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

- European Bioinformatics Institute (Senior Data Scientist)
  - Led the Service Oriented Architecture
  - Genomics databases, web services, semantic web

- Navarrabiomed (Head of bioinformatics)
  - Led the genome sequencing bioinformatics infrastructure (NAGEN, PharmaNAGEN)
  - Multiomics data integration

- IOMED (Head of data)
  - Led data normalization efforts in more than 20 hospitals (OMOP-CDM, NLP)
  - 10M patients, 200million health records

- Barcelona Supercomputing Center (Head of Biomedical Data Hub)
  - Lead the Biomedical Data Hub
  - Open science, FAIR data

Descripción del curso

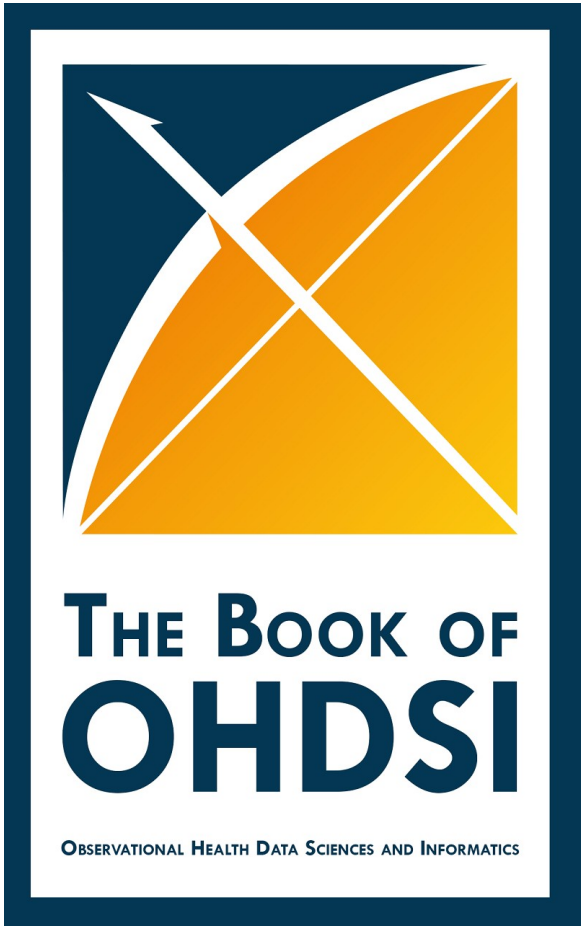
El curso tiene como objetivo proporcionar una comprensión profunda del modelo de datos **OMOP-CDM de la OHDSI**, sus vocabularios y cómo usar la **aplicación web Atlas** para crear cohortes y extraer datos para **estudios observacionales**. El curso cubrirá los conceptos y principios básicos del modelo 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 consultas de **cohortes**.

Esquema

- Introducción al modelo de datos OMOP-CDM
- Vocabularios OMOP-CDM
- Consultas habituales de interés clínico
- Introducción a las herramientas web OHDSI
- Caracterización y análisis de cohortes

Recursos

- Sitio web: https://github.com/alabarga/omop-cdm-course
- Diapositivas
- El Libro de OHDSI
- Entorno local (docker)
  - Base de datos OMOP-CDM
  - Consultas de ejemplo
  - Atlas



(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.
- 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.

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.
- 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

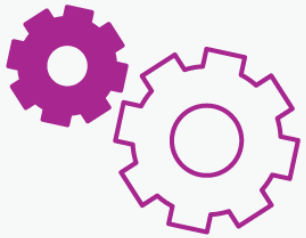
Identificadores únicos y metadatos que pueden usarse para encontrar los datos de manera sencilla

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.

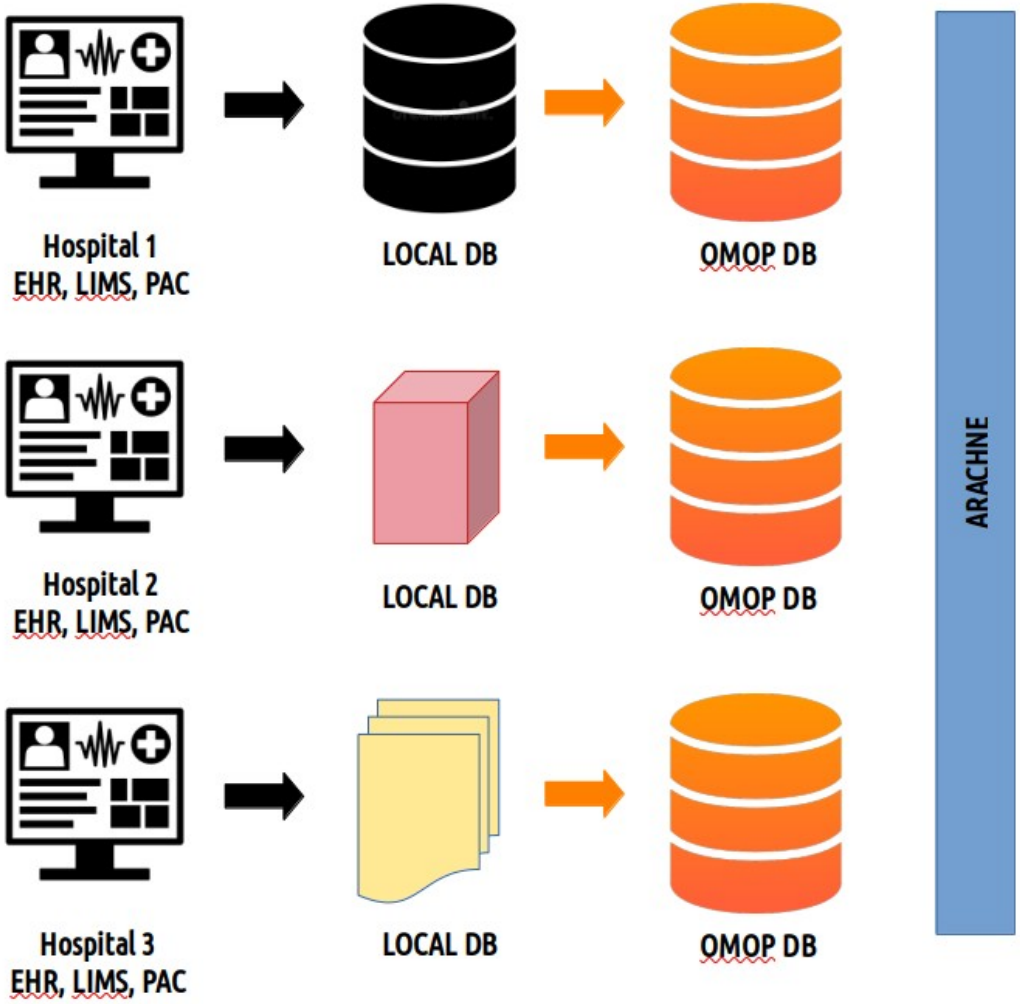
openEHR es un estándar abierto que describe la administración y almacenamiento de información sanitaria basada en arquetipos, definiciones estandarizadas para estructurar y almacenar información médica. Estos arquetipos permiten que los sistemas de salud capturen, compartan e interpreten datos de manera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

Introducción al modelo de datos OMOP-CDM

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.

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

IBSAL Instituto de Investigación Biomédica de Salamanca

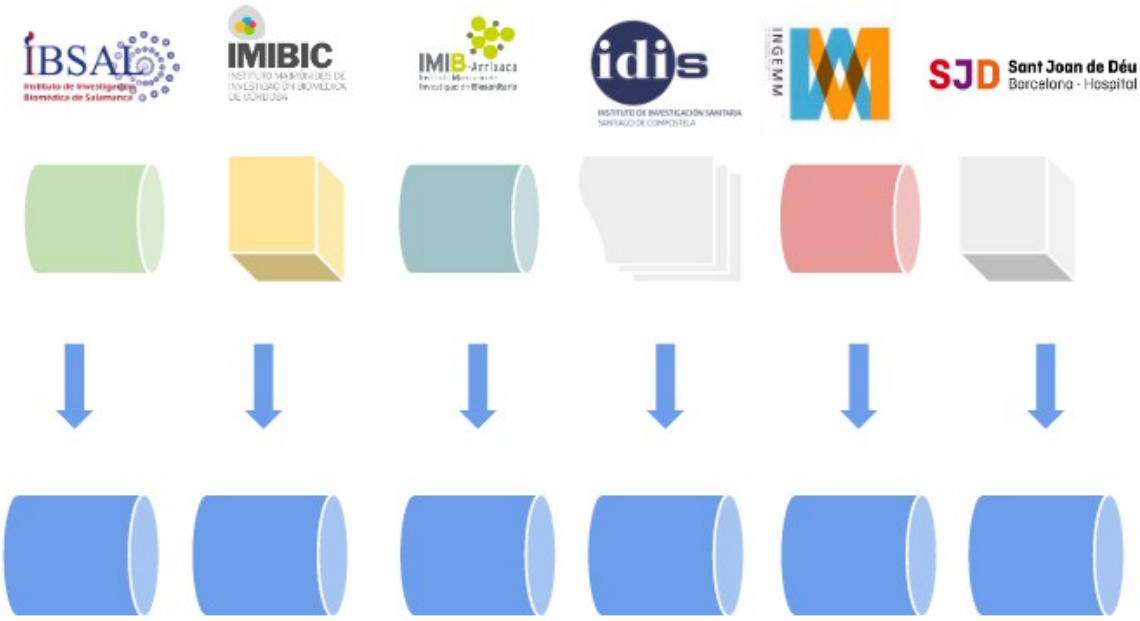
IMIBIC Instituto Maimónides de Investigación Biomédica de Córdoba

IMIB-Arrixaca Instituto Murciano de Investigación Biosanitaria

idis Instituto de Investigación Sanitaria Santiago de Compostela

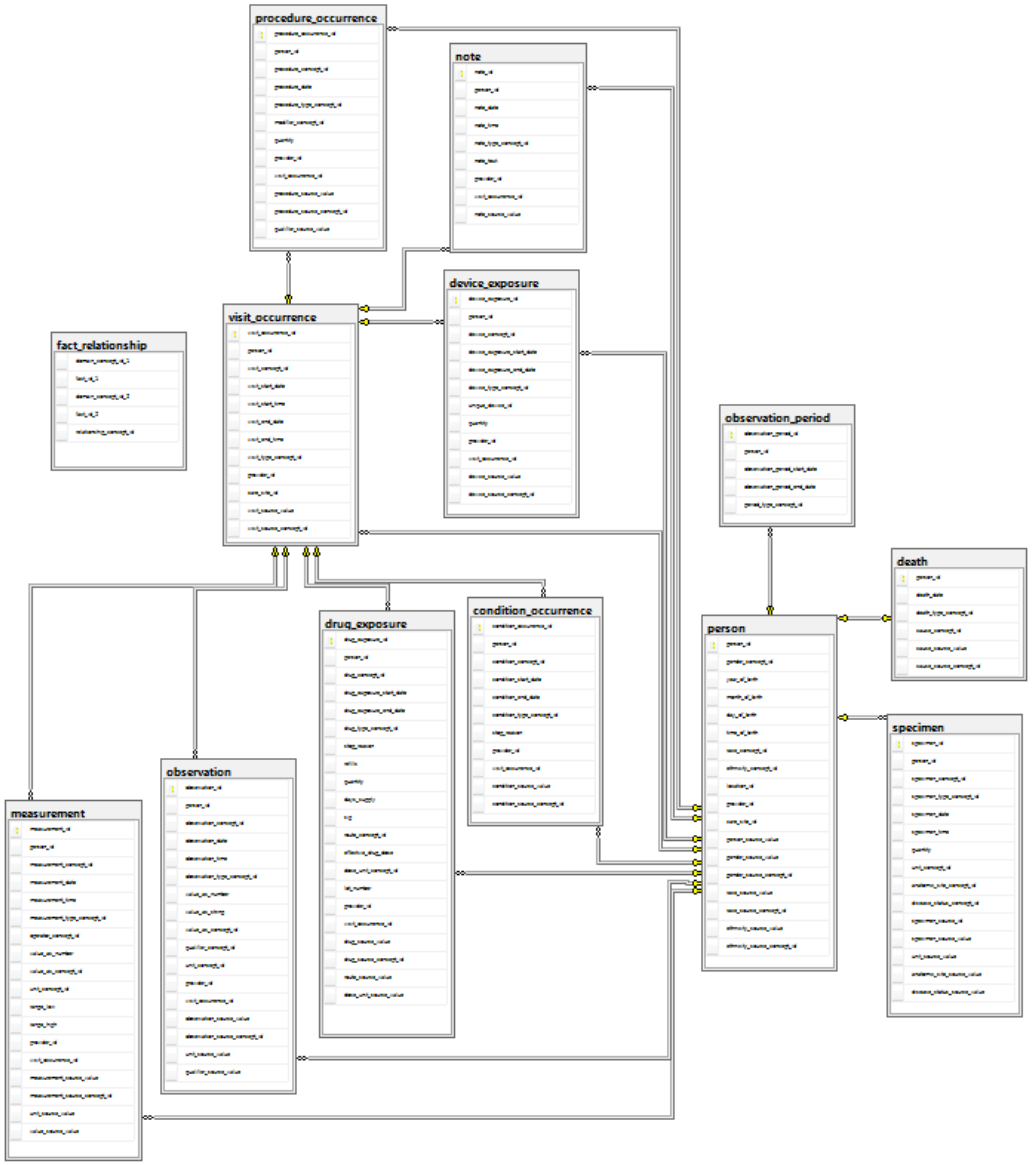
INGEMM

SJD Sant Joan de Déu Barcelona - Hospital


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

OHDSI

OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS



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

ATHENA

SEARCH | DOWNLOAD | Alberto Labarga

SEARCH BY KEYWORD: cancer

cancer ×
Standard ×

- DOMAIN
- CONCEPT
  - Classification (346)
  - Non-standard (5079)
  - Standard (8007)
- CLASS
- VOCAB
- VALIDITY

