

dsOMOP: Integrating OMOP CDM Databases into DataSHIELD

David Sarrat González
Juan R González

ISGlobal
Barcelona
Institute for
Global Health

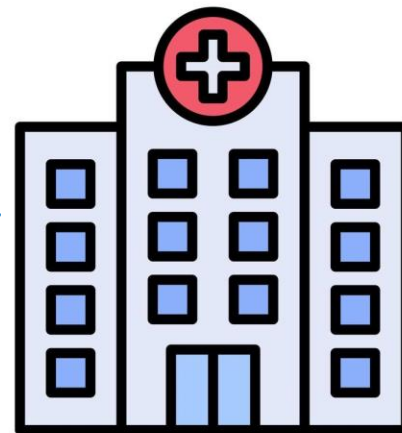


Common Data Models (CDM)

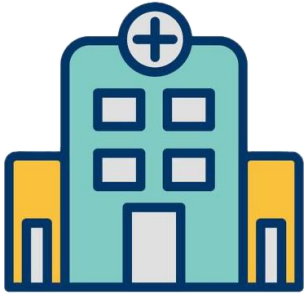


Let's share our
data with
DataSHIELD!

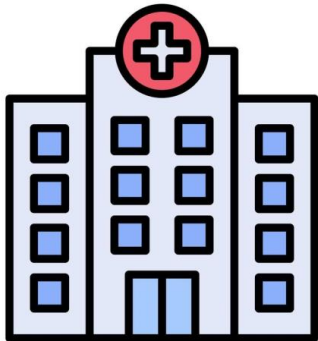
OK!



Common Data Models

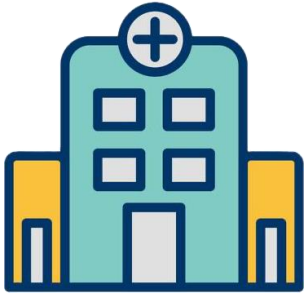


ID	Age	Sex	Current_Smoker	Conditions
1A	65	M	Y	Severe COPD
2A	45	F	N	Mild COPD
3A	55	M	Y	Moderate COPD, Diabetes Type 2

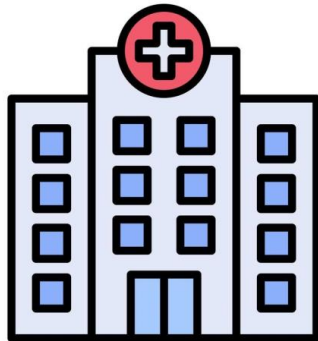
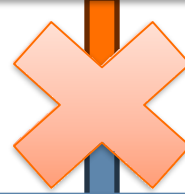


PatientCode	DateOfBirth	Gender	Observations	COPD_Level	Diabetes
B001	12/1/1945	Male	Smoker	High	
B002	5/15/1980	Female			Type 2
B003	11/23/1955	Male	Smoker	Medium	

Common Data Models

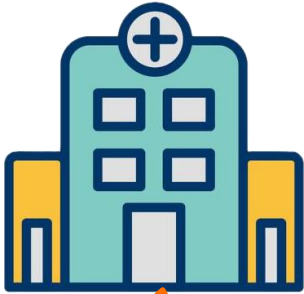


ID	Age	Sex	Current_Smoker	Conditions
1A	65	M	Y	Severe COPD
2A	45	F	N	Mild COPD
3A	55	M	Y	Moderate COPD, Diabetes Type 2

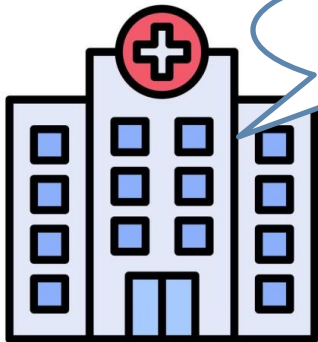


PatientCode	DateOfBirth	Gender	Observations	COPD_Level	Diabetes
B001	12/1/1945	Male	Smoker	High	
B002	5/15/1980	Female			Type 2
B003	11/23/1955	Male	Smoker	Medium	

Common Data Models



How do we fix this?

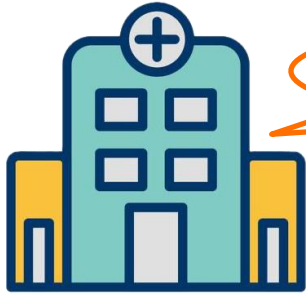


Let's use a CDM!

ID	Age	Sex	Current_Smoker	Conditions
1A	65	M	Y	Severe COPD
2A	45	F	N	Mild COPD
3A	55	M	Y	Moderate COPD, Diabetes Type 2



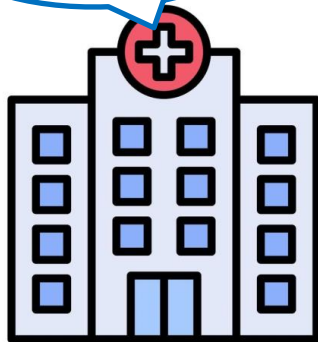
PatientCode	DateOfBirth	Gender	Observations	COPD_Level	Diabetes
B001	12/1/1945	Male	Smoker	High	
B002	5/15/1980	Female			Type 2
B003	11/23/1955	Male	Smoker	Medium	



These will be our CDM rules:

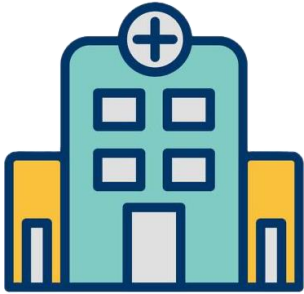


OK!

