ETL in the Time of Data Privacy

Kristin Kostka, MPH
Talita Duarte-Salles, PhD

Who We Are

Kristin Kostka, MPH
IQVIA (United States)
Data steward: IQVIA data
Co-lead Project Charybdis

Talita Duarte-Salles, PhD
IDIAPJGol (Barcelona, Spain)
Data steward: SIDIAP, HM Hospitales
Co-lead Project Charybdis

I've been seeing all this cool COVID work that OHDSI's doing… how can I get my data into OMOP?

Process to become an OMOP site

Joining the OMOP Global Data Network

Becoming an OHDSI site involves a variety of decisions across the following dimensions (in no particular order):

Who writes ETL scripts and executes the conversion?
Who builds and maintains the infrastructure?
Who leads OHDSI / OMOP training and support?
Who installs the OHDSI Tools (e.g. ATLAS)?
Who certifies the CDM data quality?
How do receive, review and conduct network studies?

Sites have different rules about patient data.

In Europe:

General Data Protection Regulation (GDPR) is a regulation in EU law on data protection and privacy in the European Union (EU) and the European Economic Area (EEA)

Implemented 25 May 2018

4.5.2016 EN Official Journal of the European Union L 119/1

I

(Legislative acts)

REGULATIONS

REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL
of 27 April 2016
he protection of natural persons with regard to the processing of personal data and on the free
ovement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)
(Text with EEA relevance)

EAN PARLIAMENT AND THE COUNCIL OF THE EUROPEAN UNION,

nd to the Treaty on the Functioning of the European Union, and in particular Article 16 thereof,

nd to the proposal from the European Commission,

ansmission of the draft legislative act to the national parliaments,

nd to the opinion of the European Economic and Social Committee (1),

Having regard to the opinion of the Committee of the Regions (2),

Acting in accordance with the ordinary legislative procedure (3),

Whereas:

(1) The protection of natural persons in relation to the processing of personal data is a fundamental right. Article 8(1) of the Charter of Fundamental Rights of the European Union (the 'Charter') and Article 16(1) of the Treaty on the Functioning of the European Union (TFEU) provide that everyone has the right to the protection of personal data concerning him or her.

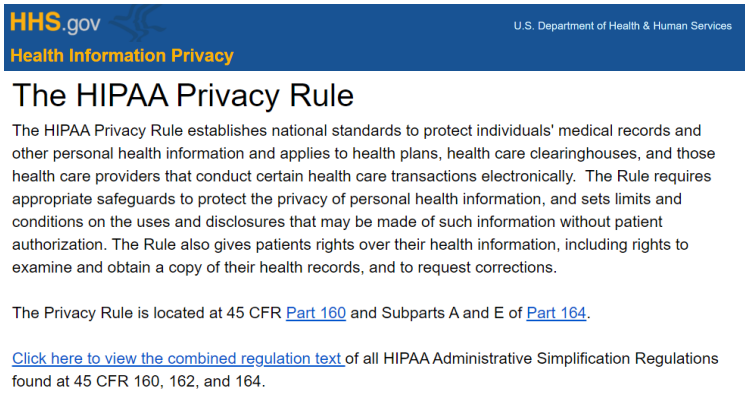
(2) The principles of, and rules on the protection of natural persons with regard to the processing of their personal data should, whatever their nationality or residence, respect their fundamental rights and freedoms, in particular their right to the protection of personal data. This Regulation is intended to contribute to the accomplishment of an area of freedom, security and justice and of an economic union, to economic and social progress, to the strengthening and the convergence of the economies within the internal market, and to the well-being of natural persons.

(3) Directive 95/46/EC of the European Parliament and of the Council (4) seeks to harmonise the protection of fundamental rights and freedoms of natural persons in respect of processing activities and to ensure the free flow of personal data between Member States.

(1) OJ C 229, 31.7.2012, p. 90.
(2) OJ C 391, 18.12.2012, p. 127.
(3) Position of the European Parliament of 12 March 2014 (not yet published in the Official Journal) and position of the Council at first reading of 8 April 2016 (not yet published in the Official Journal). Position of the European Parliament of 14 April 2016.
(4) Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data (OJ L 281, 23.11.1995, p. 31).

