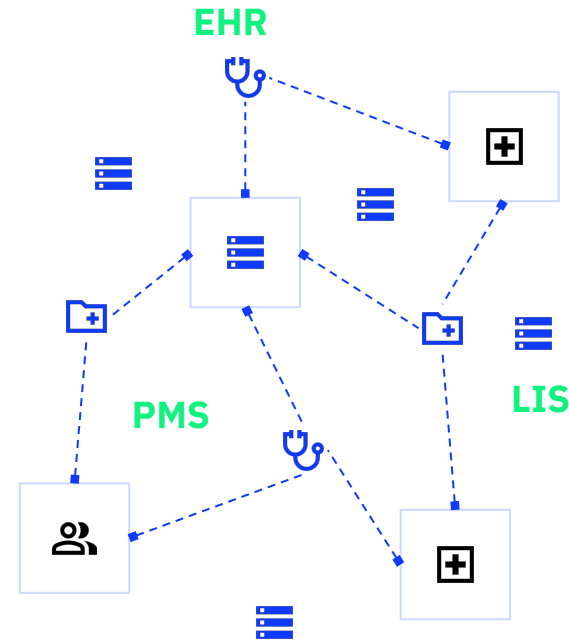


Mapeo de Datos & Aseguramiento de la Calidad

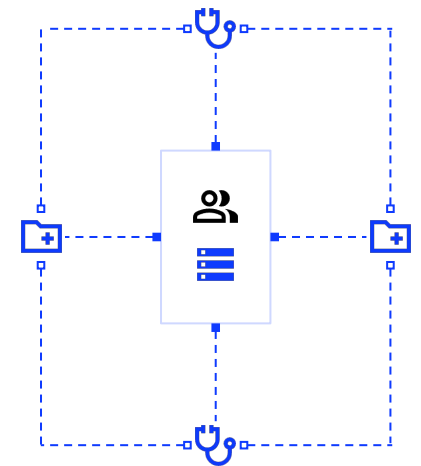
Gabriel Maeztu M.D. - Cofundador y Director Médico @ IOMED

→ Creando datos paciente-céntricos

- La **digitalización** en los hospitales se ha **centrado en los procesos del hospital** y no en representar pacientes.
- En cada centro tiende a haber **varios sistemas de información** que dan soporte a procesos diferentes:
 - Historia Clínica Electrónica
 - Sistemas de gestión de Laboratorio
 - Farmacia y dispensación
 - Sistema de gestión de visitas
 - [...]
- El proceso de mapear los datos al OHDSI OMOP CDM se trata de construir **una versión digital del paciente** sumando los datos de múltiples fuentes, convirtiendo datos proceso-céntricos en datos paciente-céntricos.



Operational Data Architecture



OMOP

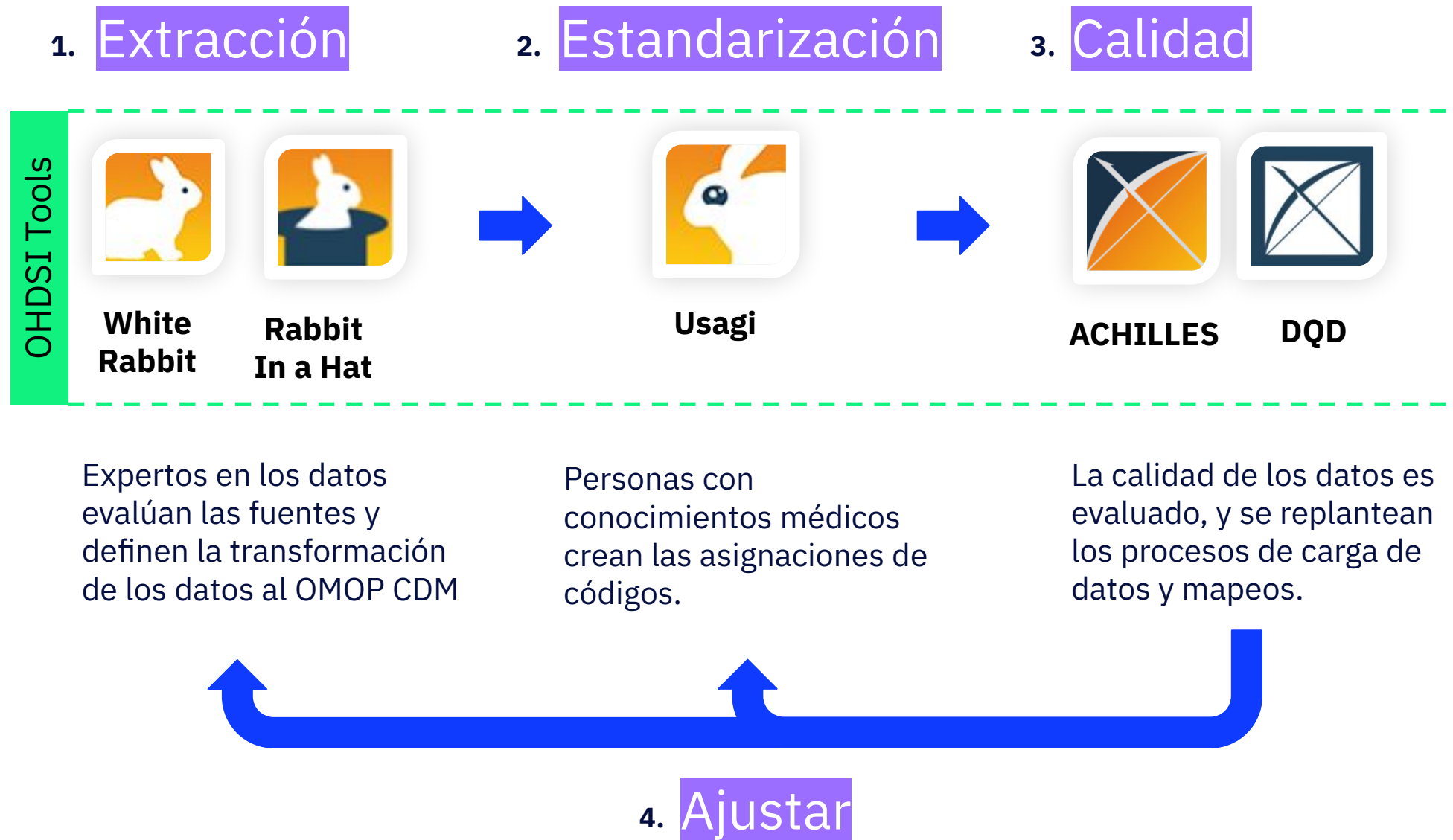
Patient-centric Data Architecture

Proceso General

↳ Extracción y carga de Datos

≡ Mapeo y Estandarización

⚙️ Aseguramiento de la Calidad



→ Proceso General

↳ Extracción y carga de Datos

- El objetivo principal es **identificar, obtener y preparar** los datos del centro.
- Esto requiere de **equipos multidisciplinarios** y de la colaboración de múltiples departamentos y proveedores de sistemas

→ Los centros disponen de media de 5 sistemas de información diferentes puestos en producción

- Se ha de comenzar el proceso con la **identificación de los sistemas de información** en producción en el centro.
- Hemos de determinar la **cobertura temporal** en cada uno de los sistemas.
- Establecer una **estrategia de extracción y actualización** de los datos de los sistemas productivos.
- Es importante **evaluar** las diferentes fuentes de datos.



→ Para que los datos extraídos sean de confianza, los procesos han de ser reproducibles, replicables y estar documentados

- Para realizar un proceso de evaluación e **implementación replicable y reproducible** la OHDSI establece una manera sistemática de extraer los datos y mapearlos al OMOP CDM.
- Se recomienda el uso del programa **White Rabbit** para establecer el linaje de los datos desde el origen hasta el OMOP CDM.



La herramienta **White Rabbit**¹

- Realiza un **escaneo de los datos de origen**, proporcionando información detallada sobre las tablas, campos y valores que aparecen en un campo.

