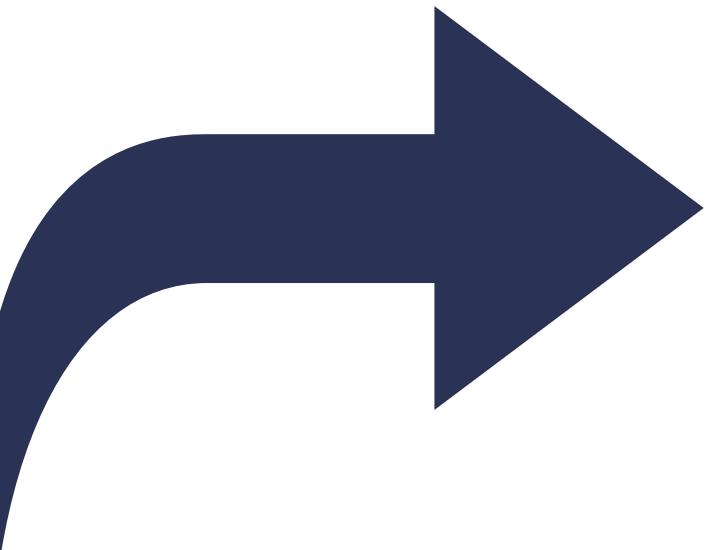




Introducción al modelo OMOP-CDM

Alberto Labarga – Febrero 2026



I am Alberto

 alabarga

 alabarga

 /in/albertolabarga

 alberto.labarga@gmail.com

Febrero
2026 

OMOP-CDM

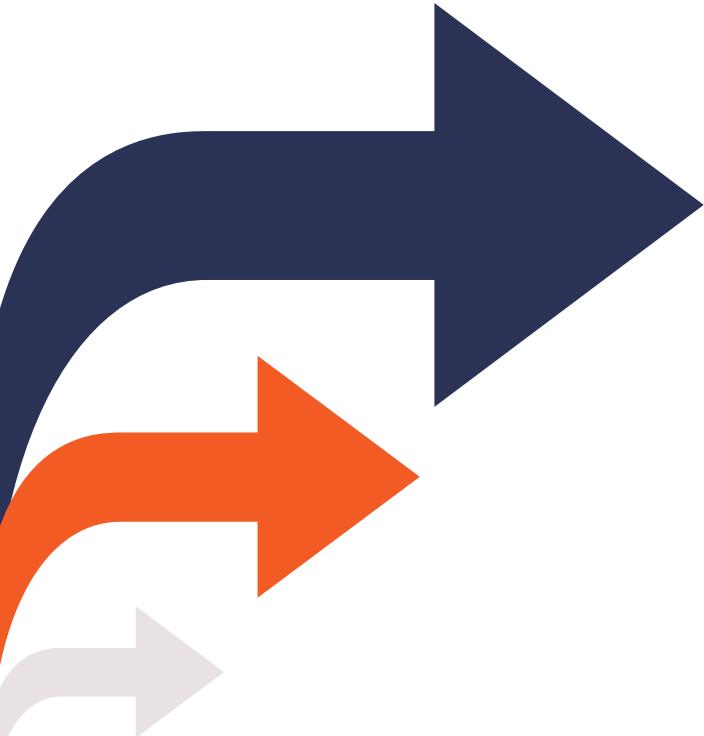
El curso

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

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

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

Objetivos del curso



01

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

02

