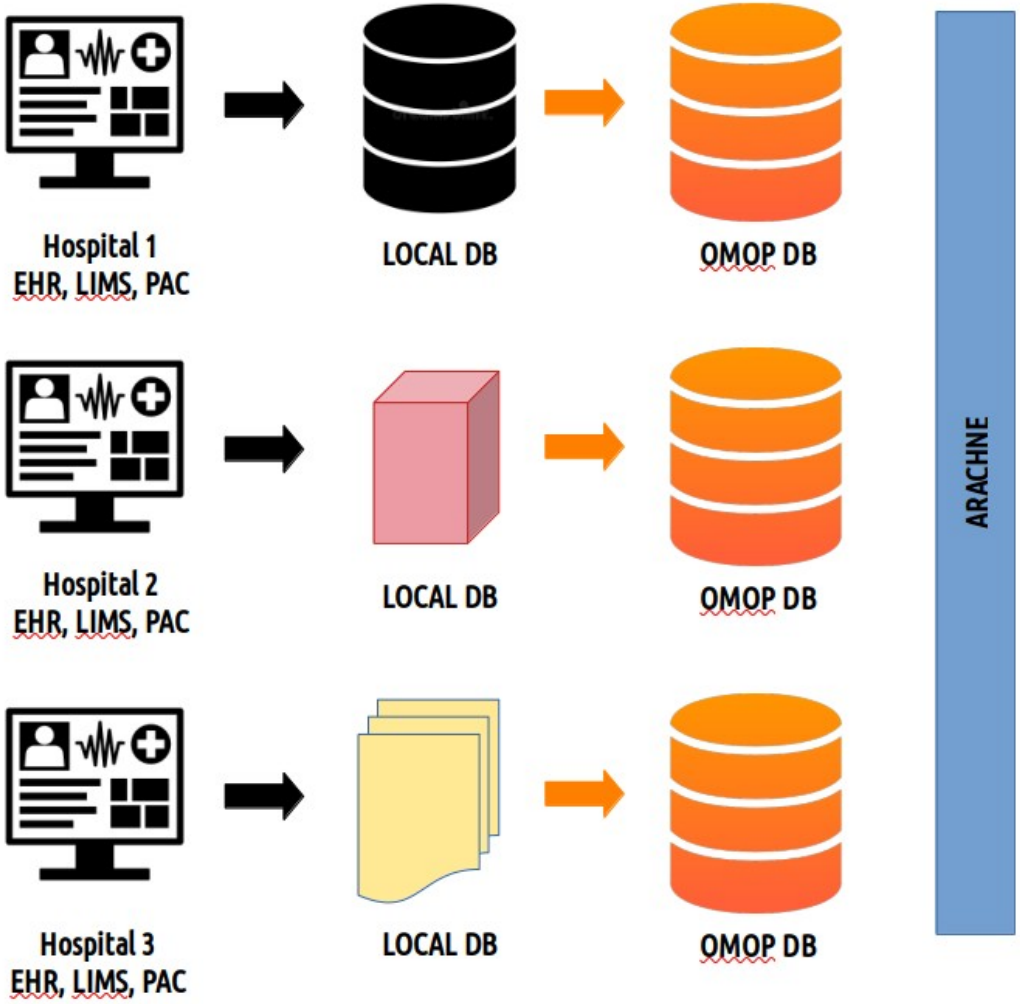


Geographic spread of data partners. The shade of blue indicates the # of data partners in that country (darker = more)



EHDEN
EUROPEAN HEALTH DATA & EVIDENCE NETWORK

OMOP-CDM



Una comunidad internacional de desarrollo de herramientas, estándares y buenas prácticas en estandarizaciónde datos de salud

Herramientas libres para el análisis y explotación de los datos (generación de cohortes, caracterización, incidencia, predicción)

Red europea para la estandarización y federación de datos para la generación de evidencia clinica basada en datos del mundo real

OMOP-CDM

Global Alliance for Genomics & Health
Collaborate. Innovate. Accelerate.

International Cancer Genome Consortium

ARGO Accelerating Research in Genomic Oncology
International Cancer Genome Consortium

European Commission | European Health Data Space
European Cancer Mission

eosc

1+MG

B1MG

elixir

iMPaCT | Data

IBSAL Instituto de Investigación Biomédica de Salamanca

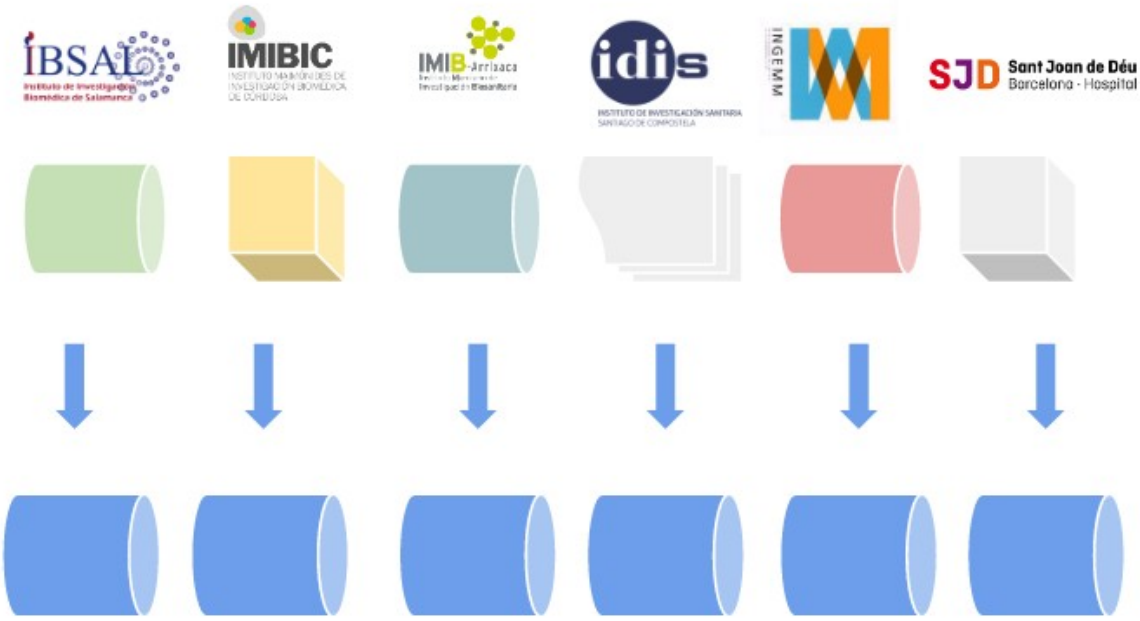
IMIBIC Instituto Maimónides de Investigación Biomédica de Córdoba

IMIB-Arrixaca Instituto Murciano de Investigación Biosanitaria

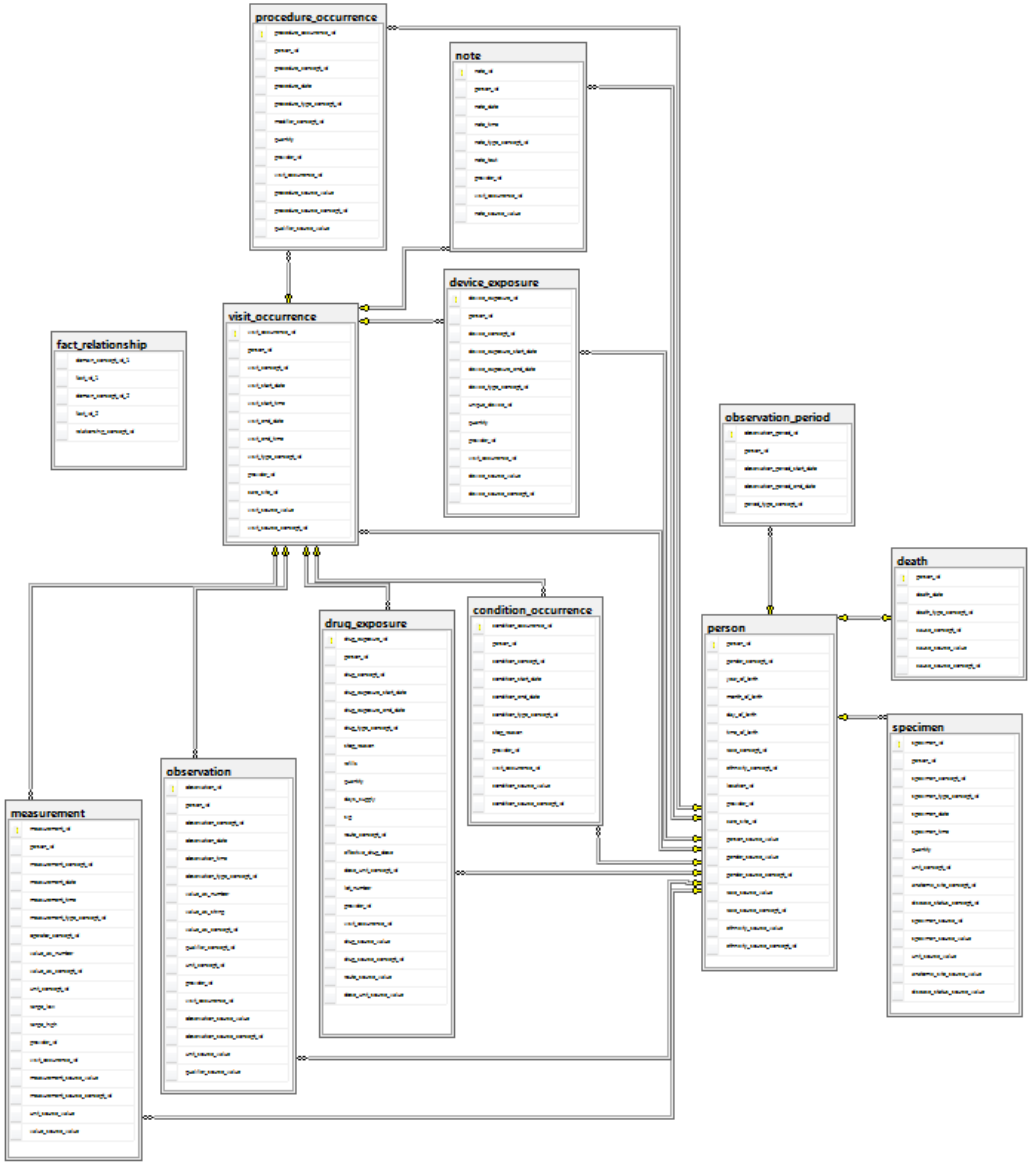
idis Instituto de Investigación Sanitaria Santiago de Compostela

INGEMM

SJD Sant Joan de Déu Barcelona - Hospital

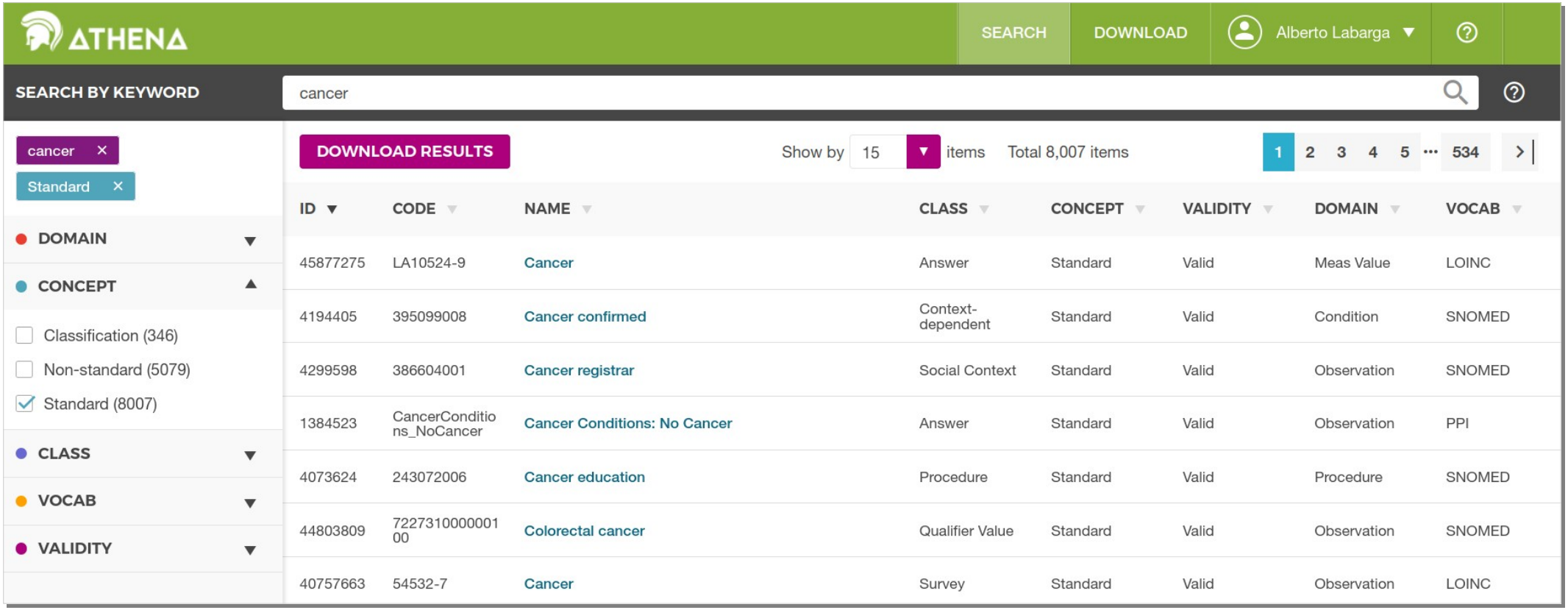

IBSAL
IMIBIC
IMIB-Arrixaca
idis
INGEMM
SJD Sant Joan de Déu Barcelona · Hospital

Interoperabilidad sintáctica



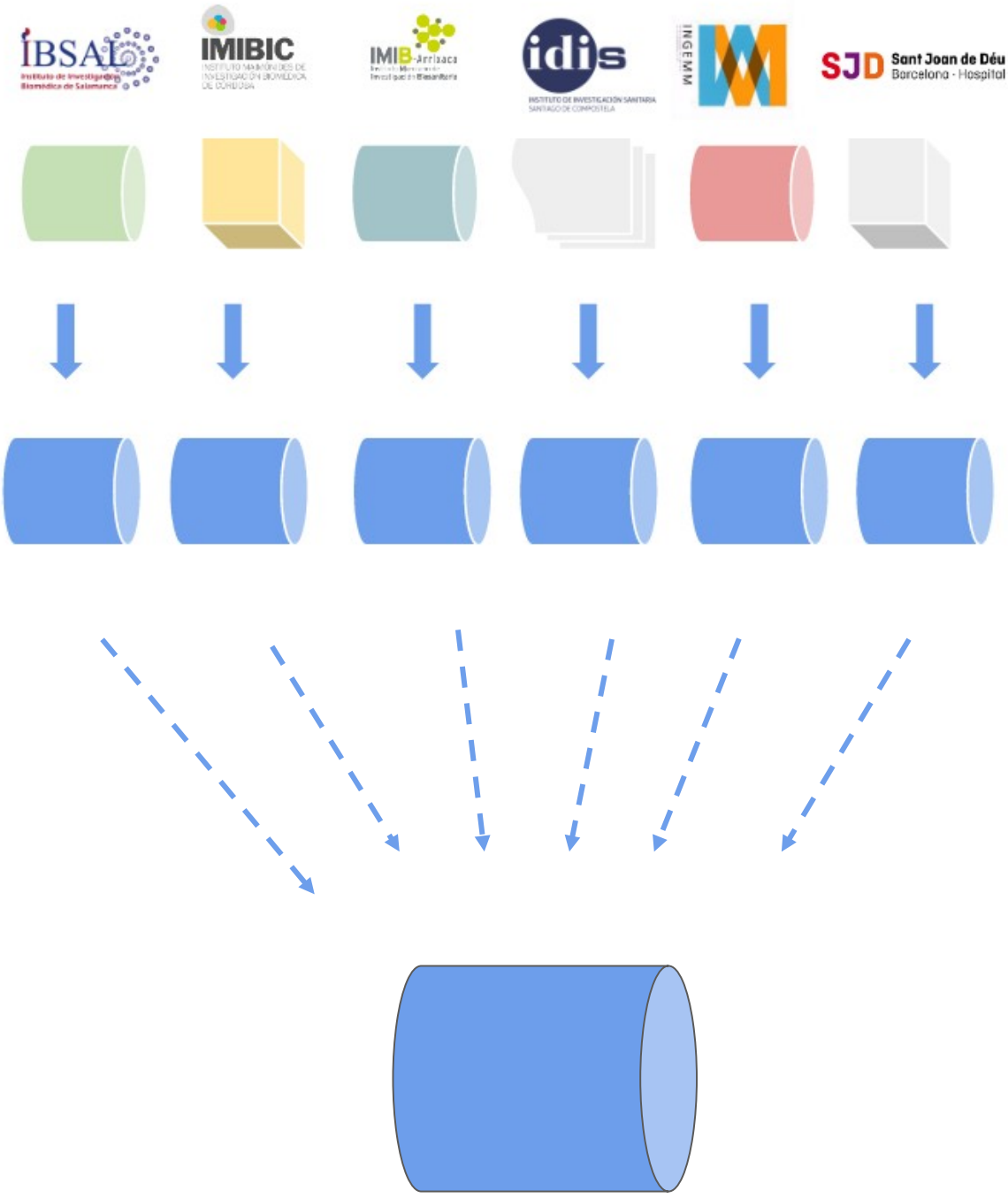
https://ohdsi.github.io/CommonDataModel/

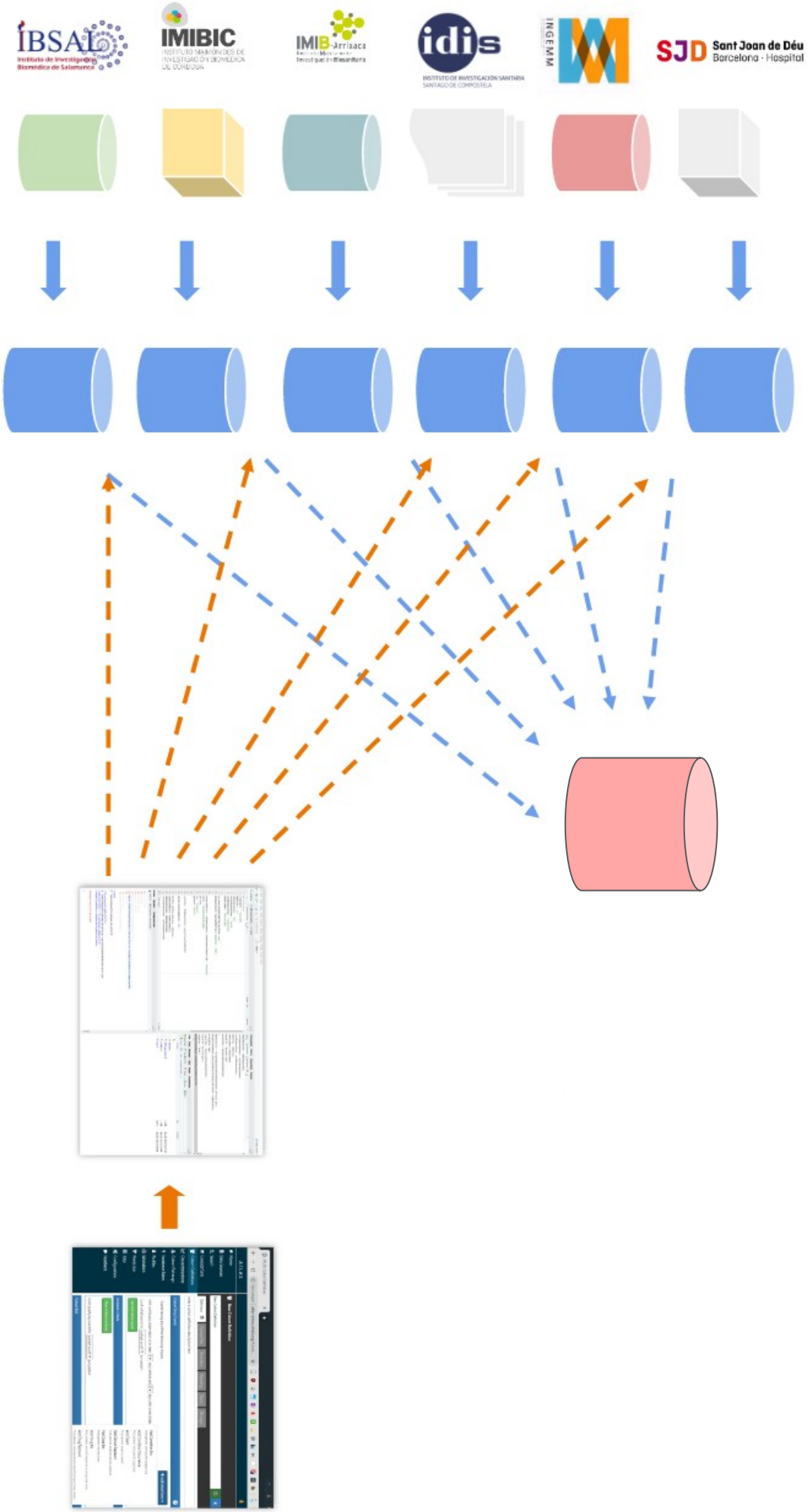
Interoperabilidad semántica



https://athena.ohdsi.org

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


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


IBSAL
Instituto de Investigación Biomédica de Salamanca
IMIBIC
INSTITUTO MAIMÓNIDES DE INVESTIGACIÓN BIOMÉDICA DE CÓRDOBA
IMIB-Arrixaca
Instituto Murciano de Investigación Biosanitaria
idis
INSTITUTO DE INVESTIGACIÓN SANITARIA SANTIAGO DE COMPOSTELA
INGEMM
SJD Sant Joan de Déu
Barcelona · Hospital

OMOP-CDM

- OMOP 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
- 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.

OMOP-CDM

- **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.
- Se trata de 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, que se encuentran enlazadas a los distintos dominios u orígenes.
- **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.
- **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. Es importante destacar que 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).
- **Justificación de los dominios**: los dominios se encuentran definidos aparte, en un modelo entidad-relación, en caso de que su uso sea analítico y el dominio tenga incorporados atributos específicos que no sean aplicables de ninguna otra forma. El resto de los datos pueden almacenarse en tablas con estructura entidad-atributo-valor. Hay un total de 30 dominios.