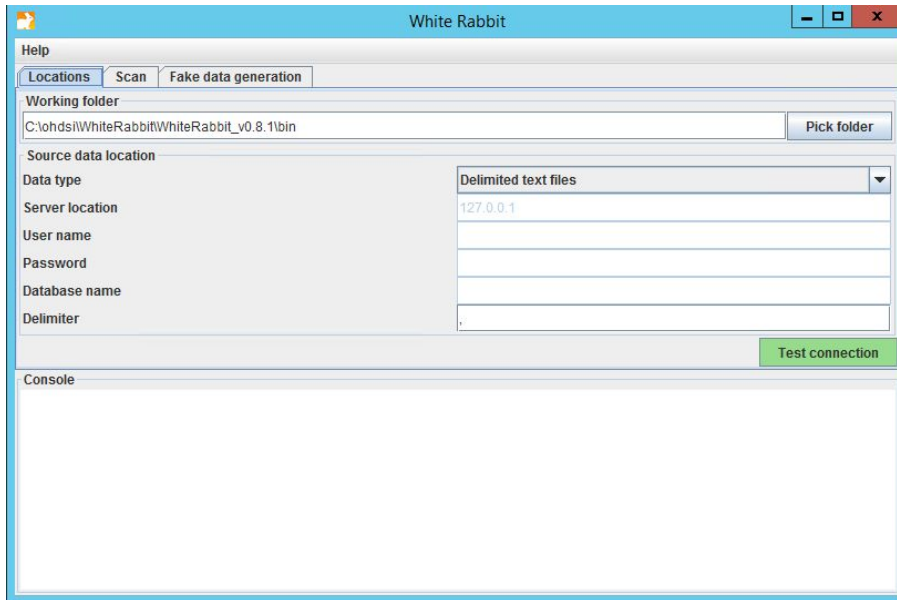
[1]: White Rabbit - <https://ohdsi.github.io/WhiteRabbit/index.html>

→ Para que los datos extraídos sean de confianza, los procesos han de ser reproducibles, replicables y estar documentados



La herramienta White Rabbit¹

- Realiza un **escaneo de los datos de origen**, proporcionando información detallada sobre las tablas, campos y valores que aparecen en un campo.



[1]: White Rabbit - <https://ohdsi.github.io/WhiteRabbit/index.html>

→ Para que los datos extraídos sean de confianza, los procesos han de ser reproducibles, replicables y estar documentados

- Partiendo del análisis de las fuentes de datos que hemos realizado con White Rabbit, esta herramienta ayuda a **generar la documentación** de la transformación de los datos.
- No genera el código para la transformación.

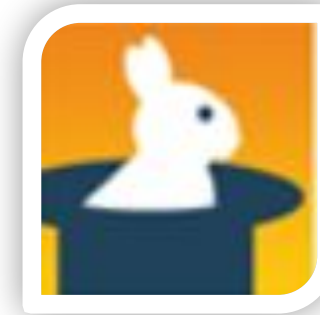
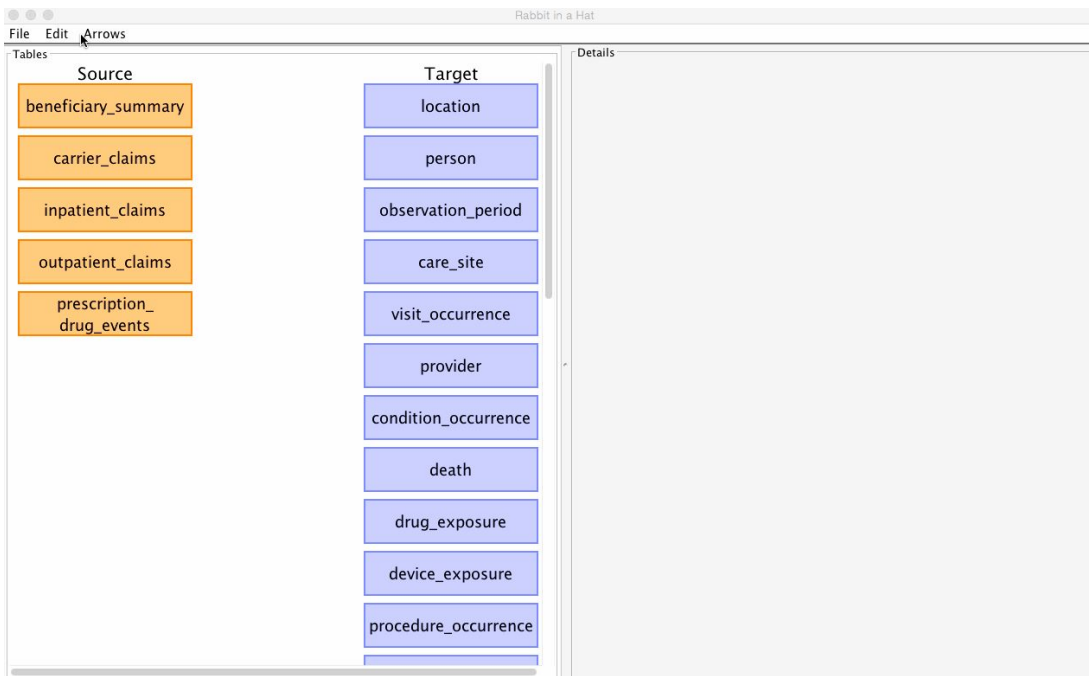


La herramienta Rabbit in a Hat¹

- Proporcionar una interfaz gráfica de usuario que ayuda a **crear la documentación de la extracción** y transformación de datos.

[1]: RabbitInAHat- <https://ohdsi.github.io/WhiteRabbit/RabbitInAHat.html>

→ Para que los datos extraídos sean de confianza, los procesos han de ser reproducibles, replicables y estar documentados

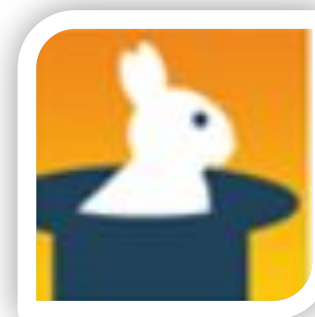
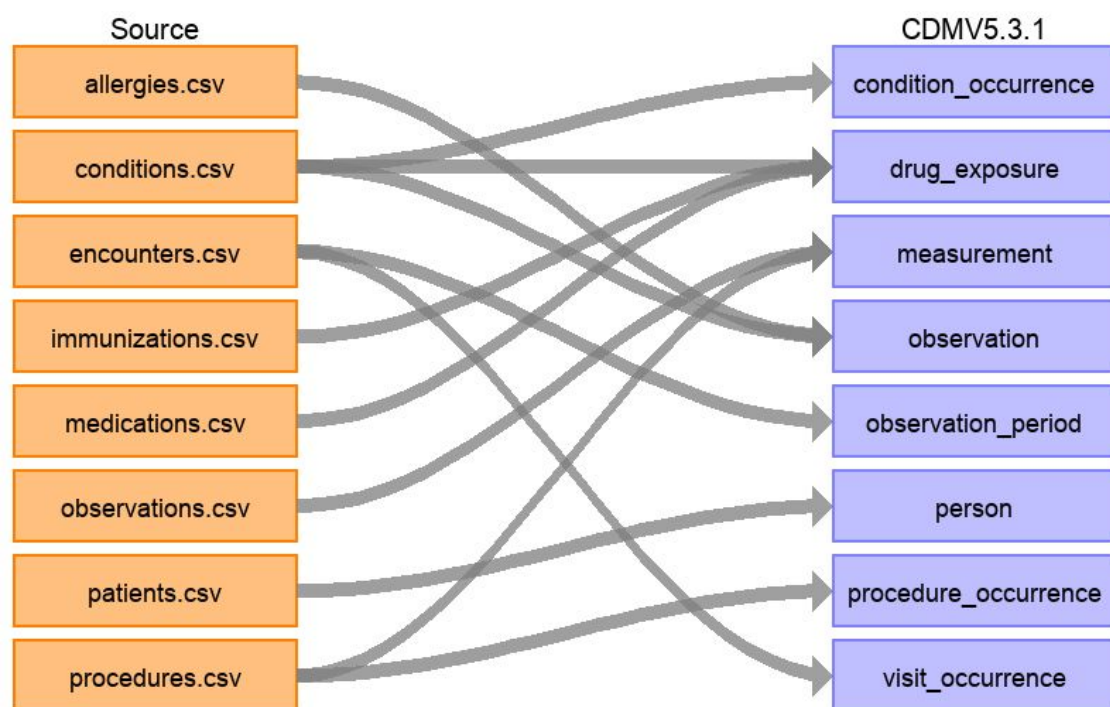