DOWNLOAD RESULTS

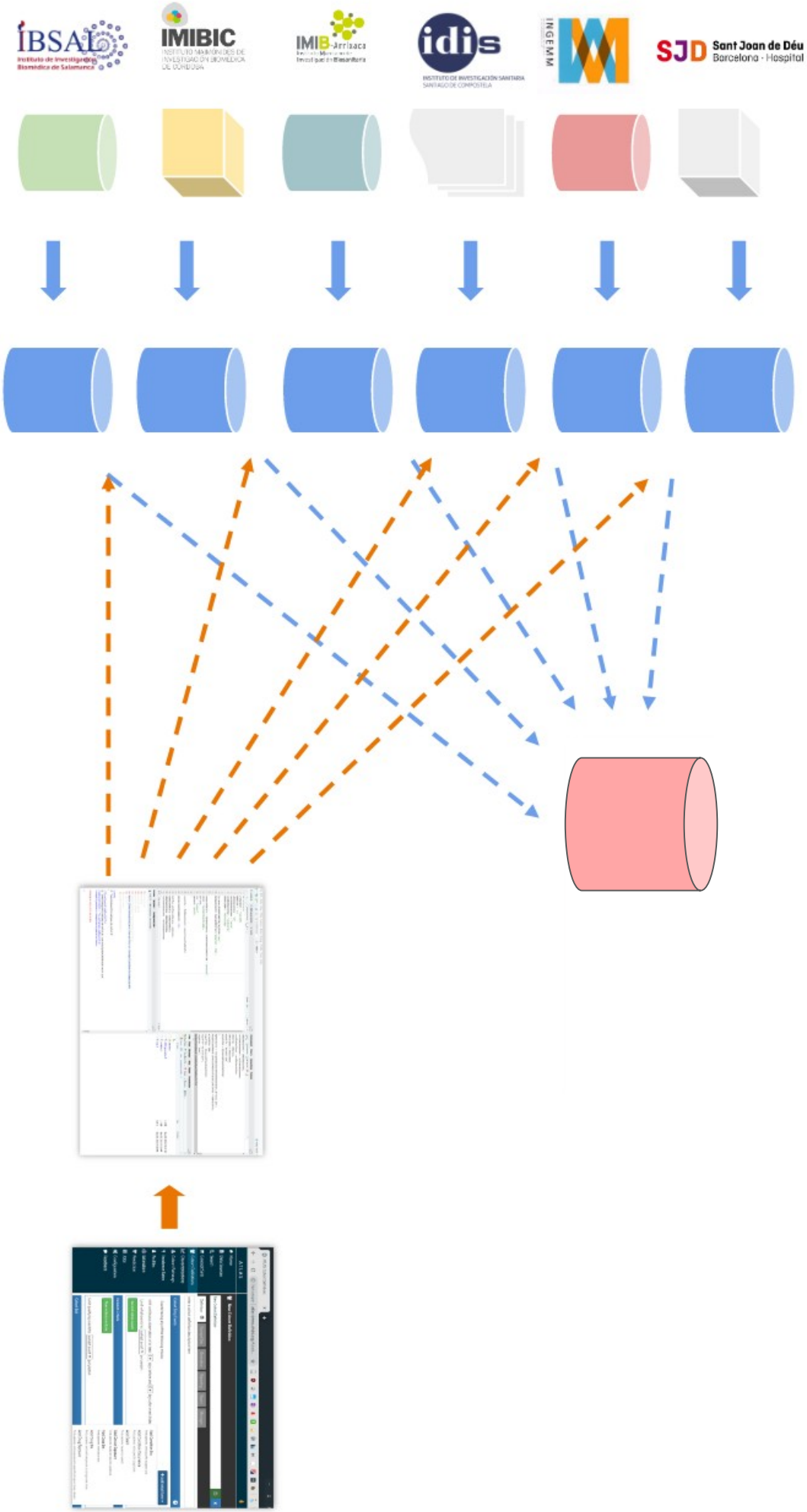
Show by 15 items Total 8,007 items

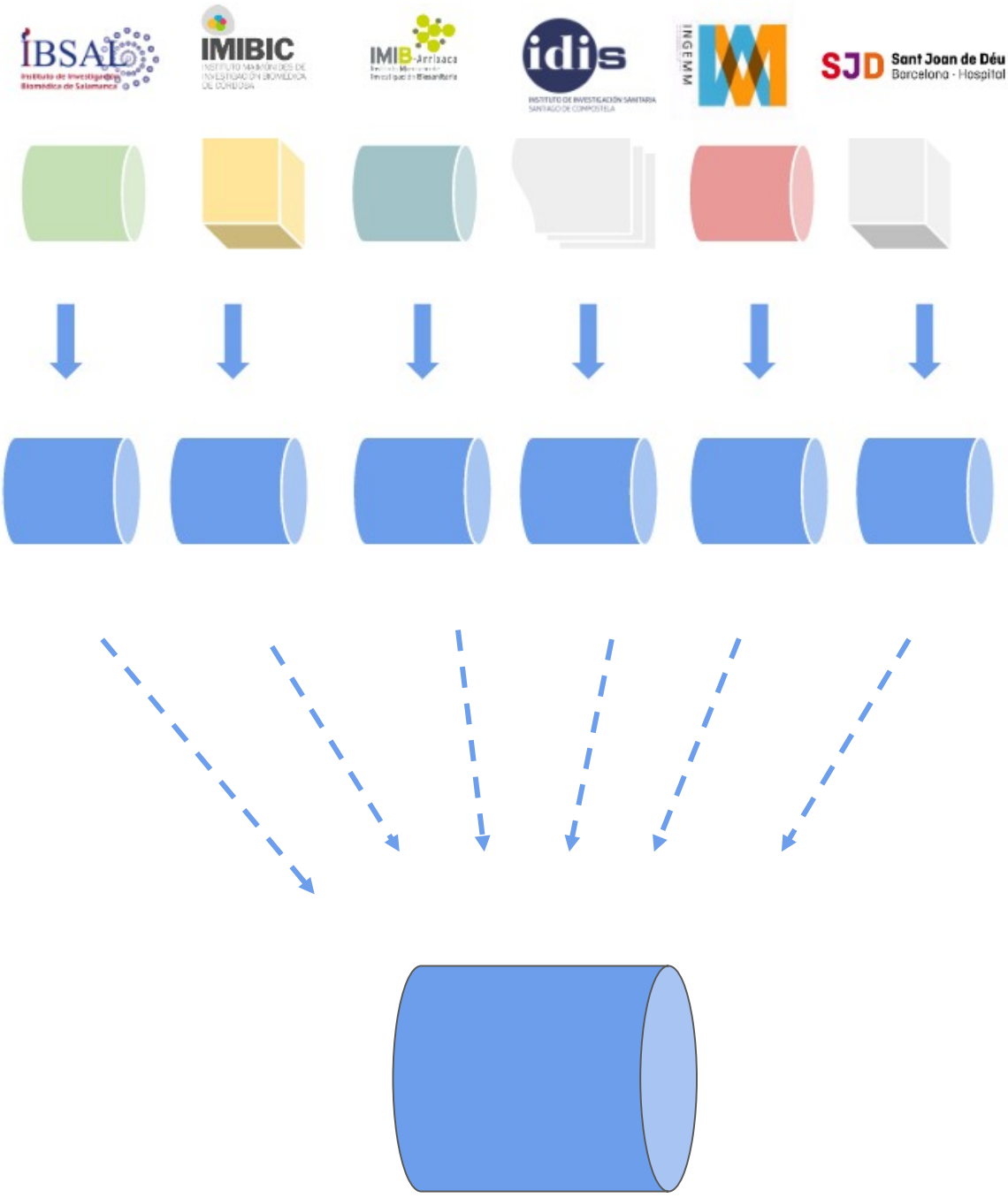
1 2 3 4 5 ··· 534

ID	CODE	NAME	CLASS	CONCEPT	VALIDITY	DOMAIN	VOCAB
45877275	LA10524-9	Cancer	Answer	Standard	Valid	Meas Value	LOINC
4194405	395099008	Cancer confirmed	Context-dependent	Standard	Valid	Condition	SNOMED
4299598	386604001	Cancer registrar	Social Context	Standard	Valid	Observation	SNOMED
1384523	CancerConditions_NoCancer	Cancer Conditions: No Cancer	Answer	Standard	Valid	Observation	PPI
4073624	243072006	Cancer education	Procedure	Standard	Valid	Procedure	SNOMED
44803809	722731000000100	Colorectal cancer	Qualifier Value	Standard	Valid	Observation	SNOMED
40757663	54532-7	Cancer	Survey	Standard	Valid	Observation	LOINC

https://athena.ohdsi.org

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


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

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

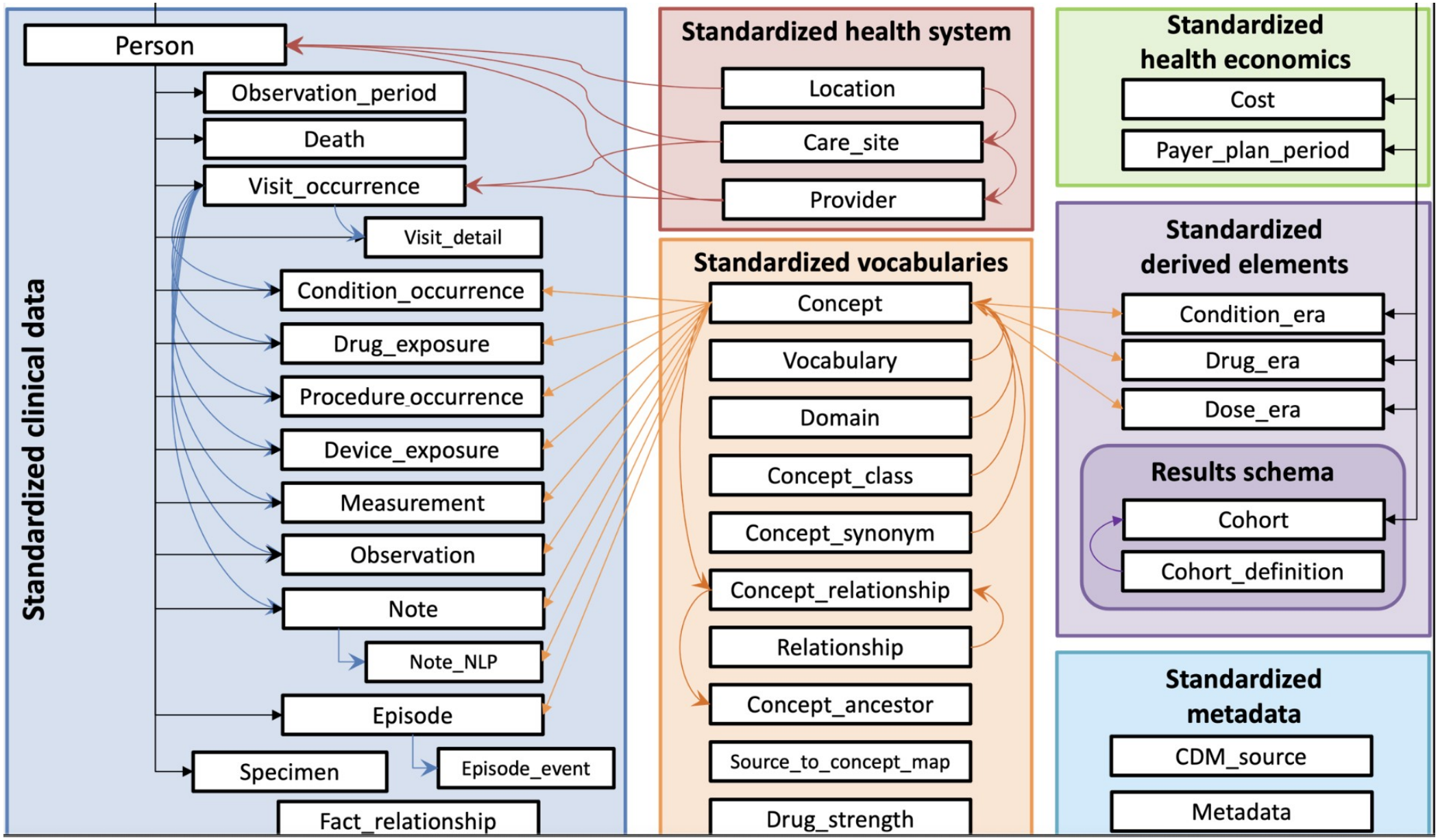
All records in this table are independent Persons.