OMOP-CDM

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

- **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.
- **Escalabilidad**: este CDM está optimizado para el uso de orígenes de datos con distintos tamaños, ya que existen bases de datos que pueden llegar a contener millones o billones de registros clínicos.
- **Compatibilidad** con versiones anteriores
- Los derechos 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”.
- Los **tipos de datos** normalmente están definidos por el estándar ANSI SQL, entre los cuales se encuentran VARCHAR, INTEGER, FLOAT, DATE, DATE-TIME y CLOB

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

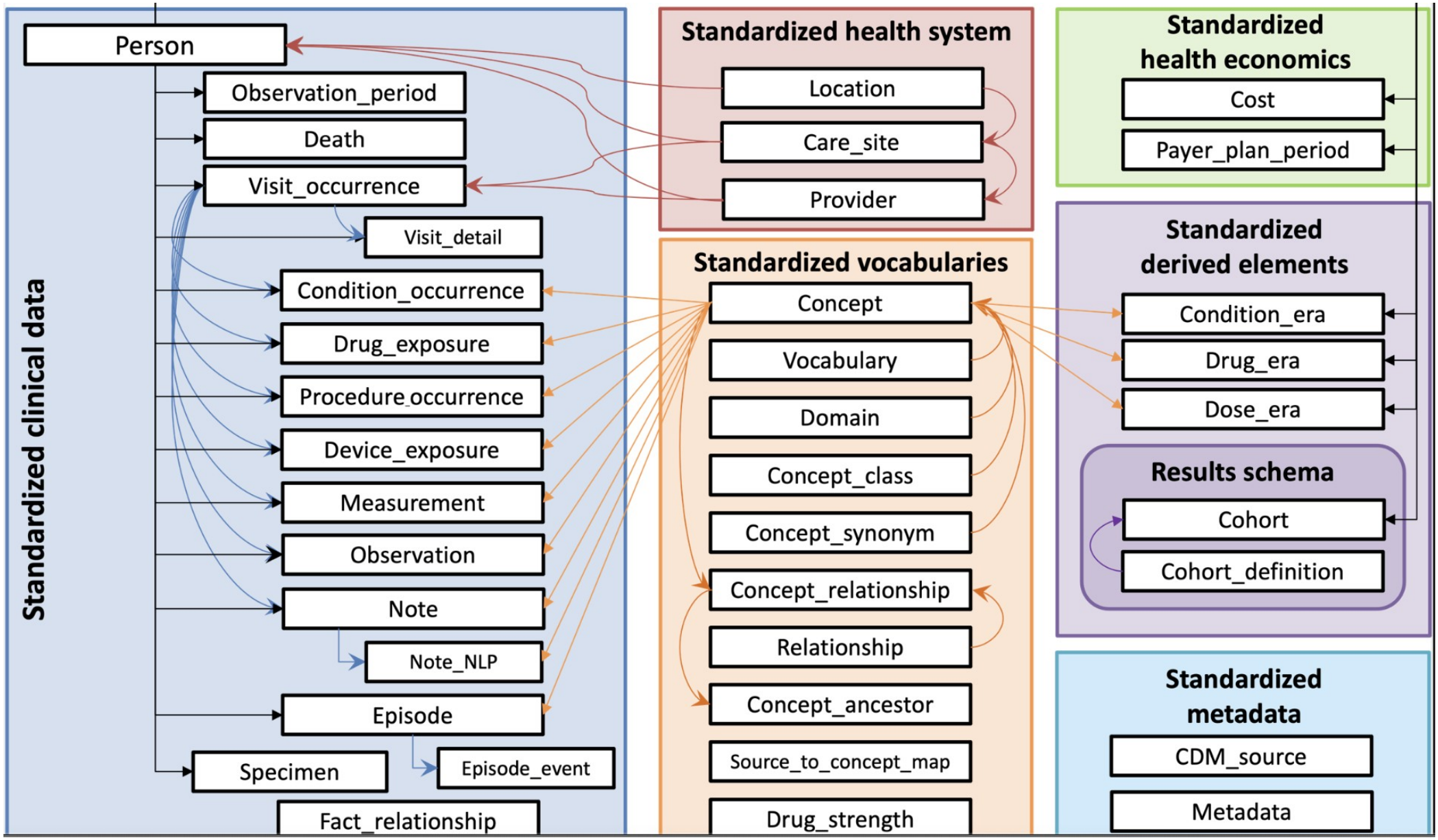
All records in this table are independent Persons.

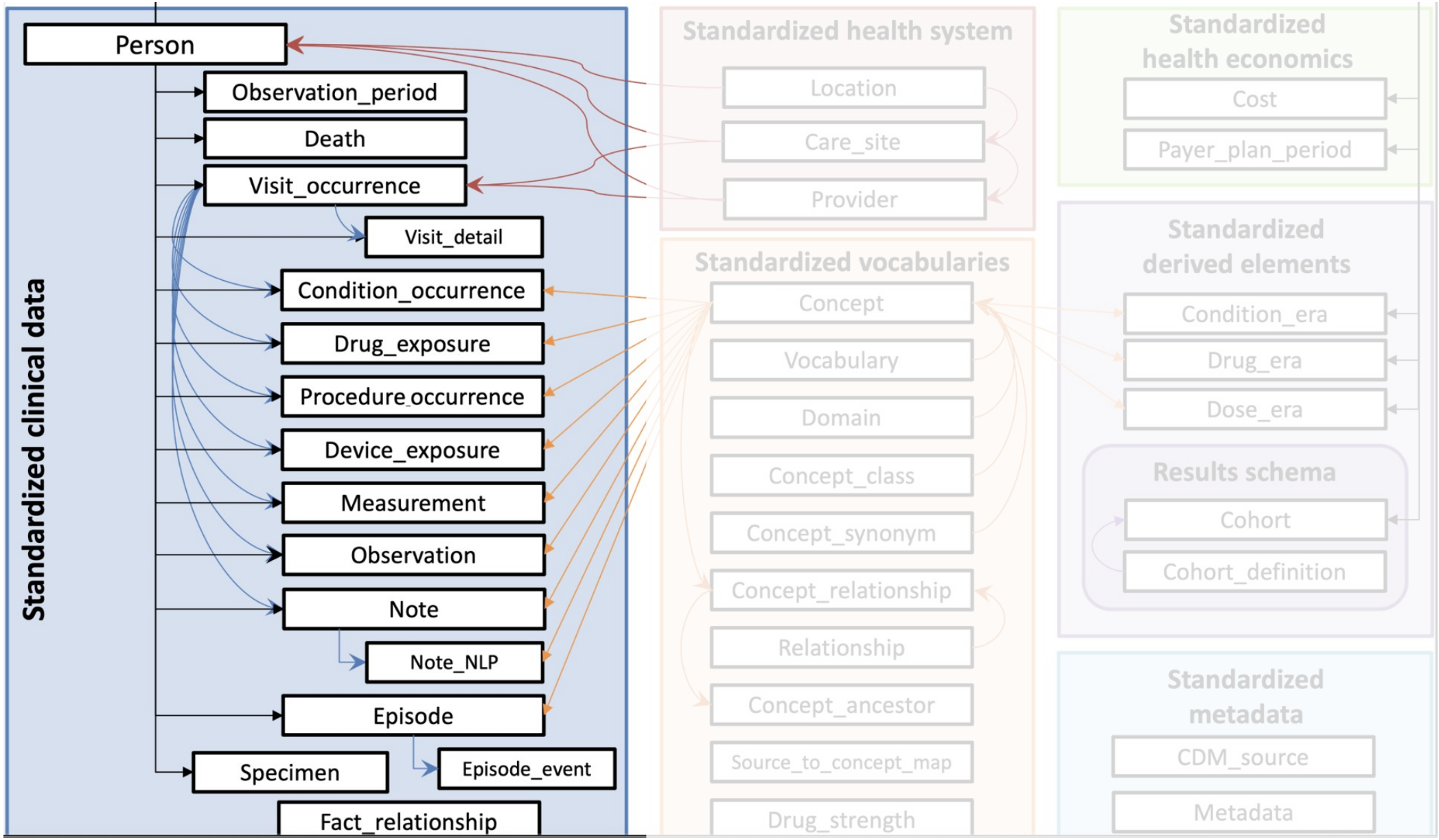
ETL Conventions

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	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	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


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

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

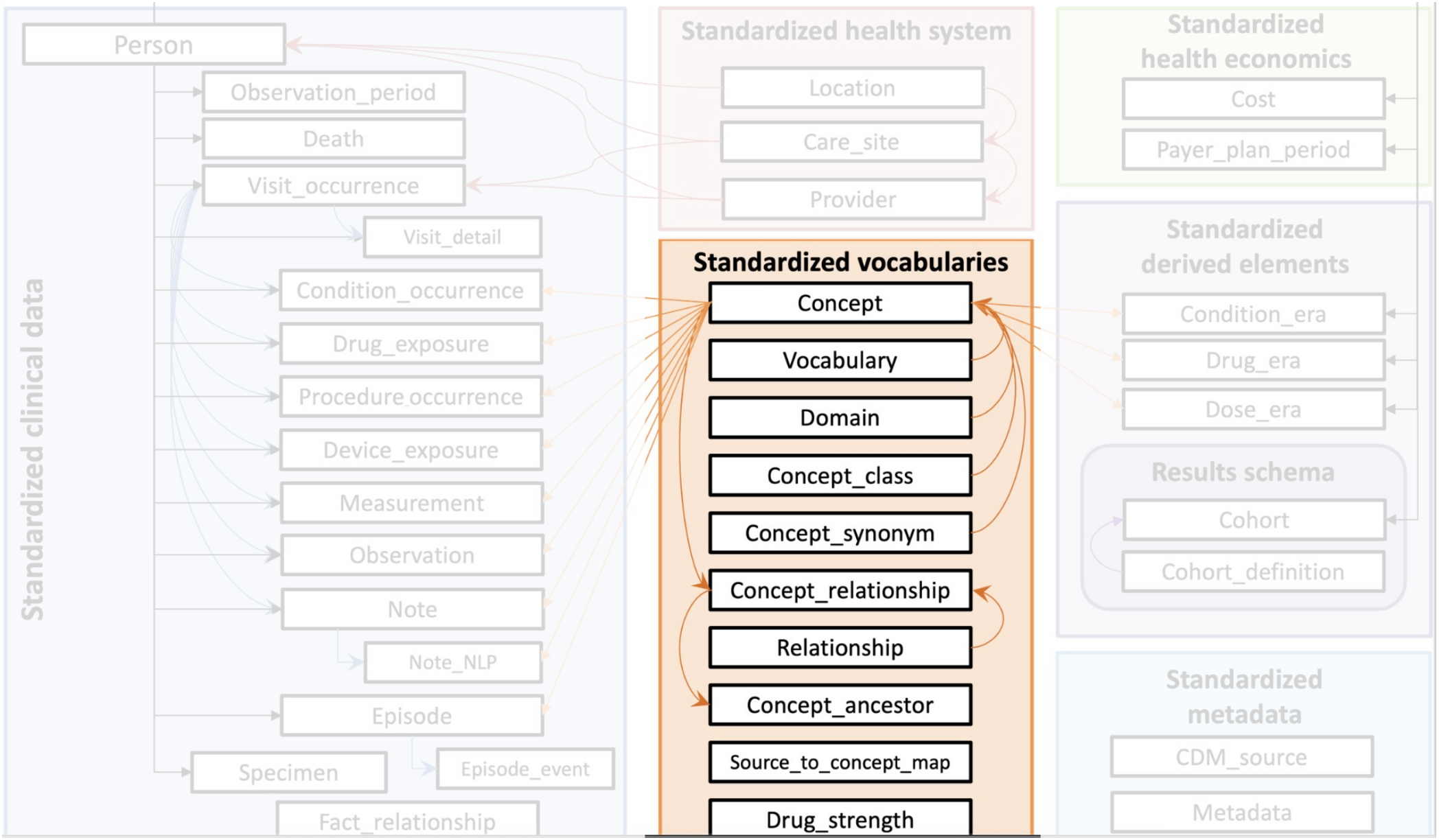
death	
person_id	bigint NN
death_datetime	timestamp
death_date	date NN
death_type_concept_id	integer NN
cause_concept_id	integer
cause_source_concept_id	integer
cause_source_value	text

Principios generales

- [Event]_ID es una clave primaria de las tablas de eventos. Por ejemplo, el campo PERSON_ID identifica a una persona concreta en la tabla PERSON,
- [Event]_CONCEPT_ID es una clave foránea a un concepto estándar de la tabla CONCEPT, por ejemplo gender_concept_id 8532 corresponde a ‘Female’

Conceptos estándar

- Los conceptos estándar son la representación principal del evento, ya que definen el significado de una entidad clínica de la misma forma independientemente de su valor en el origen.


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

Conceptos

- 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.

Conceptos

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

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_ID	313217
CONCEPT_NAME	Atrial fibrillation
DOMAIN_ID	Condition
VOCABULARY_ID	SNOMED
CONCEPT_CLASS_ID	Clinical Finding
STANDARD_CONCEPT	S
CONCEPT_CODE	49436004
VALID_START_DATE	01-Jan-1970
VALID_END_DATE	31-Dec-2099
INVALID_REASON	

For use in CDM

English description

Domain

Vocabulary

Class in SNOMED

Concept in data

Code in SNOMED

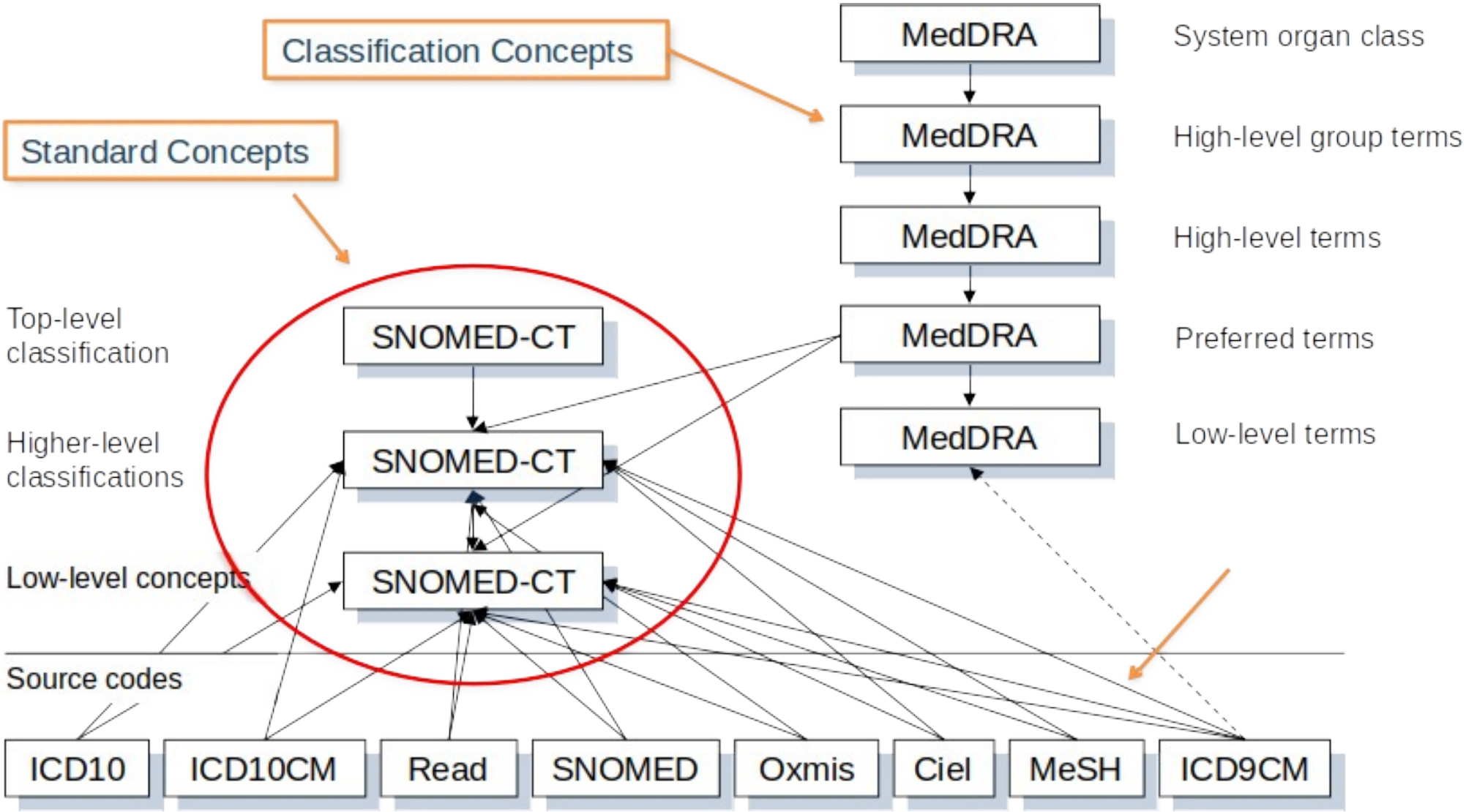
Valid during time interval

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)

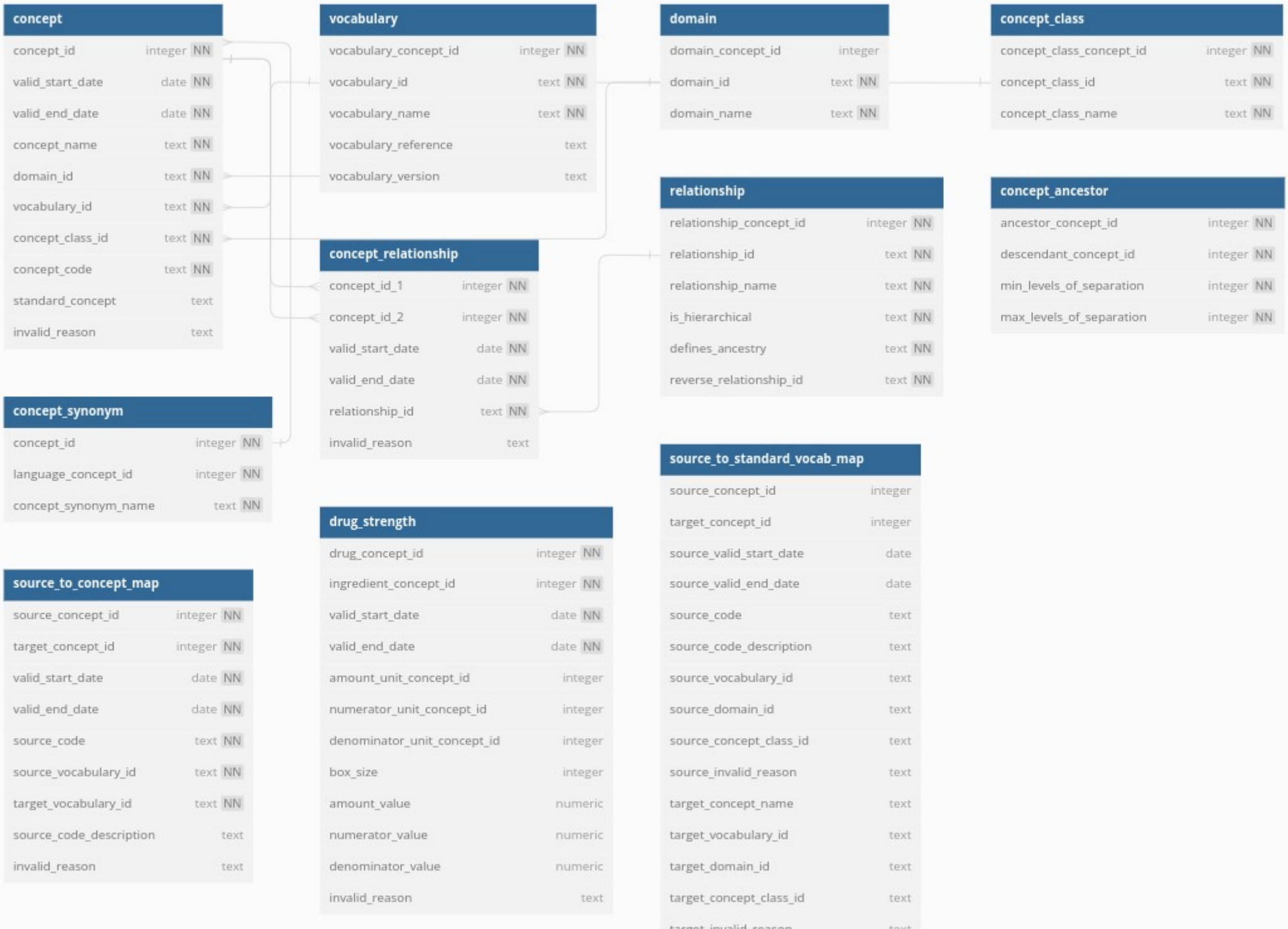
Conceptos

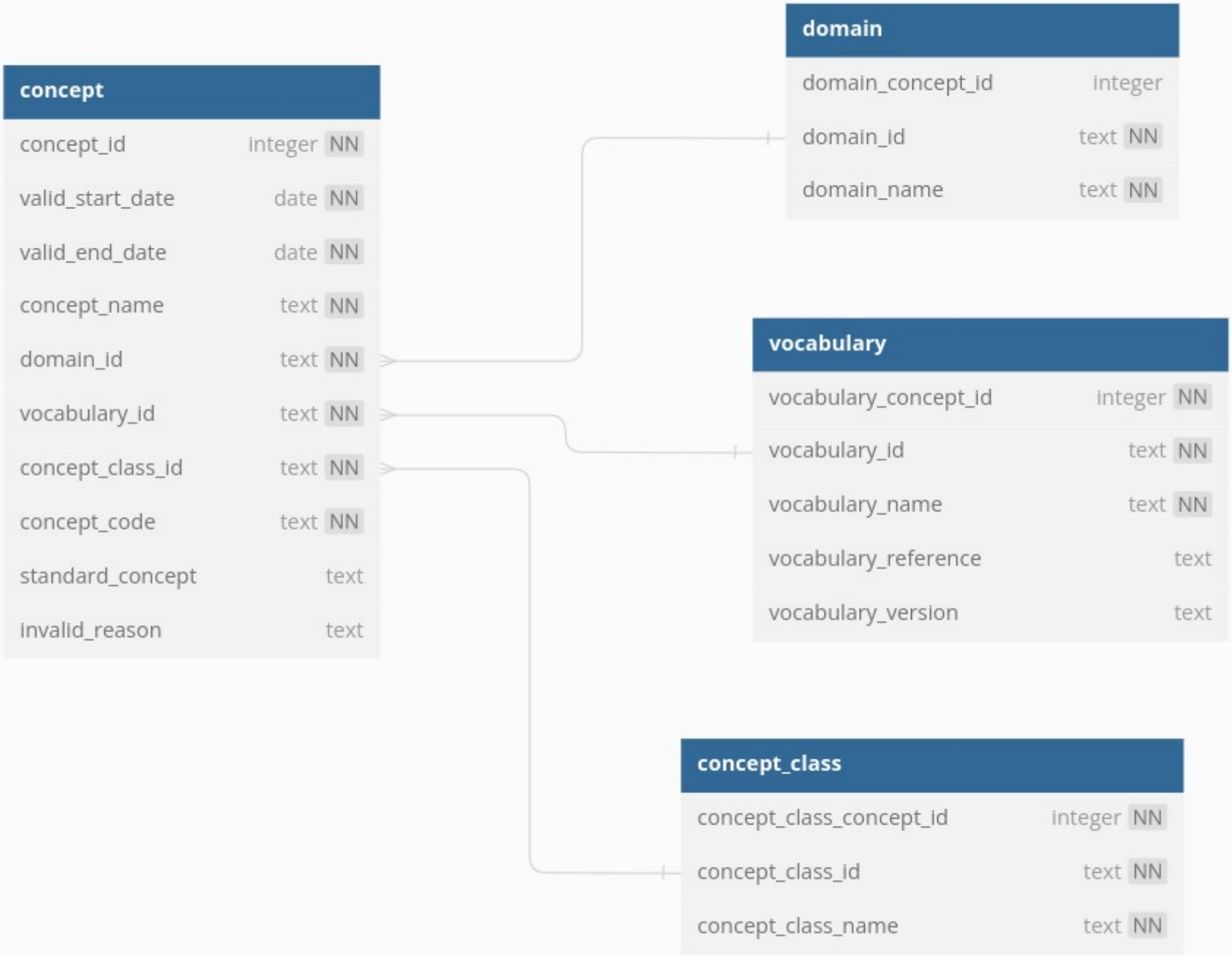
- Hay tres categorías de conceptos:
  - **Non-standard Concepts**: representaciones unívocas de un código existente
  - **Standard Concept**: Conceptos de un vocabulario considerado estándar para un determinado dominio para facilitar los análisis y estudios usando las herramientas de la OHDSI
  - **Classification Concepts**: No se utilizan directamente para codig¡ficar datos de origen sino que se utilizan para ejecutar consultas jerárquicas o definir un **concept set**