La herramienta Rabbit in a Hat¹

- Proporcionar una interfaz gráfica de usuario que ayuda a **crear la documentación de la extracción y transformación de datos.**

[1]: White Rabbit - <https://ohdsi.github.io/WhiteRabbit/index.html>

→ Para que los datos extraídos sean de confianza, los procesos han de ser reproducibles, replicables y estar documentados



La herramienta Rabbit in a Hat¹

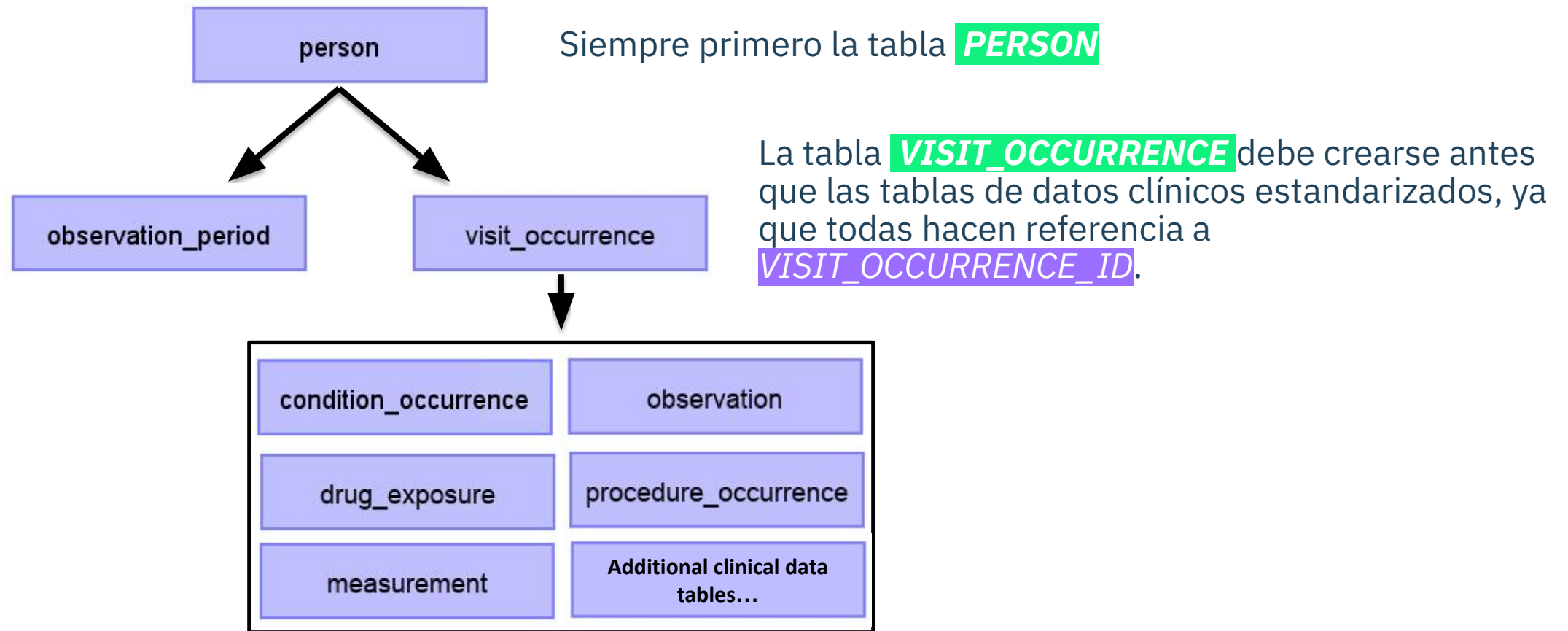
- Proporcionar una interfaz gráfica de usuario que ayuda a **crear la documentación de la extracción** y transformación de datos.

[1]: White Rabbit - <https://ohdsi.github.io/WhiteRabbit/index.html>

→ Para que los datos extraídos sean de confianza, los procesos han de ser reproducibles, replicables y estar documentados

		Target	Source	
person	Gender:	gender_concept_id	gender	When gender = 'M' then set gender_concept_id to 8507, when gender = 'F' then set to 8532
				Drop any rows with missing/unknown gender.
	Birthdate:	year_of_birth	birthdate	Take year from birthdate
		month_of_birth	birthdate	Take month from birthdate
		day_of_birth	birthdate	Take day from birthdate
		birth_datetime	birthdate	With midnight as time 00:00:00
	Race:	race_concept_id	race	When race = 'WHITE' then set as 8527, when race = 'BLACK' then set as 8516, when race = 'ASIAN' then set as 8515, otherwise set as 0

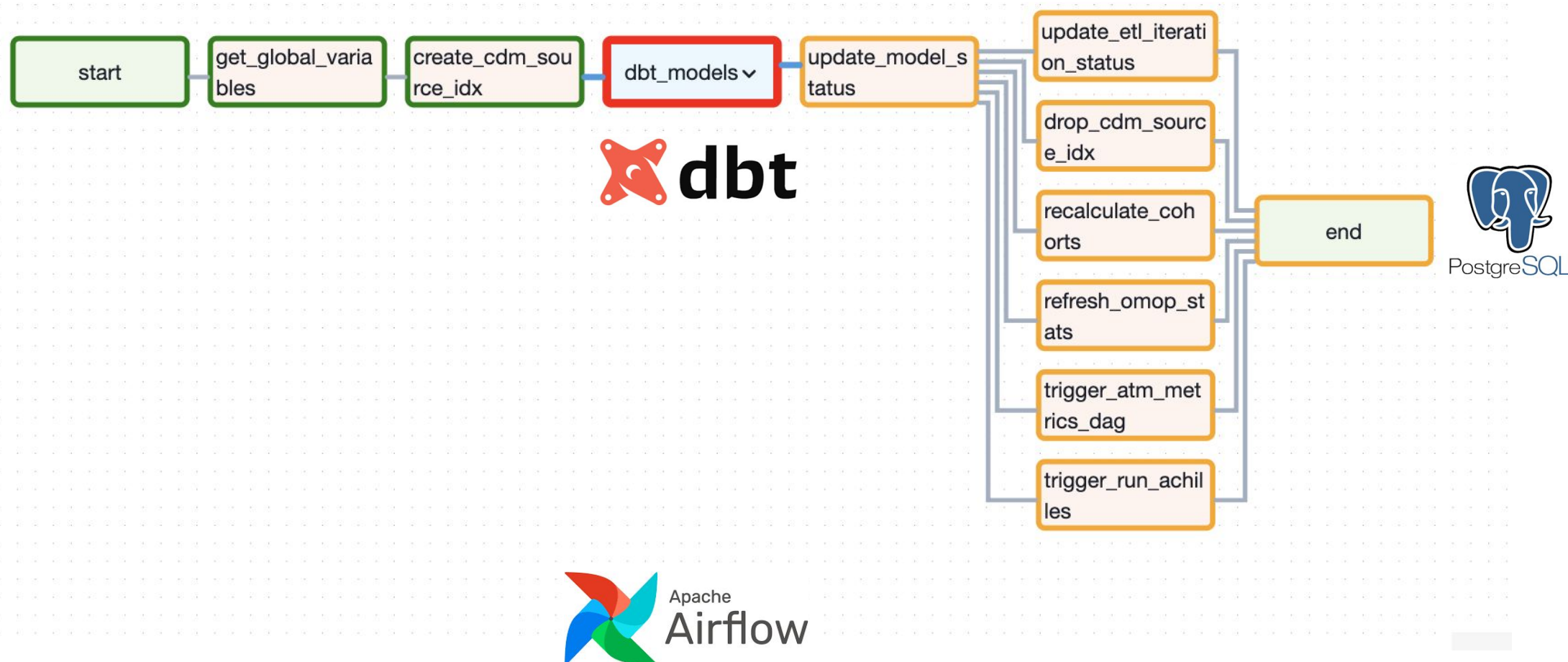
→ La carga de datos en el OMOP CDM tiene un flujo específico