In the US:

The Health Insurance Portability and Accountability Act (HIPAA) of 1996 establishes national standards to protect individuals' medical records and other personal health information



HHS.gov U.S. Department of Health & Human Services

Health Information Privacy

The HIPAA Privacy Rule

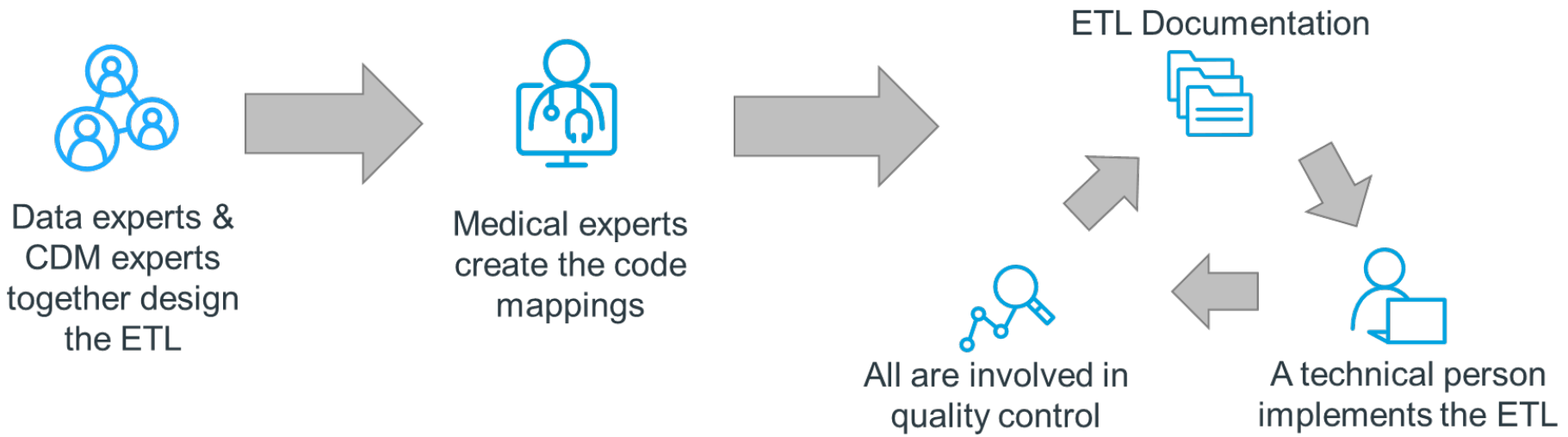
The HIPAA Privacy Rule establishes national standards to protect individuals' medical records and other personal health information and applies to health plans, health care clearinghouses, and those health care providers that conduct certain health care transactions electronically. The Rule requires appropriate safeguards to protect the privacy of personal health information, and sets limits and conditions on the uses and disclosures that may be made of such information without patient authorization. The Rule also gives patients rights over their health information, including rights to examine and obtain a copy of their health records, and to request corrections.

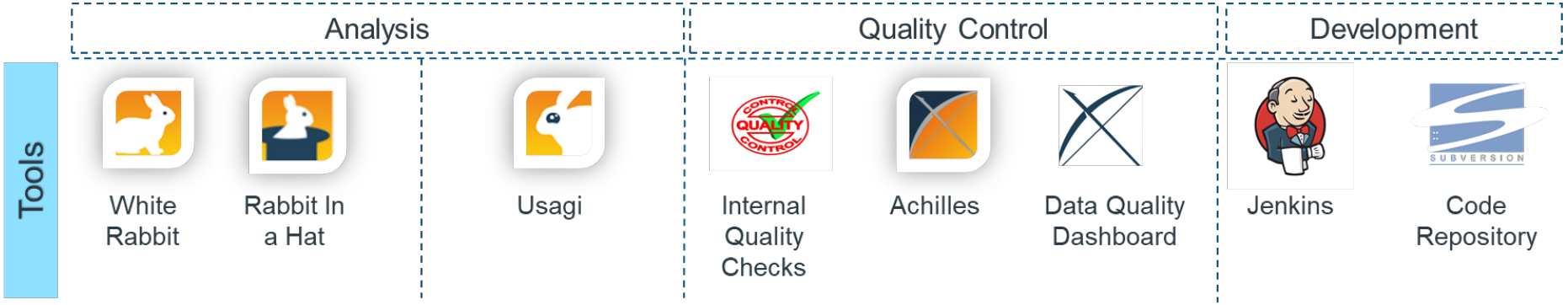
The Privacy Rule is located at 45 CFR Part 160 and Subparts A and E of Part 164.