Classification Concepts
Standard Concepts
MedDRA
System organ class
MedDRA
High-level group terms
MedDRA
High-level terms
MedDRA
Preferred terms
MedDRA
Low-level terms
Top-level classification
SNOMED-CT
Higher-level classifications
SNOMED-CT
Low-level concepts
SNOMED-CT
Source codes
ICD10
ICD10CM
Read
SNOMED
Oxmis
Ciel
MeSH
ICD9CM

Conceptos

- [Event]_SOURCE_CONCEPT_ID es otra clave foránea de la tabla CONCEPT, que contiene el valor de origen. Siguiendo el ejemplo anterior, en la Figura 3.5 este campo tendría el valor ‘70070008’, que es el valor del identificador de ‘Torticollis’ en el origen, y que se corresponde con el valor ‘440814’ en el estándar SNOMED.
- [Event]_SOURCE_VALUE contiene la representación del valor de origen. Es importante tener en cuenta que este dato no se encuentra armonizado dentro del modelo de datos. Para el ejemplo de la Figura 3.6 el valor original es ‘72166-2’
- [Event]_TYPE_CONCEPT_ID es una clave foránea de la tabla CONCEPT que representa el origen del dato de forma estandarizada. Este campo declara el mecanismo de captura que creó el registro. En el caso del ejemplo correspondiente la Figura 3.6, este valor es ‘38000280’


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


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
domain
domain_concept_id integer
domain_id text NN
domain_name text NN
vocabulary
vocabulary_concept_id integer NN
vocabulary_id text NN
vocabulary_name text NN
vocabulary_reference text
vocabulary_version text
concept_class
concept_class_concept_id integer NN
concept_class_id text NN
concept_class_name text NN

Concept Count	Domain ID	Concept Count	Domain ID
1731378	Drug	183	Route
477597	Device	180	Currency
257000	Procedure	158	Payer
163807	Condition	123	Visit
145898	Observation	51	Cost
89645	Measurement	50	Race
33759	Spec Anatomic Site	13	Plan Stop Reason
17302	Meas Value	11	Plan
1799	Specimen	6	Episode
1215	Provider Specialty	6	Sponsor
1046	Unit	5	Meas Value Operator
944	Metadata	3	Spec Disease Status
538	Revenue Code	2	Gender
336	Type Concept	2	Ethnicity
194	Relationship	1	Observation Type

```
select domain_id, count(*)
from concept
group by domain_id
```

ATHENA

SEARCH | DOWNLOAD | Alberto Labarga

SEARCH BY KEYWORD: neoplasm

neoplasm ×
Clinical Finding ×
Condition ×
Standard ×

DOMAIN
CONCEPT
CLASS
VOCAB

filter

- [ ] Nebraska Lexicon (6)
- [x] SNOMED (4275)
- [ ] SNOMED Veterinary (7)
- [ ] ABMS (0)

DOWNLOAD RESULTS

Show by 15 items Total 4,275 items

1 2 3 4 5 ... 285

4290837	399956005	Melanocytic neoplasm	Clinical Finding	Standard	Valid	Condition	SNOMED
4294432	402877008	Rhabdomyomatous neoplasm	Clinical Finding	Standard	Valid	Condition	SNOMED
4344389	240204004	Synovial-like neoplasm	Clinical Finding	Standard	Valid	Condition	SNOMED
4129890	126871001	Neoplasm of omentum	Clinical Finding	Standard	Valid	Condition	SNOMED
4113639	286904004	Benign genital neoplasm	Clinical Finding	Standard	Valid	Condition	SNOMED
42537690	737223000	Dendritic cell neoplasm	Clinical Finding	Standard	Valid	Condition	SNOMED
443736	414653009	Mast cell neoplasm	Clinical Finding	Standard	Valid	Condition	SNOMED
4054416	126567006	Neoplasm of acromion	Clinical Finding	Standard	Valid	Condition	SNOMED
4129245	126639006	Neoplasm of axilla	Clinical Finding	Standard	Valid	Condition	SNOMED

https://athena.ohdsi.org/

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)		28-Jan-2022
☑	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	01-May-2023
☑	5	HCPCS	Healthcare Common Procedure Coding System (CMS)		01-Jul-2023
☑	6	LOINC	Logical Observation Identifiers Names and Codes (Regenstrief Institute)		15-Aug-2023
☐	7	NDFRT	National Drug File - Reference Terminology (VA)		06-Aug-2018
☑	8	RxNorm	RxNorm (NLM)		03-Jul-2023
☑	9	NDC	National Drug Code (FDA and manufacturers)		27-Aug-2023
	10	GPI	Medi-Span Generic Product Identifier (Wolters Kluwer Health)	License required	14-Dec-2017
☑	12	Gender	OMOP Gender		

ATHENA

SEARCH | DOWNLOAD | Alberto Labarga

DOWNLOAD HISTORY

REFRESH | NOTIFICATIONS

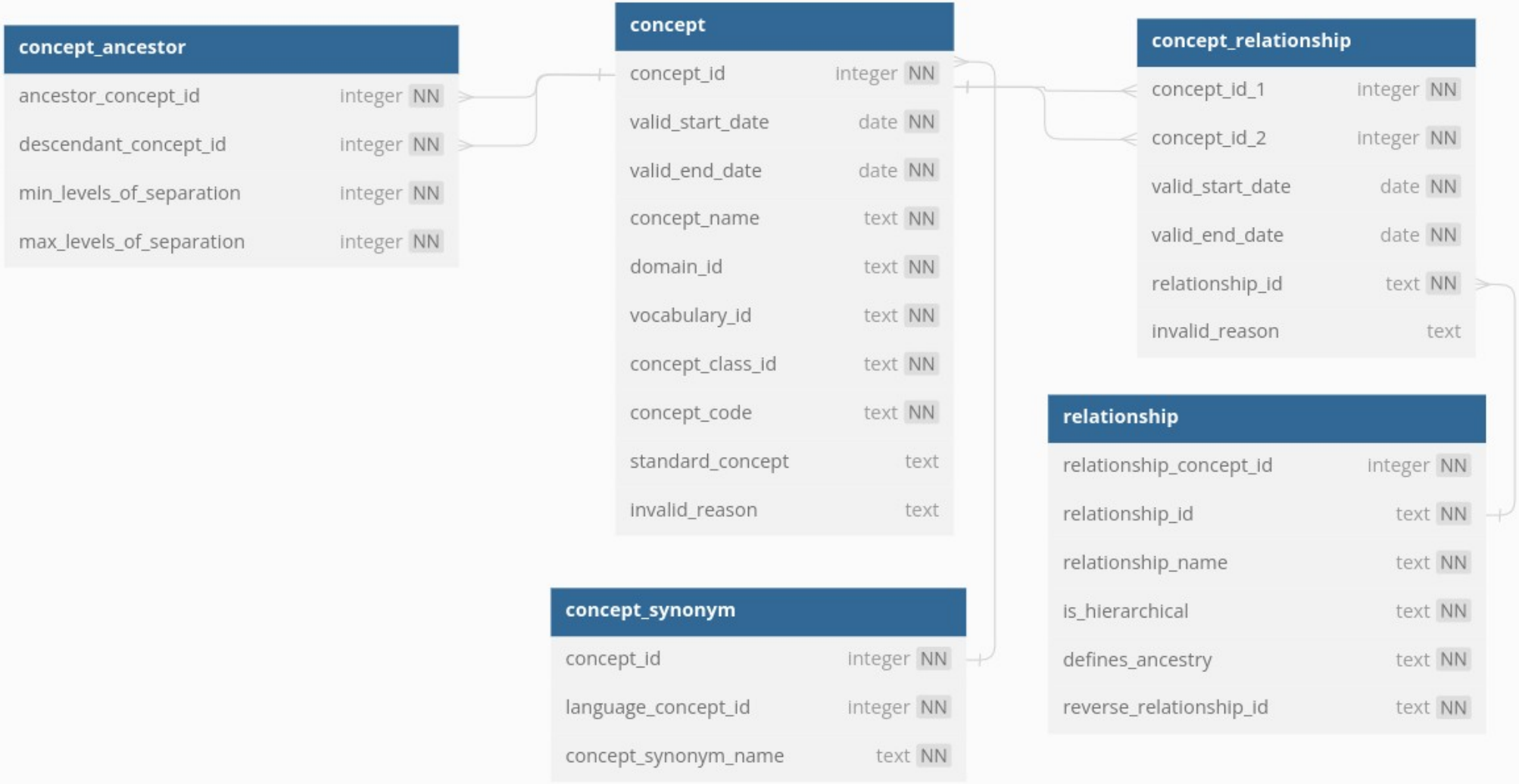
BSC_OMOP_VOCABULARIES

09:56 AM | 13-FEB-2023

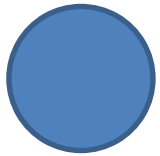
V5.0 23-JAN-23

ARCHIVED | RESTORE

Jerarquía de conceptos

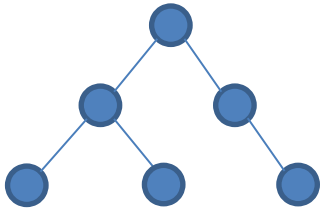


Jerarquía de conceptos



conceptos médicos en **concept**

Relaciones directas entre conceptos **concept_relationship**



Relaciones jerárquicas **concept_ancestor**

concept_ancestor

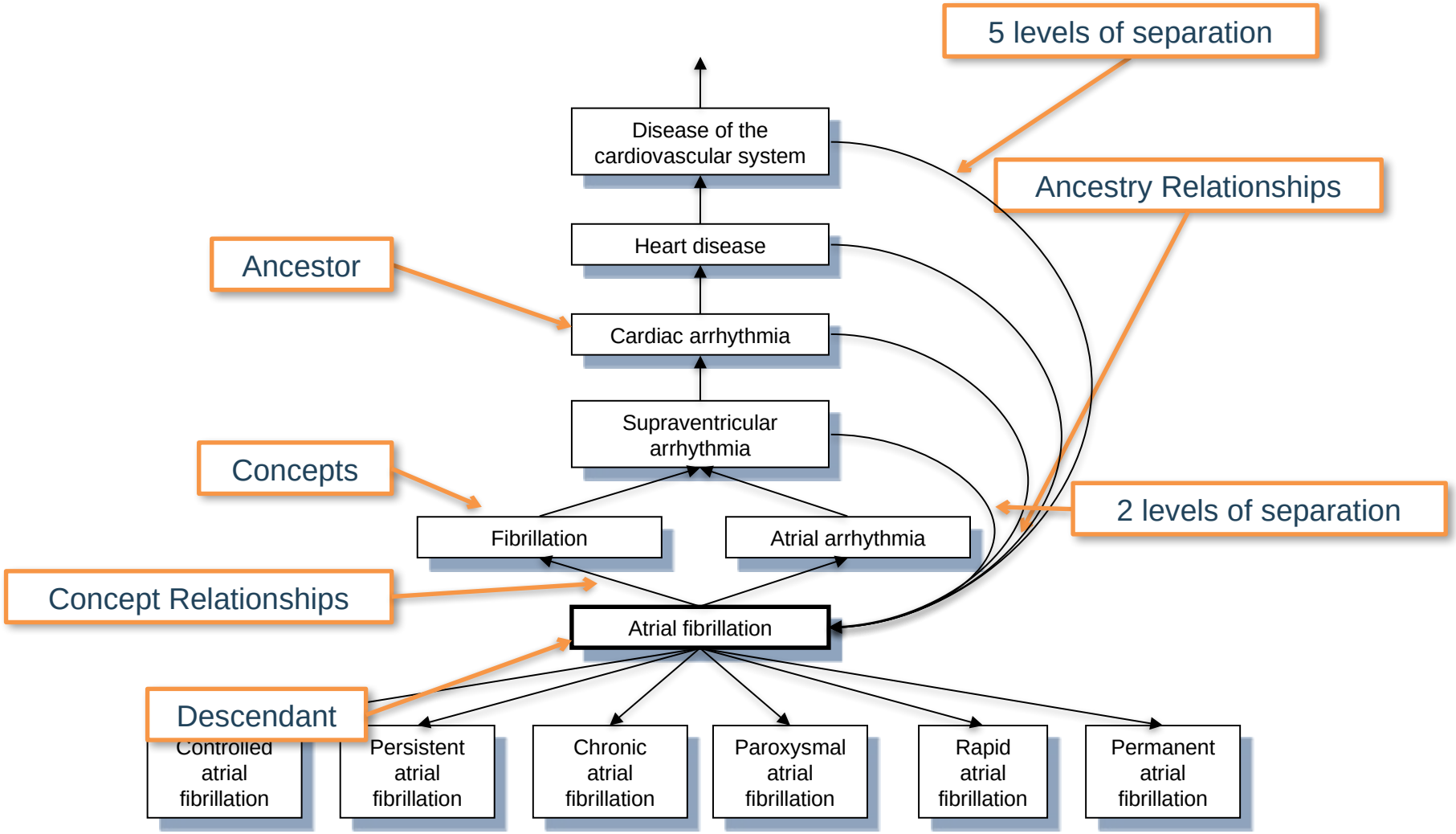
concept_ancestor		
ancestor_concept_id	integer	NN
descendant_concept_id	integer	NN
min_levels_of_separation	integer	NN
max_levels_of_separation	integer	NN

```
WITH RECURSIVE related_concepts AS (
    -- Base case
    SELECT concept_id_1, concept_id_2
    FROM concept_relationship
    WHERE concept_relatinoship_id = 'is_a'

    UNION ALL

    -- Recursive step
    SELECT rc.concept_id_1, cr.concept_id_2
    FROM concept_relationship cr
    JOIN related_concepts rc
    ON cr.concept_id_1 = rc.concept_id_2
    WHERE cr.concept_relatinoship_id = 'is_a'
)

SELECT DISTINCT concept_id_1, concept_id_2
FROM related_concepts
ORDER BY concept_id_1, concept_id_2;
```


5 levels of separation
Ancestry Relationships
Ancestor
Concepts
2 levels of separation
Concept Relationships
Descendant
Disease of the cardiovascular system
Heart disease
Cardiac arrhythmia
Supraventricular arrhythmia
Fibrillation
Atrial arrhythmia
Atrial fibrillation
Controlled atrial fibrillation
Persistent atrial fibrillation
Chronic atrial fibrillation
Paroxysmal atrial fibrillation
Rapid atrial fibrillation
Permanent atrial fibrillation

Mapeo de vocabularios

- Si los códigos utilizados en el EHR corresponden a un vocabulario existente en OMOP se pueden traducir a códigos standard
- Por ejemplo:
  - ICD9 a SNOMED
  - ATC a RxNORM

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
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
target_invalid_reason text
target_standard_concept 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

Source to Source

```
SELECT c.concept_code AS SOURCE_CODE, c.concept_id AS
SOURCE_CONCEPT_ID, c.CONCEPT_NAME AS SOURCE_CODE_DESCRIPTION,
c.vocabulary_id AS SOURCE_VOCABULARY_ID, c.domain_id AS
SOURCE_DOMAIN_ID, c.concept_class_id AS SOURCE_CONCEPT_CLASS_ID,
c.VALID_START_DATE AS SOURCE_VALID_START_DATE, c.VALID_END_DATE AS
SOURCE_VALID_END_DATE, c.invalid_reason AS SOURCE_INVALID_REASON,
c.concept_ID as TARGET_CONCEPT_ID, c.concept_name AS
TARGET_CONCEPT_NAME, c.vocabulary_id AS TARGET_VOCABULARY_ID,
c.domain_id AS TARGET_DOMAIN_ID, c.concept_class_id AS
TARGET_CONCEPT_CLASS_ID, c.INVALID_REASON AS TARGET_INVALID_REASON,
c.STANDARD_CONCEPT AS TARGET_STANDARD_CONCEPT
       FROM CONCEPT c
       UNION
       SELECT source_code, SOURCE_CONCEPT_ID, SOURCE_CODE_DESCRIPTION,
source_vocabulary_id, c1.domain_id AS SOURCE_DOMAIN_ID,
c2.CONCEPT_CLASS_ID AS SOURCE_CONCEPT_CLASS_ID, c1.VALID_START_DATE AS
SOURCE_VALID_START_DATE, c1.VALID_END_DATE AS
SOURCE_VALID_END_DATE,stcm.INVALID_REASON AS SOURCE_INVALID_REASON,
target_concept_id, c2.CONCEPT_NAME AS TARGET_CONCEPT_NAME,
target_vocabulary_id, c2.domain_id AS TARGET_DOMAIN_ID,
c2.concept_class_id AS TARGET_CONCEPT_CLASS_ID, c2.INVALID_REASON AS
TARGET_INVALID_REASON, c2.standard_concept AS TARGET_STANDARD_CONCEPT
       FROM source_to_concept_map stcm
              LEFT OUTER JOIN CONCEPT c1
                     ON c1.concept_id = stcm.source_concept_id
              LEFT OUTER JOIN CONCEPT c2
                     ON c2.CONCEPT_ID = stcm.target_concept_id
       WHERE stcm.INVALID_REASON IS NULL
```

Esta consulta genera una tabla con todos los códigos origen y sus correspondientes concept_id y se utiliza para rellenar los campos *_SOURCE_CONCEPT_ID