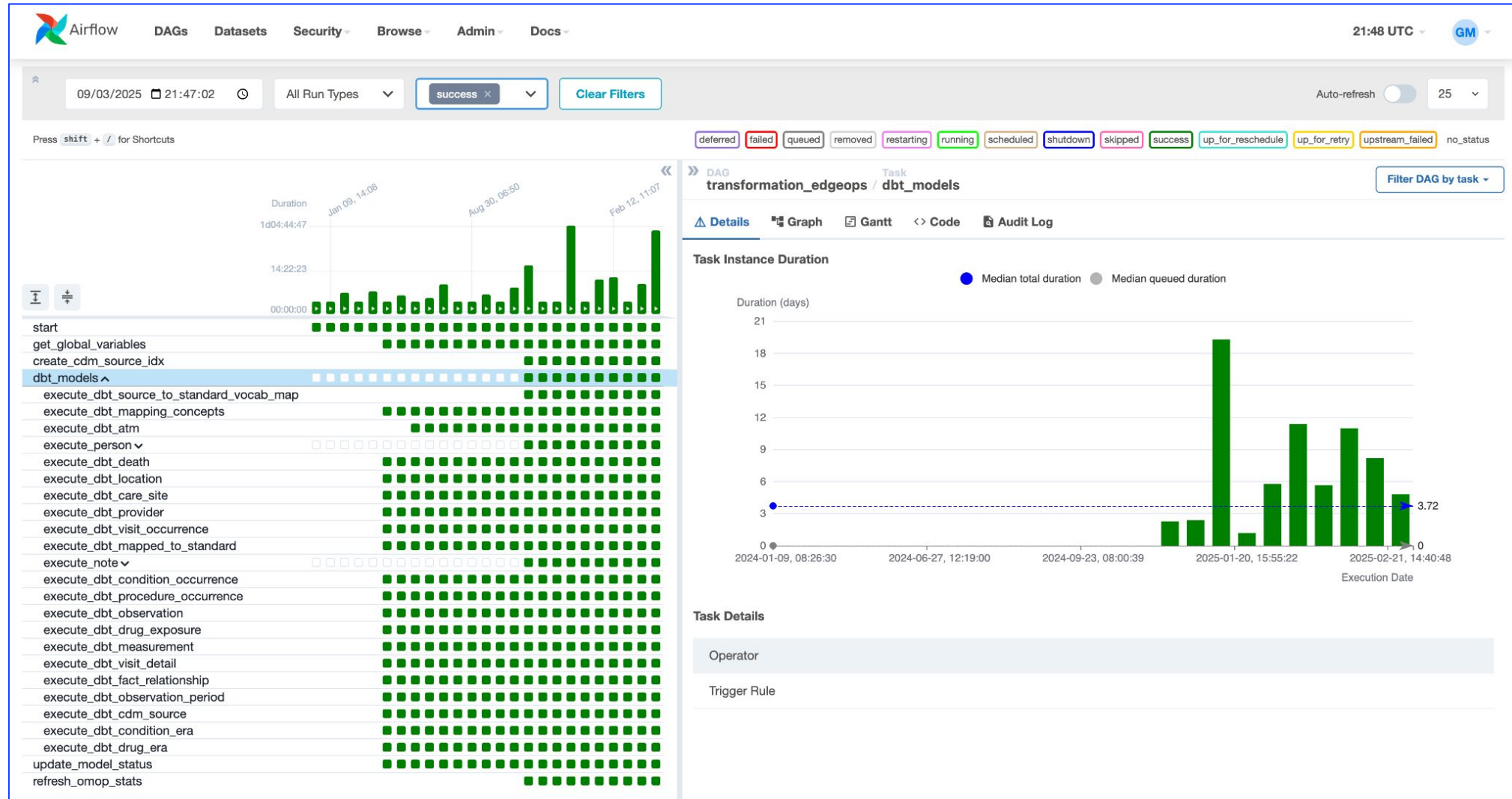
→ La carga de datos se puede realizar con diversas herramientas

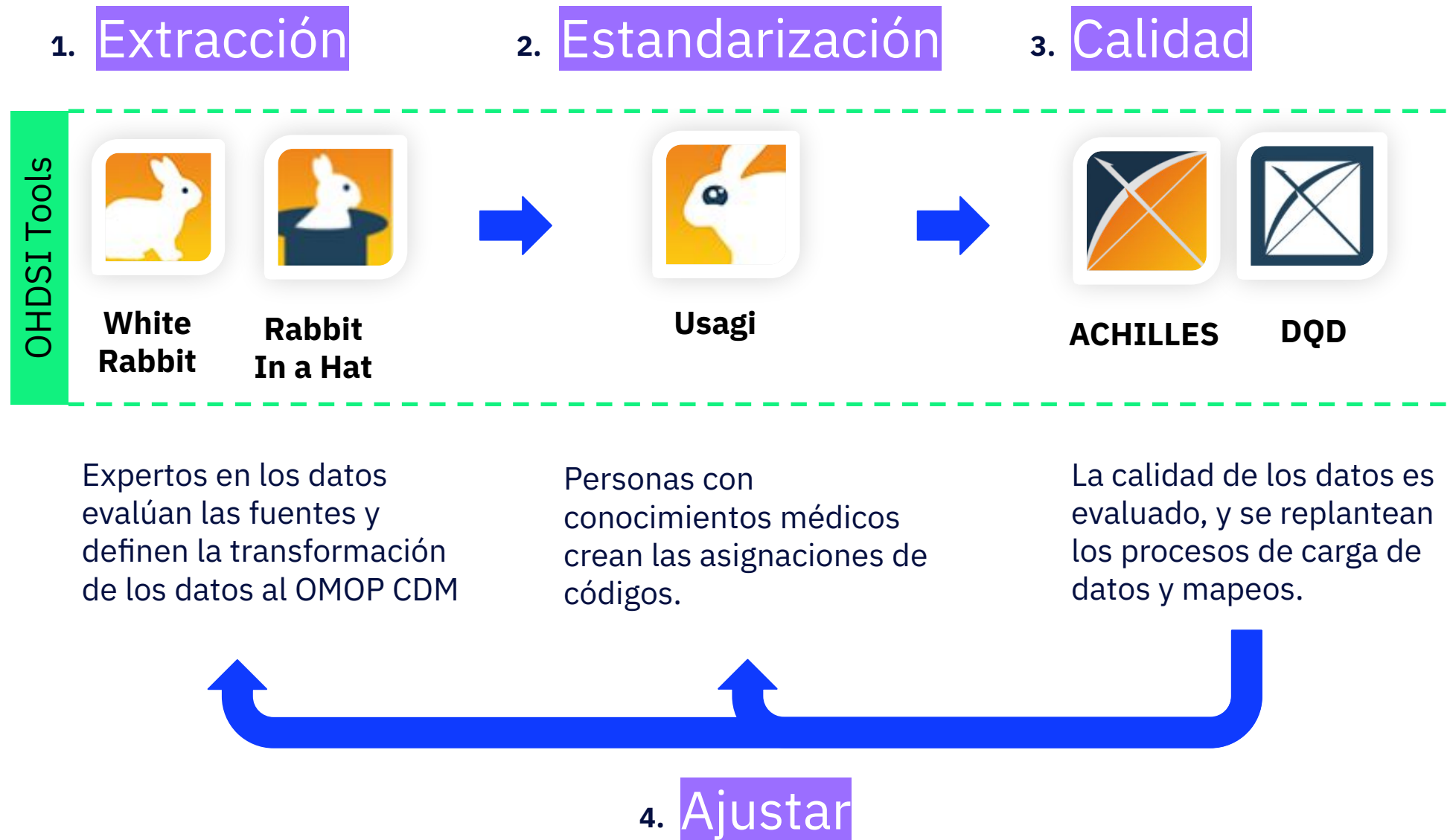


→ La carga de datos se puede realizar con diversas herramientas



→ La carga de datos se puede realizar con diversas herramientas





→ Proceso General

☰ Mapeo y Estandarización

- El objetivo principal es que los **identificadores sean transformados** a los estándares esperados por el OMOP CDM.
- Esto requiere de **equipos clínicos** y de la colaboración de múltiples departamentos para asegurarse de que los datos que los IDs mapeados mantienen la **misma semántica** que aquellos en los datos de origen.