Click here to view the combined regulation text of all HIPAA Administrative Simplification Regulations found at 45 CFR 160, 162, and 164.

What happens if you need help during developing site extract-transform-load processes but underlying data cannot be shared?

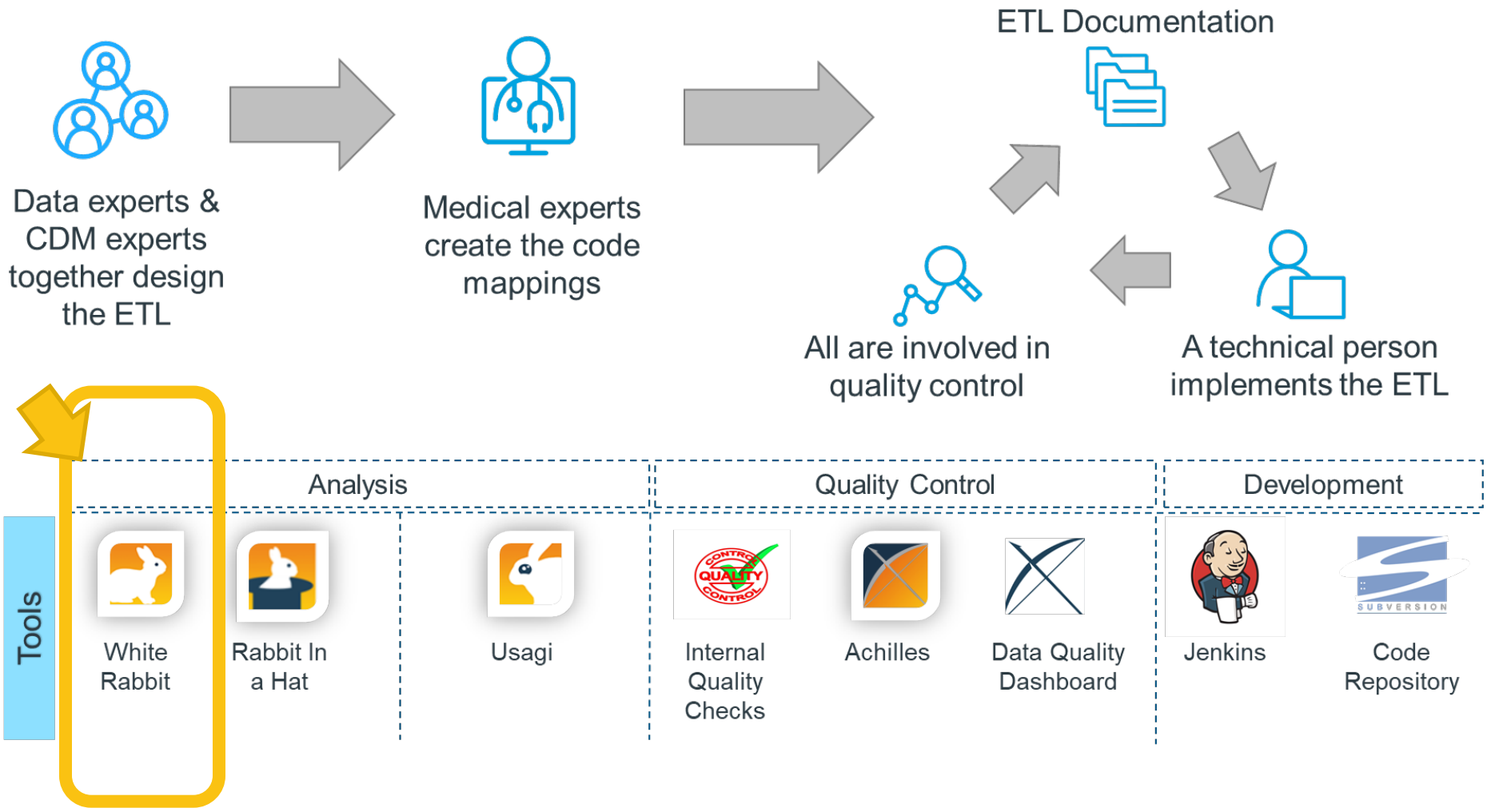
OMOP Conversion Process Flow





Take advantage of your OHDSI toolkit.

OMOP Conversion Process Flow



Leveraging the power of WhiteRabbit