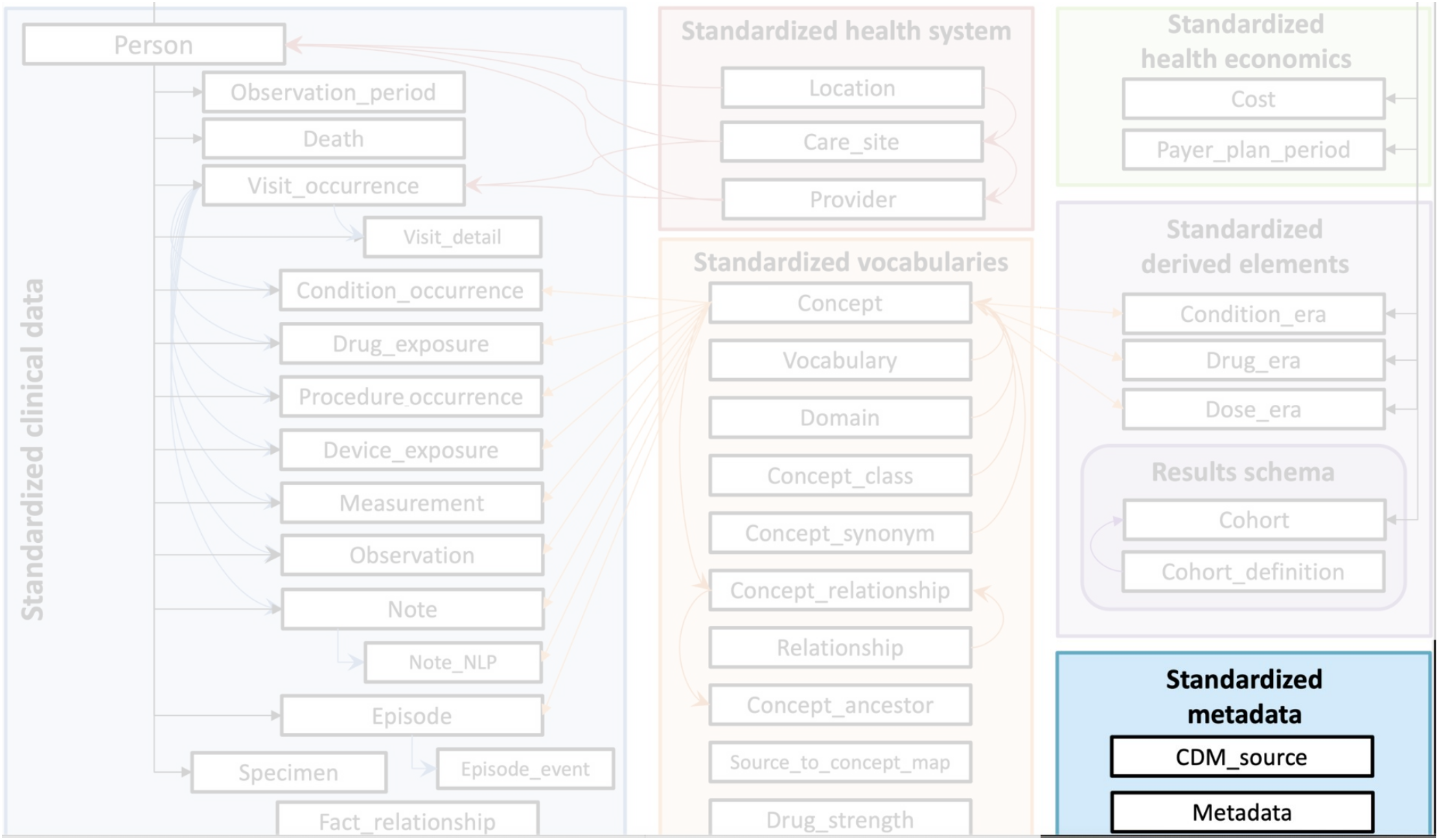
ETL Conventions

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

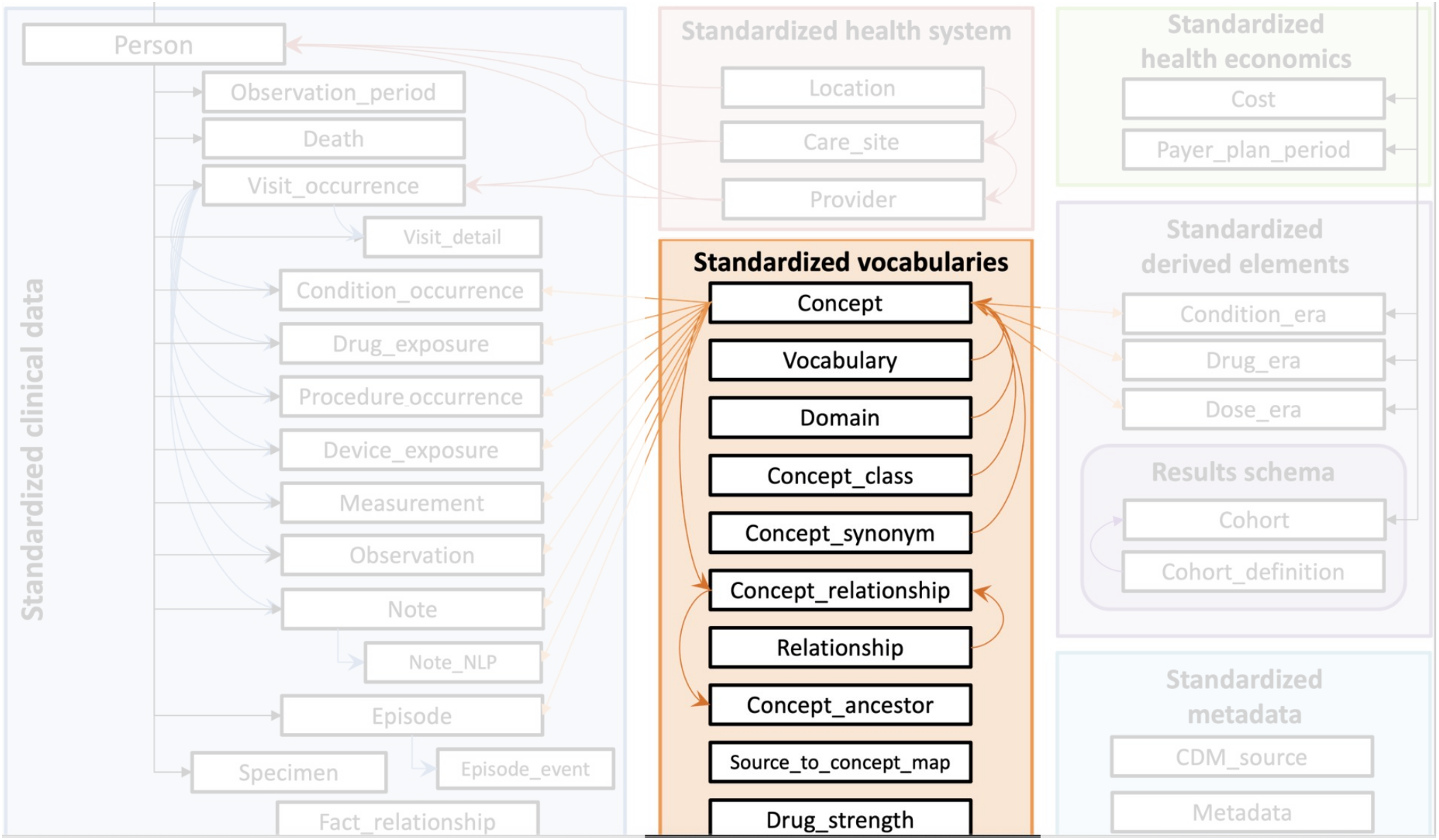

Standardized clinical data
Person
Observation_period
Death
Visit_occurrence
Visit_detail
Condition_occurrence
Drug_exposure
Procedure_occurrence
Device_exposure
Measurement
Observation
Note
Note_NLP
Episode
Specimen
Episode_event
Fact_relationship
Standardized health system
Location
Care_site
Provider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

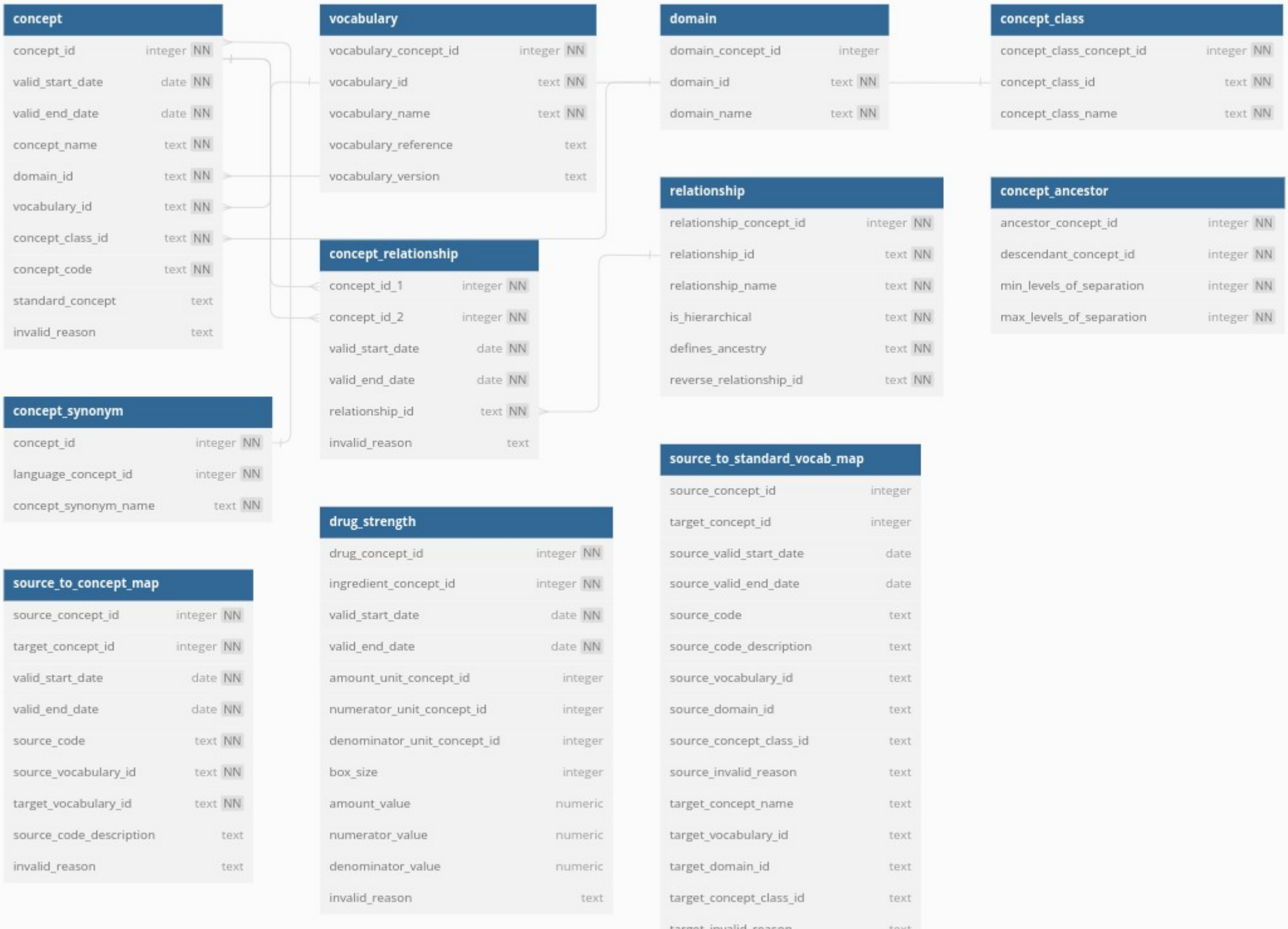
Información sobre esta instalación de OMOP-CDM

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


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	

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

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

- Proporciona una visión estandarizada de los conceptos clínicos en distintos dominios
- Reutiliza vocabularios existentes
- Se actualiza constantemente
- NO es cerrado (podemos añadir conceptos usando concept_id > 200000000)