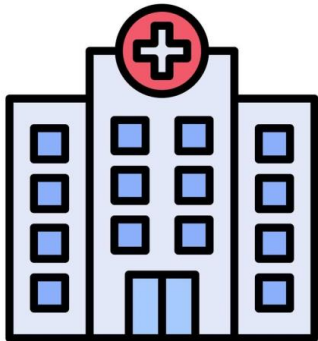


Attribute	Description
Patient_ID	A unique identifier for each patient; numeric string.
Age	Age of the patient in years; integer.
Gender	Gender of the patient; categorical string (" Male " or " Female ").
Smoking_Status	Smoking status of the patient; categorical string (" Current ", " Former ", or " Never ").
COPD_Severity	Severity level of COPD if present; categorical string (" Mild ", " Moderate ", " Severe ", or blank if not applicable).
Diabetes	Type of diabetes if present; categorical string (" Type 1 ", " Type 2 ", or blank if not applicable).

Common Data Models

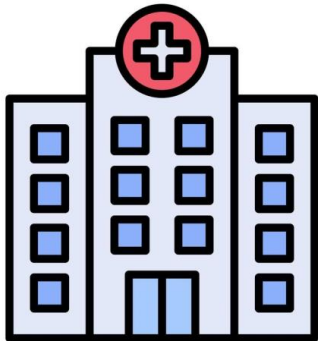
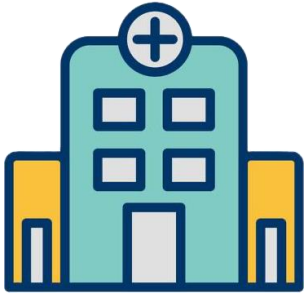


Patient_ID	Age	Gender	Smoking_Status	COPD_Severity	Diabetes
1	65	Male	Current	Severe	
2	45	Female	Never	Mild	
3	55	Male	Current	Moderate	

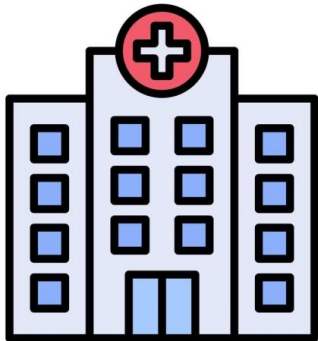
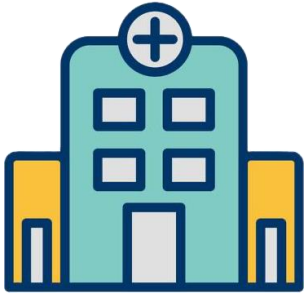


Patient_ID	Age	Gender	Smoking_Status	COPD_Severity	Diabetes
4	77	Male	Current	High	
5	43	Female	Never		Type 2
6	68	Male	Current	Medium	

Common Data Models



Patient_ID	Age	Gender	Smoking_Status	COPD_Severity	Diabetes
1	65	Male	Current	Severe	
2	45	Female	Never	Mild	
3	55	Male	Current	Moderate	
4	77	Male	Current	High	
5	43	Female	Never		Type 2
6	68	Male	Current	Medium	



Patient_ID	Age	Gender	Smoking_Status	COPD_Severity	Diabetes
1	65	Male	Current	Severe	
2	45	Female	Never	Mild	
3	55	Male	Current	Moderate	
4	77	Male	Current	High	
5	43	Female	Never		Type 2
		Male	Current	Medium	



Common Data Models enable **more robust, large-scale** research.

What are the **advantages** of adhering to a widely extended common data model?



What are the **advantages** of adhering to a widely extended common data model?



Universal
interoperability

What are the **advantages** of adhering to a widely extended common data model?