Home › Software › WhiteRabbit for ETL design

WhiteRabbit for ETL design

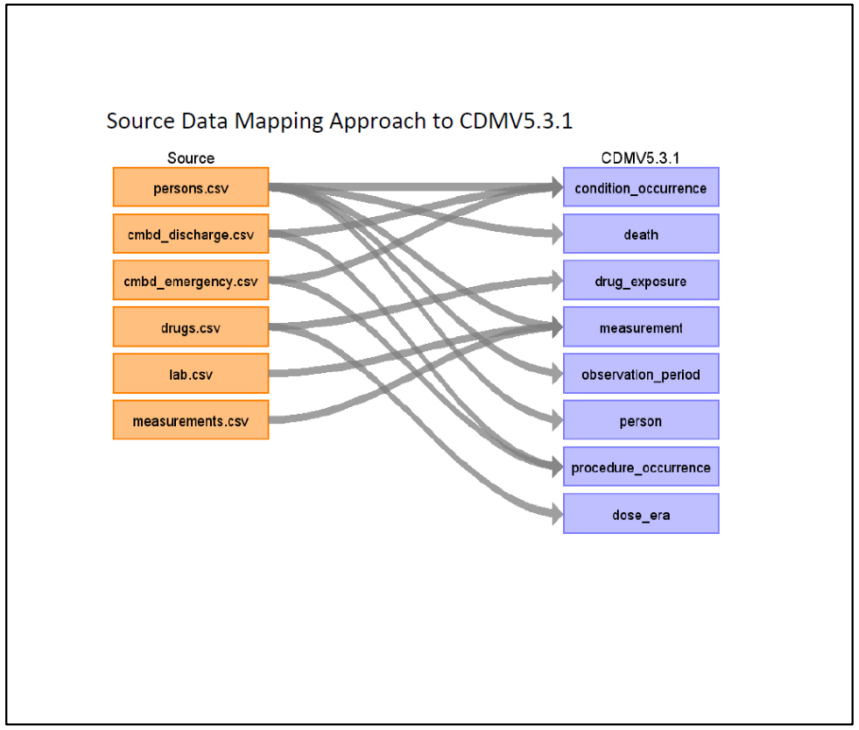
WhiteRabbit is a software tool to help prepare for ETLs (Extraction, Transformation, Loading) of longitudinal healthcare databases into the OMOP Common Data Model (CDM). The source data can be in comma-separated text files, or in a database (MySQL, SQL Server, ORACLE, PostgreSQL); the CDM will be in a database (MySQL, SQL Server, PostgreSQL).