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

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

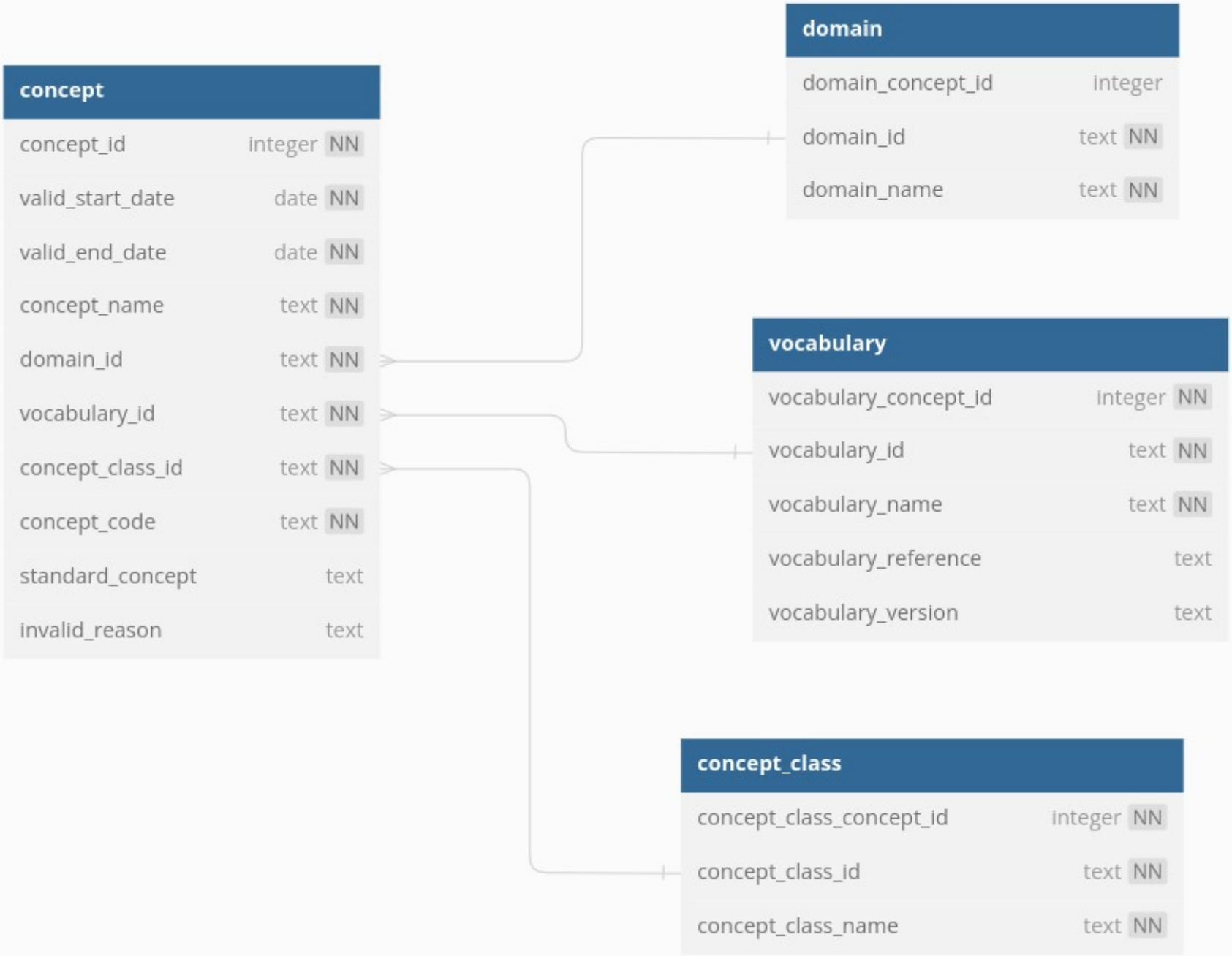
REFRESH | NOTIFICATIONS

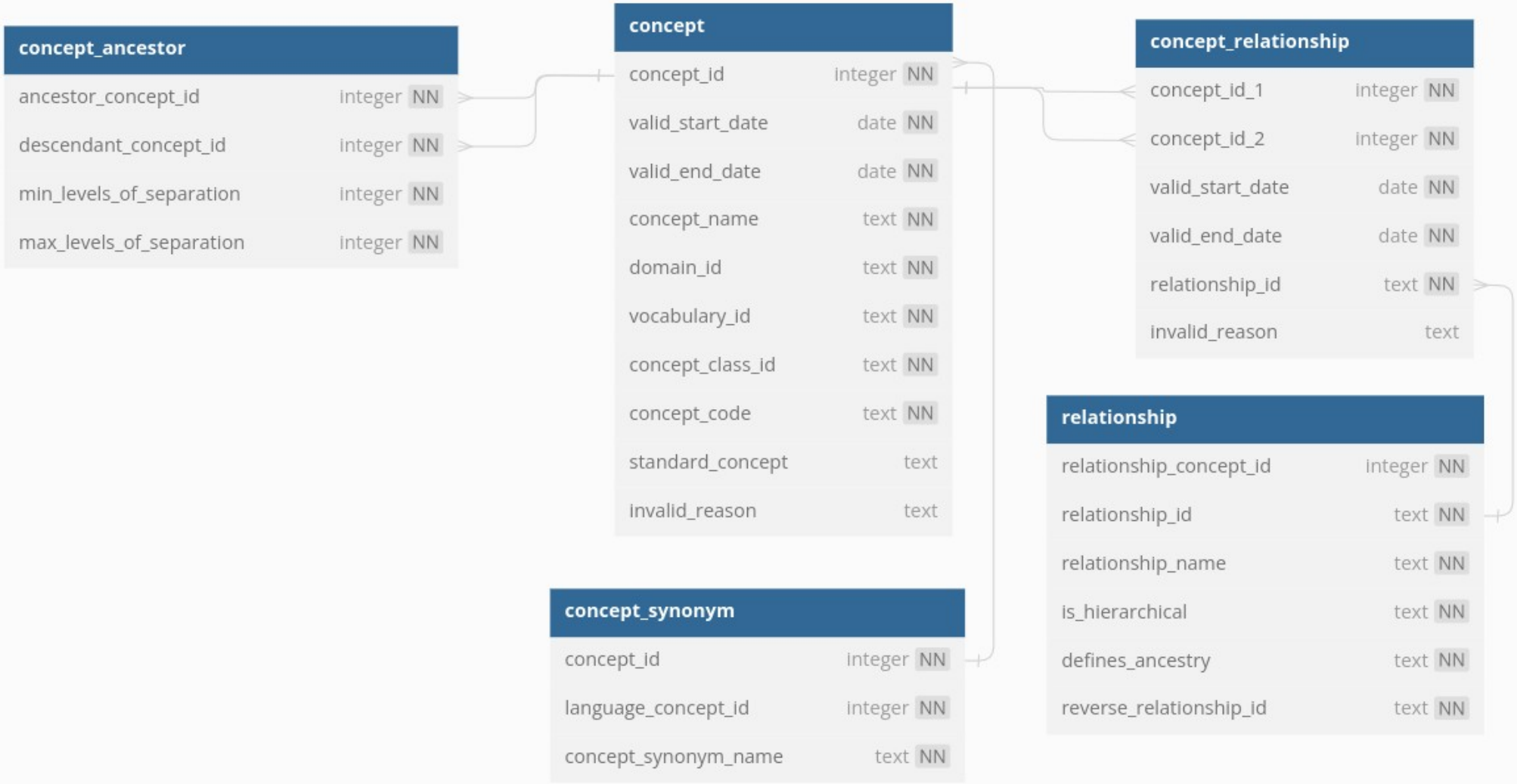
BSC_OMOP_VOCABULARIES

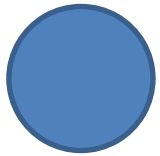
09:56 AM | 13-FEB-2023

V5.0 23-JAN-23

ARCHIVED | RESTORE

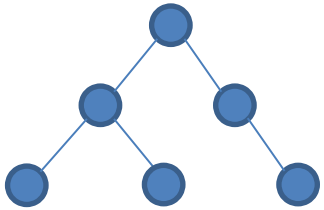

concept
concept_id integer NN
valid_start_date date NN
valid_end_date date NN
concept_name text NN
domain_id text NN
vocabulary_id text NN
concept_class_id text NN
concept_code text NN
standard_concept text
invalid_reason text
domain
domain_concept_id integer
domain_id text NN
domain_name text NN
vocabulary
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_ancestor
ancestor_concept_id integer NN
descendant_concept_id integer NN
min_levels_of_separation integer NN
max_levels_of_separation integer NN
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
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
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_synonym
concept_id integer NN
language_concept_id integer NN
concept_synonym_name text NN



conceptos médicos en **`concept`**

Relaciones directas entre conceptos **`concept_relationship`**

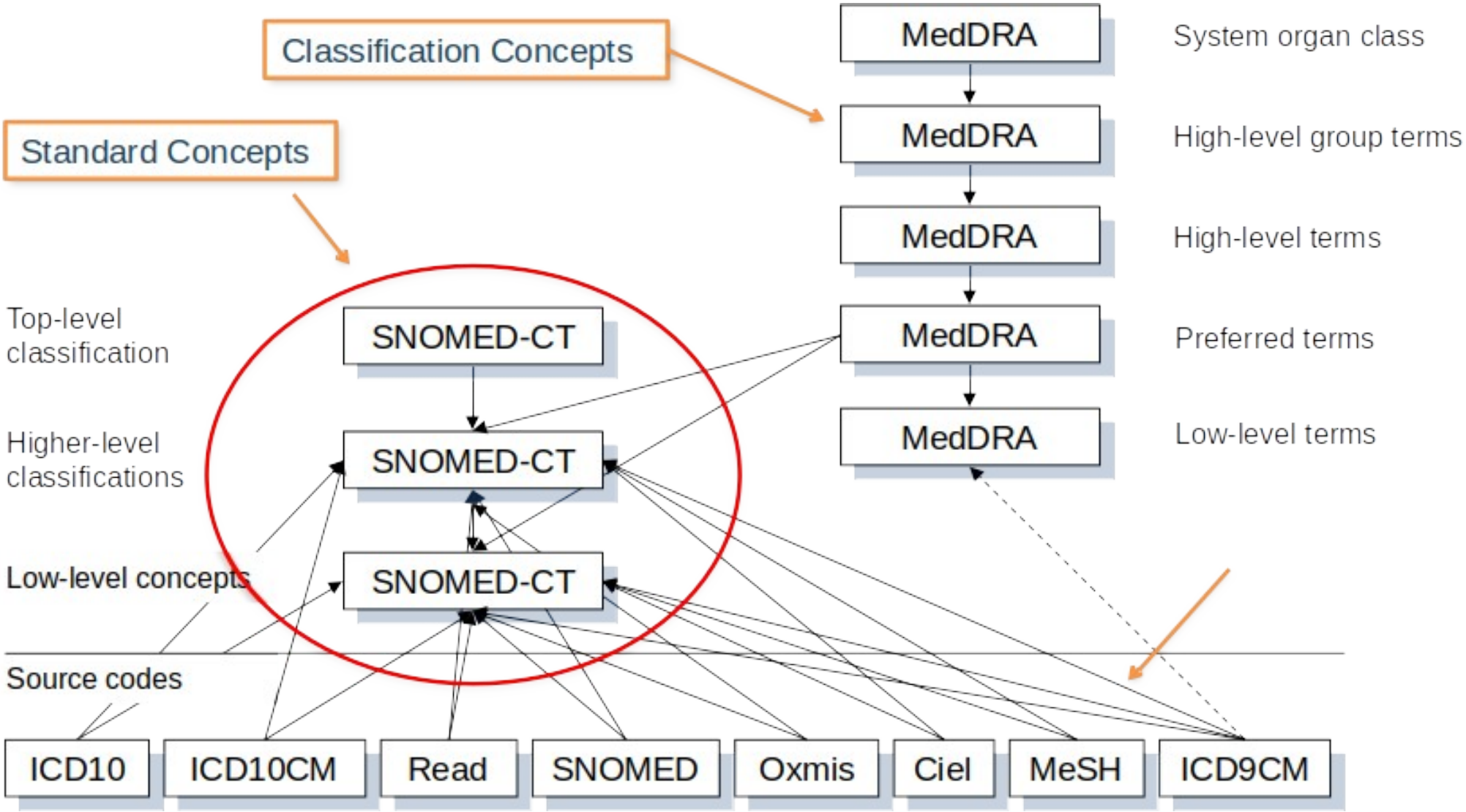


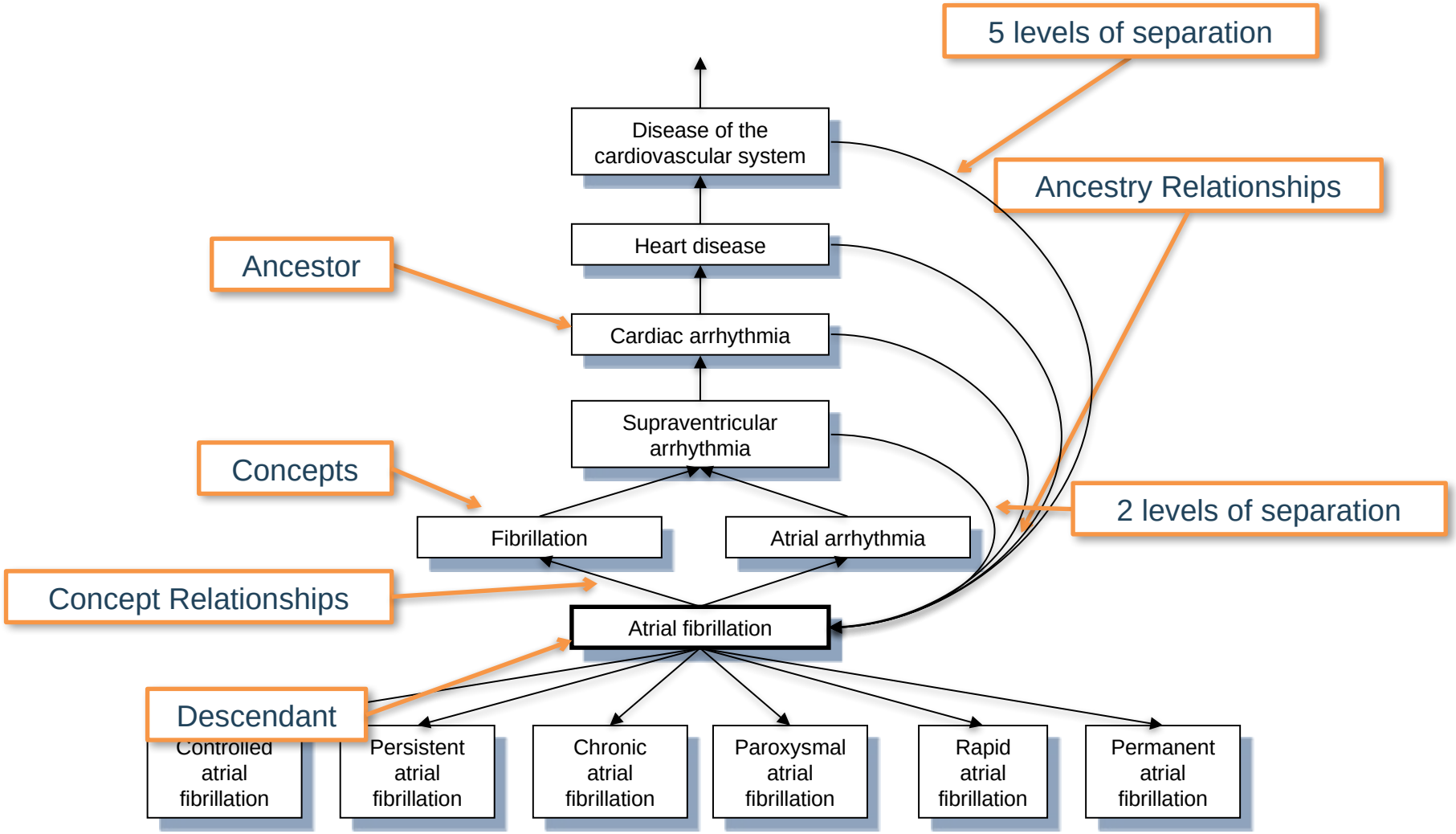
Relaciones jerárquicas **`concept_ancestor`**