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

03

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

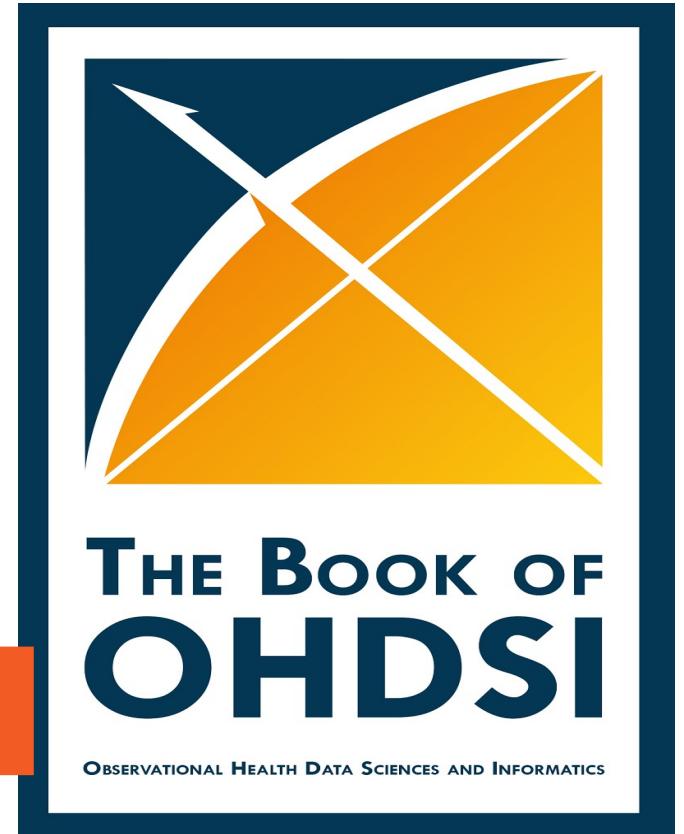
Recursos de aprendizaje



Sitio web: <https://github.com/alabarga/curso-omop-cdm>

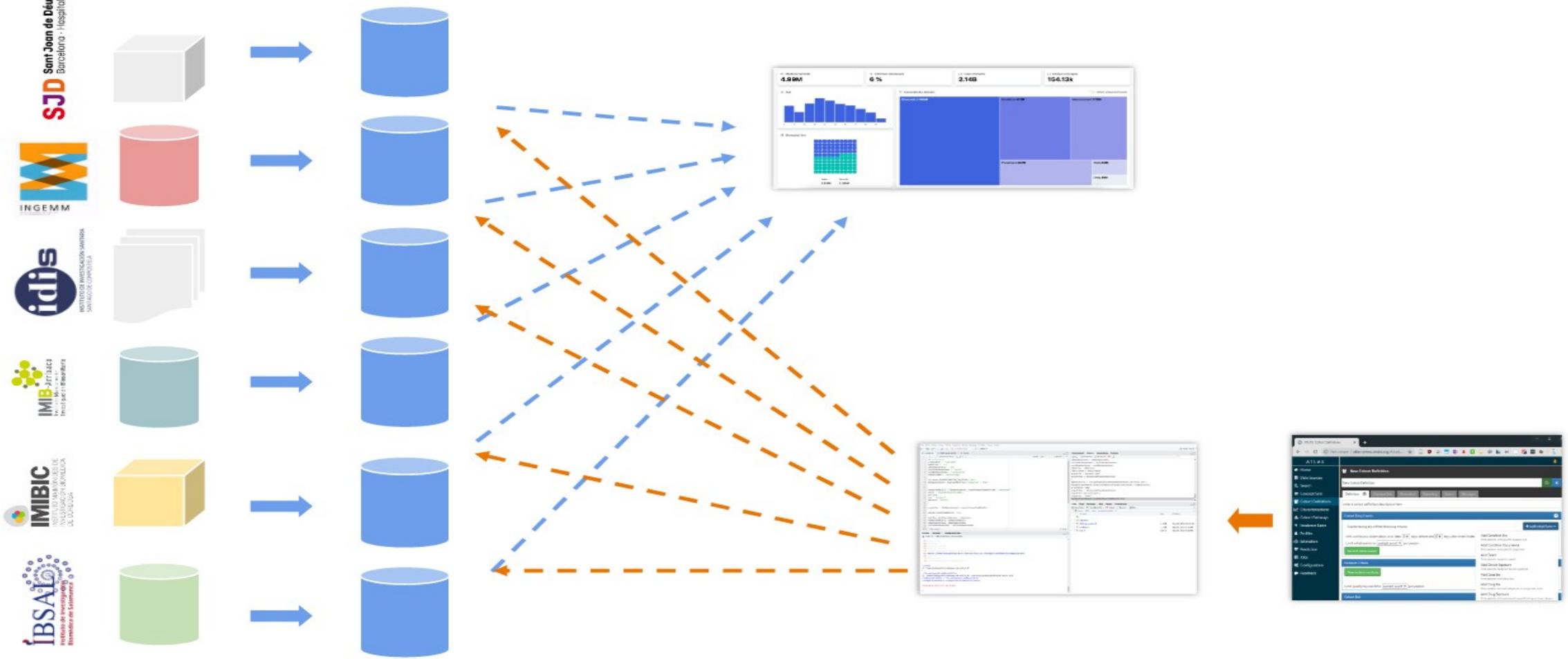


Libro <https://ohdsi.github.io/TheBookOfOhdsi/>





The logo consists of the text "OMOP-CDM" in a large, white, sans-serif font. A small, five-pointed white star is positioned above the letter "M". Below "OMOP-CDM" is the word "VOCABULARIOS" in a smaller, white, sans-serif font.