Universal
interoperability



Analytical tools and
scripts designed for
that CDM

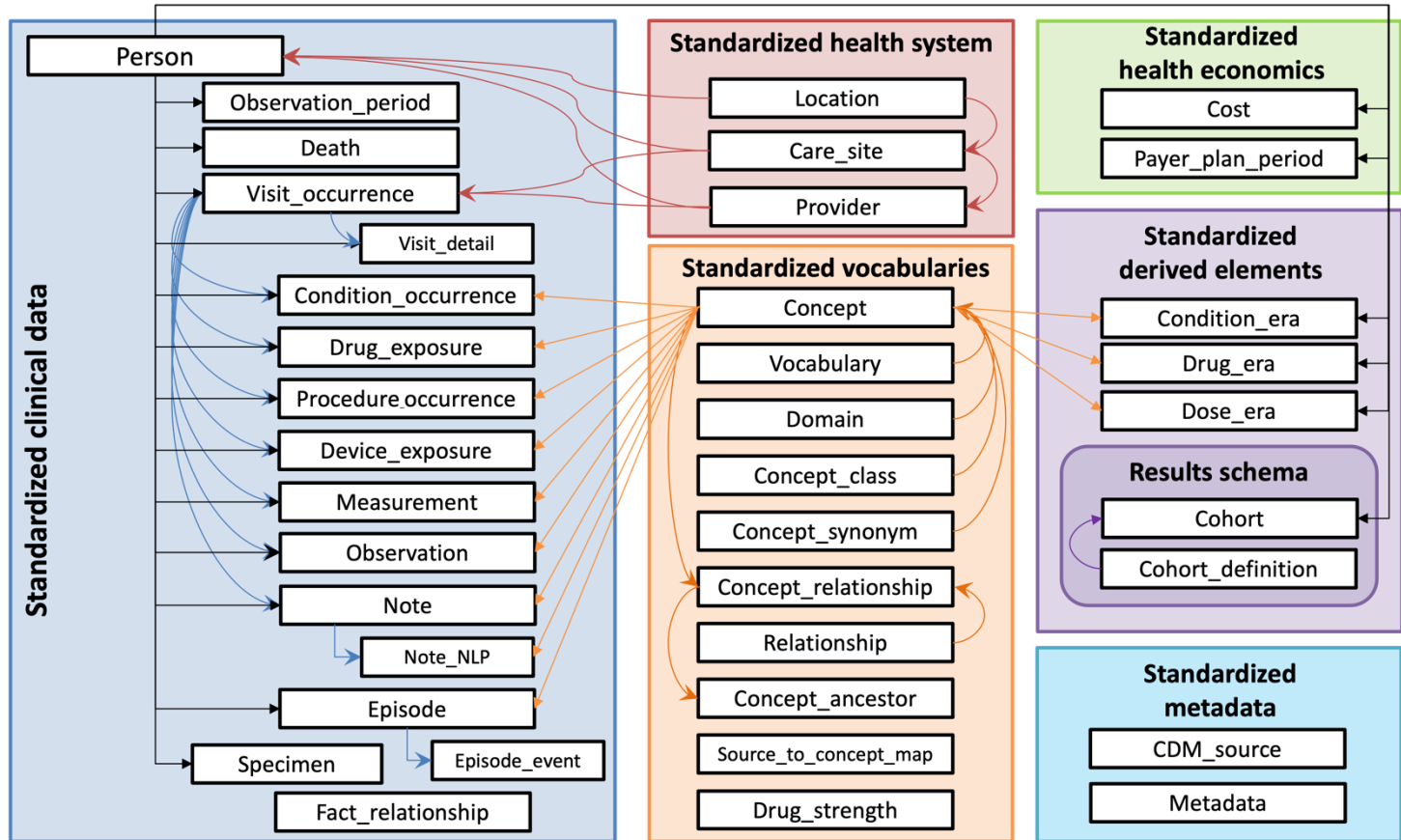
What is the OMOP CDM?

Observational Medical Outcomes Partnership Common Data Model (OMOP CDM)



- ✓ All kinds of clinical research
- ✓ Interdisciplinary collaborative
- ✓ Public, open-source
- ✓ Community developed tools

OMOP CDM structure



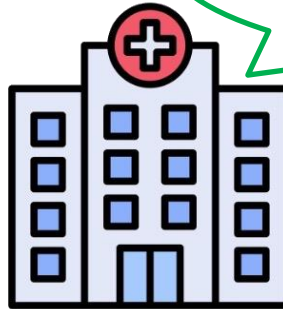
Do you have records
of **Varicella** in your
database?



No, we only have
Chickenpox



No, we only have
水痘



Original DB



Standardized
tables



Drug exposure

Atorvastatin

Loratadine

Acetylsalicylic acid

...