PROBLEMS

hyperlipidemia

essential hypertension

diabetes mellitus

aspirin

hydrochlorothiazide

STRUCTURED DATA

Vitals

Ht 6 ft in

Wt 190 lbs With clothes

BMI 25.8

Last Visit with Family Medicine

follow up, 02-17-2015

usual f/up

Follow-Up: Diabetes mellitus, Follow-Up: Essential hypertension, Follow-Up: Hyperlipidemia

Performed by Tess Morton-Trask, (555) 595-5106

hyperlipidemia

The lipid panel much improved this visit HDL is up and out the LDL's are down to 118 w on starting a statin, and encourage continued physical activity and low fat diet. Will ge again in 3 months.

Will check lipid panel, serum on 05-20-2015

essential hypertension

Blood pressure is better today than has men and several years swelling is and down with addition of hydrochlorothiazide. Seems to be tolerating lisinopril well.

Will check BMP, serum or plasma on 05-20-2015

ID	patient_id	data_id	date	value	unit	visit
643	243	weight	2021-10-12	190	lbs	367
2891	243	weight	2021-11-21	170	lbs	458

→ Los identificadores de cada fuente se tienen que mapear a conceptos

hyperlipidemia

essential hypertension

diabetes mellitus

hydrochlorothiazide

190

Ht 6 ft 0 in

STRUCTURED DATA

Vitals

Ht

6

ft

in

☐ Not Performed

-select reason-

Wt

190

lbs

With clothes

☐ Not Performed

-select reason-

☐ Out of Range

BMI

25.8

Last Visit

follow up, (usual f/up)

Follow-Up: Performed

hyperlipidemia

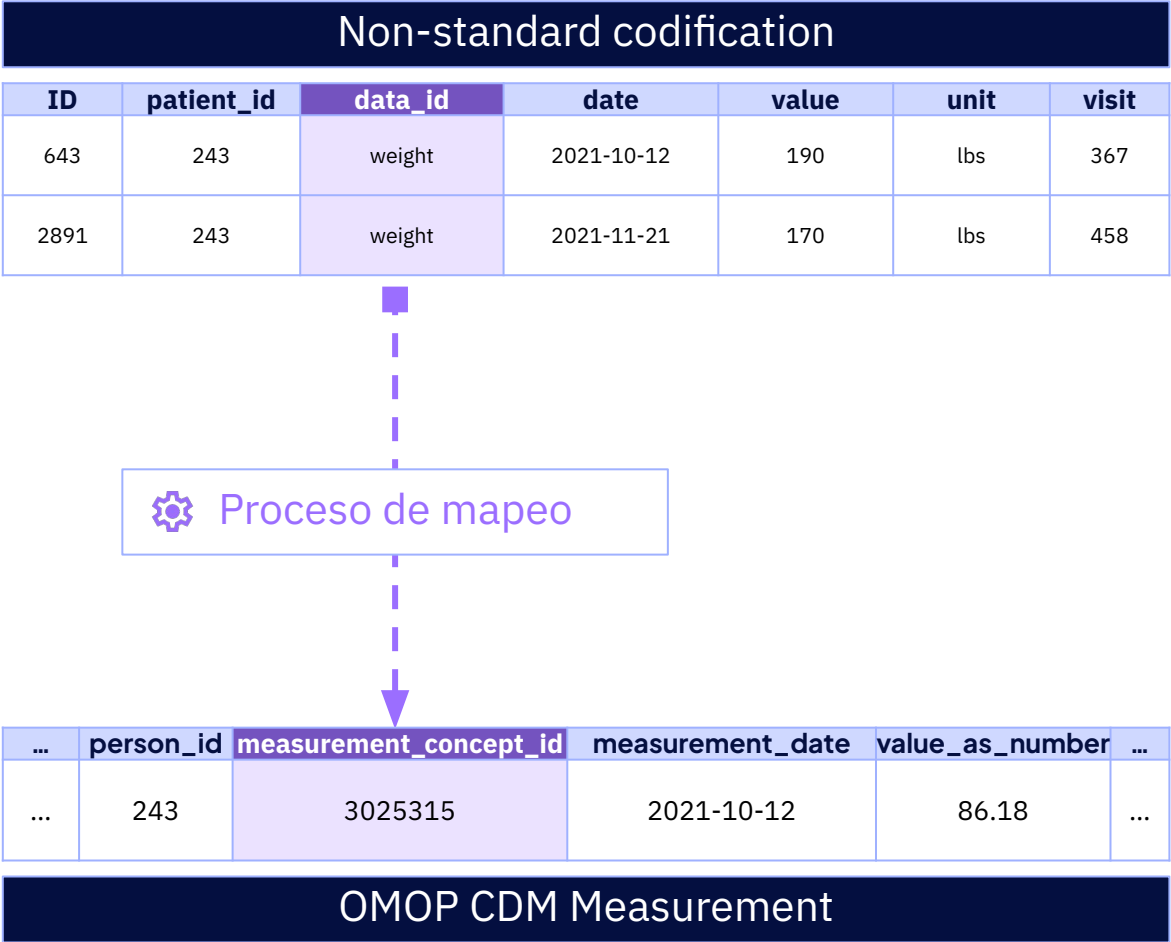
The lipid panel on starting again in 3 months

Will check

essential hypertension

Blood pressure addition of

Will check



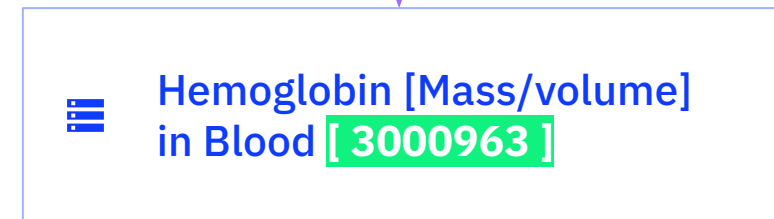
→ Todas las instancias de OMOP han de tener los mismos identificadores para poder colaborar de forma efectiva

Hospital A — Other Standards

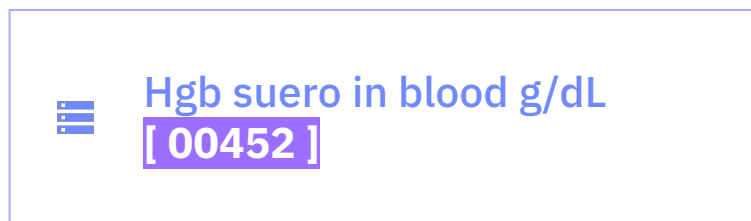


⚙ Mapeo del dato

OMOP CDM



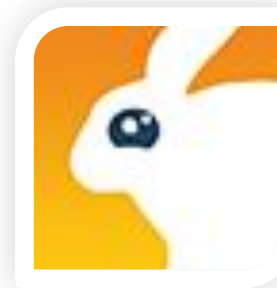
Hospital B — Internal Source Code



⚙ Mapeo del dato

→ El proceso de revisar los códigos internos y mapearlos a un código estándar ha de ser un proceso revisable y reproducible

- Se ha de cargar los códigos de tu sistema de origen (“source_codes”) que desees mapear a los conceptos del vocabulario de OMOP.
- Utiliza la interfaz de Usagi para **revisar las sugerencias de mapeo o crear nuevos mapeos**. Preferiblemente, la persona que realice esta revisión debe tener experiencia con el sistema de codificación y la terminología médica.
- Exporta el mapa final generado por Usagi al SOURCE_TO_CONCEPT_MAP del vocabulario de OMOP.