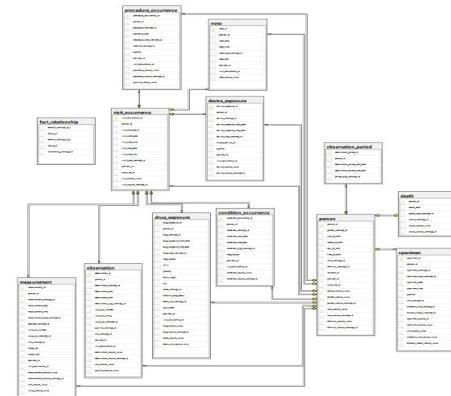


OMOP-CDM

Common Data Model

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

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

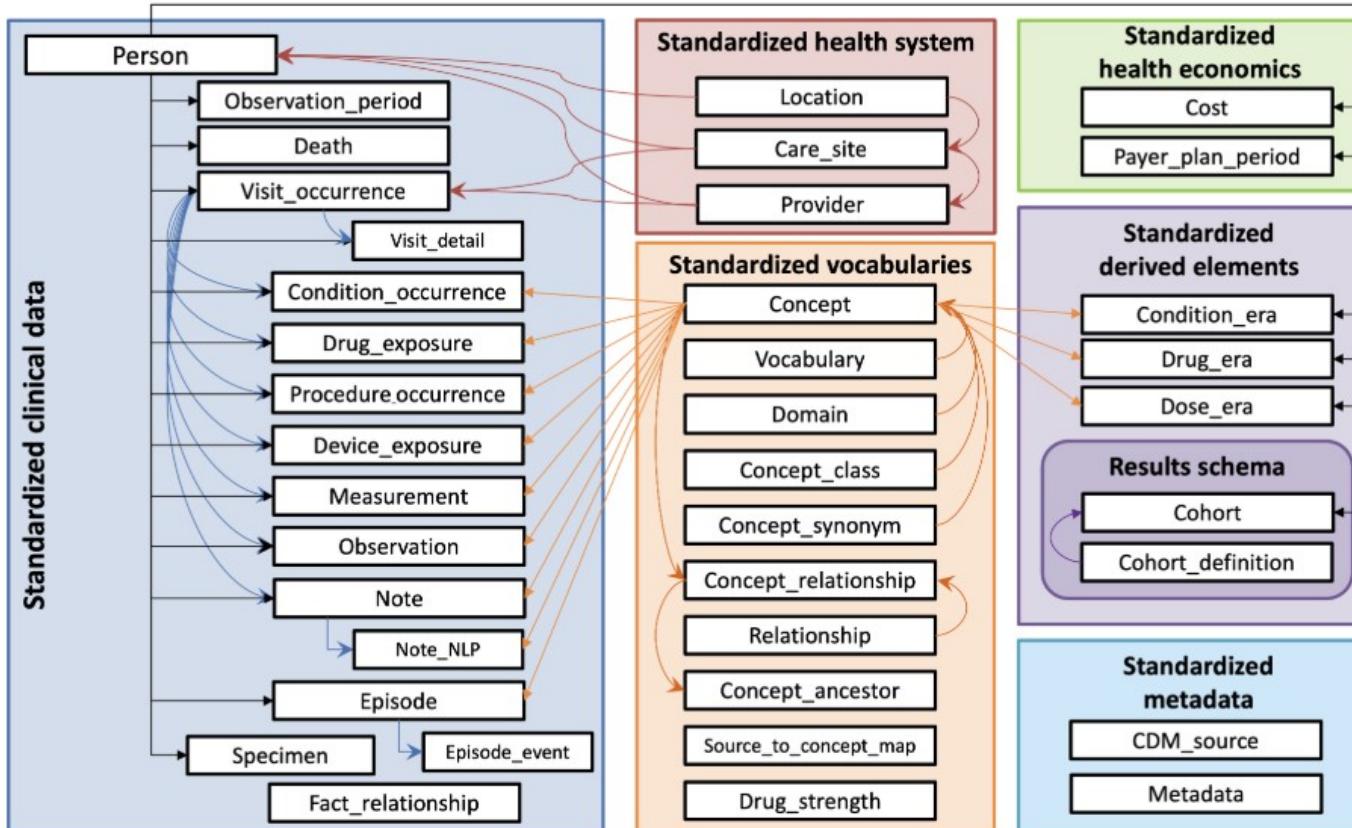


A screenshot of the ATHENA platform interface. The top navigation bar includes 'SEARCH', 'DOWNLOAD', and 'Album Lateral'. The search bar contains the word 'cancer'. The main area is titled 'DOWNLOAD RESULTS' and shows a list of 8,007 items. The columns in the table are ID, CODE, NAME, CLASS, CONCEPT, VALIDITY, DOMAIN, and VOCAB. The first few rows listed are:

ID	CODE	NAME	CLASS	CONCEPT	VALIDITY	DOMAIN	VOCAB
4587725	LA10524-9	Cancer	Answer	Standard	Valid	Measure Value	LINC
419405	350000008	Cancer confirmed	Control dependent	Standard	Valid	Condition	SNOMED
429508	386000101	Cancer registrar	Social Context	Standard	Valid	Observation	SNOMED
1384553	240307206	Cancer Conditions: No Cancer	Answer	Standard	Valid	Observation	PRI
4073624	240307206	Cancer education	Procedure	Standard	Valid	Procedure	SNOMED
4480309	T027310000001	Colorectal cancer	Qualifier Value	Standard	Valid	Observation	SNOMED
4075763	54532-7	Cancer	Survey	Standard	Valid	Observation	LINC