Clinical
vocabularies

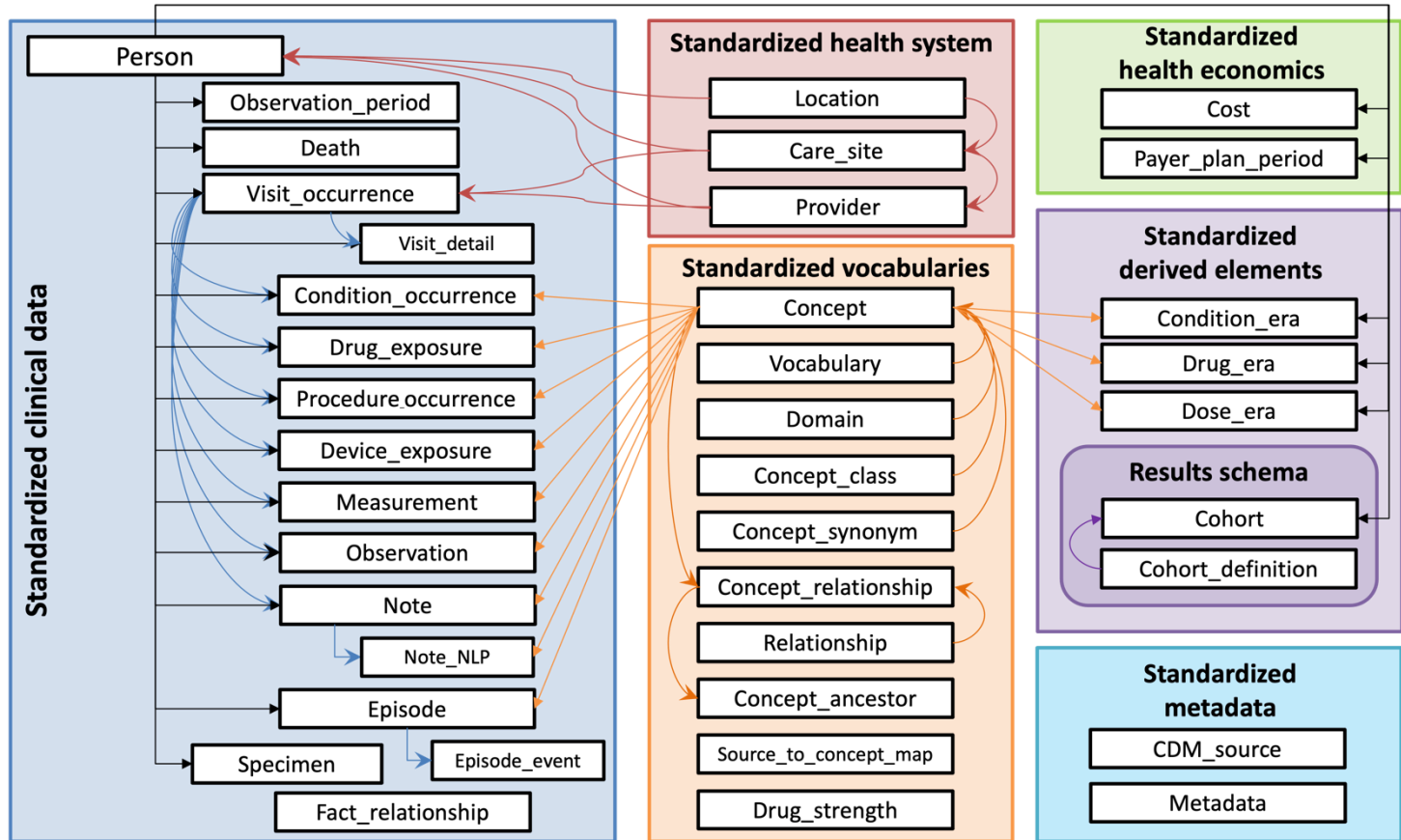
21601860

21603514

21600041

...

OMOP CDM structure



Person table:

person_id	gender_concept_id	year_of_birth
11312	8507 (<i>male</i>)	1985



Person ID: 11312

Person table:

person_id	gender_concept_id
11312	8507 (<i>male</i>)



Person ID: 11312

Measurement table:

measurement_id	person_id	measurement_concept_id	value_as_number
64534	11312	4152194 (<i>SBP</i>)	120

Person table:

person_id	gender_concept_id
11312	8507 (<i>male</i>)



Person ID: 11312

Measurement table:

measurement_id	person_id	measurement_concept_id	value_as_number
64534	11312	4152194 (<i>SBP</i>)	120
76857	11312	4245997 (<i>BMI</i>)	25

Person table:

person_id	gender_concept_id
11312	8507 (<i>male</i>)



Person ID: 11312

Measurement table:

measurement_id	person_id	measurement_concept_id	value_as_number
64534	11312	4152194 (<i>SBP</i>)	120
76857	11312	4245997 (<i>BMI</i>)	25

Observation table:

observation_id	person_id	observation_concept_id	value_as_concept_id
63453453	11312	4005823 (<i>Smoking status</i>)	8515 (<i>Current</i>)

Person table:

person_id	gender_concept_id
11312	8507 (<i>male</i>)



Person ID: 11312

Measurement table:

measurement_id	person_id	measurement_concept_id	value_as_number
64534	11312	4152194 (<i>SBP</i>)	120
76857	11312	4245997 (<i>BMI</i>)	25

Observation table:

observation_id	person_id	observation_concept_id	value_as_concept_id
63453453	11312	4005823 (<i>Smoking status</i>)	8515 (<i>Current</i>)

Condition occurrence table:

condition_occurrence_id	person_id	condition_concept_id
423483	11312	317009 (<i>Asthma</i>)

Person table:

person_id	gender_concept_id	year_of_birth
11312	8507 (<i>male</i>)	1985



Person ID: 11312

Person table:

person_id	gender_concept_id
11312	8507 (male)



Person ID: 11312

Measurement table:

measurement_id	person_id	measurement_concept_id	value_as_number	measurement_date
64534	11312	4152194 (SBP)	120	2022-01-01

Person table:

person_id	gender_concept_id
11312	8507 (male)



Person ID: 11312

Measurement table:

measurement_id	person_id	measurement_concept_id	value_as_number	measurement_date
64534	11312	4152194 (SBP)	120	2022-01-01
89851	11312	4152194 (SBP)	125	2022-04-14

Person table:

person_id	gender_concept_id
11312	8507 (male)



Person ID: 11312

Measurement table:

measurement_id	person_id	measurement_concept_id	value_as_number	measurement_date
64534	11312	4152194 (SBP)	120	2022-01-01
89851	11312	4152194 (SBP)	125	2022-04-14
124352	11312	4152194 (SBP)	130	2022-07-22

Person table:

person_id	gender_concept_id
11312	8507 (male)



Person ID: 11312

Measurement table:

measurement_id	person_id	measurement_concept_id	value_as_number	measurement_date
64534	11312	4152194 (SBP)	120	2022-01-01
89851	11312	4152194 (SBP)	125	2022-04-14
124352	11312	4152194 (SBP)	130	2022-07-22
138176	11312	4152194 (SBP)	128	2022-11-15