concept_ancestor

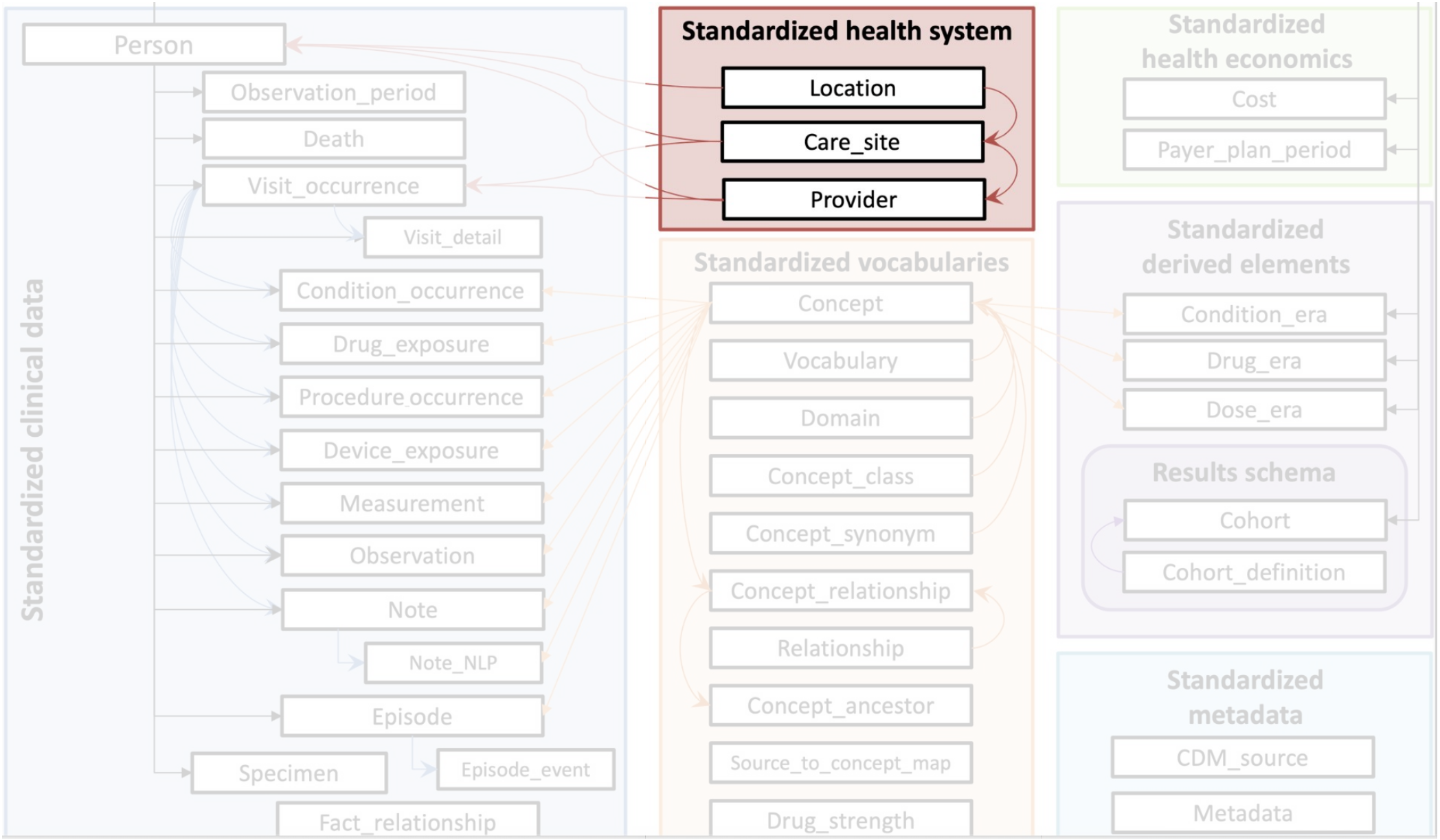
concept_ancestor		
ancestor_concept_id	integer	NN
descendant_concept_id	integer	NN
min_levels_of_separation	integer	NN
max_levels_of_separation	integer	NN

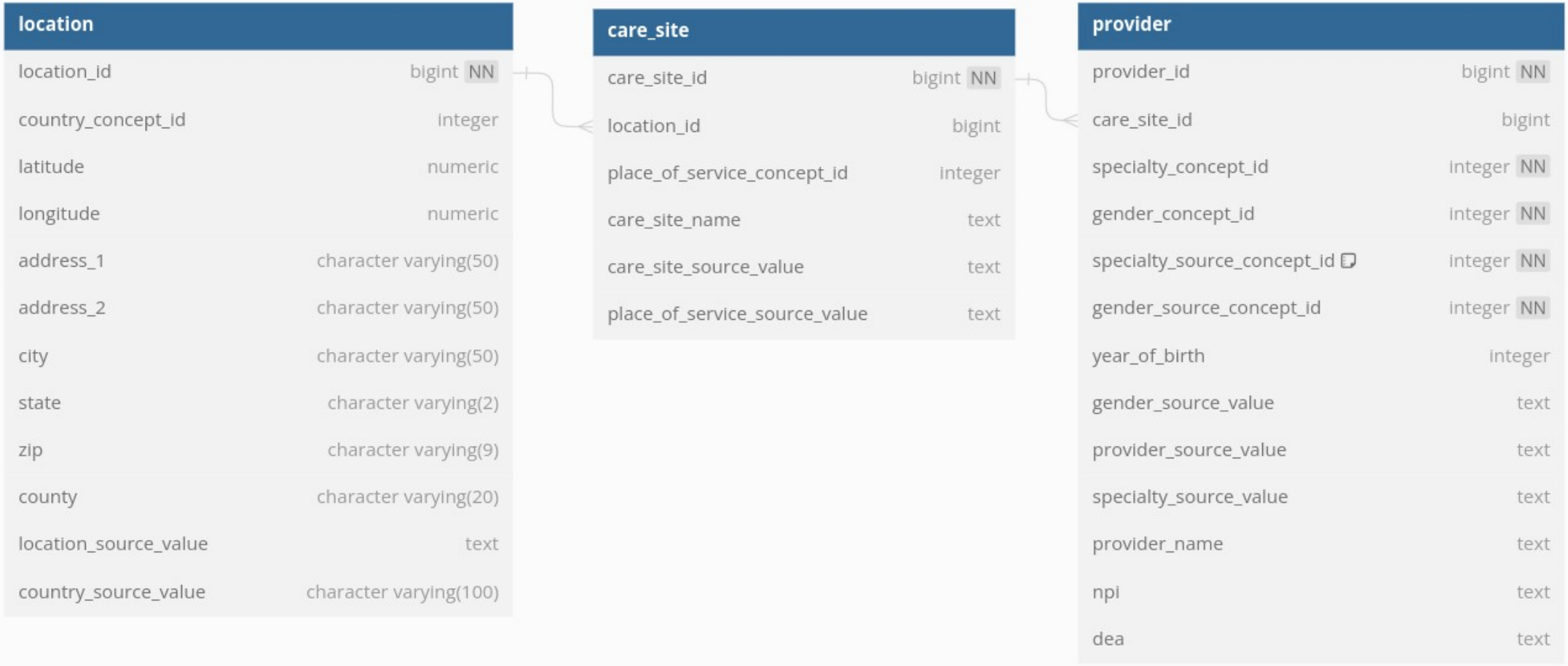
```
WITH RECURSIVE related_concepts AS (
    -- Base case
    SELECT concept_id_1, concept_id_2
    FROM concept_relationship
    WHERE concept_relatinoship_id = 'is_a'

    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;
```

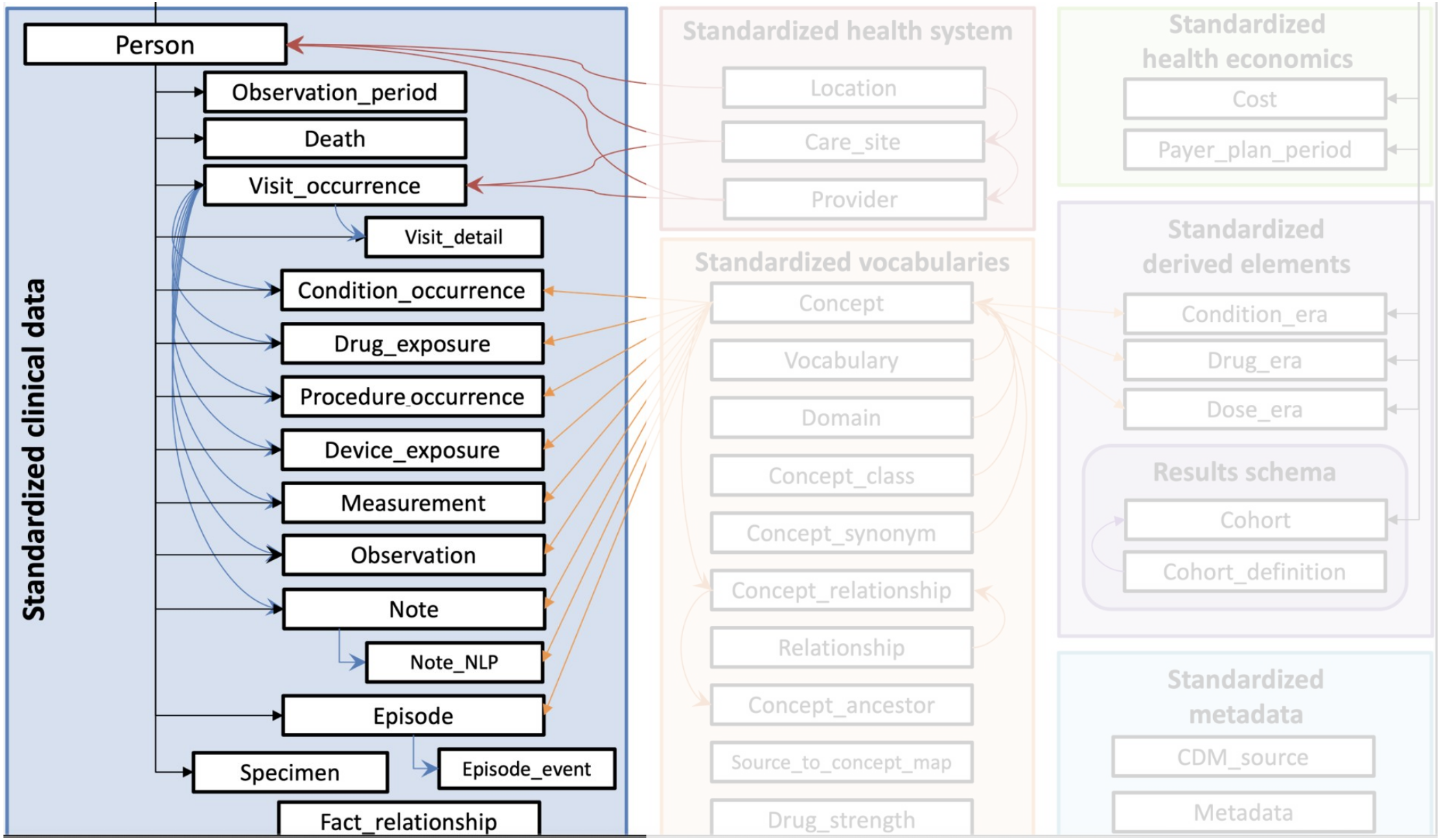

Classification Concepts
Standard Concepts
MedDRA
System organ class
MedDRA
High-level group terms
MedDRA
High-level terms
MedDRA
Preferred terms
MedDRA
Low-level terms
Top-level classification
SNOMED-CT
Higher-level classifications
SNOMED-CT
Low-level concepts
SNOMED-CT
Source codes
ICD10
ICD10CM
Read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


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

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


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

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

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	

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

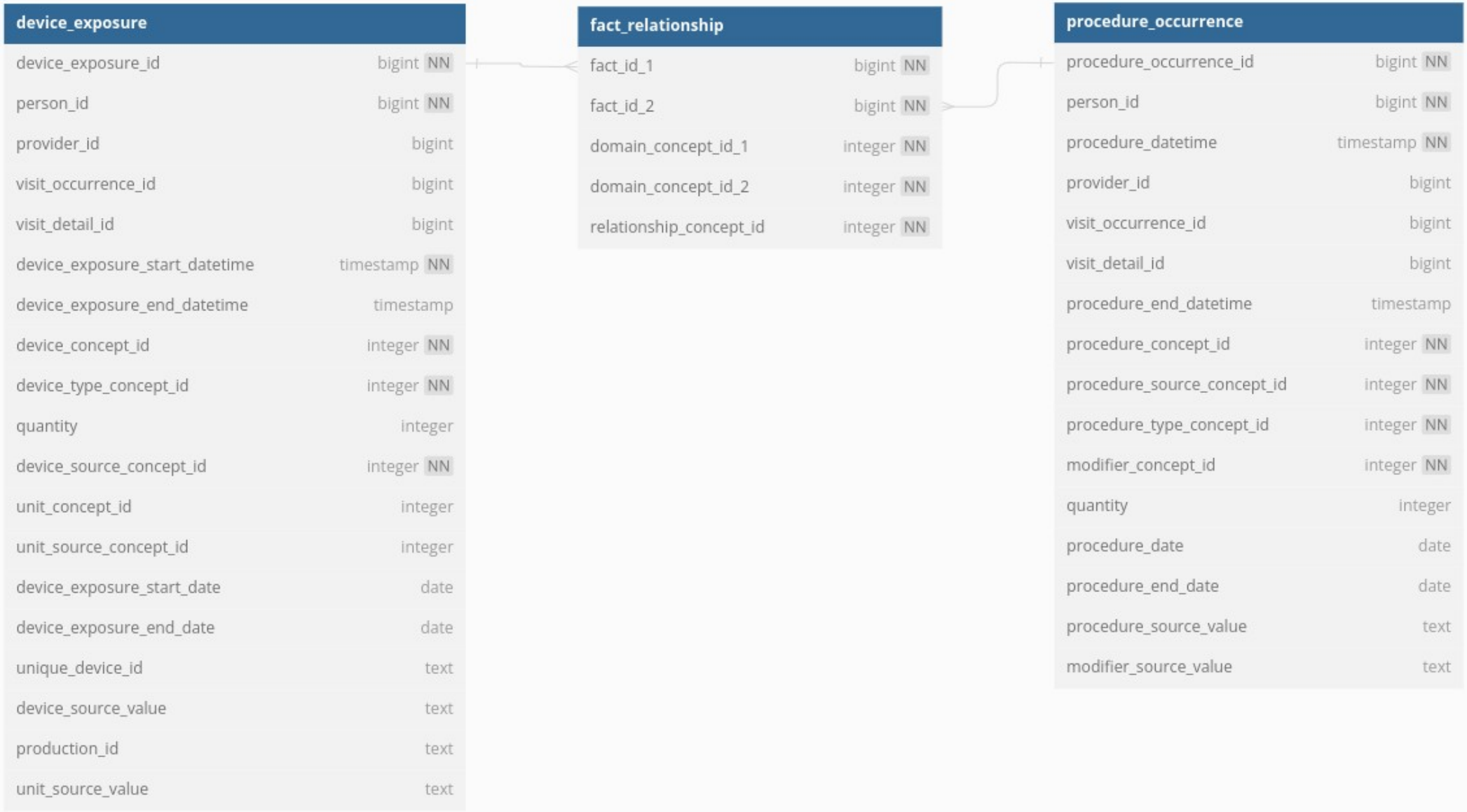
condition_era	
condition_era_id	INTEGER
person_id	INTEGER
condition_concept_id	INTEGER
condition_era_start_date	DATE
condition_era_end_date	DATE
condition_occurrence_count	INTEGER