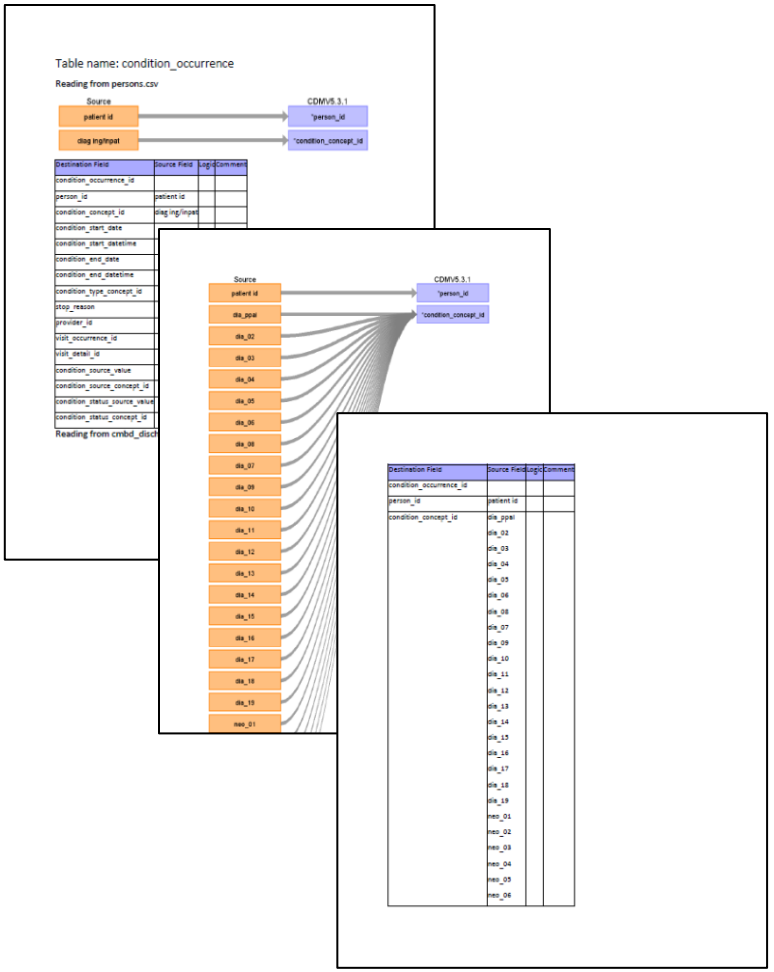
The main function of WhiteRabbit is to perform a scan of the source data, providing detailed information on the tables, fields, and values that appear in a field. This scan will generate a report that can be used as a reference when designing the ETL, for instance when using the Rabbit-In-a-Hat tool. Rabbit-In-a-Hat uses the scan document and displays source data information through a graphical user interface to allow a user to connect source data structure to the CDM data structure. The function of Rabbit-In-a-Hat is to generate documentation for the ETL process, not generate code to create an ETL.

Download WhiteRabbit: https://github.com/OHDSI/WhiteRabbit

Leveraging the power of WhiteRabbit

White Rabbit





Scan reports are our friends

White Rabbit

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Table	Field	Type	Max lengt	N rows	N rows ch	Fraction e	N unique v	Fraction unique values				
2	CMBD_dis	PATIENT II	int	4	-1	1775	0%	1775	100%				
3	CMBD_dis	DIA_PPAL	varchar	8	-1	1775	0%	99	6%				
4	CMBD_dis	DIA_02	varchar	8	-1	1775	0%	151	9%				
5	CMBD_dis	DIA_03	varchar	8	-1	1775	8%	330	19%				
6	CMBD_dis	DIA_04	varchar	8	-1	1775	19%	443	25%				
7	CMBD_dis	DIA_05	varchar	8	-1	1775	30%	431	24%				
8	CMBD_dis	DIA_06	varchar	8	-1	1775	40%	391	22%				
9	CMBD_dis	DIA_08	varchar	8	-1	1775	59%	335	19%				
10	CMBD_dis	DIA_07	varchar	8	-1	1775	50%	360	20%				
11	CMBD_dis	DIA_09	varchar										
12	CMBD_dis	DIA_10	varchar										
13	CMBD_dis	DIA_11	varchar										
14	CMBD_dis	DIA_12	varchar										
15	CMBD_dis	DIA_13	varchar										
16	CMBD_dis	DIA_14	varchar										
17	CMBD_dis	DIA_15	varchar										
18	CMBD_dis	DIA_16	varchar										
19	CMBD_dis	DIA_17	varchar										