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

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



CommonDataModel / inst / ddl / 5.4 /

 clairblacketer Removing codespell and older dll folder for CRAN submission

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



Show all

SHOW HISTORY

DOWNLOAD VOCABULARIES

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

TRANSFORMACIÓN AL MODELO OMOP-CDM

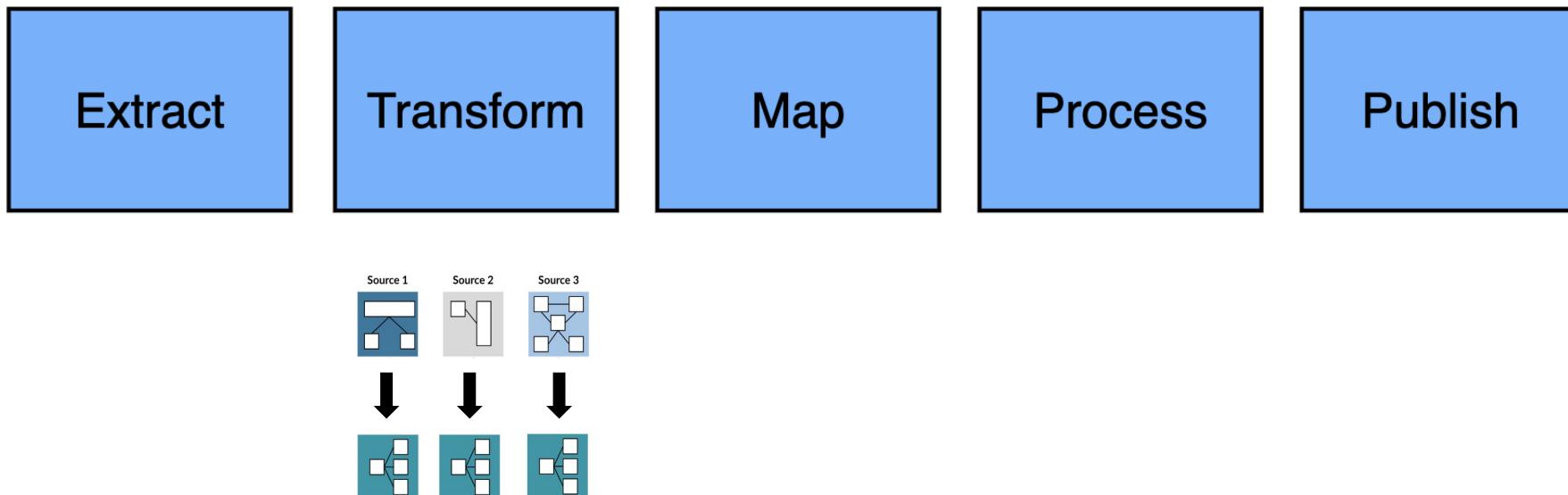


TRANSFORMACIÓN

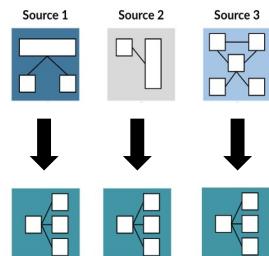
AL MODELO OMOP-CDM



TRANSFORMACIÓN AL MODELO OMOP-CDM



TRANSFORMACIÓN AL MODELO OMOP-CDM



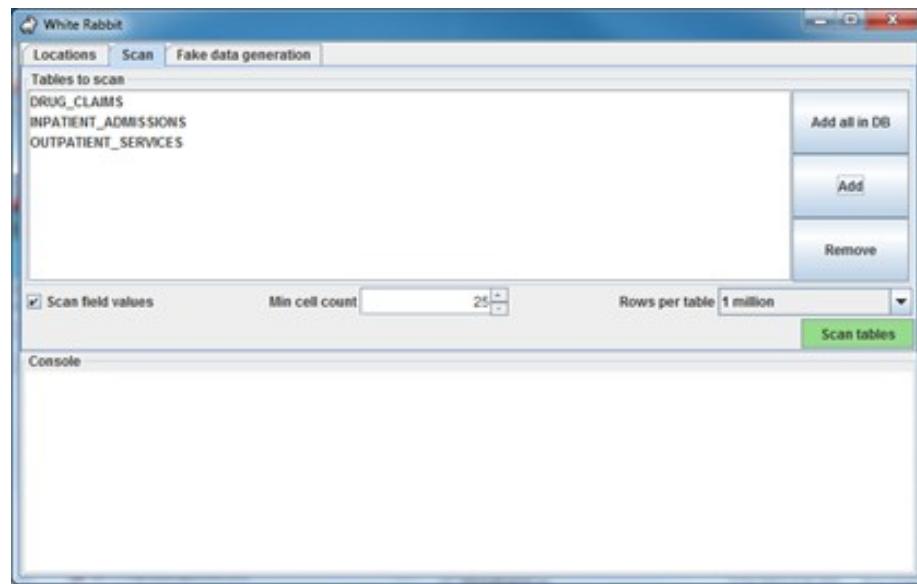
WHITE RABBIT



RABBIT IN A HAT

WhiteRabbit

DATA PROFILING

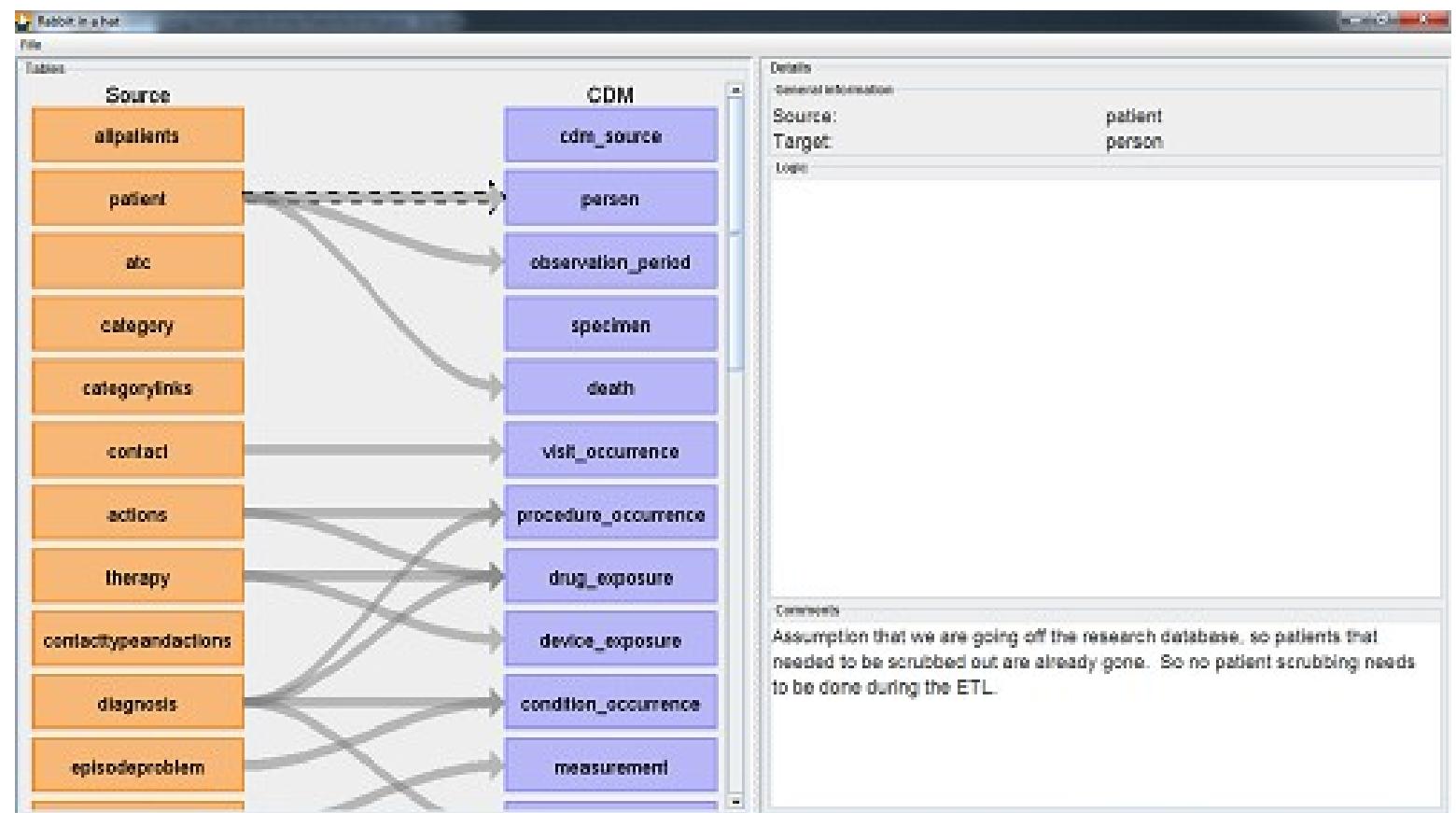


Liberation Sans 10 pt							
	A	B	C	D	E	F	G
1	Table	Field	Type	Max length	N rows	N rows checked	Fraction empty
2	atc	atc	text	7	6231	6231	0
3	atc	txt	text	58	6231	6231	0
4	atc	dddhpk	double precision	8	6231	6231	0.951693147167389
5							
6	contact	exppatidx	numeric	7	100000	100000	0
7	contact	prakid	numeric	9	100000	100000	0
8	contact	patid	numeric	9	100000	100000	0
9	contact	contactid	numeric	5	100000	100000	0
10	contact	contactd	date	10	100000	100000	0.00182
11	contact	dateperiod	text	1	100000	100000	0
12	contact	rawcodsys	text	6	100000	100000	0.61758
13	contact	rawcod	text	9	100000	100000	0.30036
14	contact	rawcodtxt	text	62	100000	100000	0.38449
15	contact	codsystxt	text	6	100000	100000	0.39153
16	contact	cod	text	5	100000	100000	0.39153
17	contact	codtxt	text	151	100000	100000	0.39153
18	contact	employeeid	numeric	9	100000	100000	0.50622
19							