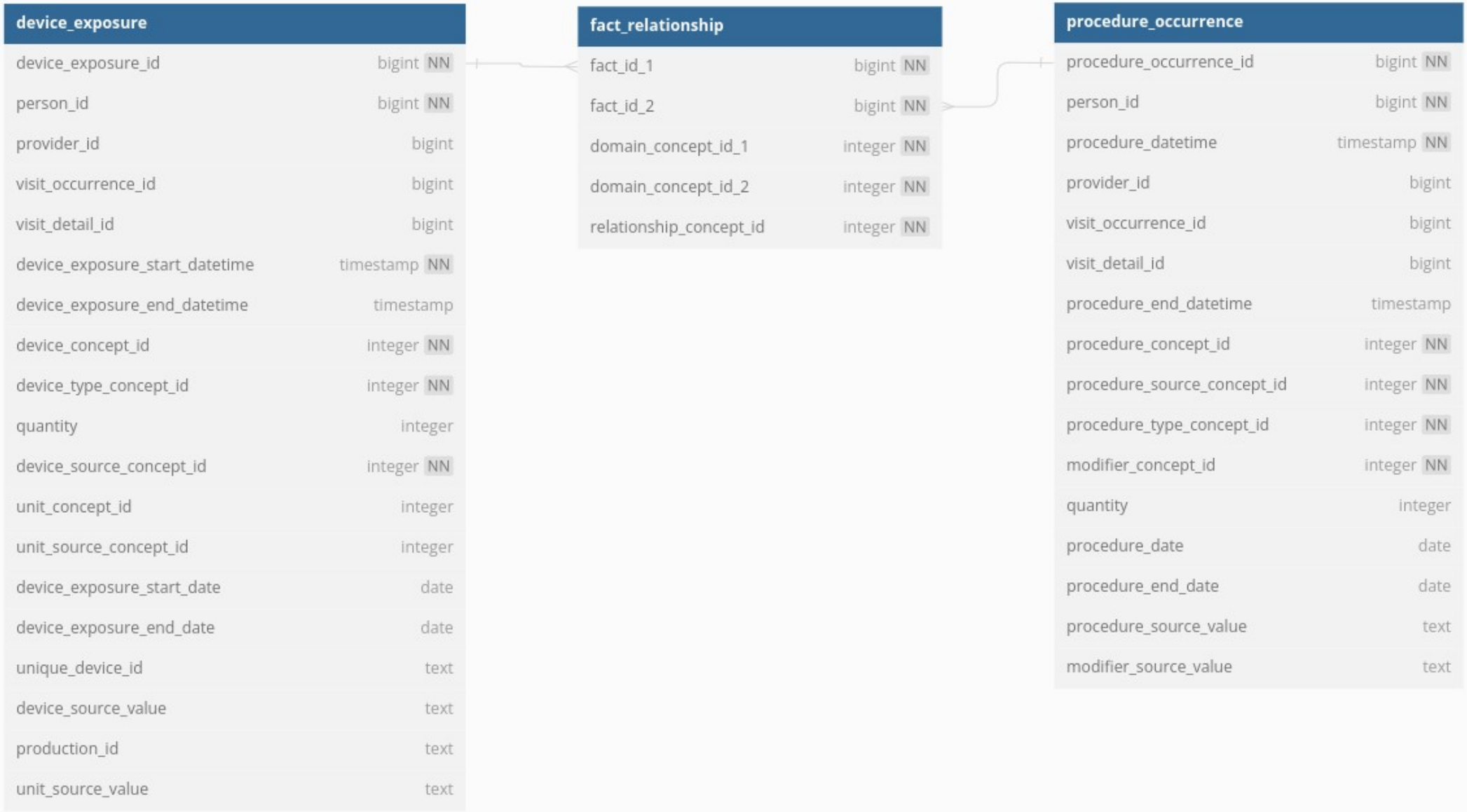
observation_period		
observation_period_id	**INTEGER**	NN
person_id	INTEGER	NN
observation_period_start_date	DATE	NN
observation_period_end_date	DATE	NN
period_type_concept_id	INTEGER	

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

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

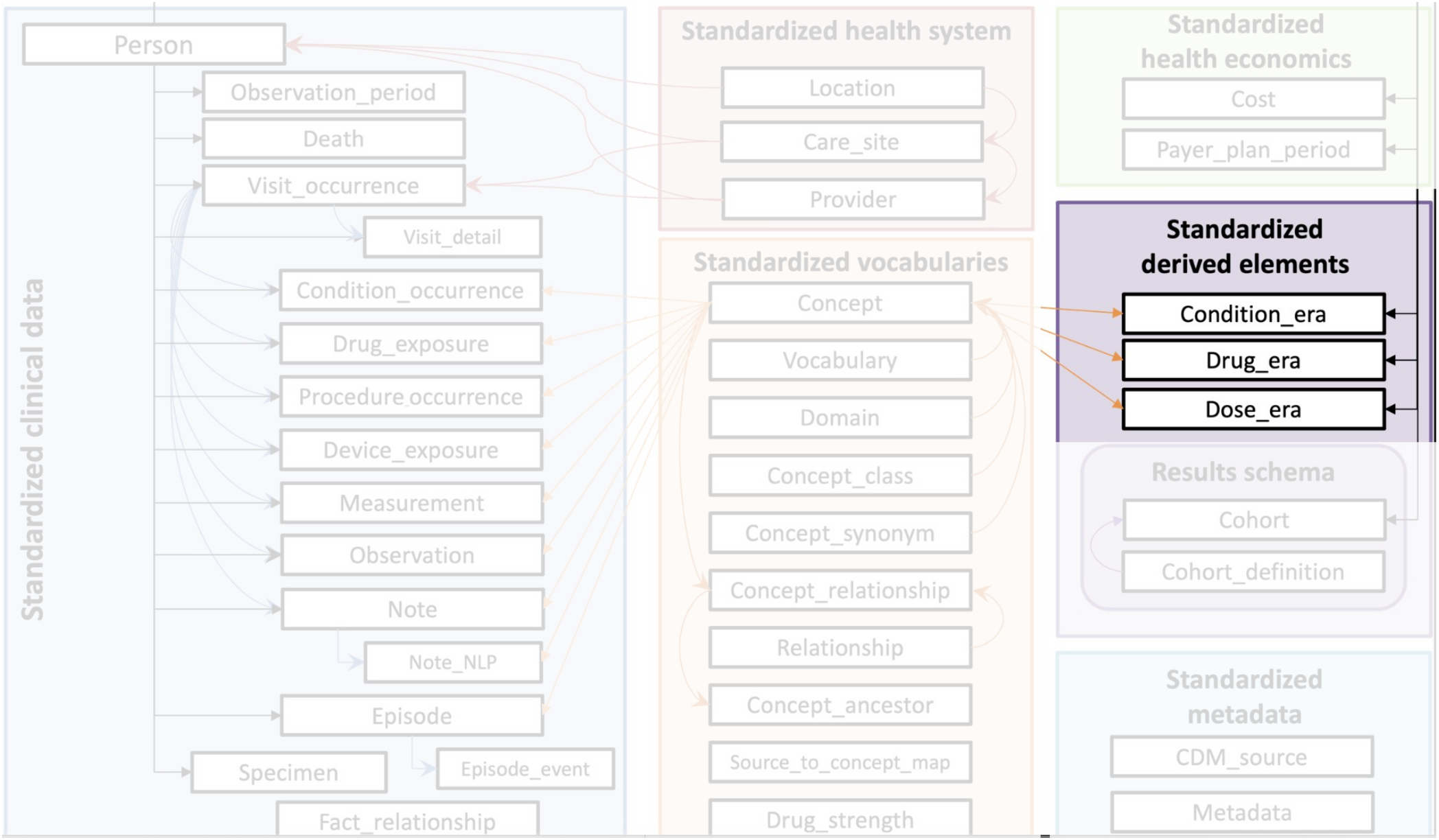
device_exposure	
device_exposure_id	bigint NN
person_id	bigint NN
provider_id	bigint
visit_occurrence_id	bigint
visit_detail_id	bigint
device_exposure_start_datetime	timestamp NN
device_exposure_end_datetime	timestamp
device_concept_id	integer NN
device_type_concept_id	integer NN
quantity	integer
device_source_concept_id	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


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



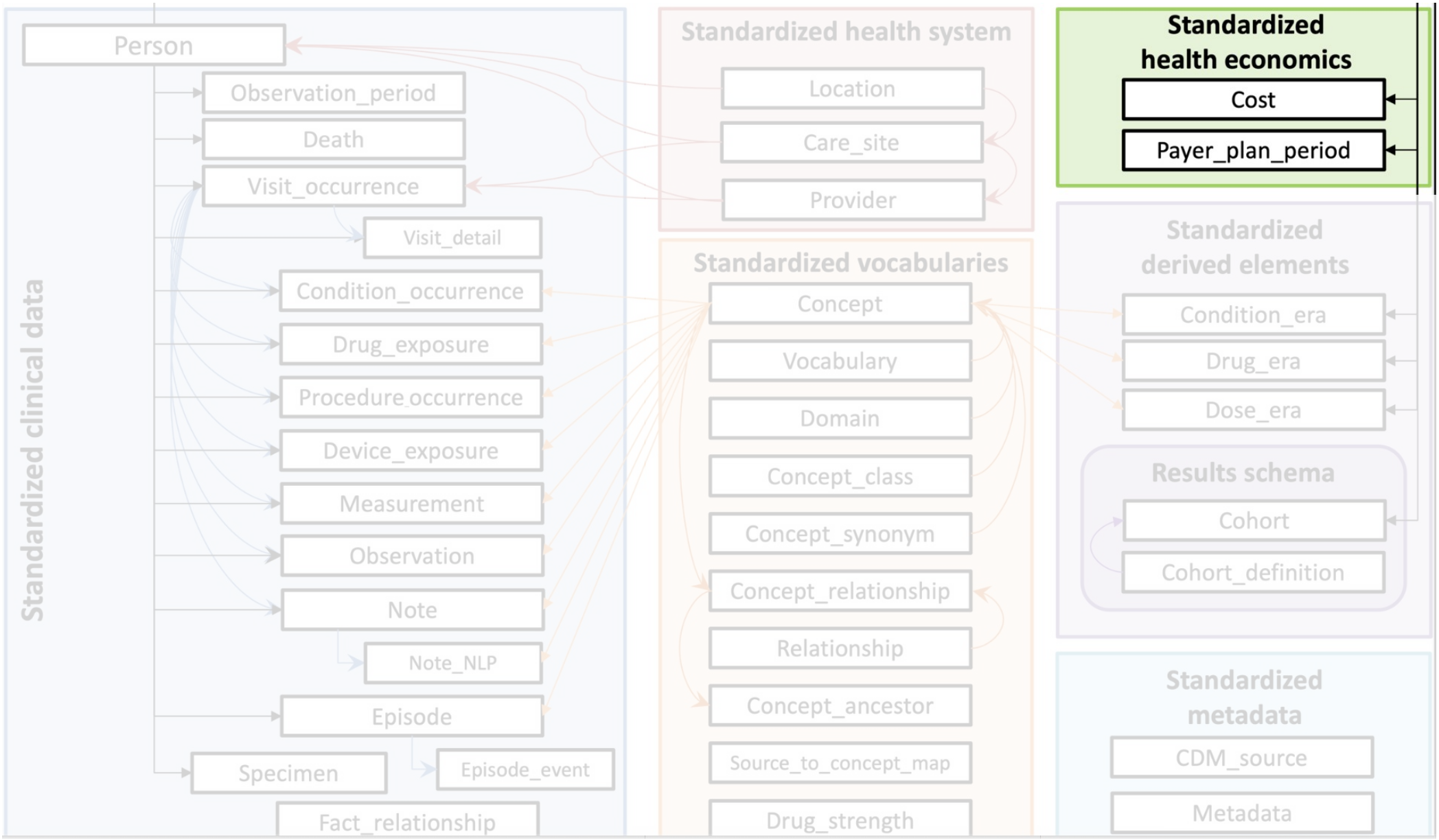
44818892	OMOP generated	Device used by (SNOMED)	Relationship	Non-standard	Valid	Metadata	Relationship

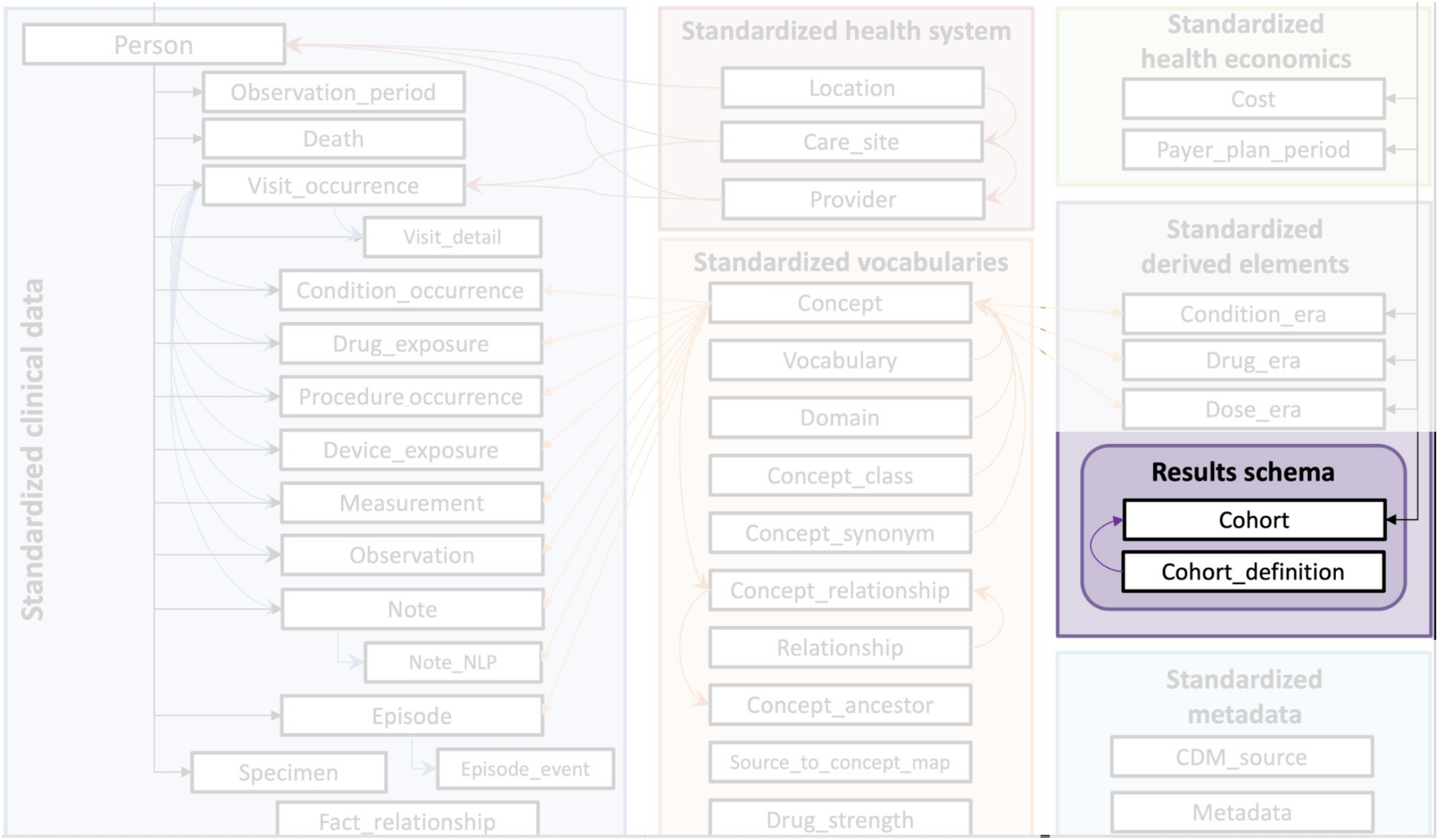
drug_exposure	
drug_exposure_id	bigint NN
person_id	bigint NN
provider_id	bigint
visit_occurrence_id	bigint
visit_detail_id	bigint
drug_exposure_start_datetime	timestamp
drug_exposure_end_datetime	timestamp
drug_concept_id	integer NN
drug_type_concept_id	integer NN
drug_source_concept_id	integer NN
route_concept_id	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