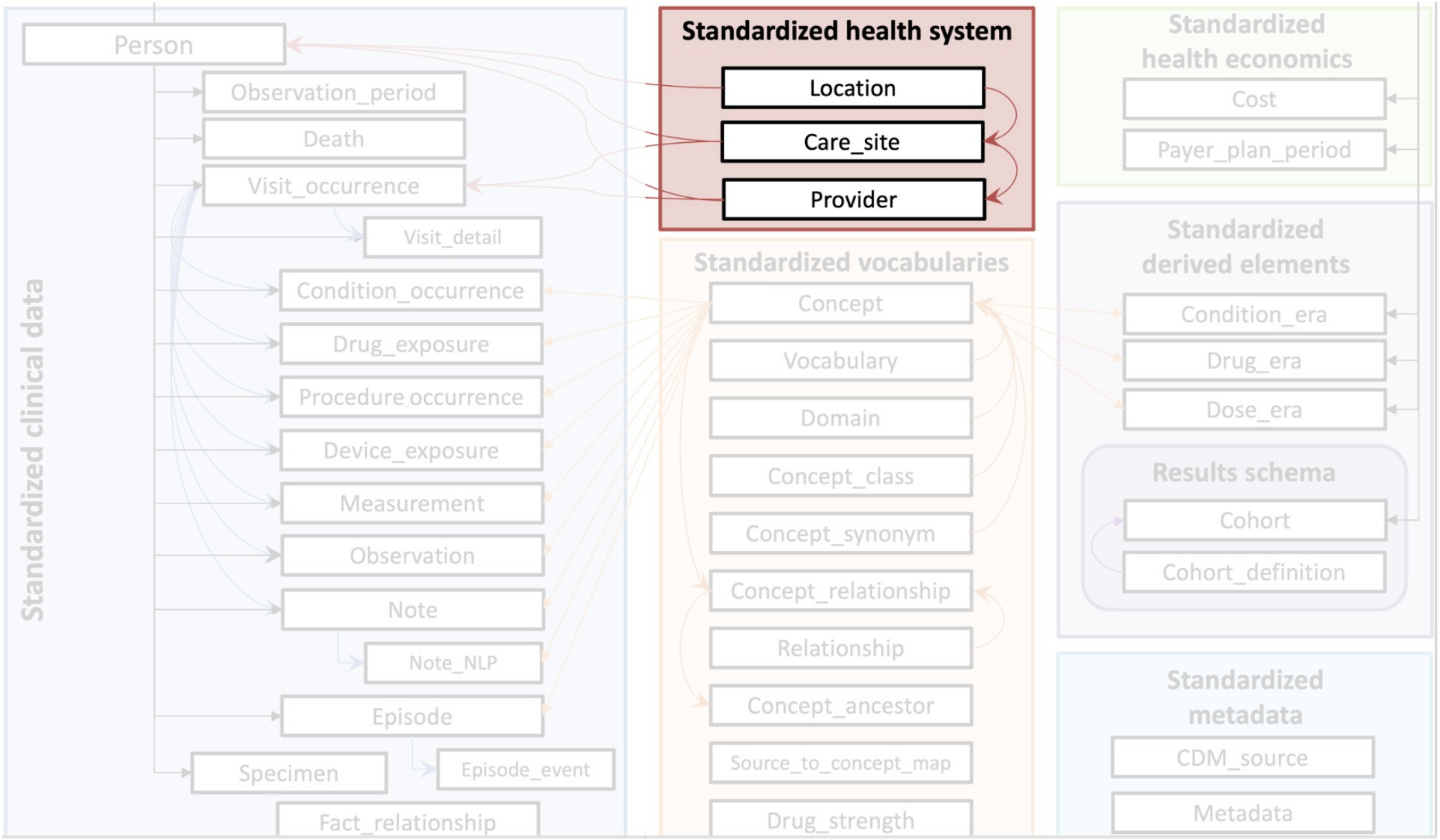
Source to Standard

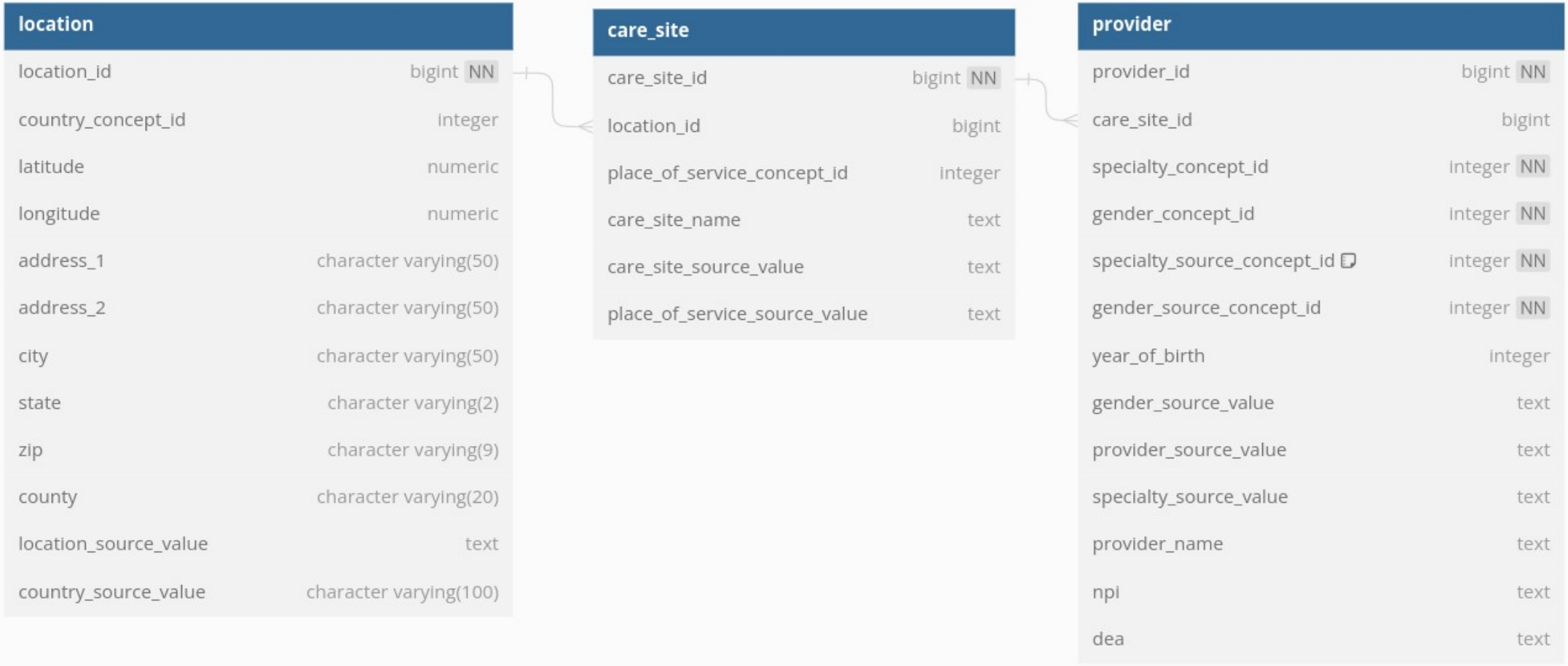
```
SELECT c.concept_code AS SOURCE_CODE, c.concept_id AS SOURCE_CONCEPT_ID, c.concept_name
AS SOURCE_CODE_DESCRIPTION, c.vocabulary_id AS SOURCE_VOCABULARY_ID, c.domain_id AS
SOURCE_DOMAIN_ID, c.CONCEPT_CLASS_ID AS SOURCE_CONCEPT_CLASS_ID, c.VALID_START_DATE
AS SOURCE_VALID_START_DATE, c.VALID_END_DATE AS SOURCE_VALID_END_DATE,
c.INVALID_REASON AS SOURCE_INVALID_REASON, c1.concept_id AS TARGET_CONCEPT_ID,
c1.concept_name AS TARGET_CONCEPT_NAME, c1.VOCABULARY_ID AS TARGET_VOCABUALRY_ID,
c1.domain_id AS TARGET_DOMAIN_ID, c1.concept_class_id AS TARGET_CONCEPT_CLASS_ID,
c1.INVALID_REASON AS TARGET_INVALID_REASON, c1.standard_concept AS
TARGET_STANDARD_CONCEPT
    FROM CONCEPT C
        JOIN CONCEPT_RELATIONSHIP CR
            ON C.CONCEPT_ID = CR.CONCEPT_ID_1
            AND CR.invalid_reason IS NULL
            AND lower(cr.relationship_id) = 'maps to'
        JOIN CONCEPT C1
            ON CR.CONCEPT_ID_2 = C1.CONCEPT_ID
            AND C1.INVALID_REASON IS NULL
    UNION
    SELECT source_code, SOURCE_CONCEPT_ID, SOURCE_CODE_DESCRIPTION,
source_vocabulary_id, c1.domain_id AS SOURCE_DOMAIN_ID, c2.CONCEPT_CLASS_ID AS
SOURCE_CONCEPT_CLASS_ID, c1.VALID_START_DATE AS SOURCE_VALID_START_DATE,
c1.VALID_END_DATE AS SOURCE_VALID_END_DATE, stcm.INVALID_REASON AS
SOURCE_INVALID_REASON,target_concept_id, c2.CONCEPT_NAME AS TARGET_CONCEPT_NAME,
target_vocabulary_id, c2.domain_id AS TARGET_DOMAIN_ID, c2.concept_class_id AS
TARGET_CONCEPT_CLASS_ID, c2.INVALID_REASON AS TARGET_INVALID_REASON,
c2.standard_concept AS TARGET_STANDARD_CONCEPT
    FROM source_to_concept_map stcm
        LEFT OUTER JOIN CONCEPT c1
            ON c1.concept_id = stcm.source_concept_id
        LEFT OUTER JOIN CONCEPT c2
            ON c2.CONCEPT_ID = stcm.target_concept_id
    WHERE stcm.INVALID_REASON IS NULL
```

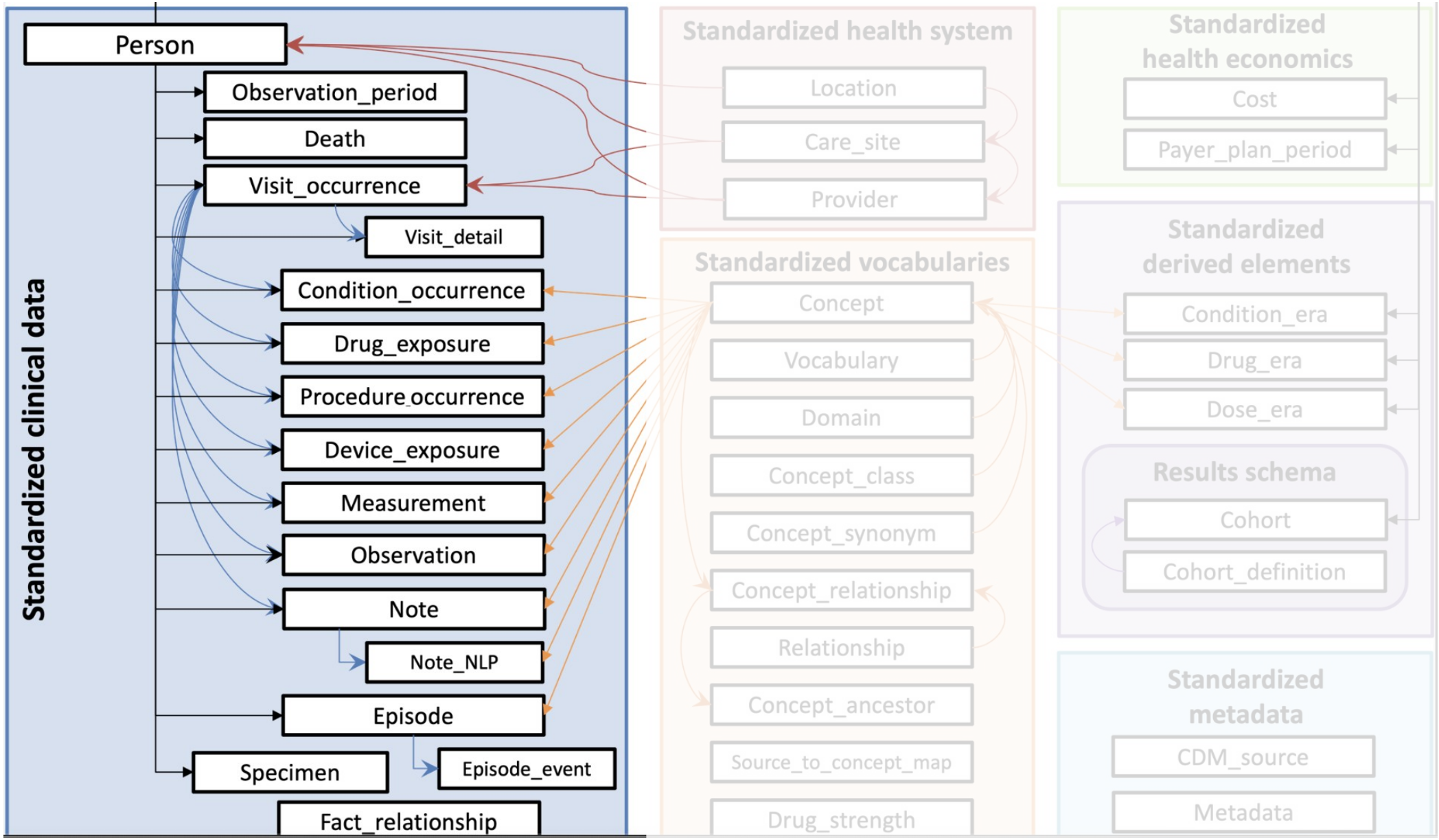
Esta consulta genera una tabla con todos los códigos origen y sus correspondientes concept_id estándar y se utiliza para rellenar los campos *_CONCEPT_ID

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


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
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

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 Type

- 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
- Telehealth Visit: Patient engages with Provider through communication media
- Pharmacy Visit: Person visiting pharmacy for dispensing of Drug, at a Care Site, within one day
- Laboratory Visit: Patient visiting dedicated institution, at a Care Site, within one day, for the purpose of a Measurement.
- Ambulance Visit: Person using transportation service for the purpose of initiating one of the other Visits, without a Care Site, within one day, potentially with Providers accompanying the Visit and 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

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 genral, cuando la visita tiene múltiples PROVIDER or muúltiples CARE_SITE

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.

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 en el modelo OMOP-CDM (Observational Medical Outcomes Partnership - Common Data Model) de OHDSI se utiliza para registrar eventos individuales de procedimientos médicos realizados en pacientes durante su atención médica.

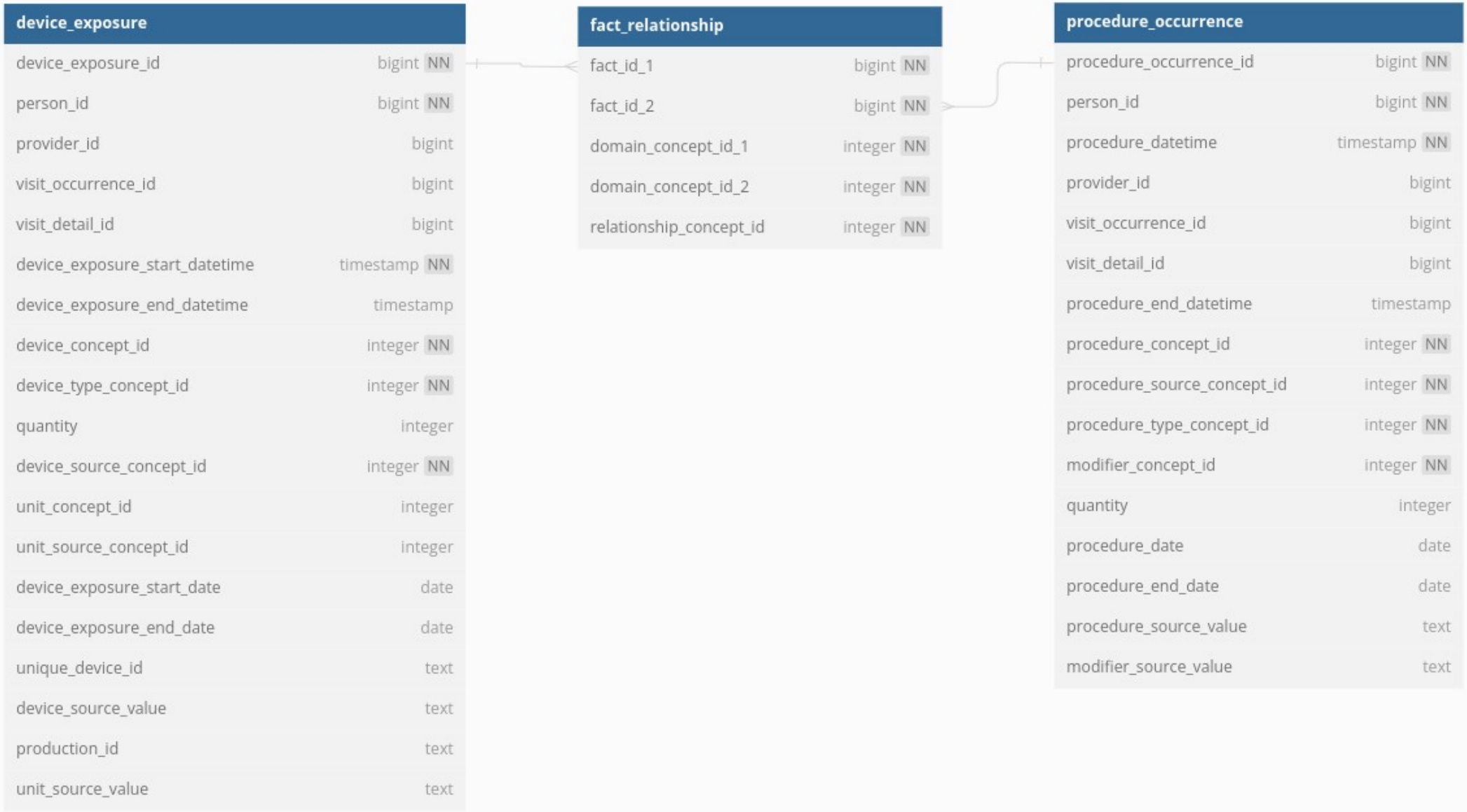
procedure_concept_id: Representa el concepto del procedimiento médico utilizando un identificador de concepto estándar, que está vinculado a un vocabulario médico controlado, como SNOMED-CT u otro.

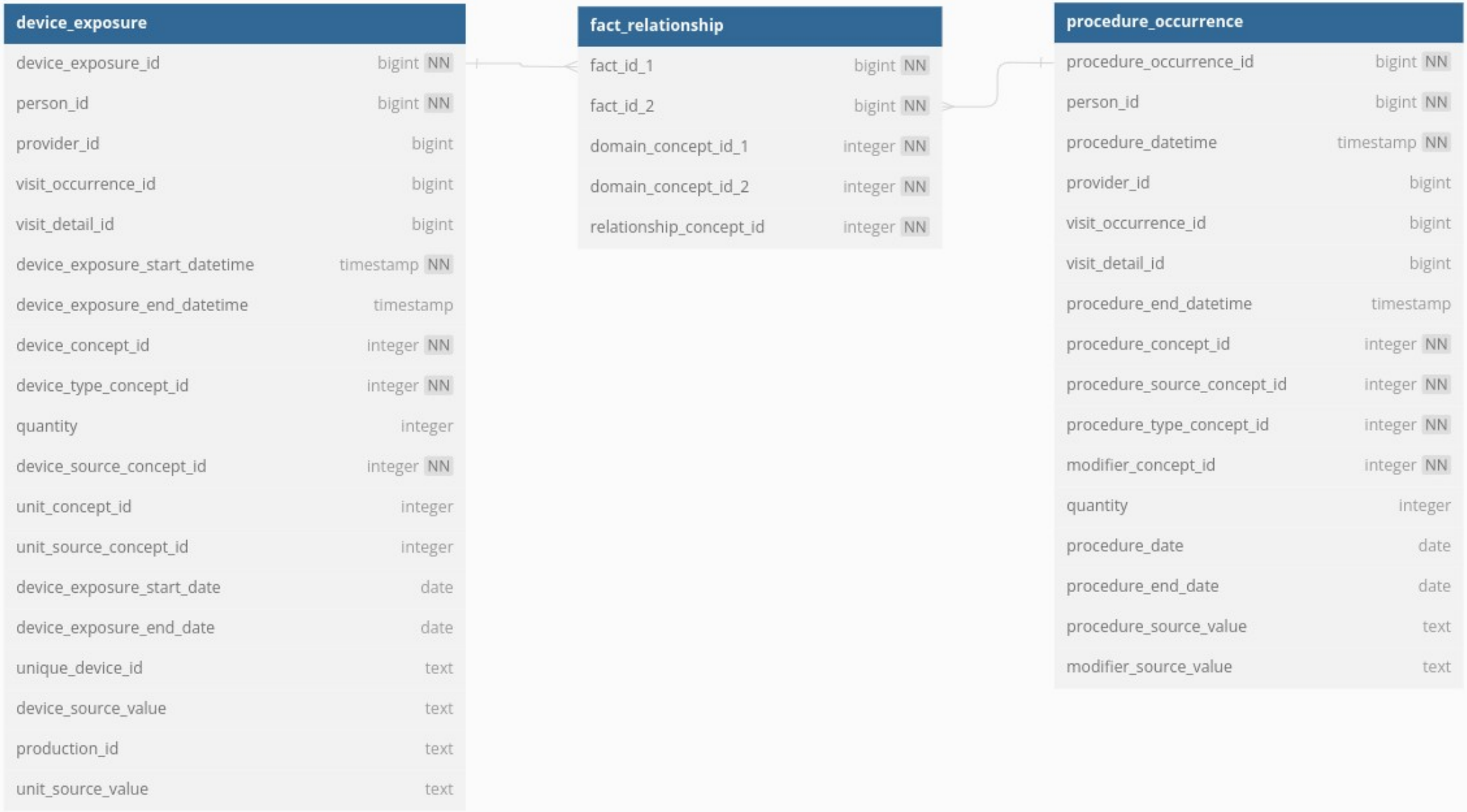
modifier_concept_id: Opcionalmente, se puede registrar un modificador que proporciona información adicional sobre el procedimiento, como el enfoque o la técnica utilizada.

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

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

El propósito de los registros en esta tabla es indicar una exposición a un cierto medicamento de la mejor manera posible. En este contexto, un medicamento se define como un principio activo. El dominio DRUG, forma una jerarquía complejapor 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**

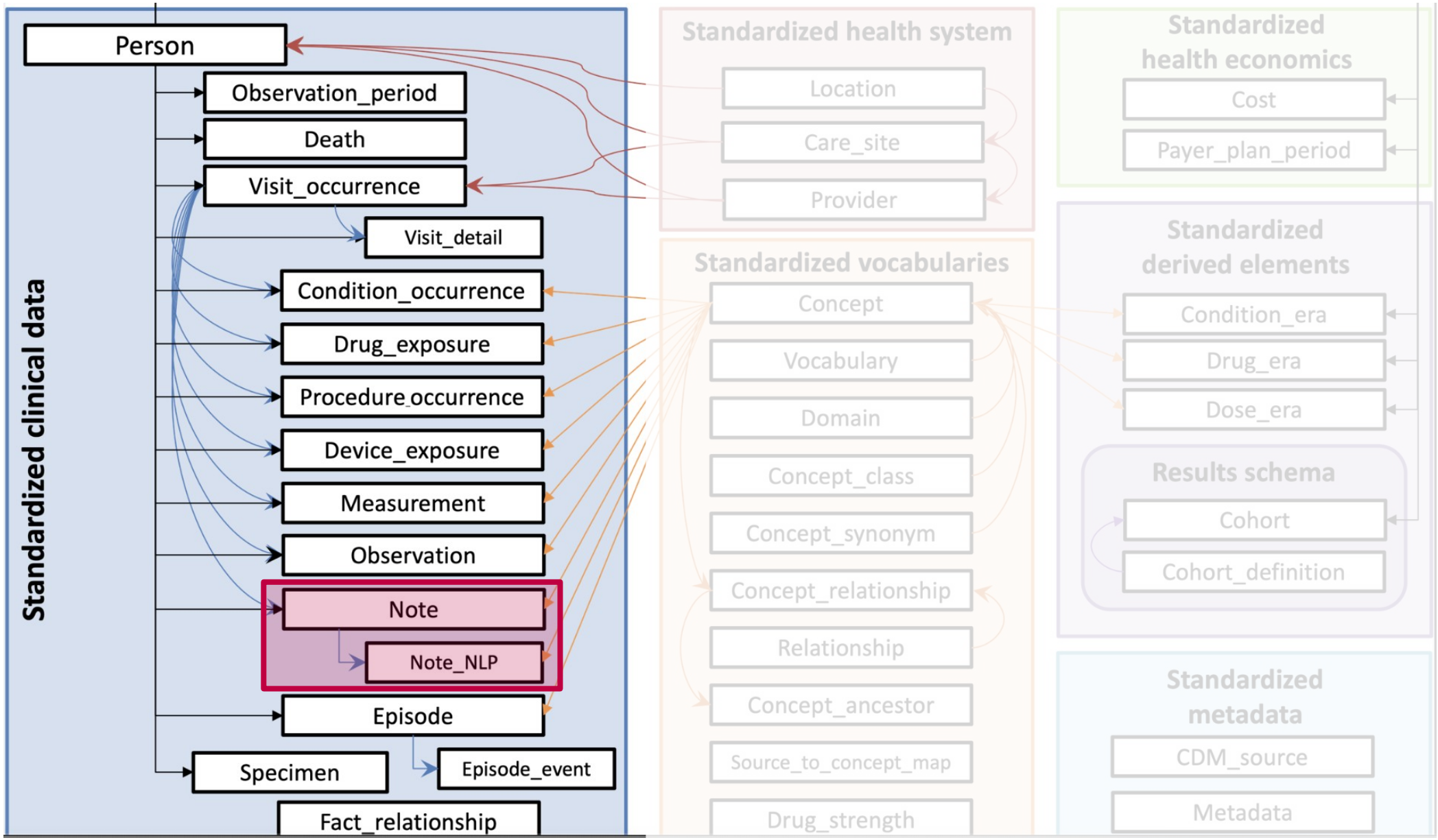
observation		
observation_id	bigint	NN
person_id	bigint	NN
observation_datetime	timestamp	NN
provider_id	bigint	
visit_occurrence_id	bigint	
visit_detail_id	bigint	
observation_event_id	bigint	
observation_date	date	
observation_concept_id	integer	NN
observation_type_concept_id	integer	NN
qualifier_concept_id	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
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** en el modelo OMOP-CDM (Observational Medical Outcomes Partnership - Common Data Model) de OHDSI 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 HL7/LOINC CDO, 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. CDO 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.