**The goal: combining
DataSHIELD and OMOP CDM**

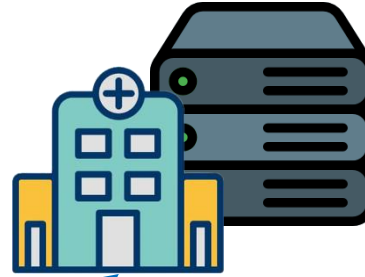
Distributed analysis



DataSHIELD
client



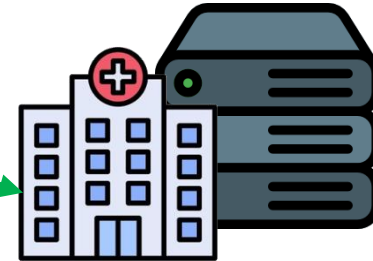
Researcher



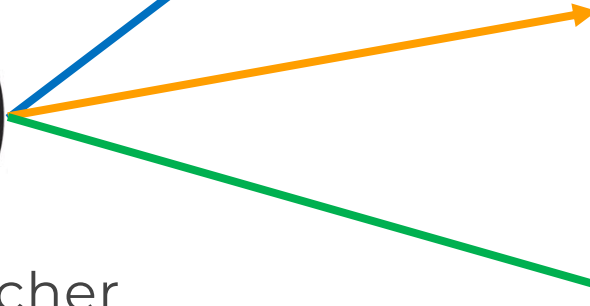
Server 1

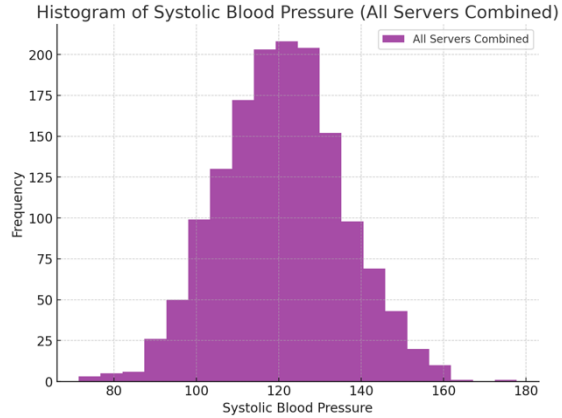


Server 2



Server 3

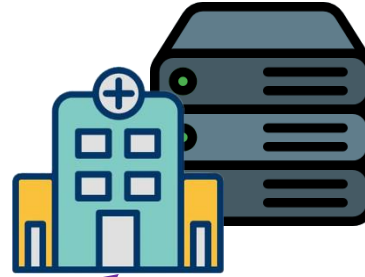




DataSHIELD
client



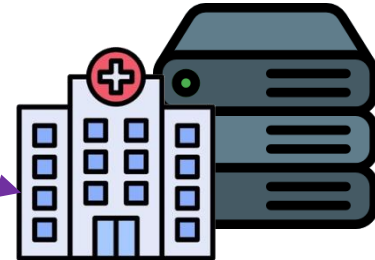
Researcher



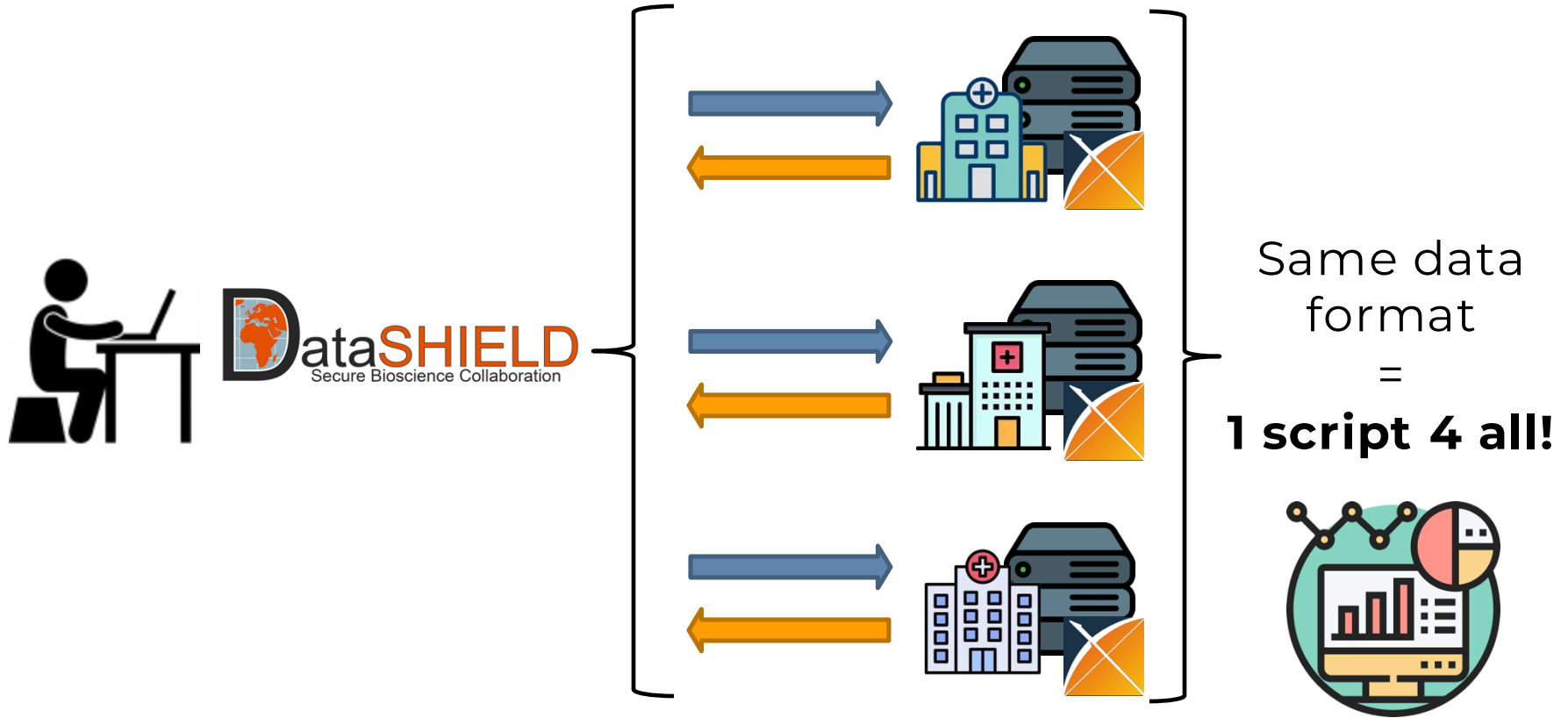
Server 1



Server 2



Server 3



Pushing the boundaries of DataSHIELD



DataSHIELD
server

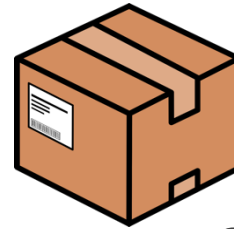
**Let's say we want to incorporate a
new functionality to DataSHIELD**