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

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


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


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

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

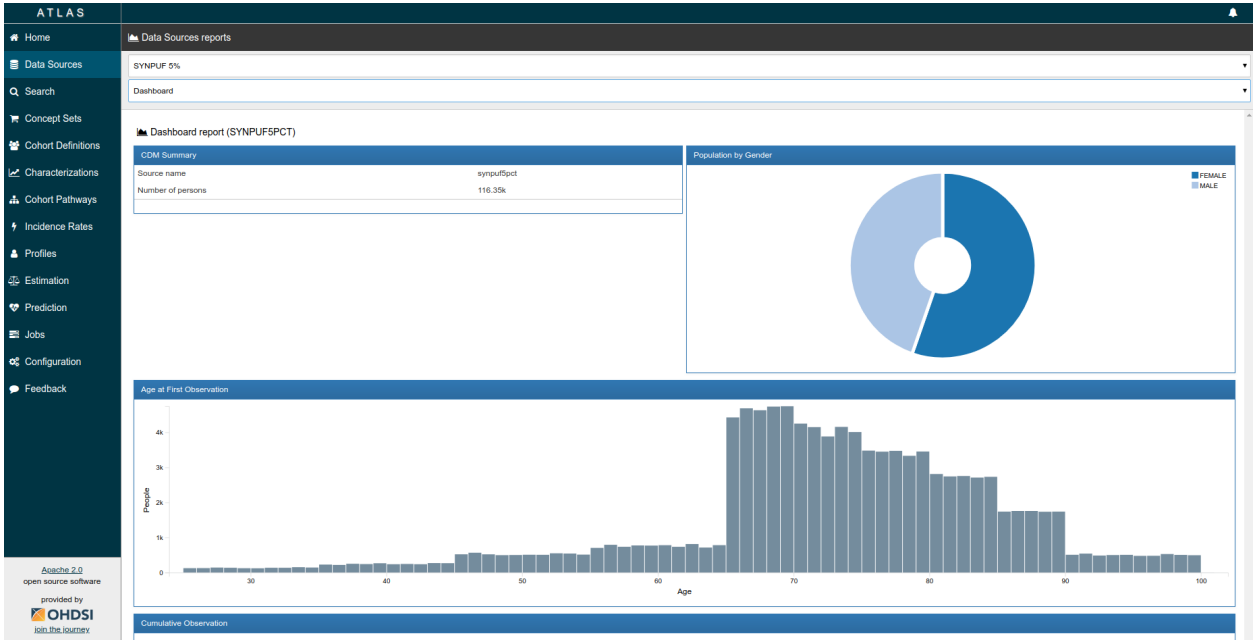
- .github
- R
- docs
- extras
- inst
  - csv
  - doc
  - shinyApps
  - sql/sql_server
    - sqlOnly
    - concept_plausible_gender.sql
    - concept_plausible_unit_concept…
    - concept_plausible_value_high.sql
    - concept_plausible_value_low.sql
    - field_cdm_datatype.sql
    - field_cdm_field.sql

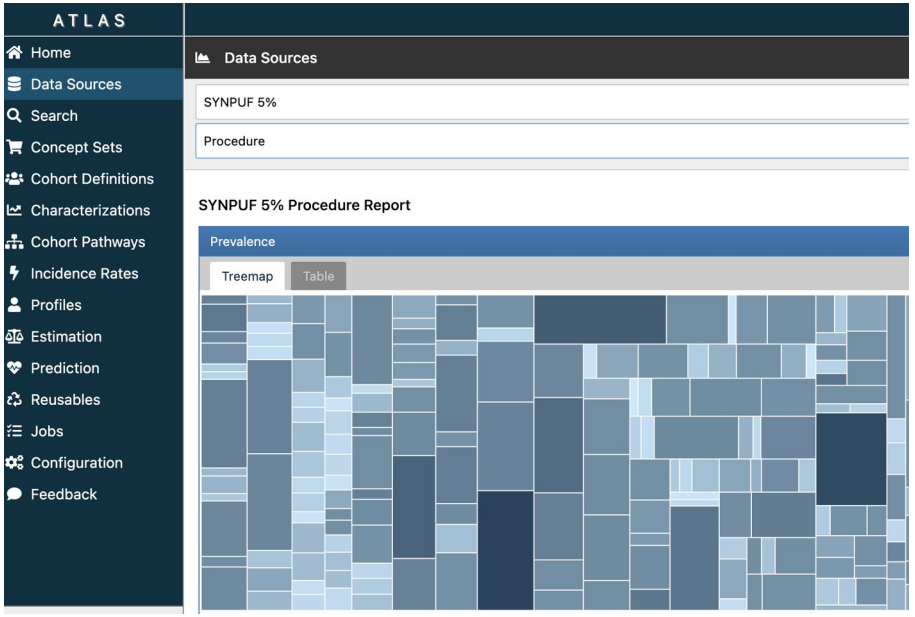
DataQualityDashboard / inst / sql / sql_server / **concept_plausible_gender.sql**

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()

What about genomic data?

1620918	LA31701-8	SARS-CoV-2 A.1.1 lineage	Answer	Standard	Valid	Meas Value	LOINC
1620971	LA31704-2	SARS-CoV-2 B.1.1 lineage	Answer	Standard	Valid	Meas Value	LOINC
1620719	LA31620-0	SARS-CoV-2 B.1.1.238 lineage	Answer	Standard	Valid	Meas Value	LOINC
1620866	LA31705-9	SARS-CoV-2 B.1.1.7 lineage	Answer	Standard	Valid	Meas Value	LOINC
1620654	LA31703-4	SARS-CoV-2 B.1.2 lineage	Answer	Standard	Valid	Meas Value	LOINC
1620888	LA31706-7	SARS-CoV-2 B.1.29 lineage	Answer	Standard	Valid	Meas Value	LOINC
1621088	LA32553-2	SARS-CoV-2 B.1.351 lineage	Answer	Standard	Valid	Meas Value	LOINC
1620416	LA32556-5	SARS-CoV-2 B.1.617.2 lineage	Answer	Standard	Valid	Meas Value	LOINC

What about genomic data?

36767827	25389	F8, 1-BP DEL	Variant	Non-standard	Valid	Measurement	ClinVar
36765625	24749	m.11696G>A	Variant	Non-standard	Valid	Measurement	ClinVar
36765670	24750	m.11777C>A	Variant	Non-standard	Valid	Measurement	ClinVar
36766737	24649	m.12147G>A	Variant	Non-standard	Valid	Measurement	ClinVar
36767268	24648	m.12183G>A	Variant	Non-standard	Valid	Measurement	ClinVar
36767617	24734	m.14319T>C	Variant	Non-standard	Valid	Measurement	ClinVar
36767433	24732	m.14482C>A	Variant	Non-standard	Valid	Measurement	ClinVar
36767752	24639	m.3242G>A	Variant	Non-standard	Valid	Measurement	ClinVar

What about genomic data?

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

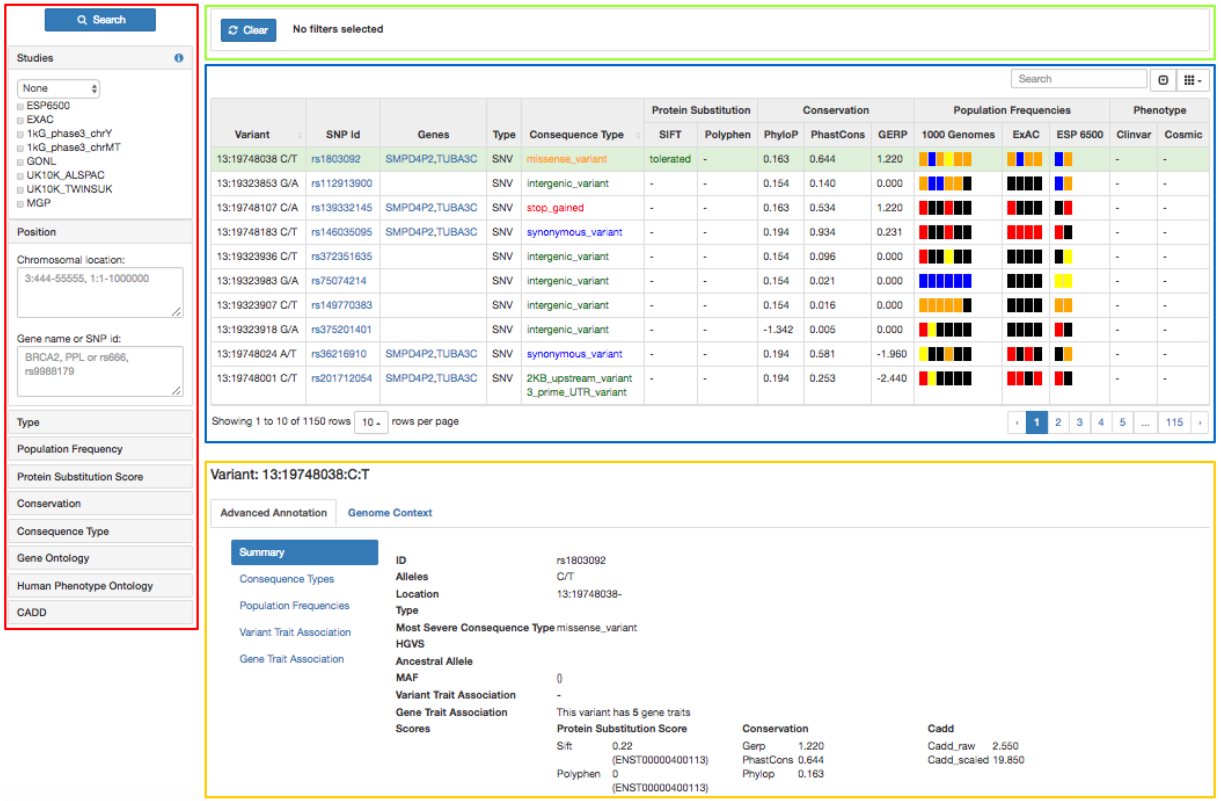
Measurement table

ID	CODE	NAME	CLASS	CONCEPT	VALIDITY	DOMAIN	VOCAB
40776139	LP19695-3	EGFR gene	LOINC Hierarchy	Classification	Valid	Measurement	LOINC
40758886	55765-2	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