LOINC Document Ontology (DO)

- Type of Service (ToS): the kind of healthcare services provided to patients. e.g., Consultation, Evaluation and Management, Procedure
- Kind of Document (KoD): the type of clinical documents based on its structure. e.g., Note, Report, Checklist
- Setting: the location or channel where clinical care is provided. e.g., Ambulance, Birthing Center, Intensive Care Unit
- Role: people and their occupations involved in the service or authors who created the clinical note. e.g., Physicians, Nurse, Pharmacist
- Subject Matter Domain (SMD): clinical specialty relevant to the document or the main purpose of creating the document. e.g., Anesthesiology, Urology, Cardiovascular Disease

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

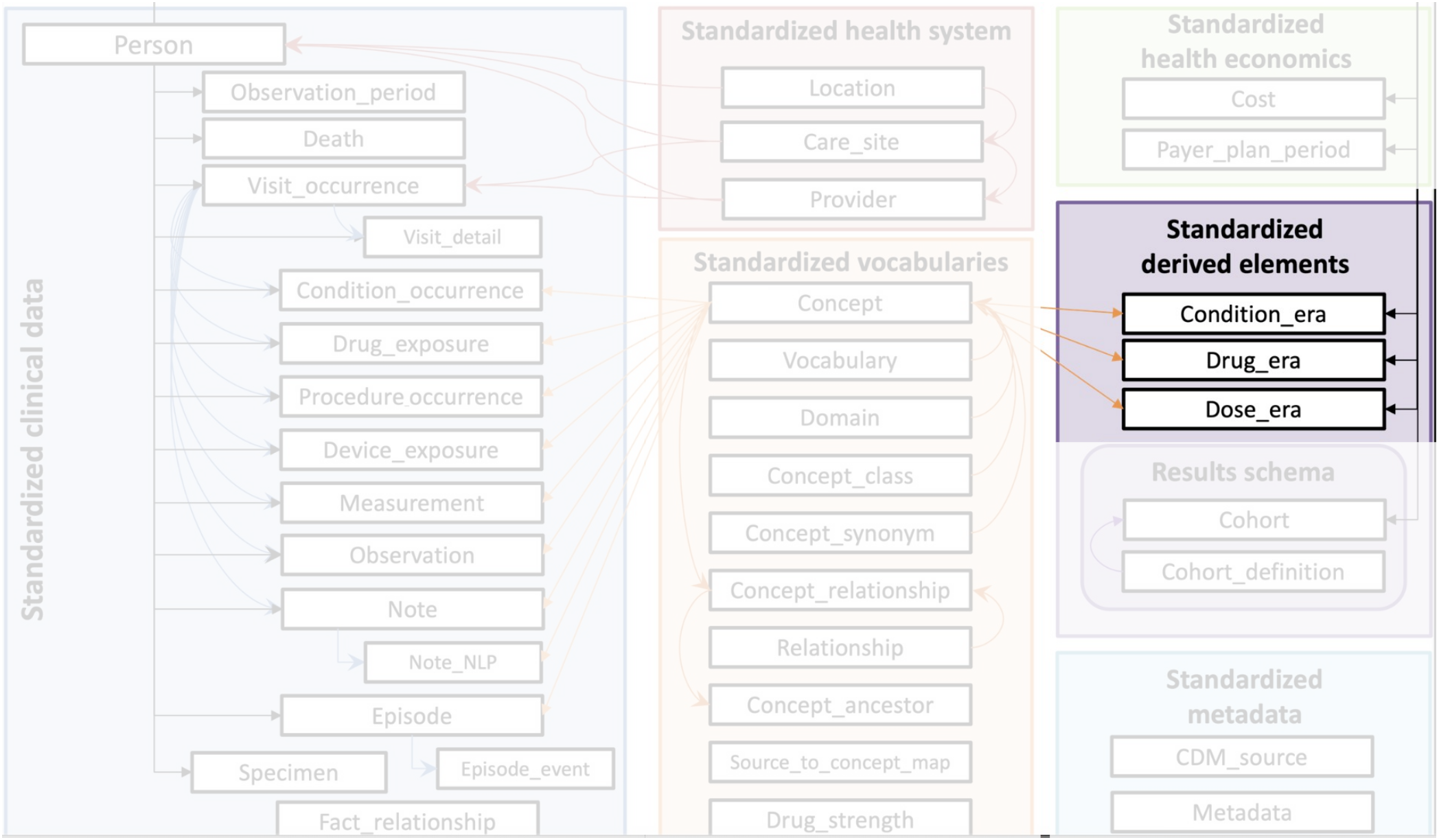
note_nlp		
note_nlp_id	bigint	NN
note_id	bigint	NN
section_concept_id	integer	NN
snippet	text	
offset	text	
lexical_variant	text	NN
note_nlp_concept_id	integer	NN
nlp_system	text	
nlp_date	date	NN
nlp_datetime	timestamp	
term_exists	text	
term_temporal	text	
term_modifiers	text	
note_nlp_source_concept_id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


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

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

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.

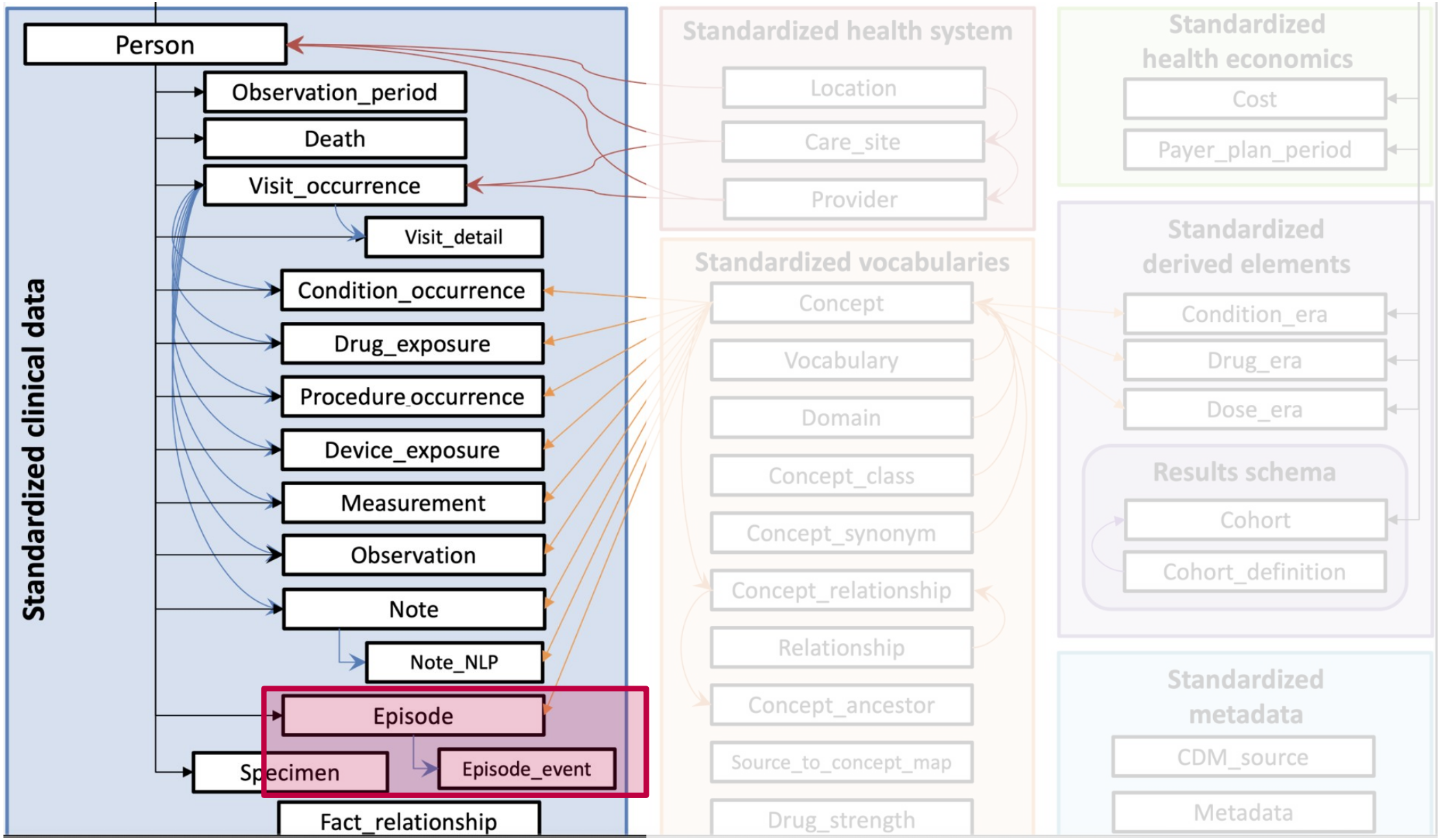
La columna drug_era_start_date de la era de uso de medicamento se establece en la fecha de inicio de la primera exposición a medicamento en el grupo, y la columna drug_era_end_date se establece en la fecha de finalización de la última exposición a medicamento en el grupo.

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

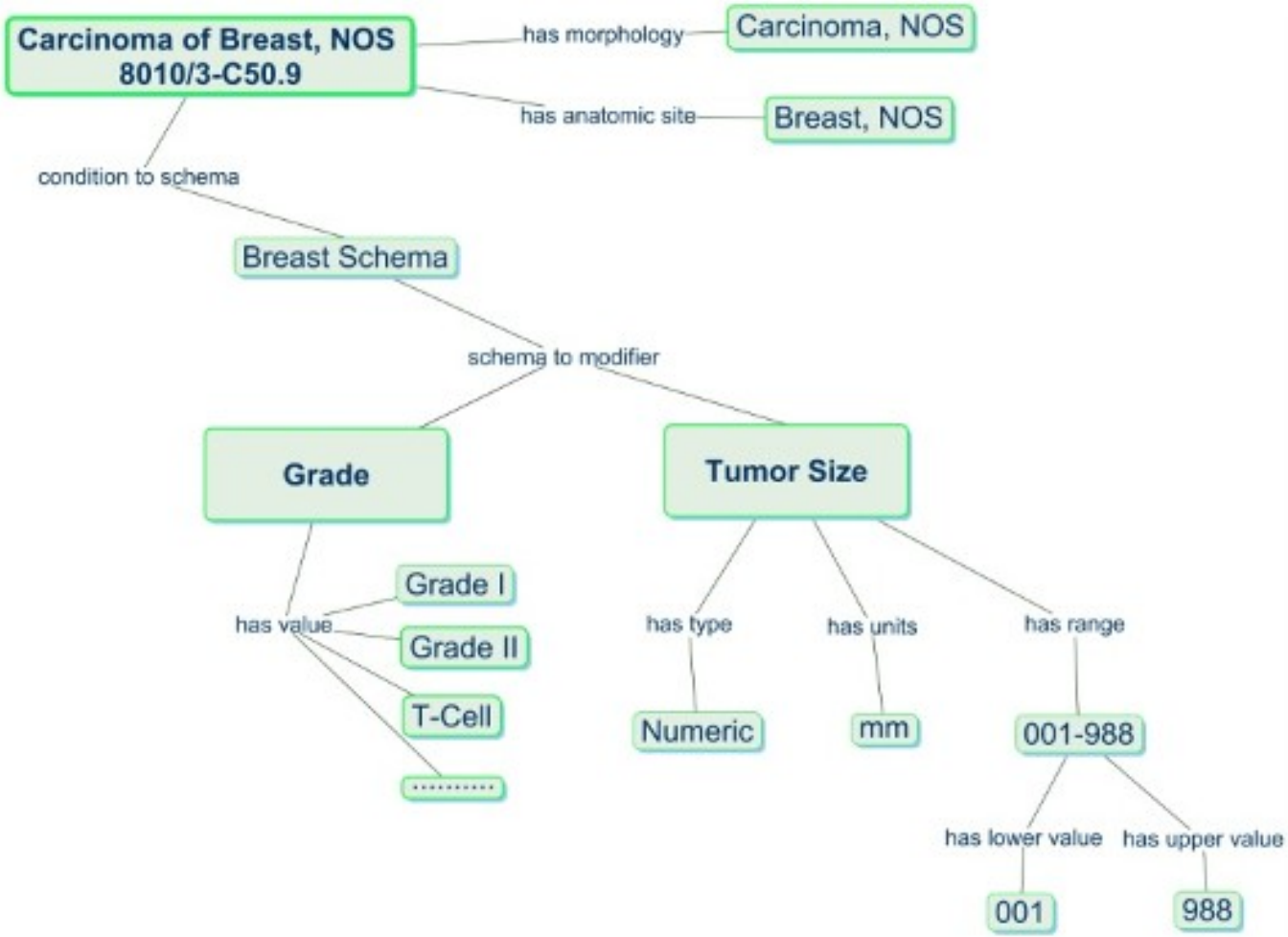
Del mismo modo **DOSE_ERA** se genera agrupando las exposiciones a dosis de medicamentos que ocurren en secuencia continua sin interrupciones significativas para un paciente específico.

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.

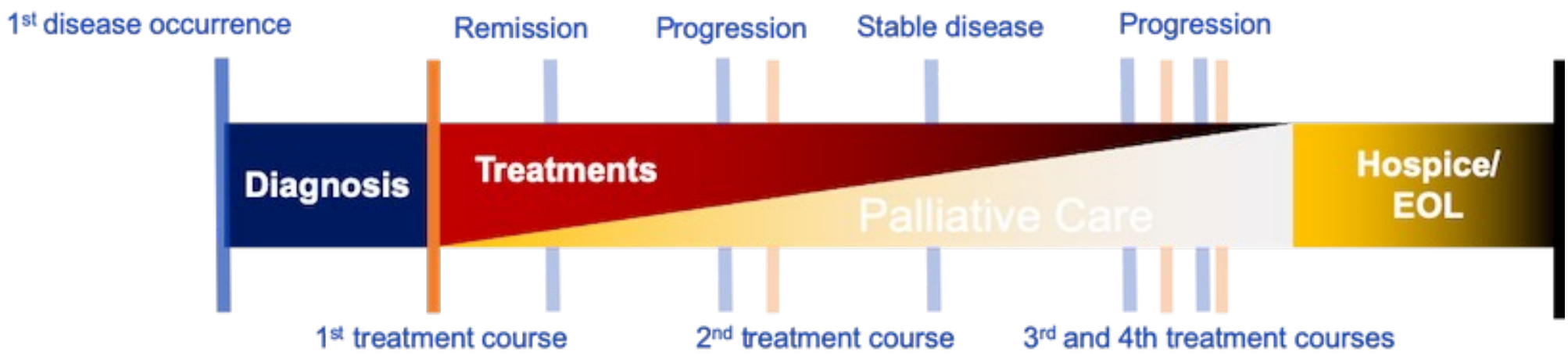
La columna dose_era_start_date de la era de exposición a dosis se establece en la fecha de inicio de la primera exposición a dosis en el grupo, y la columna dose_era_end_date se establece en la fecha de finalización de la última exposición a dosis en el grupo.


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

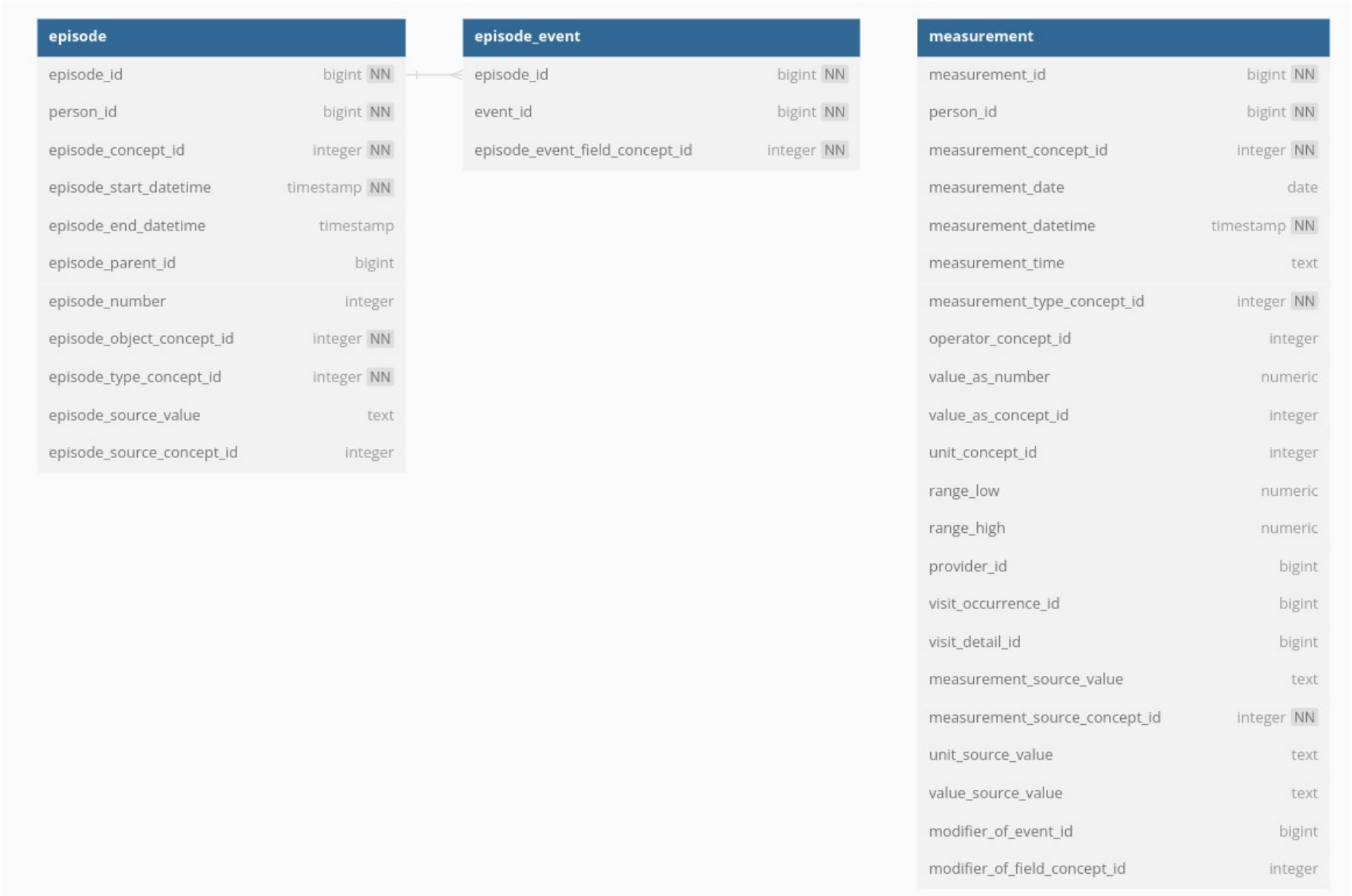
Oncology Extension



Oncology Extension



Oncology Extension



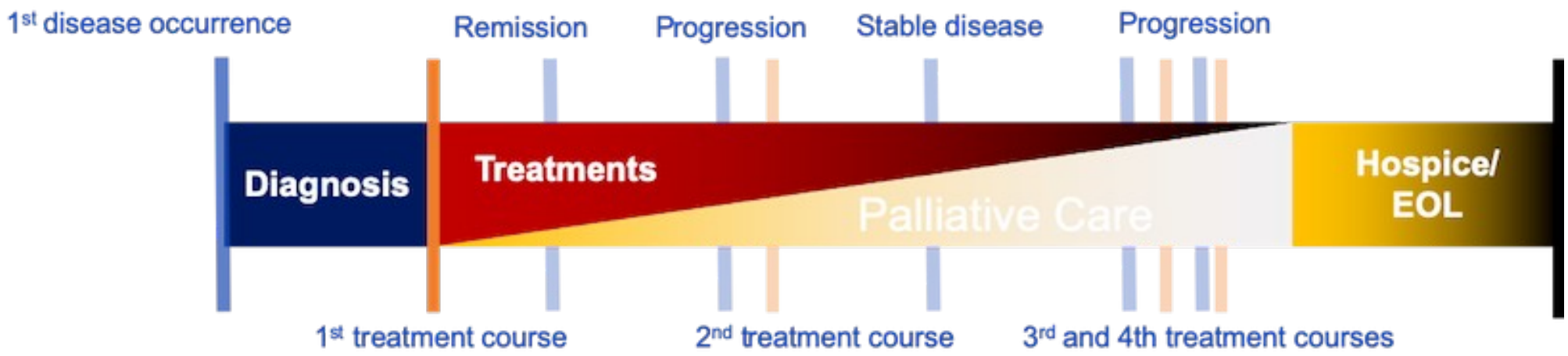
https://ohdsi.github.io/CommonDataModel/oncology.html

Oncology Extension

The **EPISODE** table aggregates lower-level clinical events (VISIT_OCCURRENCE, DRUG_EXPOSURE, PROCEDURE_OCCURRENCE, DEVICE_EXPOSURE) into a higher-level abstraction representing clinically and analytically **relevant disease phases/outcomes and treatments**. The **EPISODE_EVENT** table **connects qualifying clinical event**s (VISIT_OCCURRENCE, DRUG_EXPOSURE, PROCEDURE_OCCURRENCE, DEVICE_EXPOSURE) to the appropriate EPISODE entry.