<https://github.com/OHDSI/WhiteRabbit>

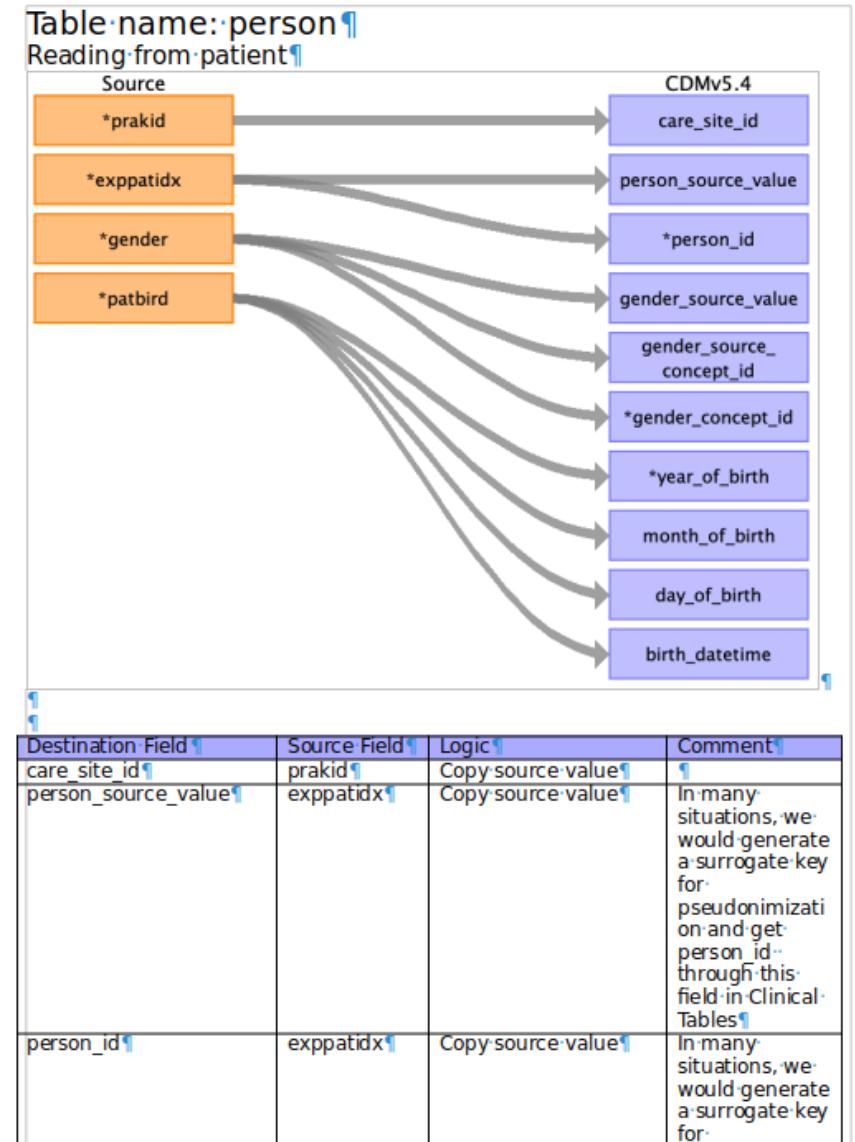
Rabbit -in-a-Hat

DATA MAPPING

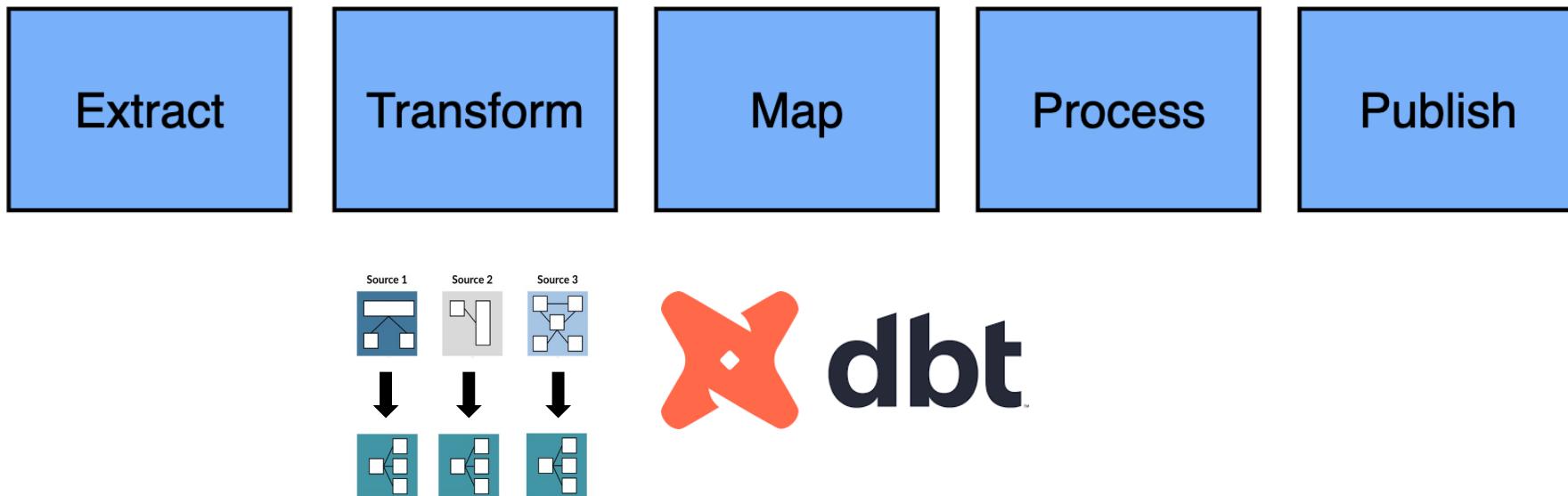


Rabbit -in-a-Hat

DATA MAPPING



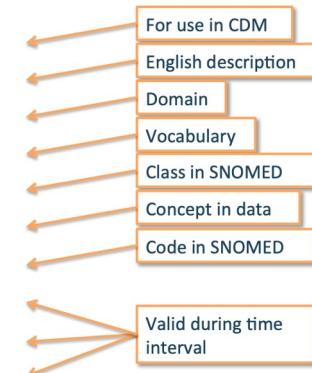
TRANSFORMACIÓN AL MODELO OMOP-CDM



TRANSFORMACIÓN AL MODELO OMOP-CDM



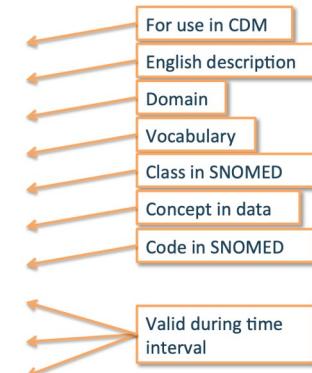
CONCEPT_ID	313217
CONCEPT_NAME	Atrial fibrillation
DOMAIN_ID	Condition
VOCABULARY_ID	SNOMED
CONCEPT_CLASS_ID	Clinical Finding
STANDARD_CONCEPT	S
CONCEPT_CODE	49436004
VALID_START_DATE	01-Jan-1970
VALID_END_DATE	31-Dec-2099
INVALID_REASON	



TRANSFORMACIÓN AL MODELO OMOP-CDM



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	



Usagi

CONCEPT MAPPING

Screenshot of the Usagi Concept Mapping application interface.

The main window displays a grid of concept mapping results with the following columns:

Status	Source code	Source term	Frequency	CodeText	Match score	ConceptID	Concept name	Domain	Concept class	Vocabulary	Concept code	Standard concept	Parents	Children	Comment