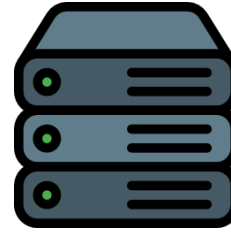
Researcher



Researcher



Server package with
new functionalities

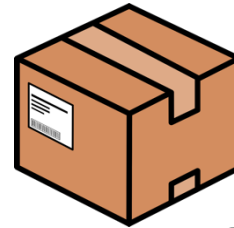


DataSHIELD
server

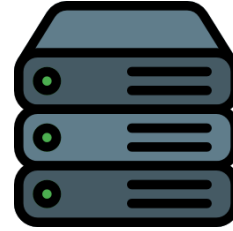




Researcher



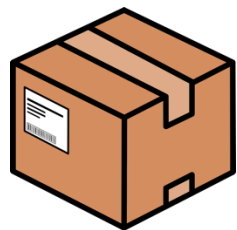
Server package with
new functionalities



DataSHIELD
server



The server can
now execute
new methods



Client package that
can call the server
package functions



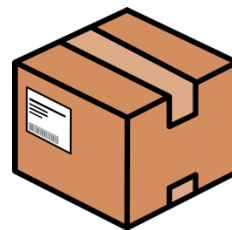
Researcher



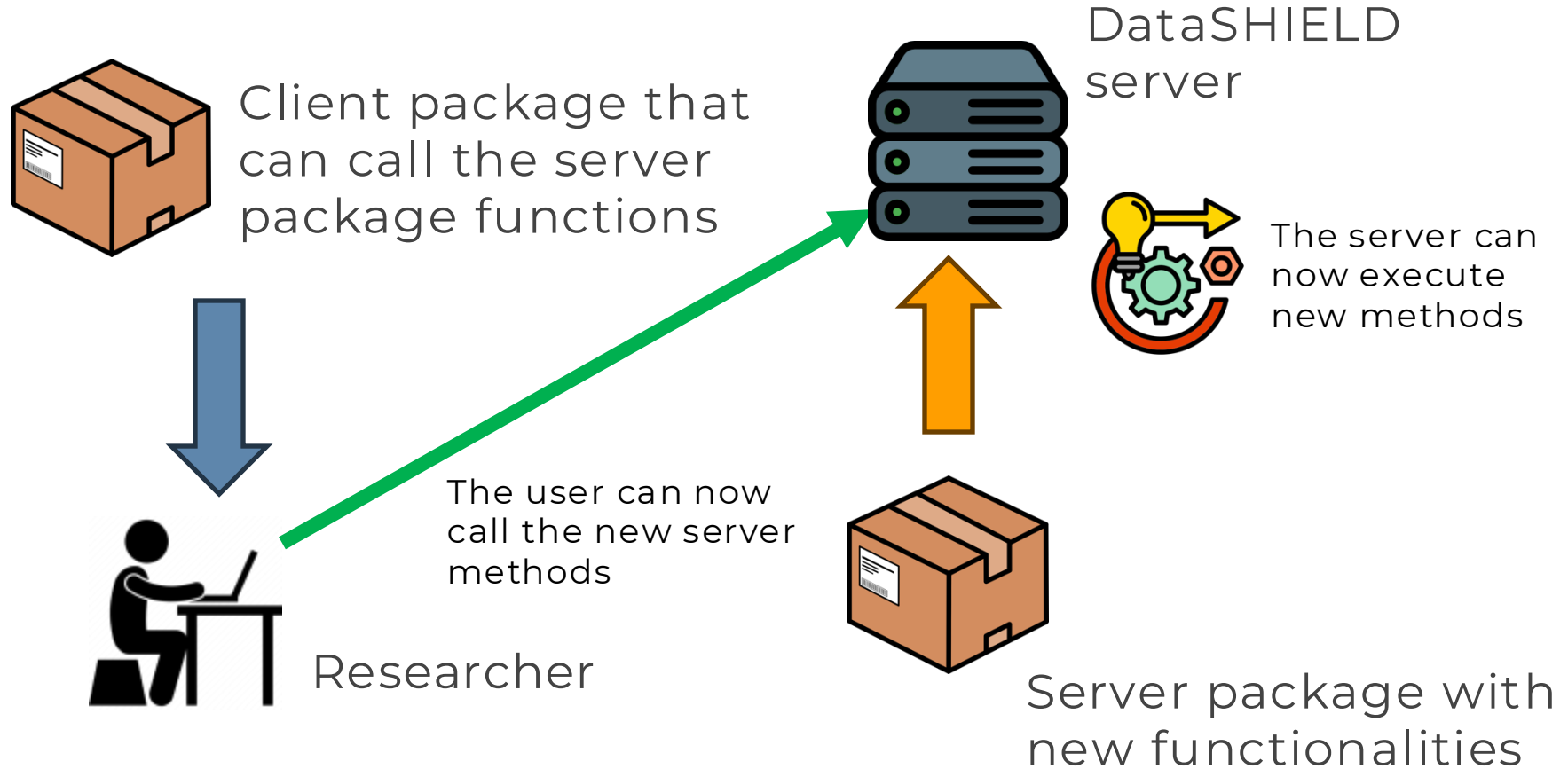
DataSHIELD
server

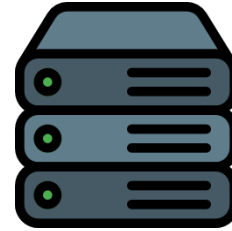


The server can
now execute
new methods

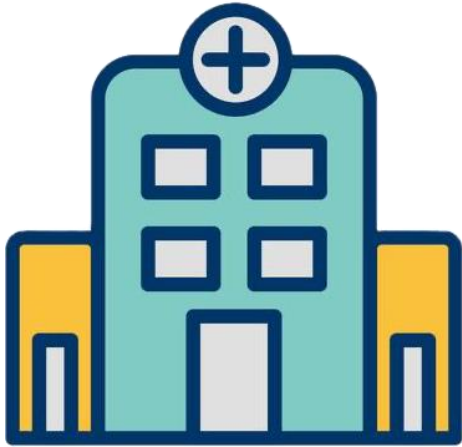


Server package with
new functionalities





DataSHIELD
server



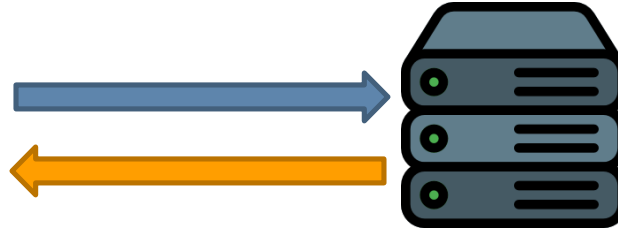