EPISODE
PK episode_id
FK person_id
FK episode_concept_id
episode_start_datetime
episode_end_datetime
FK episode_parent_id
episode_number
FK episode_object_concept_id
FK episode_type_concept_id
episode_source_value
FK episode_source_concept_id
EPISODE_EVENT
FK episode_id
FK event_id
FK episode_event_field_concept_id
CONDITION_OCCURRENCE
PK condition_occurrence_id
PROCEDURE_OCCURRENCE
PK procedure_occurrence_id
DRUG_EXPOSURE
PK procedure_occurrence_id

Oncology Extension



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

episode_concept_id: OMOP concept 'First disease occurrence'

episode_object_concept_id: SNOMED concept 'Carcinoma of breast'

episode_type_concept_id: 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

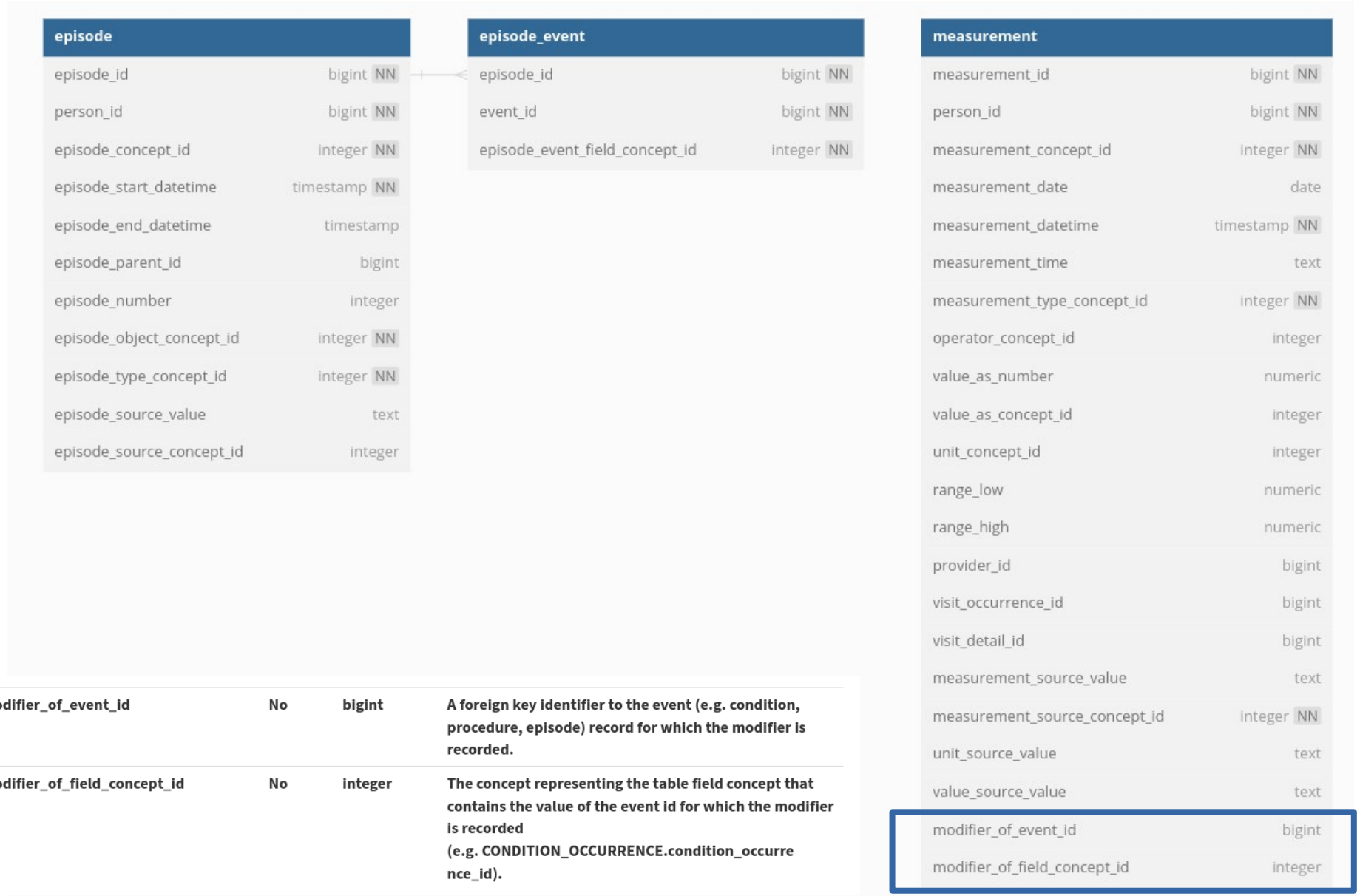
episode_concept_id: OMOP concept 'Treatment Regimen'

episode_parent_id: Foreign key to the disease Episode record

episode_object_concept_id: HemOnc concept Paclitaxel and Bevacizumab'

episode_type_concept_id: OMOP concept 'Episode algorithmically derived from EHR'

Oncology Extension



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'

Vocabularios

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

concept_numeric

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

concept_numeric

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
```

concept_numeric

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?

```
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
```

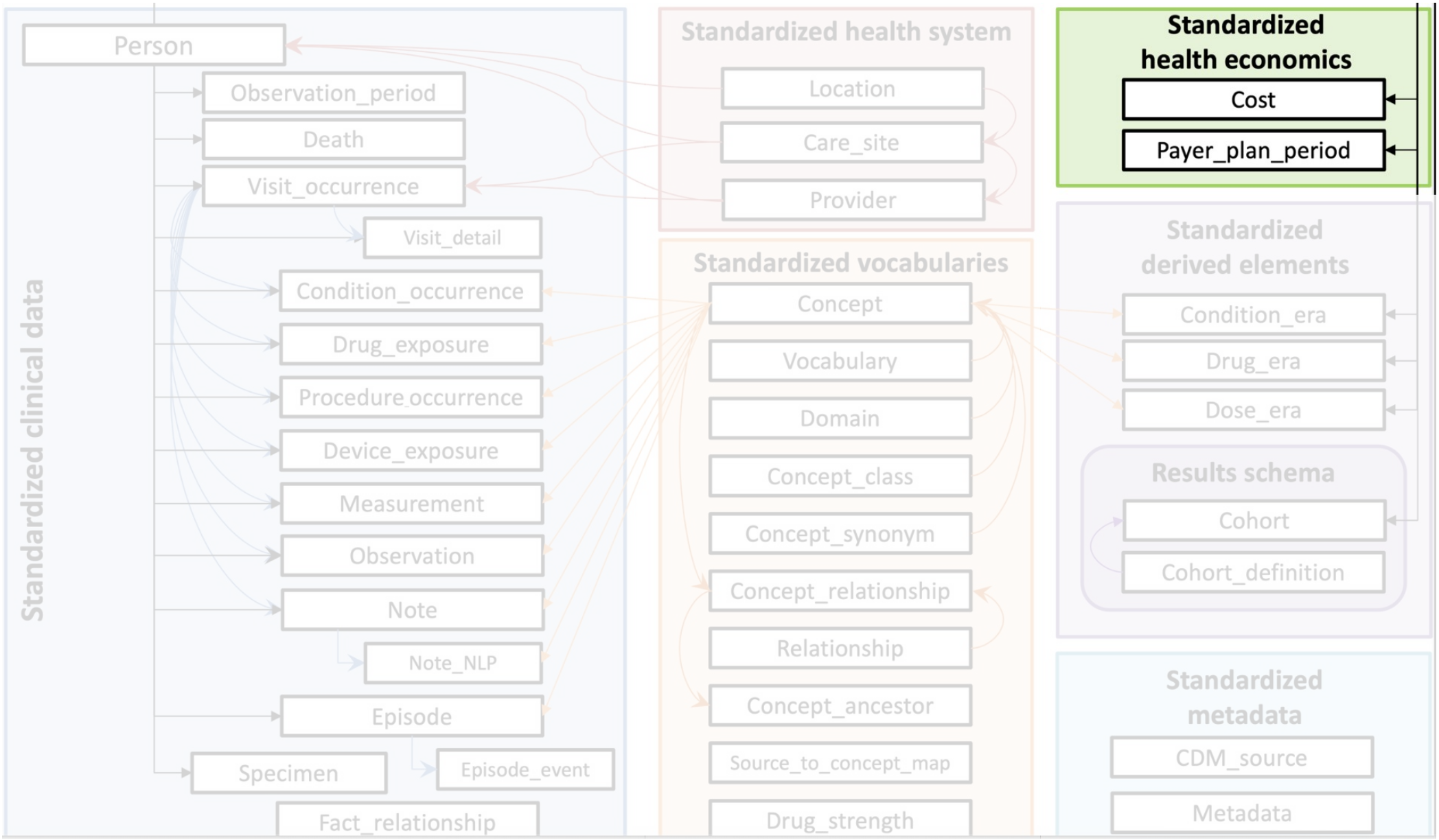
concept_numeric

concept_numeric	
concept_id	integer
value_as_number	float
unit_concept_id	integer
operator_concept_id	float

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

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’


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

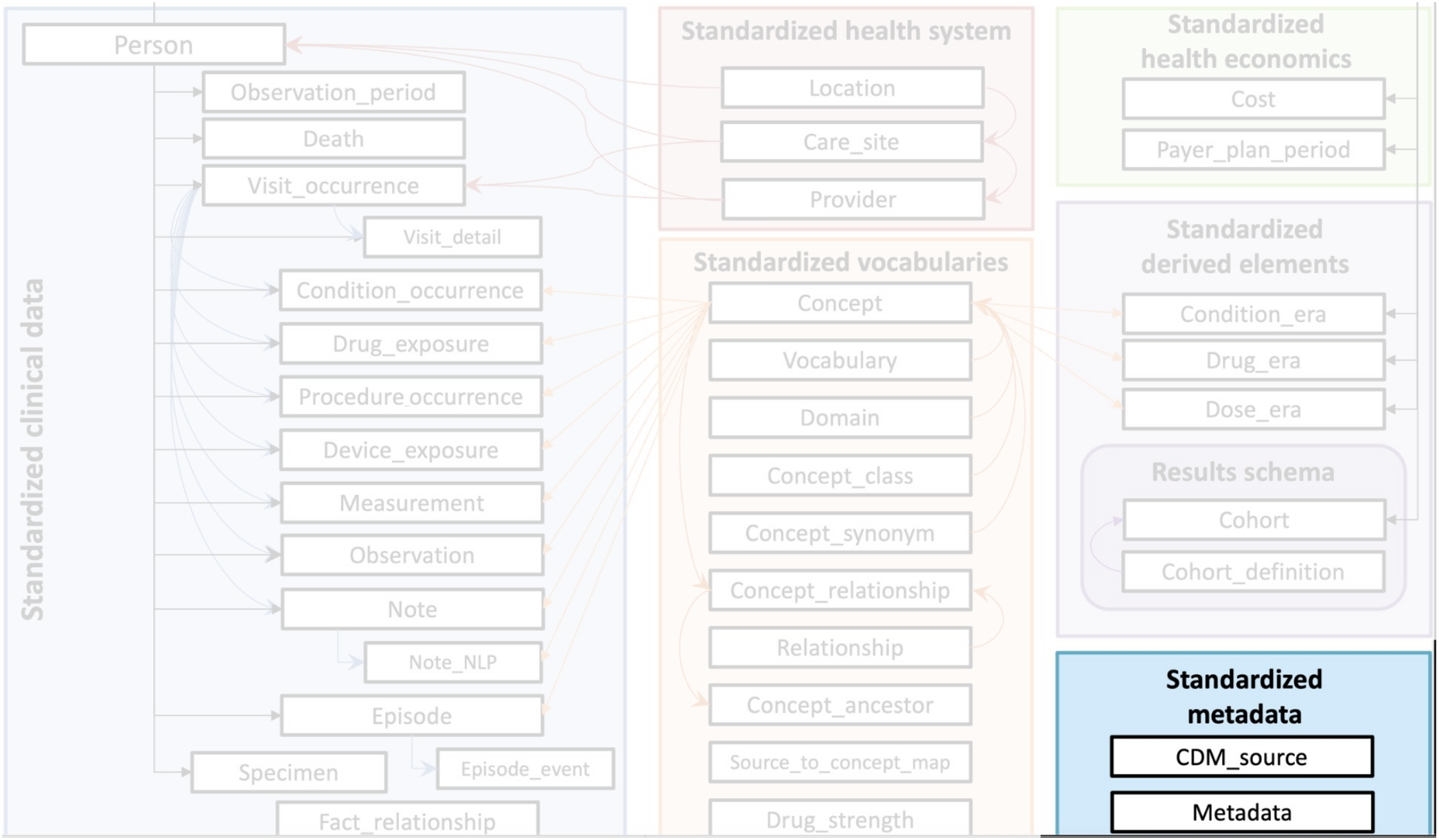
cost	
cost_id	bigint NN
cost_event_id	bigint NN
cost_domain_id	text NN
payer_plan_period_id	bigint
cost_type_concept_id	integer NN
currency_concept_id	integer
revenue_code_concept_id	integer
drg_concept_id	integer
total_charge	double
total_cost	double
total_paid	double
paid_by_payer	double
paid_by_patient	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 tecnologñias 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**

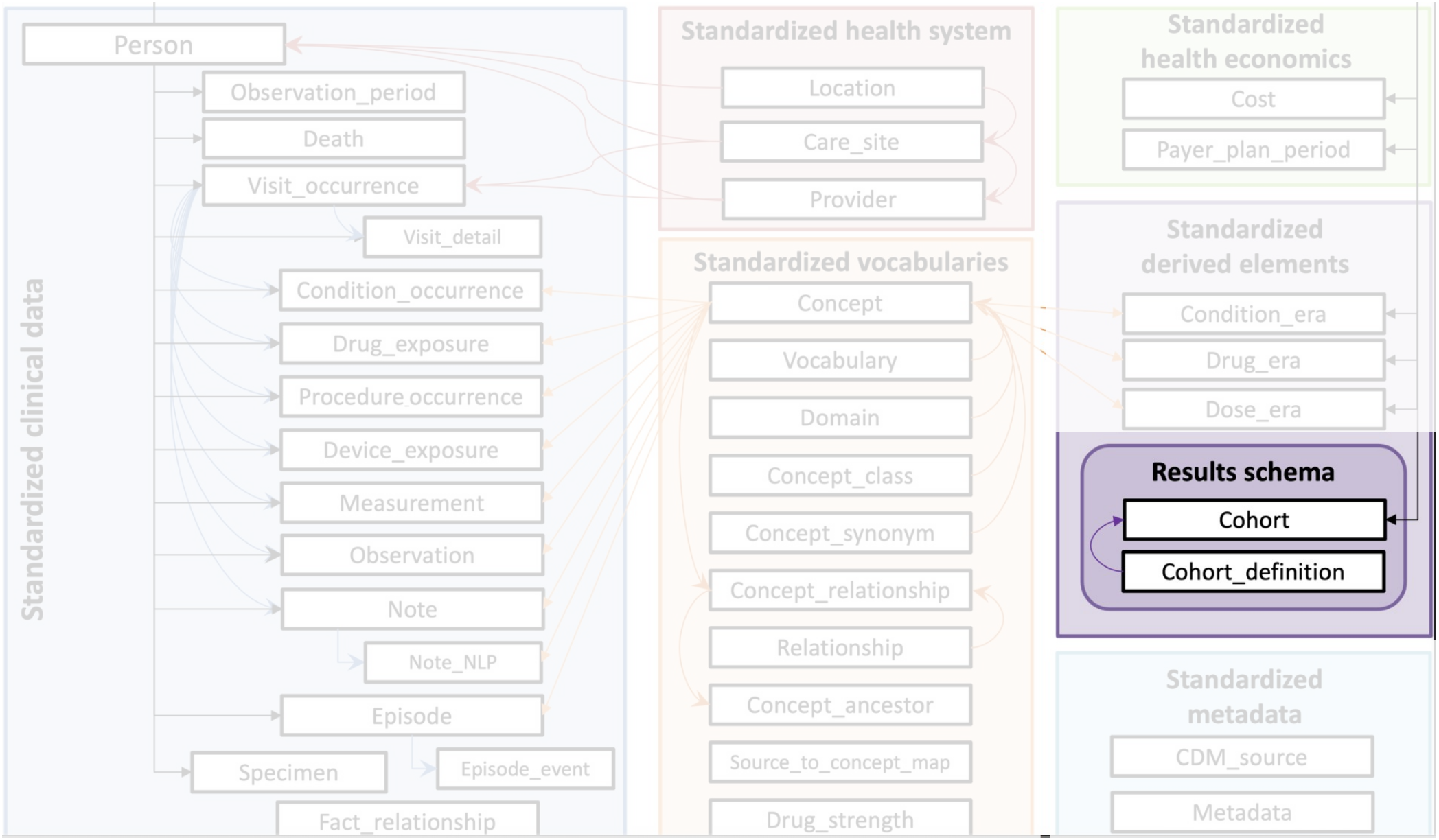

Standardized clinical data
Person
Observation_period
Death
Visit_occurrence
Visit_detail
Condition_occurrence
Drug_exposure
Procedure_occurrence
Device_exposure
Measurement
Observation
Note
Note_NLP
Episode
Specimen
Episode_event
Fact_relationship
Standardized health system
Location
Care_site
Provider
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

Información sobre esta instalación de OMOP-CDM

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

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

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

- 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

Publicación de nuestra base de datos

Data publication - Quality

OVERVIEW

METADATA

RESULTS

ABOUT

DATA QUALITY ASSESSMENT

SYNTHEA SYNTHETIC HEALTH DATABASE

Results generated at 2019-08-22 14:15:06 in 29 mins

	Verification				Validation				Total			
	Pass	Fail	Total	% Pass	Pass	Fail	Total	% Pass	Pass	Fail	Total	% Pass
Plausibility	159	21	180	88%	283	0	283	100%	442	21	463	95%
Conformance	637	34	671	95%	104	0	104	100%	741	34	775	96%
Completeness	369	17	386	96%	5	10	15	33%	374	27	401	93%
Total	1165	72	1237	94%	392	10	402	98%	1557	82	1639	**95%**

https://github.com/OHDSI/DataQualityDashboard

Files

main

Go to file t

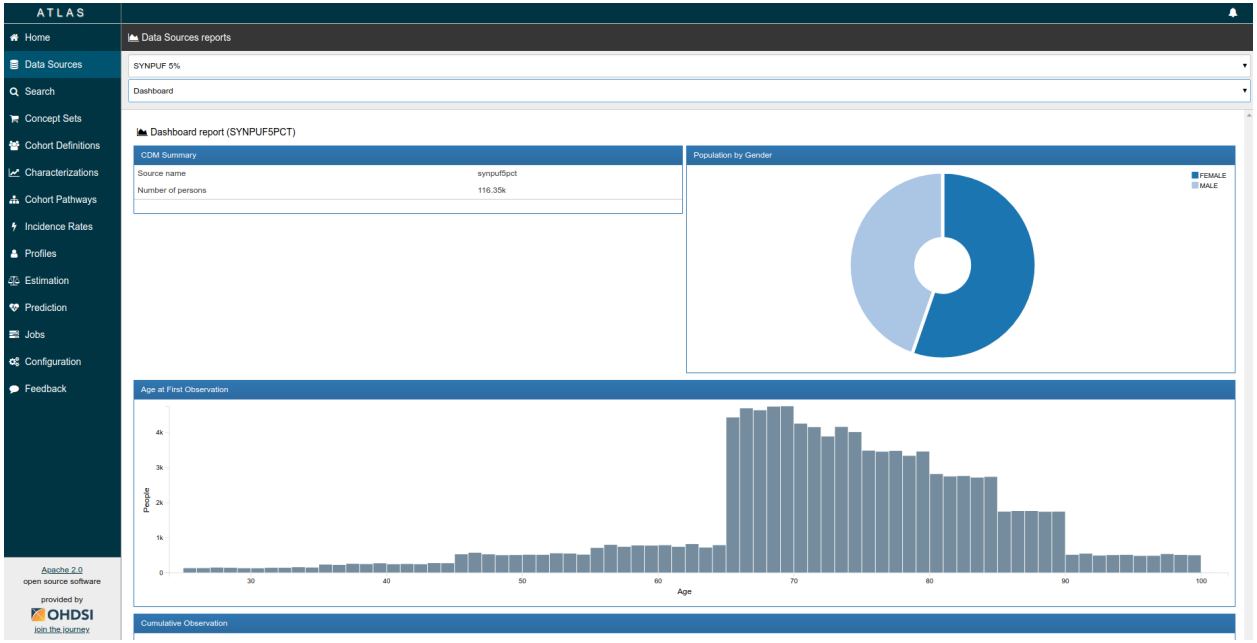
- .github
- R
- docs
- extras
- inst
  - csv
  - doc
  - shinyApps
  - sql/sql_server
    - sqlOnly
    - concept_plausible_gender.sql
    - concept_plausible_unit_concept…
    - concept_plausible_value_high.sql
    - concept_plausible_value_low.sql
    - field_cdm_datatype.sql
    - field_cdm_field.sql

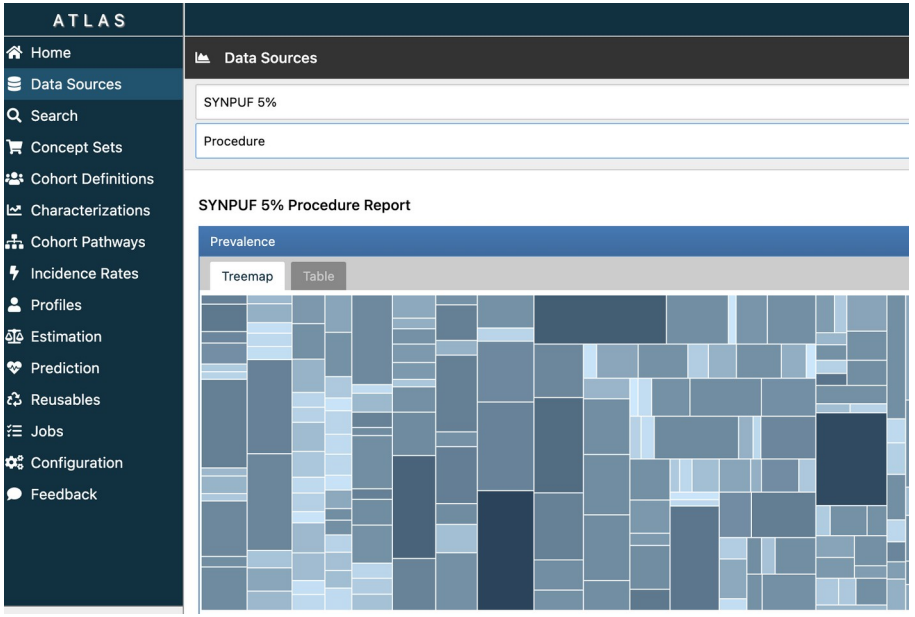
DataQualityDashboard / inst / sql / sql_server / **concept_plausible_gender.sql**

Code Blame Executable File · 59 lines (55 loc) · 1.56 KB