Overview | CMBD_

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	PATIENT II	Frequency	DIA_PPAL	Frequency	DIA_02	Frequency	DIA_03	Frequency	DIA_04	Frequency	DIA_05	Frequency	DIA_06	Frequency	DIA_08
2	List trunca		J12.89	1127	B97.29	1240	J96.00	269		335		528		718	
3			J18.9	190	R68.89	125	I10	168	I10	154	I10	107	I10	61	E78.5
4			J98.8	160	Z20.828	67		141	E78.5	72	E78.5	63	E78.5	56	Z87.891
5			J44.0	42	J12.89	48	B97.29	72	E11.9	51	E11.9	40	Z87.891	36	I10
6			J84.9	32	J18.9	21	R68.89	66	N17.9	28	Z87.891	31	E11.9	31	Z79.82
7			J22	19	J96.00	17	J96.90	57	Z87.891	25	N40.0	21	Z79.82	25	Z79.84
8			R91.8	13	J98.8	16	J96.01	54	J96.00	24	R68.89	18	R68.89	20	Z90.49
9			J43.9	12	I10	13	R09.02	47	R68.89	24	K44.9	17	N17.9	20	Z79.01
10			N39.0	10	B97.89	8	E78.5	38	J44.1	24	E03.9	17	E03.9	18	N40.0
11			I26.99	8	R09.02	8	J96.91	26	E03.9	23	E78.00	17	Z90.49	17	K44.9
12			B34.2	7	E11.65	8	J45.901	25	B97.29	16	N17.9	16	Z79.84	17	R68.89
13			J44.1	7	N17.9	5	E11.9	23	E11.65	15	T37.5X5A	15	E66.9	14	I48.91
14			R50.9	6	List trunca		J80	23	F05	15	Z90.49	14	E79.0	14	E78.00
15			J45.901	6			Z87.891	19	N40.0	15	F05	13	N18.9	14	G47.33
16			J80	5			I26.99	17	I26.99	15	J96.00	13	N40.0	12	E11.9
17			Z05.8	5			Z20.828	16	E78.00	15	B97.29	13	F32.9	11	I25.10
18			J47.0	5			N17.9	15	F17.210	14	F32.9	12	I48.2	11	Z99.89
19			List trunca				R74.0	14	J45.901	14	F17.210	11	K44.9	11	E66.9

Overview | CMBD_discharge.csv | CMBD_emergency.csv | Drugs.csv | Lab.csv | Measurements.csv | persons.csv

Don’t be afraid to ask for help!

The power of collaboration in action

White Rabbit

Talita runs WhiteRabbit to generate the mapping document + ScanReport on her new Madrid data.

Along the way she realizes…

¡Necesito ayuda!

Patrick offers to help… except Patrick *no habla español*

Talita can do that! She translates the Scan Report into English and shares her artifacts to Patrick.

Patrick uses this to mock-up a DB and helps Talita write ETL for her database. He shares back the code she can deploy in her environment.

How does this work? This is aggregate results profiling the source data. **No data moves.**

More OHDSI Resources on your ETL journey

You can always phone a friend!

Clair Blacketer, MPH, PMP
https://www.ohdsi.org/who-we-are/collaborators/clair-blacketer/

Erica Voss, MPH
https://www.ohdsi.org/who-we-are/collaborators/erica-voss/

We can help you prioritize data elements!

Timeline	***Elements***
Health pre-COVID19	*Demographics*
	Conditions
	Drugs
	Health service utilization
COVID19 presentation	*Recent health behaviour*
	Measurements/tests/lab
	Conditions
	Drugs
Test time	*COVID19 test presence*
	COVID19 test type
	COVID19 test results
Admission for COVID19	*Inpatient services*
	Complications (eg AKI, sepsis)
	ICU
	Procedures (eg tracheostomy)
	Devices (eg ECMO)
ICU Admission detail	*Date of ICU admission*
	CPAP
	Intubation/mechanical vent
	Procedures (eg tracheostomy)
	Devices (eg ECMO)
Mortality	*Date of death*
	Cause of death

Join the Journey

http://ohdsi.org