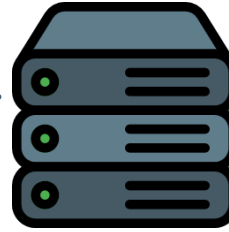
Medical data owner



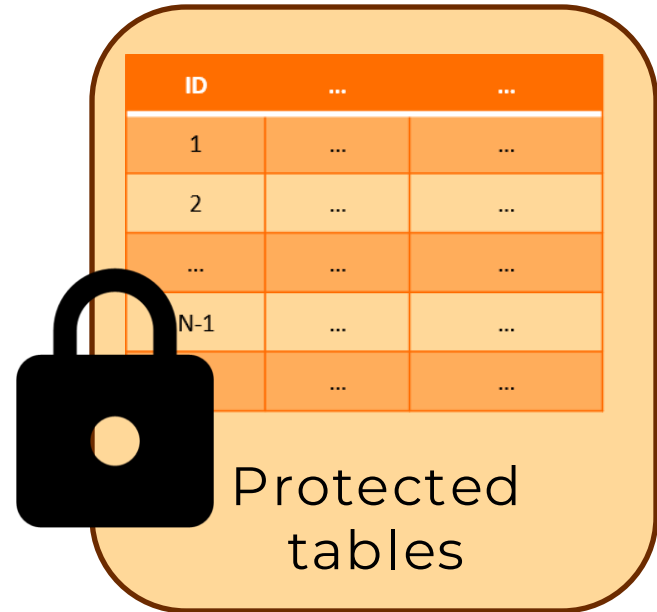
Data
uploading

ID
1
2
...
N-1
N

Data tables

DataSHIELD
clientDataSHIELD
serverServer's R
session

Researcher

Protected
tables

DataSHIELD
client



Resources

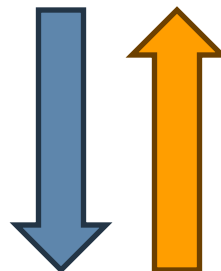
DataSHIELD
server



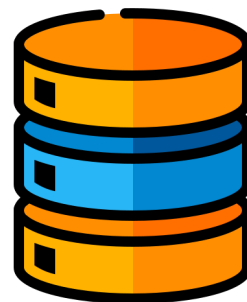
Server's R
session



Researcher



Resources



Any kind of
external data

Server-side session

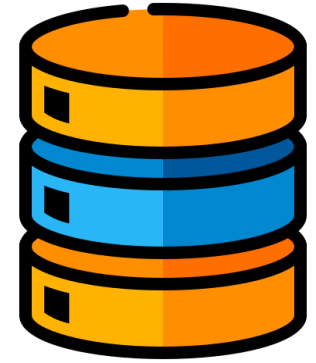


DataSHIELD's base
server-side functions

Client-side session



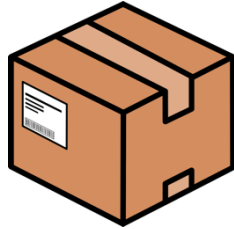
DataSHIELD's base
client-side functions



OMOP CDM
Database

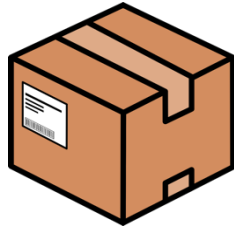
There was no native
implementation to
connect to and map
OMOP CDM databases!