```sql
SELECT
  num_violated_rows,
        CASE
                WHEN denominator.num_rows = 0 THEN 0
                ELSE 1.0*num_violated_rows/denominator.num_rows
        END AS pct_violated_rows,
        denominator.num_rows AS num_denominator_rows
FROM
(
        SELECT
          COUNT_BIG(*) AS num_violated_rows
        FROM
        (
                /*violatedRowsBegin*/
                SELECT cdmTable.*
                FROM @cdmDatabaseSchema.@cdmTableName cdmTable
                        INNER JOIN @cdmDatabaseSchema.person p
                        ON cdmTable.person_id = p.person_id
                        {@cohort}?{
        JOIN @cohortDatabaseSchema.@cohortTableName c
          ON cdmTable.person_id = c.subject_id
          AND c.cohort_definition_id = @cohortDefinitionId
        }
                WHERE cdmTable.@cdmFieldName = @conceptId
                  AND p.gender_concept_id <> {@plausibleGender == 'Male'} ? {8507} : {8532}
                /*violatedRowsEnd*/
        ) violated_rows
) violated_row_count,
```

Data publication - Visualization





https://atlas-demo.ohdsi.org/

OHDSI / **Achilles**

Type / to search

Code | Issues 30 | Pull requests 4 | Discussions | Actions | Projects | Wiki | Security | Insights

Files

main

Go to file t

- export
- summary
  - dbSourceVocabs.sql
  - dbVisitDist.sql
  - generateDbSummary.sql
- temporal
- man
- tests
- vignettes
- .Rbuildignore
- .gitattributes
- .gitignore

Achilles / inst / sql / sql_server / summary / **generateDbSummary.sql**

clairblacketer Added new folder for database summary sql. Updated generateDbSummary....

Code | Blame | 149 lines (116 loc) · 4.77 KB

```sql
-- Generates a one-line database summary table

select c.cdm_source_name,
       c.cdm_source_abbreviation,
       '@country' as country,
       '@provenance' as provenance,
             t3.attribute_name,
             round(t3.attribute_value, 2) as attribute_value,
             rank_order