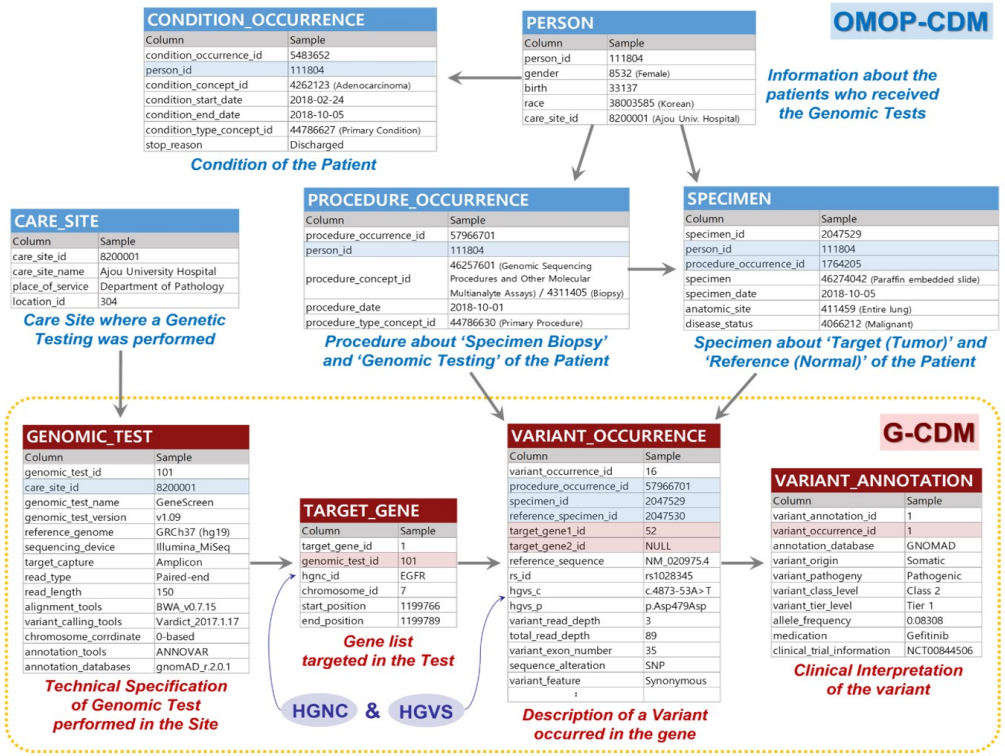
Observation table

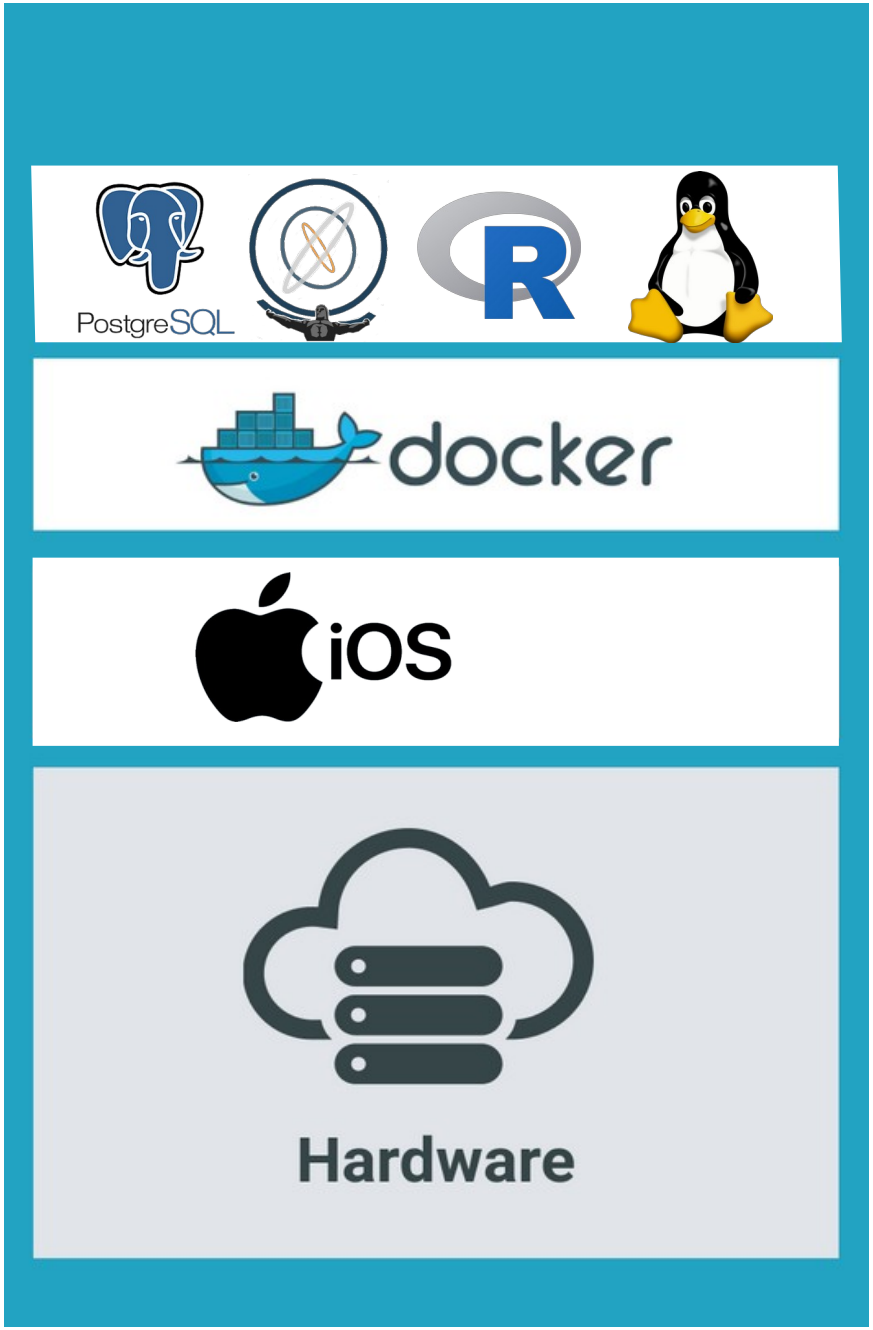
ID	CODE	NAME	CLASS	CONCEPT	VALIDITY	DOMAIN	VOCAB
40775288	LP95041-7	EGFR gene c.2573T>G	LOINC Hierarchy	Classification	Valid	Observation	LOINC
40785298	LP95040-9	EGFR gene c.2156G>C+2155G>A+2155G>T	LOINC Hierarchy	Classification	Valid	Observation	LOINC
40785299	LP95045-8	EGFR gene exon 20 insertion	LOINC Hierarchy	Classification	Valid	Observation	LOINC
40788607	LP95063-1	EGFR gene exon 19	LOINC Hierarchy	Classification	Valid	Observation	LOINC
40791895	LP95042-5	EGFR gene c.2582T>A	LOINC Hierarchy	Classification	Valid	Observation	LOINC

What about genomic data?

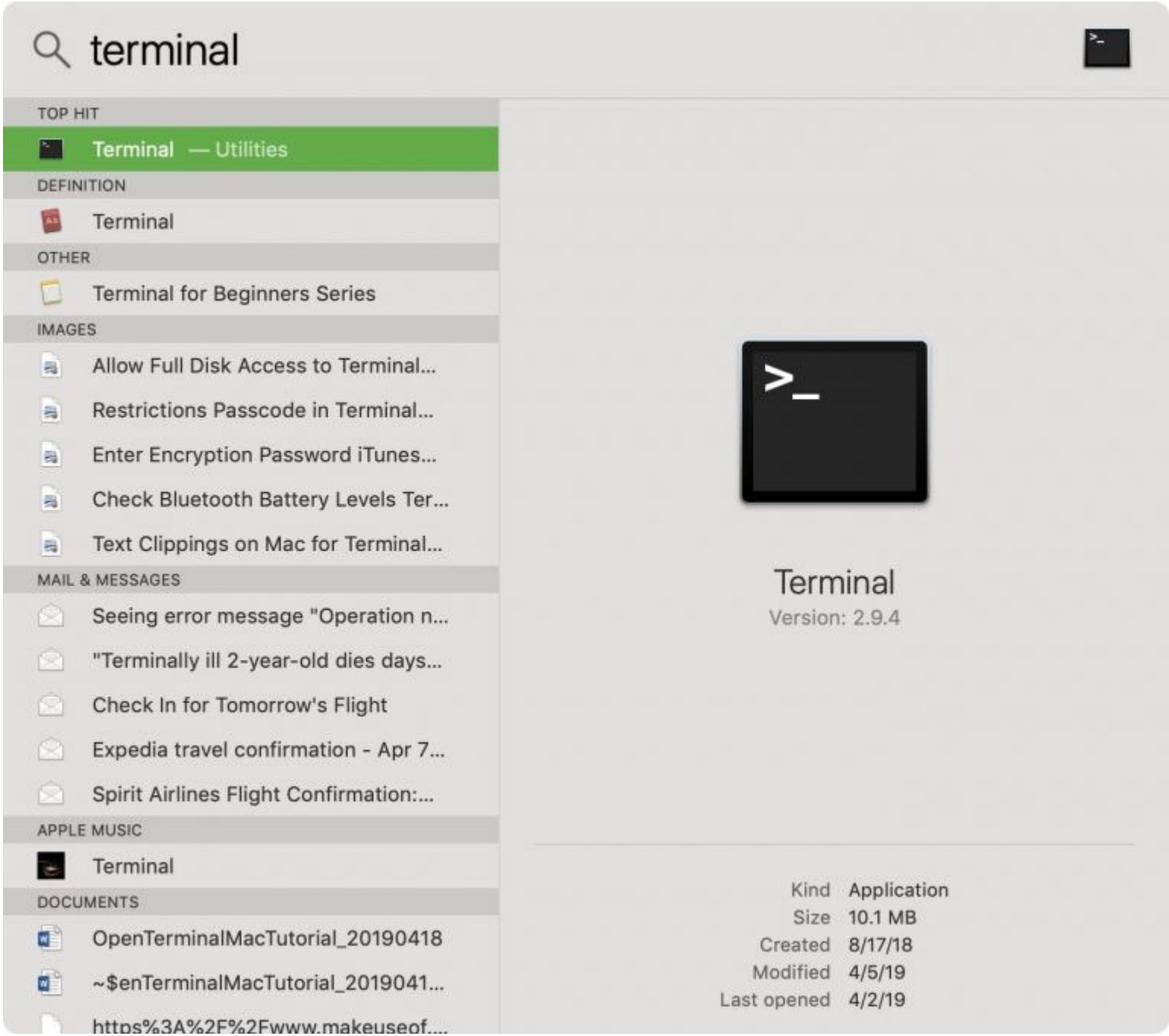


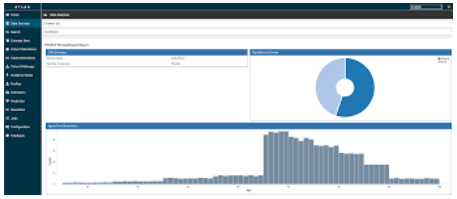
What about genomic data?




PostgreSQL
R
docker
iOS
Hardware

Spotlight Search button in your menu bar, click it. Otherwise, you can use the keyboard shortcut Command + Space.





HADES

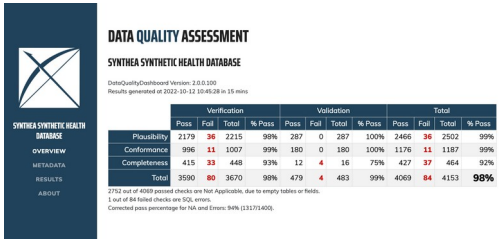
HADES

HEALTH ANALYTICS DATA-TO-EVIDENCE SUITE

Web API

Apache Tomcat

PostgreSQL



SYNTHEA SYNTHETIC HEALTH DATABASE

OVERVIEW

METADATA

RESULTS

ABOUT

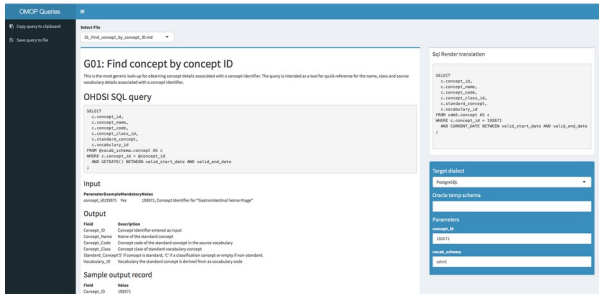
DATA QUALITY ASSESSMENT

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

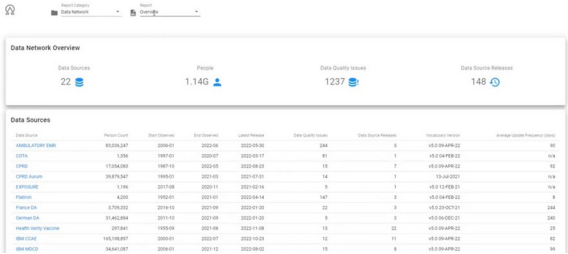
Sample output record

Sql Render translation

Target dialect

Oracle temp schema

Parameters



Data Network Overview

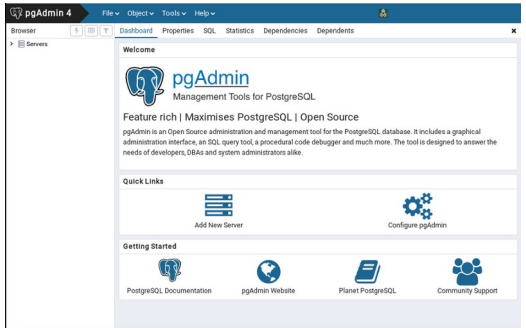
Data Sources 22

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

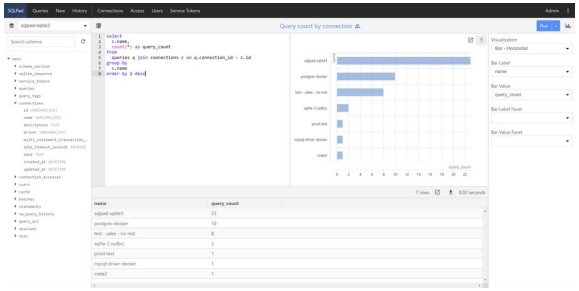
Getting Started

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

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

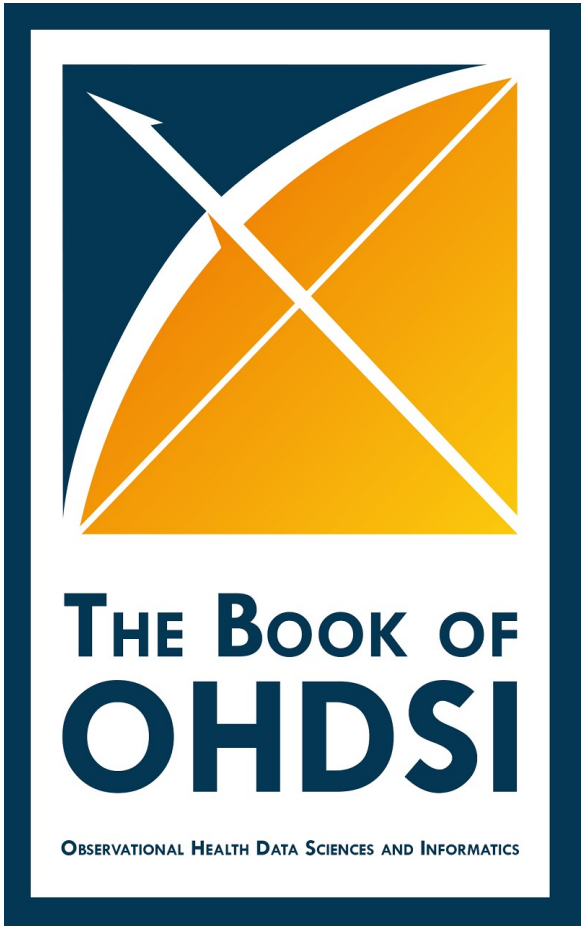
Assessments LEARN MORE

Applied Examples LEARN MORE

Maturity model LEARN MORE

Recursos

- Sitio web: https://github.com/alabarga/omop-cdm-course
- Diapositivas
- El Libro de OHDSI
- EHDEN academy
- OHDSI workshops
- Atlas tutorials
- Entorno local (docker)



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.