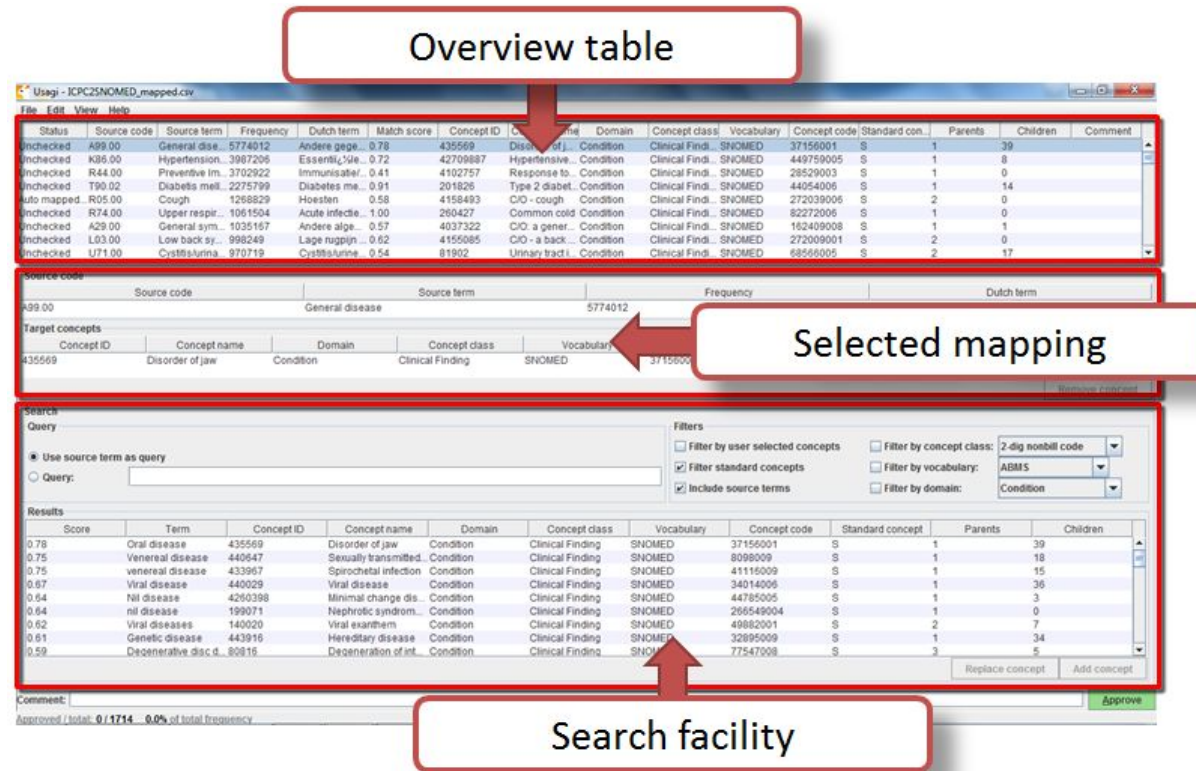
La herramienta Usagi¹

- Asiste en el proceso de **mapear los códigos de una fuente de datos** a las terminologías estándar del OMOP CDM.

[1]: Usagi - <https://ohdsi.github.io/Usagi/>

→ El proceso de revisar los códigos internos y mapearlos a un código estándar ha de ser un proceso revisable y reproducible

Overview table



Selected mapping

Search facility

The screenshot displays the Usagi application interface. At the top, a red box labeled 'Overview table' points to a table listing various source terms and their mappings to standard concepts. Below this, a red box labeled 'Selected mapping' points to a detailed view of a specific mapping, showing the source term 'Disorder of jaw' and its mapping to the standard concept 'Disorder of jaw'. At the bottom, a red box labeled 'Search facility' points to a search bar and filters section, which includes options to search by source term or query, and filters for concept class, vocabulary, and domain.

La herramienta Usagi¹

- Asiste en el proceso de **mapear los códigos de una fuente de datos** a las terminologías estándar del OMOP CDM.

[1]: Usagi - <https://ohdsi.github.io/Usagi/>

Overview table

Usagi - ICPC25NOMED_mapped.csv

Status	Source code	Source term	Frequency	Dutch term	Match score	Concept ID	Concept name	Domain	Concept class	Vocabulary	Concept code	Standard con.	Parents	Children	Comment
Unchecked	A99.00	General dise...	5774012	Andere gepe...	0.78	435569	Disorder of j...	Condition	Clinical Findi...	SNOMED	37156001	S	1	39	
Unchecked	K86.00	Hypertension...	3987206	Essentiële h...	0.72	42709887	Hypertensive...	Condition	Clinical Findi...	SNOMED	449759005	S	1	8	
Unchecked	R44.00	Preventive Im...	3702922	Immunisatie...	0.41	4102757	Response to...	Condition	Clinical Findi...	SNOMED	28529003	S	1	0	
Unchecked	T90.02	Diabetes mell...	2275799	Diabetes me...	0.91	201826	Type 2 diabet...	Condition	Clinical Findi...	SNOMED	44054006	S	1	14	
Auto mapped	R05.00	Cough	1268829	Hoesten	0.58	4158493	C/O - cough	Condition	Clinical Findi...	SNOMED	272039006	S	2	0	
Unchecked	R74.00	Upper respir...	1061504	Acute infectie...	1.00	260427	Common cold	Condition	Clinical Findi...	SNOMED	82272006	S	1	0	
Unchecked	A29.00	General sym...	1035167	Andere alge...	0.57	4037322	C/O - a gener...	Condition	Clinical Findi...	SNOMED	162409008	S	1	1	
Unchecked	L03.00	Low back sy...	998249	Lage rugpijn...	0.62	4155085	C/O - a back...	Condition	Clinical Findi...	SNOMED	272009001	S	2	0	
Unchecked	U71.00	Cystitis/urina	970719	Cystitis/urine...	0.54	81902	Urinary tract i...	Condition	Clinical Findi...	SNOMED	68566005	S	2	17	

Source code

Source code	Source term	Frequency	Dutch term
A99.00	General disease	5774012	

Target concepts

Concept ID	Concept name	Domain	Concept class	Vocabulary
435569	Disorder of jaw	Condition	Clinical Finding	SNOMED

Search

Query

☒ Use source term as query

☐ Query:

Filters

☐ Filter by user selected concepts

☒ Filter standard concepts

☒ Include source terms

☐ Filter by concept class: 2-dig nonbill code

☐ Filter by vocabulary: ABMS

☐ Filter by domain: Condition

Results

Score	Term	Concept ID	Concept name	Domain	Concept class	Vocabulary	Concept code	Standard concept	Parents	Children
0.78	Oral disease	435569	Disorder of jaw	Condition	Clinical Finding	SNOMED	37156001	S	1	39
0.75	Venereal disease	440647	Sexually transmitted	Condition	Clinical Finding	SNOMED	8098009	S	1	18
0.75	venereal disease	433967	Spirochetal infection	Condition	Clinical Finding	SNOMED	41116009	S	1	15
0.67	Viral disease	440029	Viral disease	Condition	Clinical Finding	SNOMED	34014006	S	1	36
0.64	Nil disease	4260398	Minimal change dis...	Condition	Clinical Finding	SNOMED	44785005	S	1	3
0.64	nil disease	199071	Nephrotic syndrom...	Condition	Clinical Finding	SNOMED	266549004	S	1	0
0.62	Viral diseases	140020	Viral exanthem	Condition	Clinical Finding	SNOMED	49882001	S	2	7
0.61	Genetic disease	443916	Hereditary disease	Condition	Clinical Finding	SNOMED	32895009	S	1	34
0.59	Degenerative disc d...	80816	Degeneration of int...	Condition	Clinical Finding	SNOMED	77547008	S	3	5