from @cdm_database_schema.cdm_source c,

(
  select 'Earliest date available' as attribute_name,
         min(cast(stratum_1 as numeric))*1 as attribute_value,
         1 as rank_order
  from @results_database_schema.achilles_results
  where analysis_id = 110
```

Run Achilles

- DatabaseConnector::createConnectionDetails()
- DatabaseConnector::connect()
- SQLRender::render()
- Achilles::achilles()
- Achilles::exportToAres()
- AresIndexer::buildDataQualityIndex()
- DataQualityDashboard::executeDqChecks()
- DataQualityDashboard::convertJsonResultsFileCase()

EHDEN PORTAL

Free text search EHDEN

Alberto Labarga

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

PORTAL

- ABOUT
- GET STARTED
- FEEDBACK
- PROFILE
- SIGN OUT

New Entry | Compare | Export

15 Results | Filtered by: Country: Spain | Sorted by: Database Acronym | Clear All

INCLIVA | VLC Instituto de Investigación Sanitaria	ABUCASIS	ABUCASIS Instituto de Investigación Sanitaria INCLIVA	3.88M Spain
Base de Análisis Resultados DE NAvarra	BARDENA	Base de Analisis Resultados de Navarra Navarre Health Service - SNS-O	600K Spain
BiGAN	BIGAN	BIGAN - Big Data in Healthcare, Aragon (Spain) Aragon Institute for Health Sciences - Instituto Aragonés de Ciencias de la Salud	2.24M Spain
marinaSalud	DptoSalud-DENIA	DENIA Hospital de Denia - MarinaSalud	300K Spain

https://portal.ehden.eu

EHDEN PORTAL

Free text search EHDEN

Alberto Labarga

HOME
CATALOGUE
DASHBOARD
ACADEMY
EHDEN
PUBLICATIONS
STATUS

PORTAL
ABOUT
GET STARTED
FEEDBACK
PROFILE
SIGN OUT

Databases / BARDENA

Fingerprint | Literature | Database Dashboard

Hits: 51 **Unique Views:** 27 **Filled:** 98 %

DATABASE DESCRIPTION

For enquiries about this database, please contact enquiries@ehden.eu.

Database Acronym

BARDENA

Database Name

Base de Analisis Resultados de Navarra

Institution name

Navarre Health Service - SNS-O

Department name

Health results evaluation and analysis service.

Street Address

st/ Tudela 20

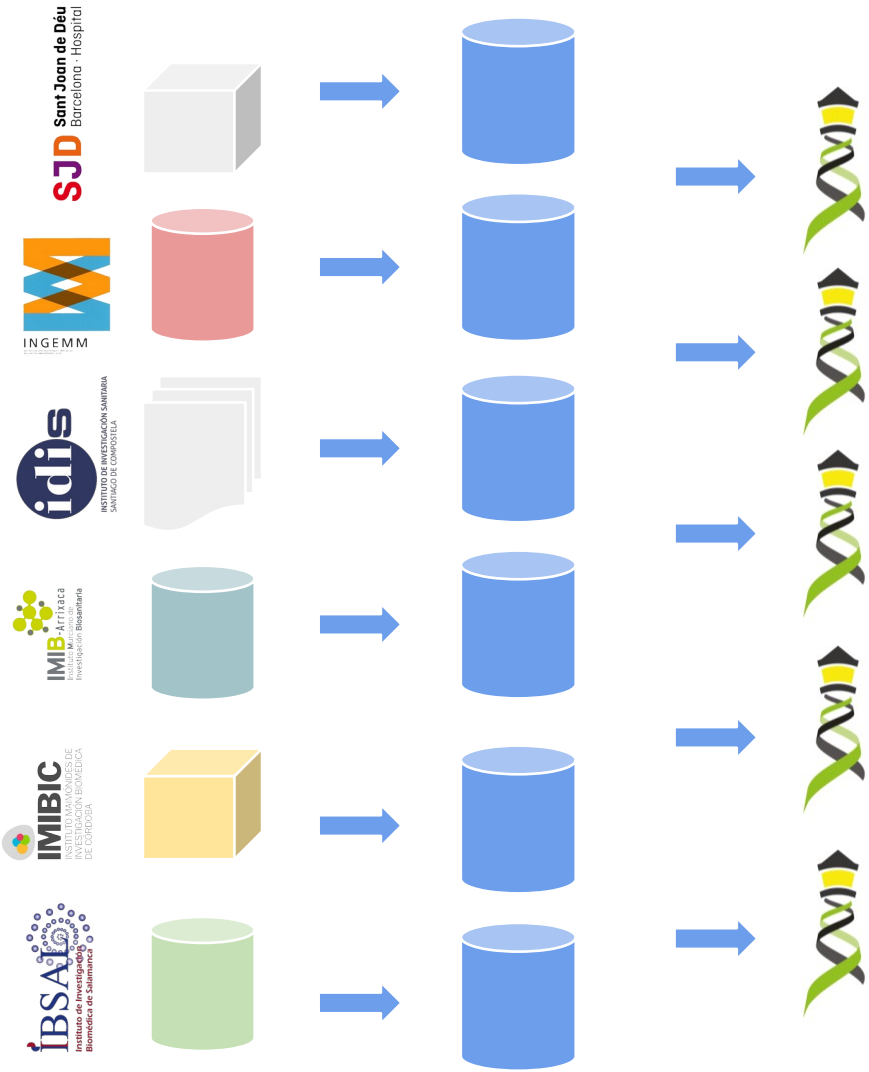
City

Pamplona

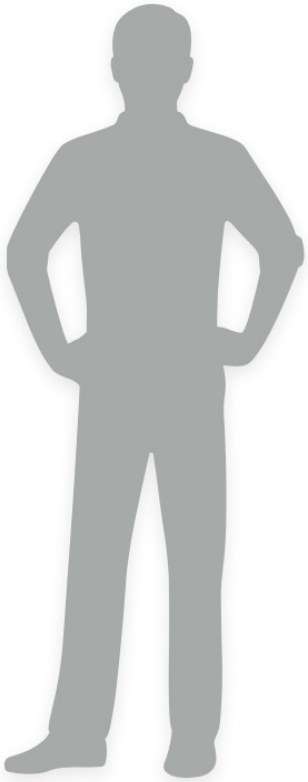
https://portal.ehden.eu

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

GA4GH Beacon



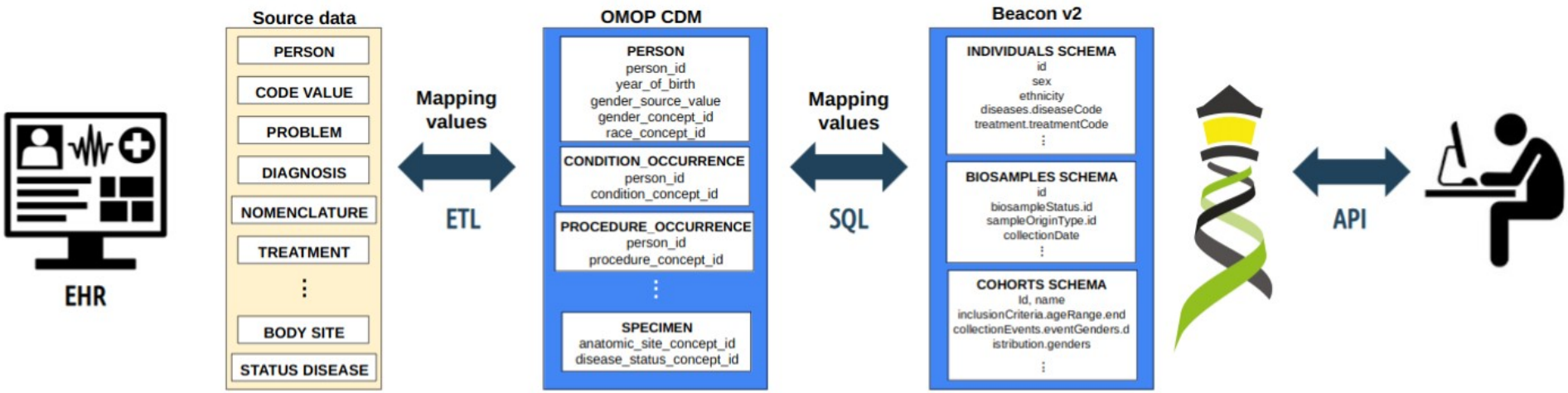
GA4GH Beacon



"Do you have a 'C' at chromosome 13 at position 32,936,732?"

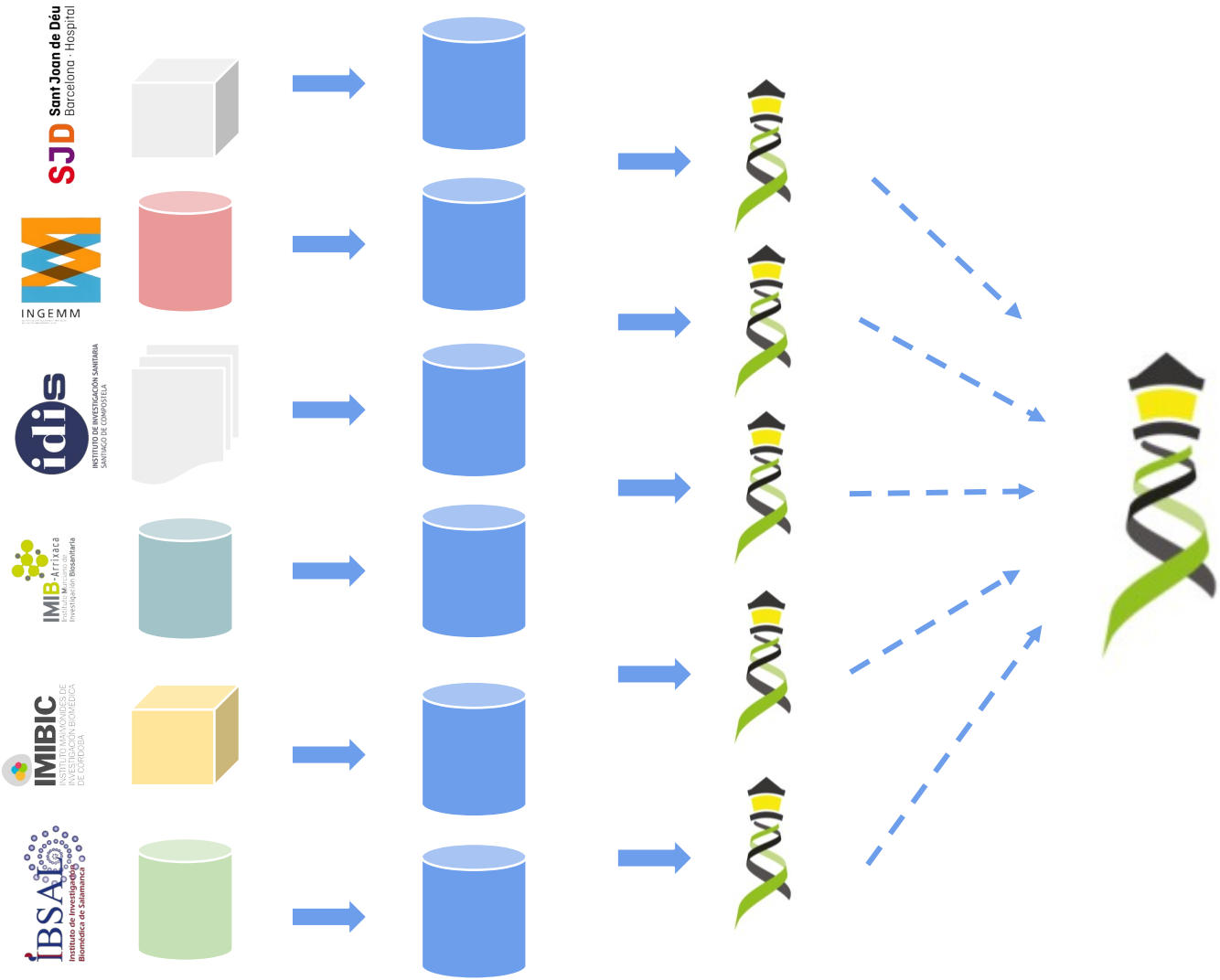
"Yes" (or "no")

GA4GH beacon



https://gitlab.bsc.es/impact-data/impd-beacon_omopcdm

GA4GH Beacon



Beacon network

Beacon Network

Search Beacons Login

Search all beacons for allele

GRCh37 | 13 : 32936732 G > C | Cancel

Log in with Science ID to search controlled access beacons

Response All None

- [x] Found 3
- [ ] Not Found 17
- [ ] Not Applicable 20

Access All None

- [x] Controlled 0
- [x] Public 40

BRCA Exchange
Hosted by BRCA Exchange
Show Metadata Found

HGMD Public
Hosted by University of California, Santa Cruz
Found

MOLGENIS EMX2
Hosted by University Medical Center Groningen
Show Metadata Found

FAIRsharing.org
standards, databases, policies

search through all content

SEARCH

ACTIONS

GENERAL INFORMATION

Minimum Information About a Bioinformatics investigation (MIABi)

10.25504/FAIRsharing.28yec8

MIABi) initiative specifies, through a
that should be provided for a
response to the growing need for
e bioinformatics and computational

DSW

Product Solutions Learn About Get Started

Data Stewardship Wizard

Create, plan, collaborate, and bring your data management plans to life with a tool trusted by thousands of people worldwide — from data management pioneers, to international research institutes.

FAIRCOOKBOOK

GITHUB Search Wizard...

Or browse the various sections

F Findability
EXEMPLAR RECIPES
- Unique, persistent identifiers
- Search engine optimization

LEARN MORE

A Accessibility
EXEMPLAR RECIPES
- Transferring data with SFTP
- Downloading data with Aspera

LEARN MORE

I Interoperability
EXEMPLAR RECIPES
- Selecting terminologies and ontologies
- Creating a metadata profile

LEARN MORE

R Reusability
EXEMPLAR RECIPES
- Data licenses
- Declaring data's permitted uses

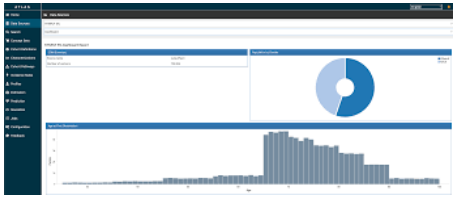
LEARN MORE

Infrastructure LEARN MORE

Assessments LEARN MORE

Applied Examples LEARN MORE

Maturity model LEARN MORE



HADES

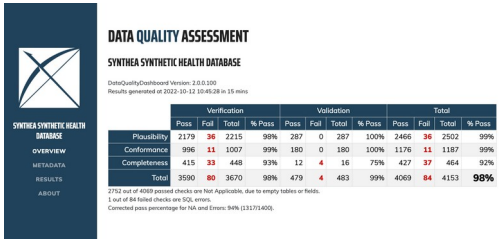
HADES

HEALTH ANALYTICS DATA-TO-EVIDENCE SUITE

Web API

Apache Tomcat

PostgreSQL



SYNTHEA SYNTHETIC HEALTH DATABASE

OVERVIEW

METADATA

RESULTS

ABOUT

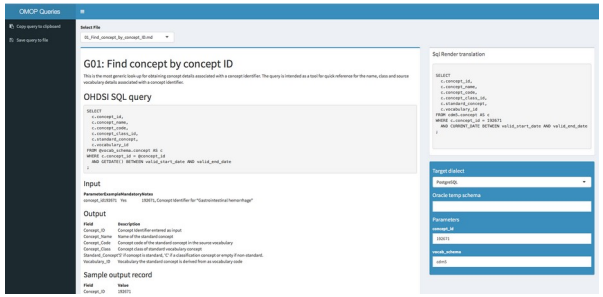
DATA QUALITY ASSESSMENT

SYNTHEA SYNTHETIC HEALTH DATABASE

DataQualityDashboard Version: 2.0.0.100
Results generated at 2022-10-12 10:45:28 in 15 mins

	Verification				Validation				Total			
	Pass	Fail	Total	% Pass	Pass	Fail	Total	% Pass	Pass	Fail	Total	% Pass
Plausibility	2179	36	2215	98%	287	0	287	100%	2466	36	2502	99%
Conformance	996	11	1007	99%	180	0	180	100%	1176	11	1187	99%
Completeness	415	33	448	93%	12	4	16	75%	427	37	464	92%
Total	3590	80	3670	98%	479	4	483	99%	4069	84	4153	**98%**

2752 out of 4069 passed checks are Not Applicable, due to empty tables or fields.
1 out of 84 failed checks are SQL errors.
Corrected pass percentage for NA and Errors: 94% (3317/3400).



OMOP Queries

G01: Find concept by concept ID

OHDSI SQL query

Input

Output

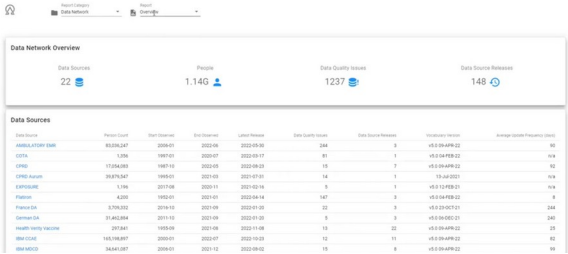
Sample output record

Sql Render translation

Target dialect

Oracle temp schema

Parameters



Data Network Overview

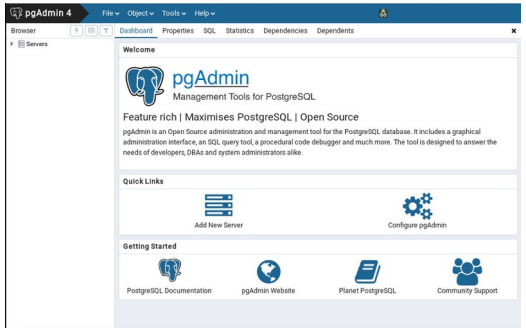
Data Sources 22

People 1.14G

Data Quality Issues 1237

Data Source Releases 148

Data Sources



pgAdmin 4

File Object Tools Help

Browser

Servers

Dashboard Properties SQL Statistics Dependencies Dependents

Welcome

pgAdmin

Management Tools for PostgreSQL

Feature rich | Maximises PostgreSQL | Open Source

pgAdmin is an Open Source administration and management tool for the PostgreSQL database. It includes a graphical administration interface, an SQL query tool, a procedural code debugger and much more. The tool is designed to answer the needs of developers, DBAs and system administrators alike.

Quick Links

Add New Server

Configure pgAdmin

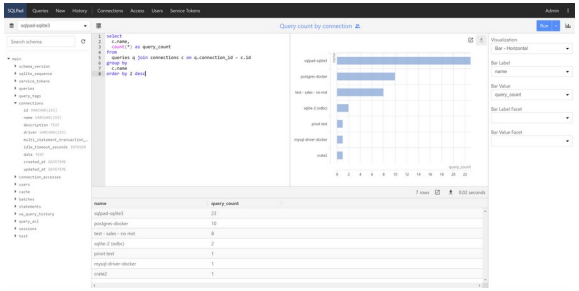
Getting Started

PostgreSQL Documentation

pgAdmin Website

Planet PostgreSQL

Community Support



SQLPad

Query count by connection

Visualization

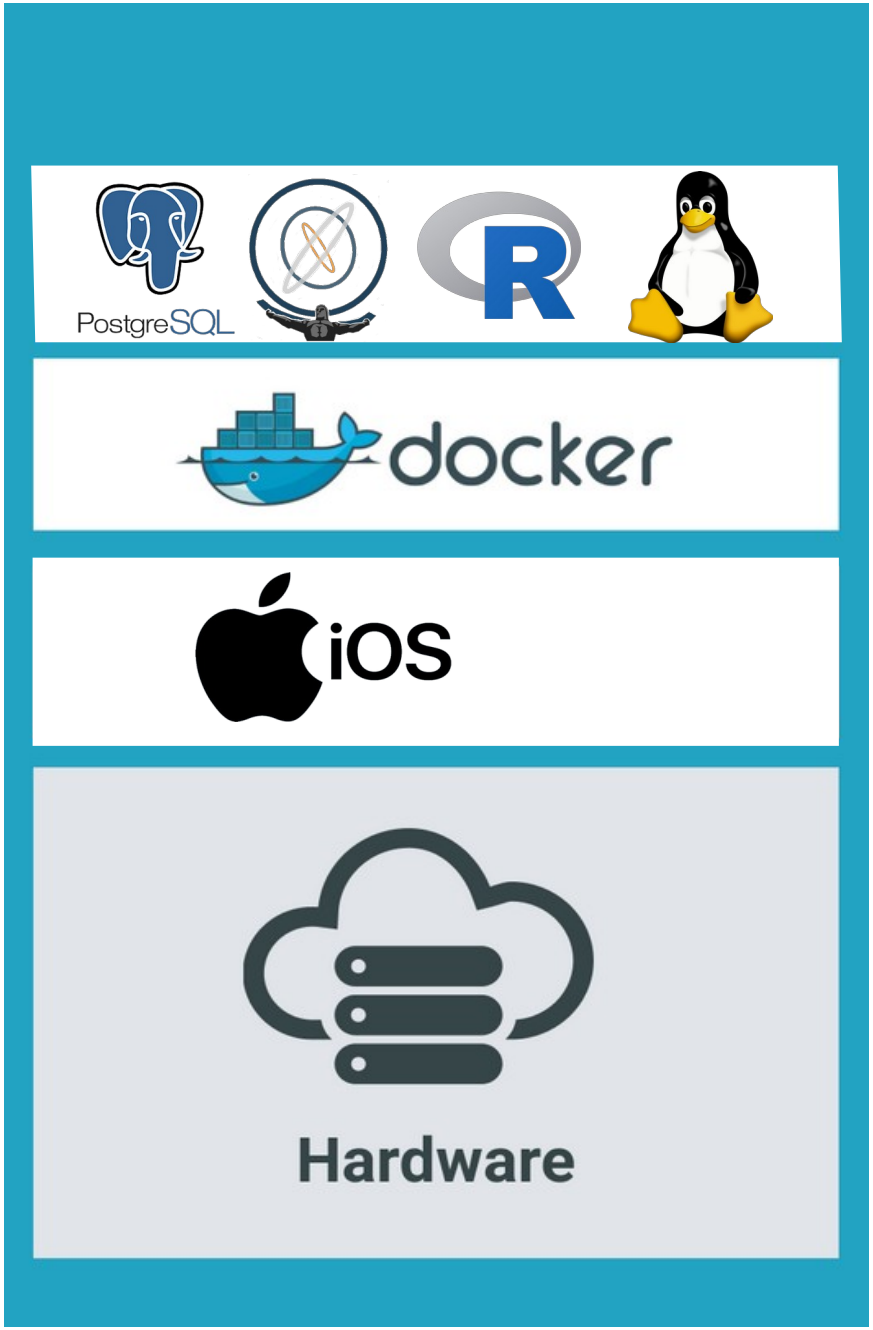
Bar - Horizontal

Bar Label

Bar Value

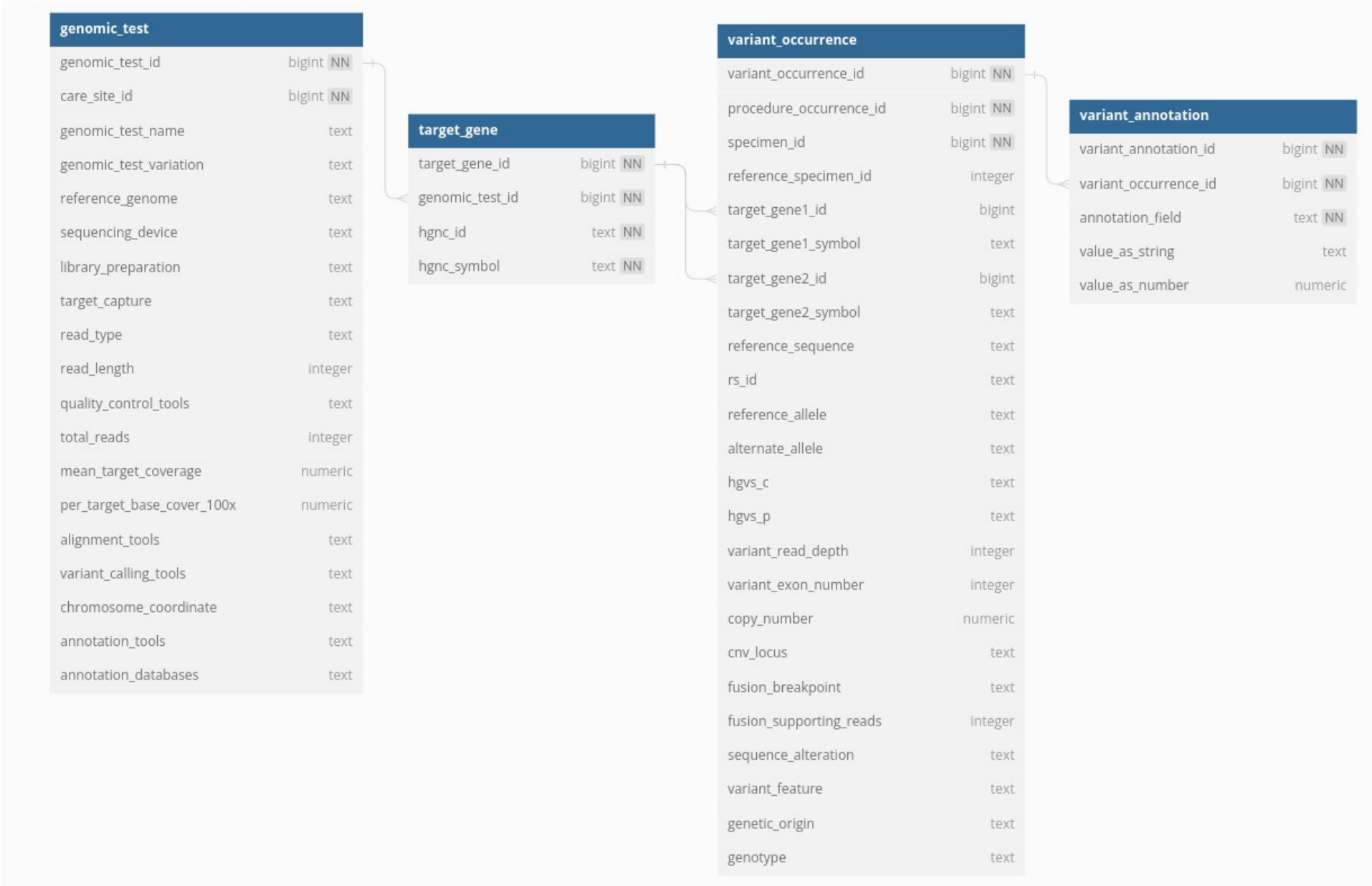
Bar Label Facet

Bar Value Facet


PostgreSQL
R
docker
iOS
Hardware

OMOP-CDM extensions

Genomics Extension



Genomic data

[gene name + sequential variant + specimen + inspection method + exon number + variant type]

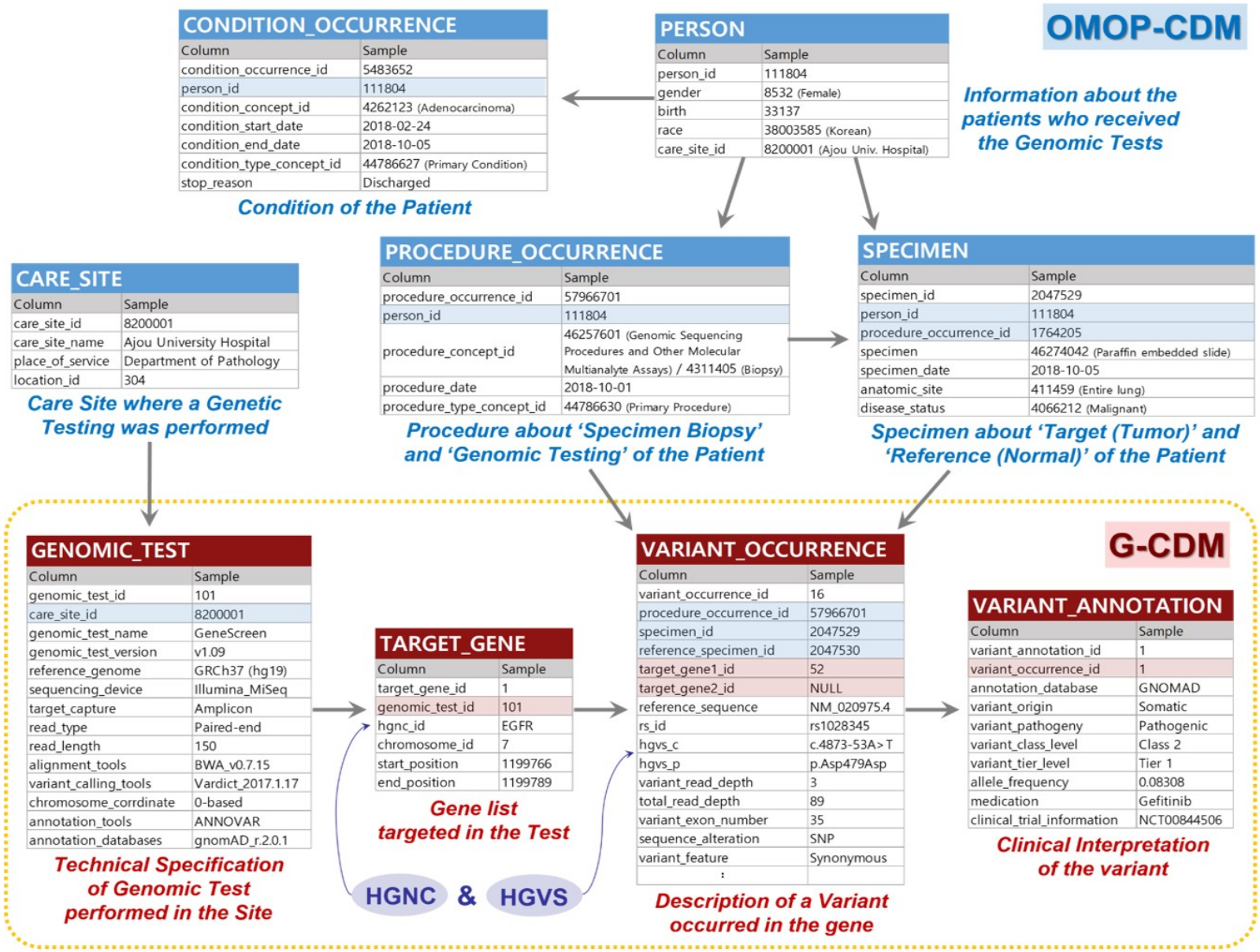
Measurement table

ID	CODE	NAME	CLASS	CONCEPT	VALIDITY	DOMAIN	VOCAB
40776139	LP19695-3	EGFR gene	LOINC Hierarchy	Classification	Valid	Measurement	LOINC
40758886	55766-2	EGFR gene c.2156G>C+2155G>A+2155G>T [Presence] in Blood or Tissue by Molecular genetics method	Lab Test	Standard	Valid	Measurement	LOINC
40758889	55768-6	EGFR gene c.2303G>T [Presence] in Blood or Tissue by Molecular genetics method	Lab Test	Standard	Valid	Measurement	LOINC
40758890	55769-4	EGFR gene c.2369C>T [Presence] in Blood or Tissue by Molecular genetics method	Lab Test	Standard	Valid	Measurement	LOINC
36203240	85383-8	EGFR gene c.2369C>T actual/normal in Plasma cell-free DNA by Molecular genetics method	Lab Test	Standard	Valid	Measurement	LOINC

Observation table

ID	CODE	NAME	CLASS	CONCEPT	VALIDITY	DOMAIN	VOCAB
40775288	LP96041-7	EGFR gene c.2573T>G	LOINC Hierarchy	Classification	Valid	Observation	LOINC
40785298	LP96040-9	EGFR gene c.2156G>C+2155G>A+2155G>T	LOINC Hierarchy	Classification	Valid	Observation	LOINC
40785299	LP96045-8	EGFR gene exon 20 insertion	LOINC Hierarchy	Classification	Valid	Observation	LOINC
40788607	LP96063-1	EGFR gene exon 19	LOINC Hierarchy	Classification	Valid	Observation	LOINC
40791895	LP96042-5	EGFR gene c.2582T>A	LOINC Hierarchy	Classification	Valid	Observation	LOINC

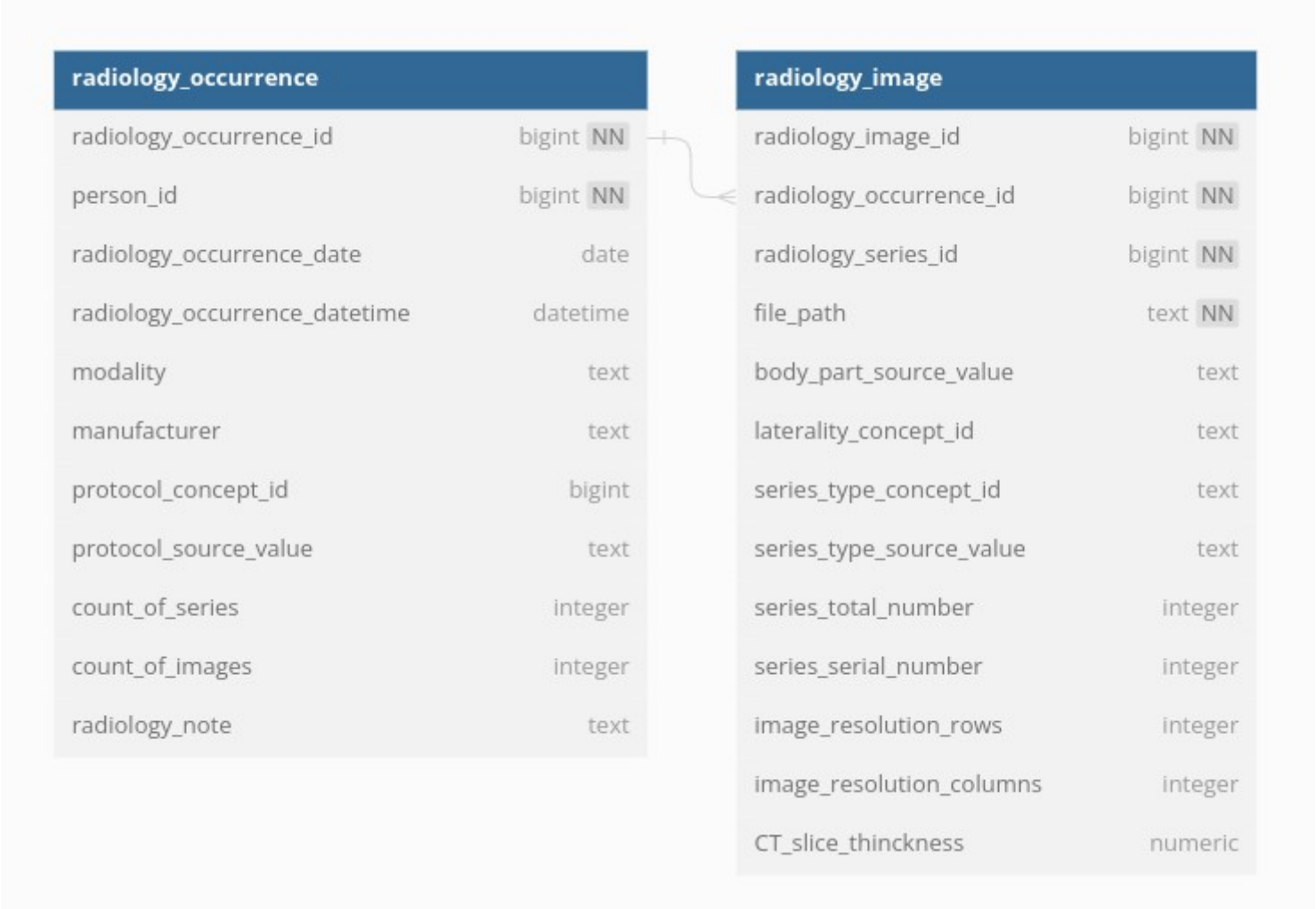
Genomic extension



Genomic Extension

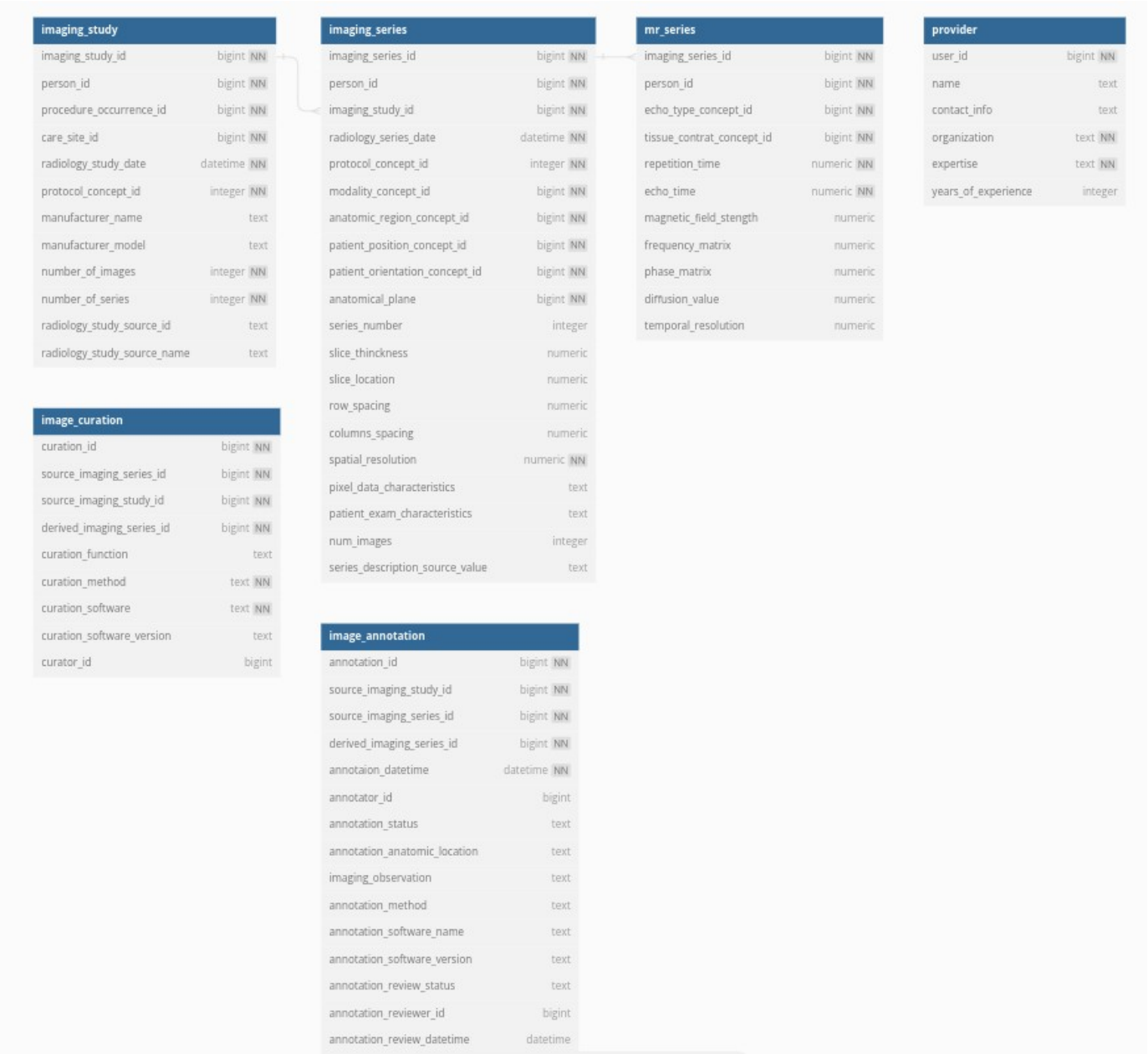
- Genomic Vocabularies
  - HGNC (HUGO Gene Nomenclature Committee)
  - ClinVar
  - CiVIC (Clinical Interpretation of Variants in Cancer)
  - NCIt (National Cancer Institute Thesaurus)
  - CGI (Cancer Genome Interpreter)
  - CAP (College of American Pathologists Cancer Checklists)
  - JAX CKB (The Clinical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

Radiology CDM



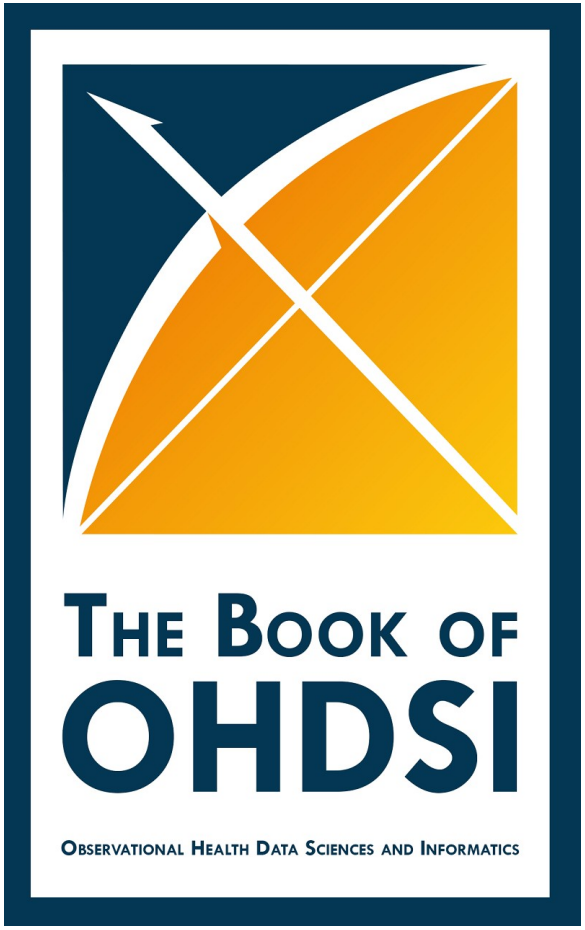
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8790584/

Medical Imaging CDM



Recursos

- Diapositivas
- El Libro de OHDSI
- EHDEN academy
- OHDSI workshops
- Atlas tutorials



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.