Approved	K87.00	Hypertension...	694195	Hypertensie ...	0.81	316866	Hypertensive... Condition	Clinical Findi...	SNOMED	38341003	S	1	27		
Approved	L99.00	Other diseas...	680422	Andere ziekte...	0.47	0	Unmapped					0	0	Too generic	
Approved	D01.00	Abdominal p...	678588	Gegenerals...	0.61	197988	Generalized ... Condition	Clinical Findi...	SNOMED	102614006	S	1	0		
Unchecked	S99.00	Skin disease...	675817	Andere ziekte...	0.75	4317258	Disorder of s...	Condition	Clinical Findi...	SNOMED	95320005	S	2	193	
Unchecked	T86.00	Hypothyroidi...	687283	Hvnothyre...A	1.00	4113642	Hvnothyroidi...	Condition	Clinical Findi...	SNOMED	286910004	S	1	0	

Below the main grid, there are sections for "Source code", "Target concepts", "Search", and "Results".

The "Source code" section shows a breakdown of the source code S99.00: Skin disease, other (Frequency: 675817). The "Target concepts" section shows a target concept for disorder of skin (Concept ID: 4317258, Concept name: Disorder of skin, Domain: Condition, Concept class: Clinical Finding, Vocabulary: SNOMED, Concept code: 95320005, Standard concept: S, Parents: 2, Children: 193).