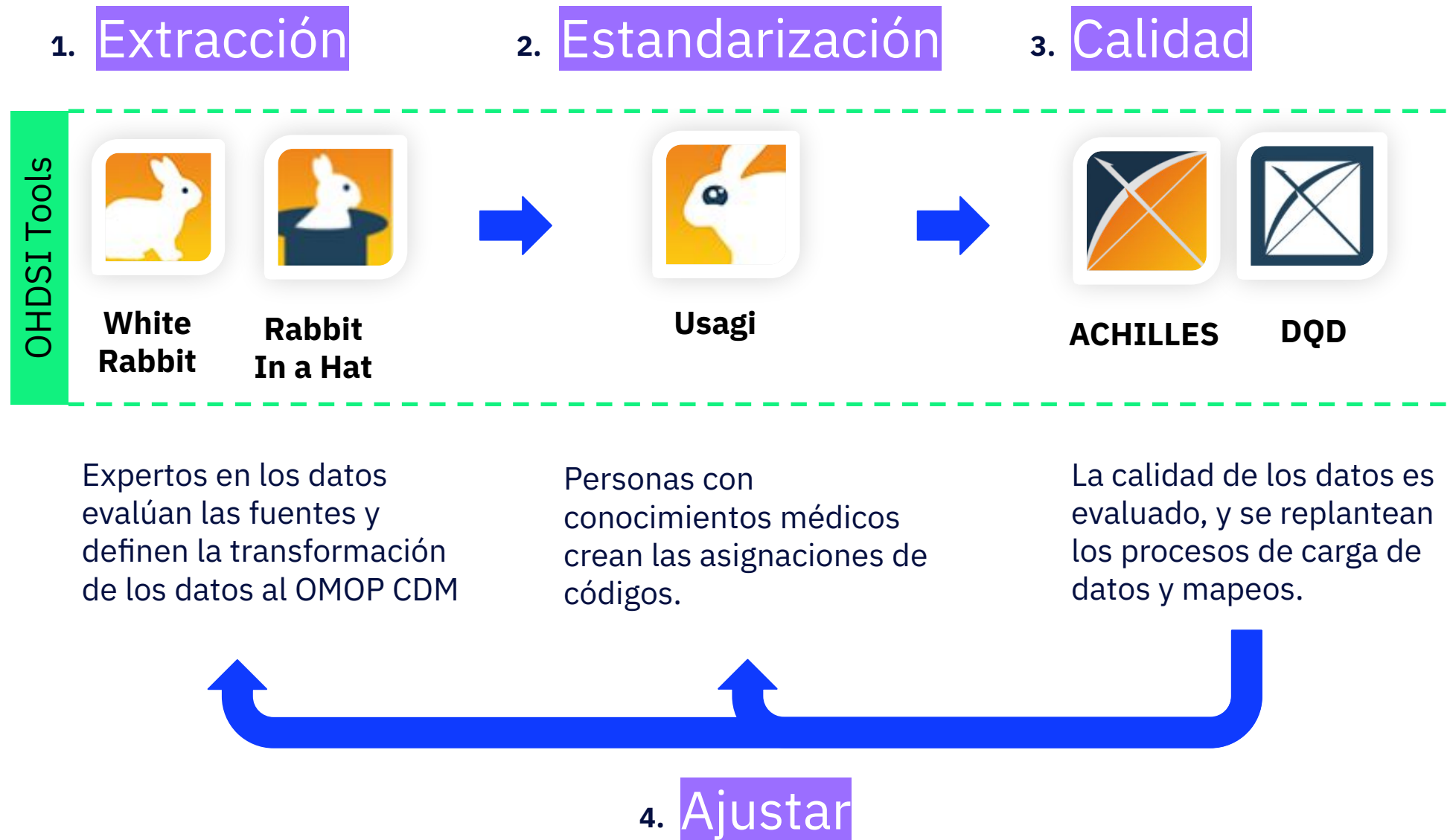
Comment:

Approved / total: 0 / 1714 0.0% of total frequency

Replace concept Add concept Approve

Selected mapping

Search facility



→ Proceso General

🏆 Aseguramiento de la Calidad

- El objetivo principal es asegurar que **los datos integrados son correctos** y que no introducirán incidencias de calidad en los estudios.
- Esto requiere de **equipos clínicos** y de la colaboración del **equipo de datos** para entender bien las posibles anomalías detectadas y gestionarlas correctamente.

→ El proceso, a menudo es complejo, lo que aumenta el riesgo de cometer errores que pasan desapercibidos

- Este paquete **ejecutará una serie de comprobaciones** de calidad de datos en una instancia de OMOP CDM.
- Ejecuta sistemáticamente las comprobaciones, **las evalúa en relación con un umbral** preestablecido y crea un reporte con los resultados de una manera transparente y fácil de entender.
- Se realizan aproximadamente 4000 comprobaciones de calidad de datos individuales en la base de datos.



El DataQualityDashboard¹

- Asiste en el proceso de evaluar la calidad del proceso completo de ETL y mapeo al OMOP CDM.

[1]: DataQualityDashboard - <https://ohdsi.github.io/DataQualityDashboard/>

→ El proceso, a menudo es complejo, lo que aumenta el riesgo de cometer errores que pasan desapercibidos

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% (1317/1400).



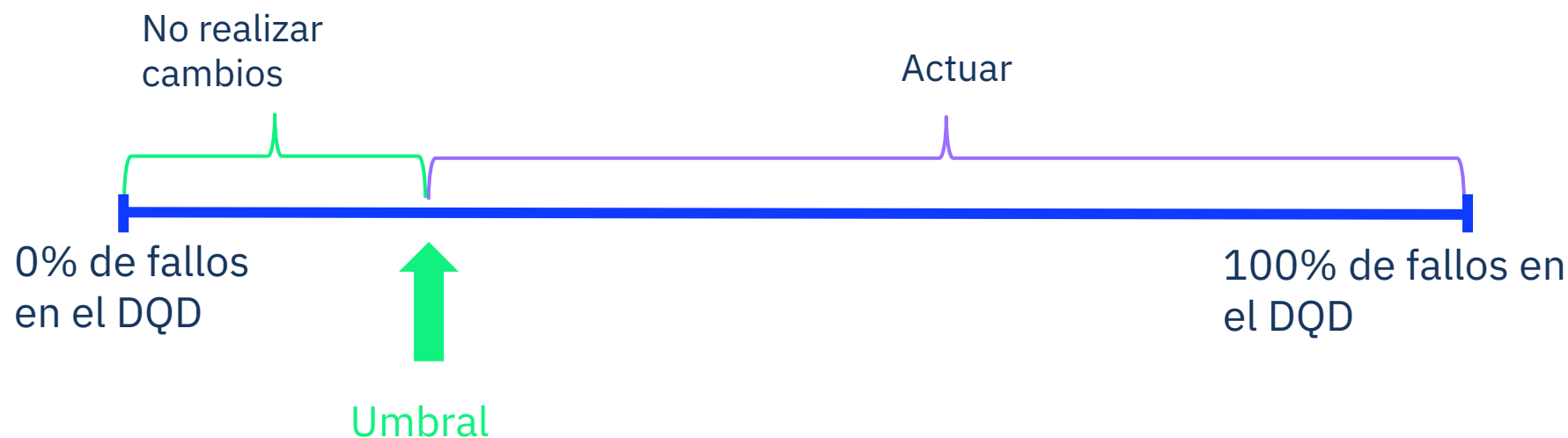
El DataQualityDashboard¹

- Asiste en el proceso de evaluar la calidad del proceso completo de ETL y mapeo al OMOP CDM.

<https://data.ohdsi.org/DataQualityDashboard/>

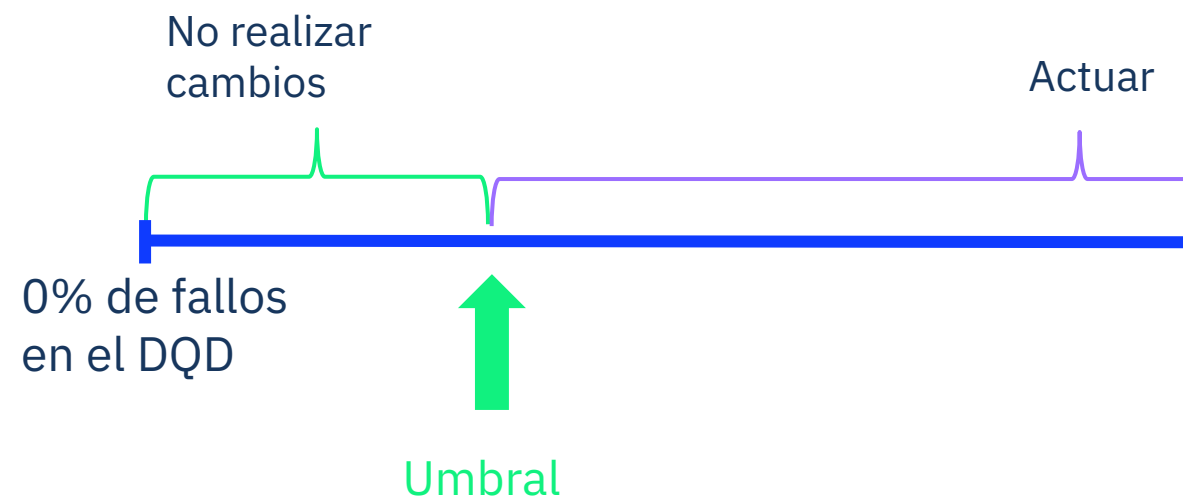
[1]: DataQualityDashboard - <https://ohdsi.github.io/DataQualityDashboard/>

→ El **estadístico** agregado computado con el DQD y un **umbral** determina la decisión de **si cumple con las expectativas**



→ El **estadístico** agregado computado con el DQD y un **umbral** determina la decisión de **si cumple con las expectativas**

- **Controles críticos:**
el umbral siempre es 0
- **Controles de la convención CDM:**
el umbral debería ser 0 en teoría, pero puede ser necesario ajustarlo en la práctica. Un 95%-98% tiende a ser mas realista
- **Controles de caracterización:**
el umbral depende de las expectativas de la fuente de datos



→ El proceso, a menudo es complejo, lo que aumenta el riesgo de cometer errores que pasan desapercibidos



El DataQualityDashboard¹

- Asiste en el proceso de evaluar la calidad del proceso completo de ETL y mapeo al OMOP CDM.

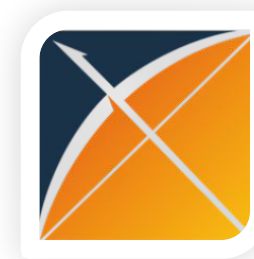
```
library(DataQualityDashboard)
```

```
DataQualityDashboard::executeDqChecks(  
  connectionDetails = connectionDetails,  
  cdmDatabaseSchema = cdmDatabaseSchema,  
  resultsDatabaseSchema = resultsDatabaseSchema,  
  cdmSourceName = cdmSourceName,  
  cdmVersion = cdmVersion,  
  numThreads = numThreads,  
  checkLevels = checkLevels,  
  checkSeverity = checkSeverity,  
  tablesToExclude = tablesToExclude,  
  checkNames = checkNames  
)
```

[1]: DataQualityDashboard - <https://ohdsi.github.io/DataQualityDashboard/>

→ Una caracterización sistematizada puede resaltar anomalías no detectadas simplemente con un vistazo

- Presenta **resultados de manera interactiva** (gráficas, tablas, dashboards), facilitando la exploración y el control de calidad de los datos.
- Ejecuta **reglas automáticas para detectar anomalías** en los datos (fechas imposibles, valores fuera de rango) y alertar sobre posibles errores de ETL o calidad.
- Funciona en conjunto con otras herramientas como ATLAS para enriquecer el análisis y la gobernanza de datos estandarizados bajo el modelo OMOP.

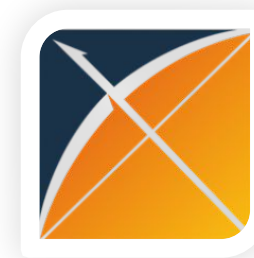
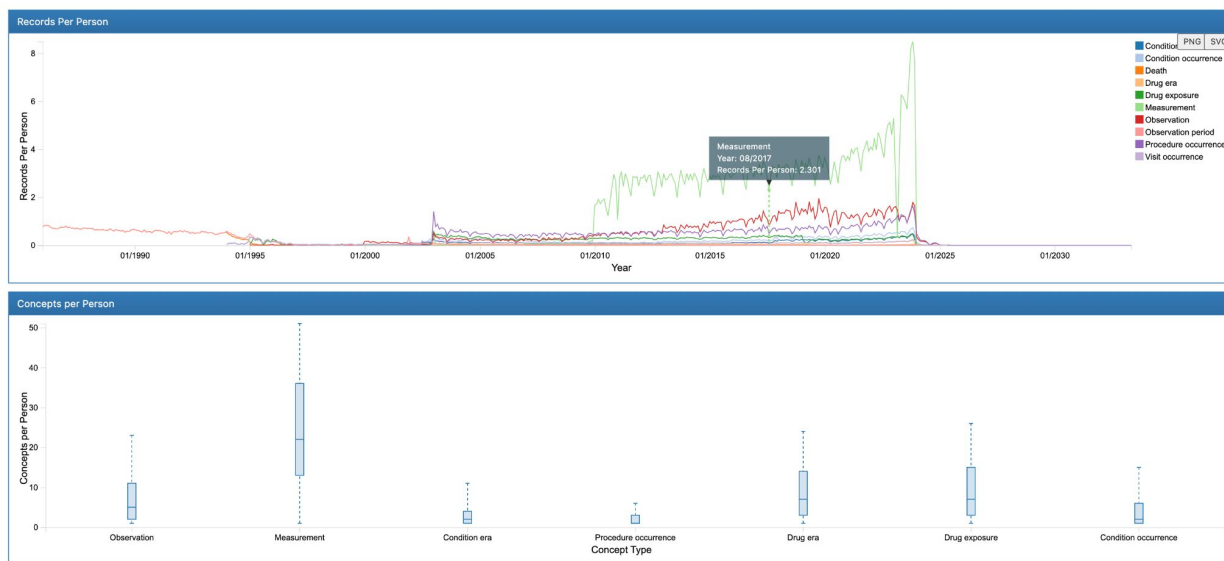


ACHILLES¹

- Genera estadísticas demográficas y clínicas (p.ej., prevalencia de condiciones, uso de fármacos) para ofrecer una visión general del contenido de la base de datos.

[1]: Achilles - <https://ohdsi.github.io/Achilles/index.html>

→ Una caracterización sistematizada puede resaltar anomalías no detectadas simplemente con un vistazo



ACHILLES¹

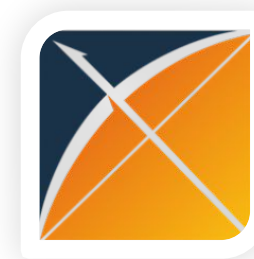
- Realiza una amplia caracterización de bases de datos que permite comprender de un vistazo los datos disponibles

[1]: Achilles - <https://ohdsi.github.io/Achilles/index.html>

→ Una caracterización sistematizada puede resaltar anomalías no detectadas simplemente con un vistazo

```
library(Achilles)
connectionDetails <- createConnectionDetails(
  dbms="redshift",
  server="server.com",
  user="secret",
  password='secret',
  port="5439")
```

```
Achilles::achilles(
  cdmVersion = "5.4",
  connectionDetails = connectionDetails,
  cdmDatabaseSchema = "yourCdmSchema",
  resultsDatabaseSchema = "yourResultsSchema"
)
```



ACHILLES¹

- Realiza una amplia caracterización de bases de datos que permite comprender de un vistazo los datos disponibles

[1]: Achilles - <https://ohdsi.github.io/Achilles/index.html>



**C/ de St. Antoni Maria Claret, 167,
Pabellón de La Mercè,
08025 — Barcelona**

info@iomed.health

<https://iomed.health> 