The "Search" section includes a "Query" field with options to use the source term as query or enter a new query, and a "Filters" section with checkboxes for filtering by user selected concepts, standard concepts, vocabulary, source terms, and domain.

The "Results" section displays a detailed list of matching concepts with columns: Score, Term, Concept ID, Concept name, Domain, Concept class, Vocabulary, Concept code, Standard concept, Parents, and Children. The results include various skin diseases and disorders, such as Skin disease, Fungal (Score: 0.75), Skin Disease, Fungal (Score: 0.65), and AIDS with skin dise... (Score: 0.57).

At the bottom of the interface, there are buttons for "Replace concept", "Add concept", "Comment", and "Approve". Status indicators at the bottom left show "Approved / total: 5 / 1037 8.8% of total frequency".

<https://github.com/OHDSI/Usagi>

TRANSFORMACIÓN AL MODELO OMOP-CDM

Extract

Transform

Map

Process

Publish

El paciente acude con su madre Pepita **PER**, con la que vive en Calle Córcega 23 **LOC**, refiriendo dolor abdominal.

El paciente acude con su madre Lucía **PER**, con la que vive en Calle Londres **LOC**, refiriendo dolor abdominal.

TRANSFORMACIÓN AL MODELO OMOP-CDM



Extract

Transform

Map

Process

Publish

Acude con dolor abdominal derecho de 2 días de duración. Pauta de vacunación completa.

Finding *dolor abdominal*.

Spatial Concept *derecho*.

Temporal Concept *2 días, duración*.

Therapeutic or Preventive Procedure *vacunación*.

Qualitative Concept *completa*.

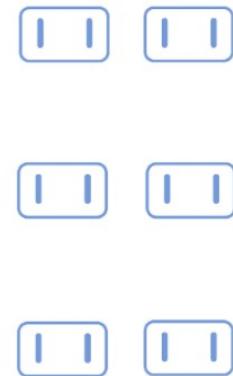
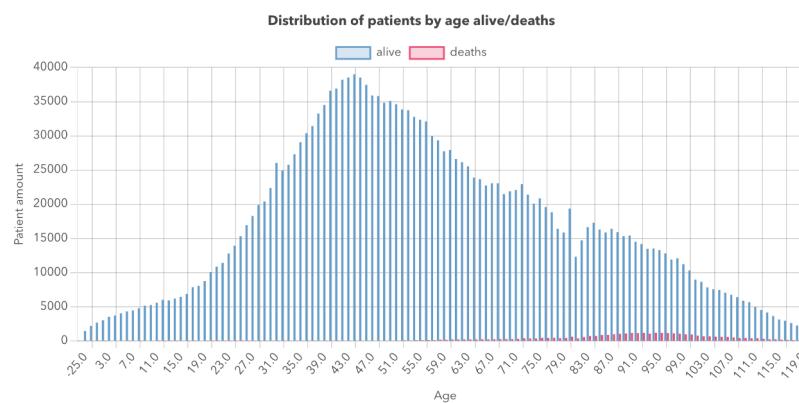
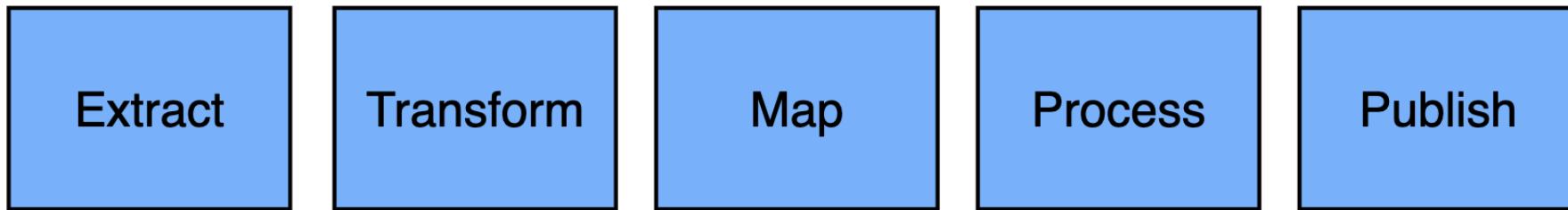
PROCESADO DEL LENGUAJE NATURAL

Estadísticas generales de los resultados obtenidos con el procesado de notas.

- Number of patients with notes: 934.000
- Number of notes: 46.430.000
- Average length: 700
- Average number of entities detected per note: 56

Domain	Events in structure data	Events retrieved by NLP
Condition	2.562.120	537.854.624
Procedure	800.717	205.344.592
Observation	335.659	1.621.526.706
Drug	5.002.550	132.818.091
Measurement	185.752.461	158.167.899
Total	194.4M	2.650M

TRANSFORMACIÓN AL MODELO OMOP-CDM



Preguntas?

GRACIAS POR VUESTRA ATENCIÓN

You are free to:

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

Under the following terms:



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



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



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