Server-side session



dsOMOP

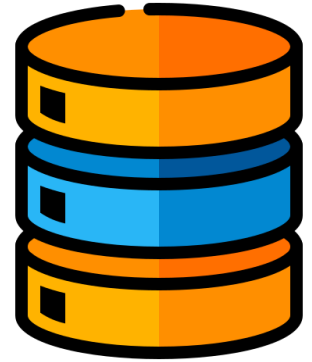
Client-side session



dsOMOP
Client



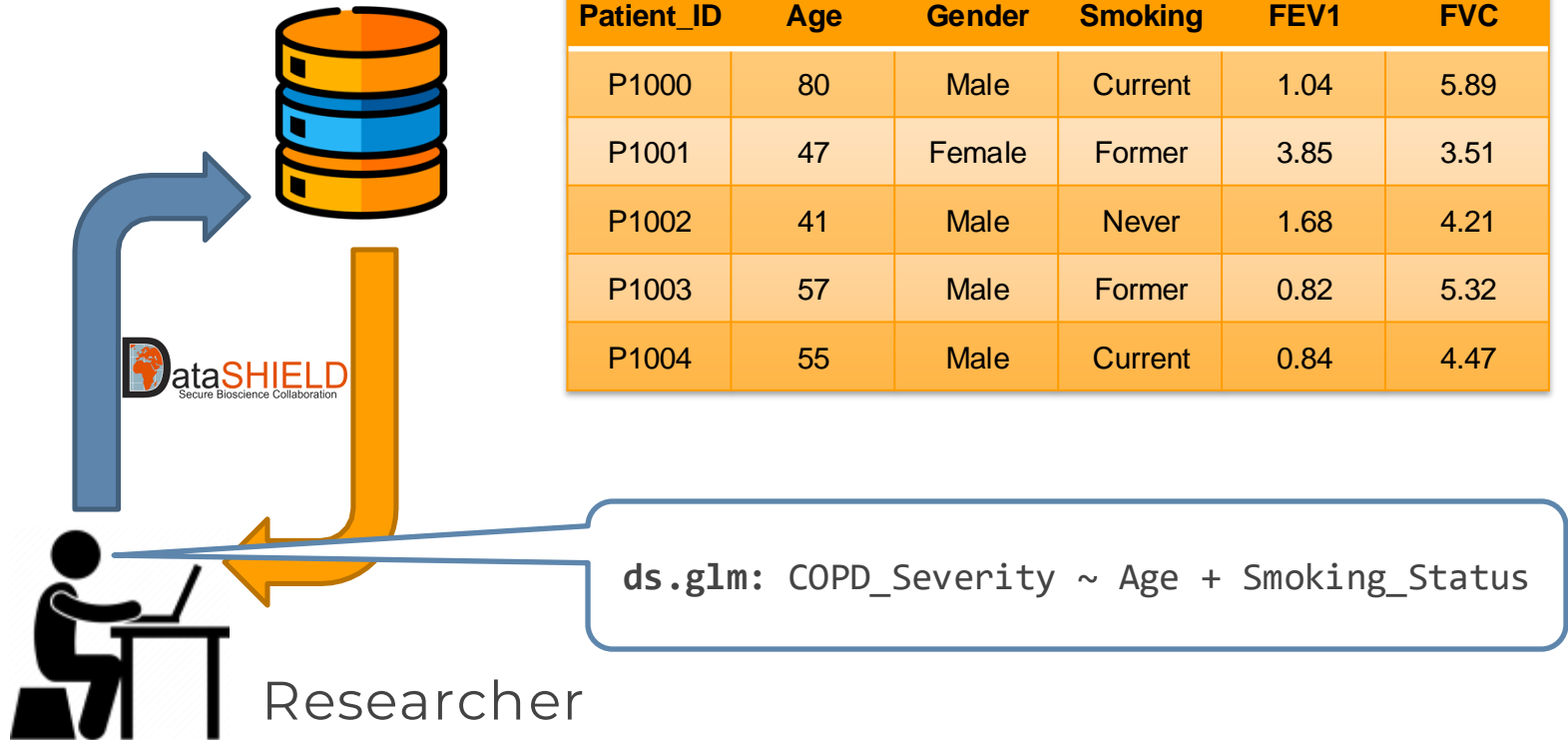
OMOP CDM
Resource



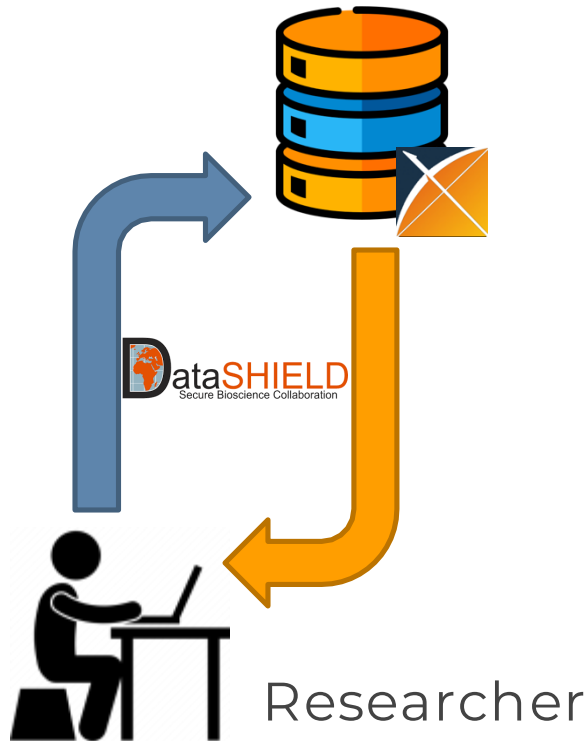
OMOP CDM
Database

**We fixed it with
dsOMOP!**

Complexity



Complexity



Person table:

person_id	year_of_birth	gender_concept_id
1	1940	8507
2	1973	8532
...

gender_concept_id: 8507 for Male, 8532 for Female

Observation table:

person_id	observation_concept_id	value_as_concept_id
1	4005823	8515
2	4005823	8516
...

observation_concept_id: 4005823 for Smoking Status

value_as_concept_id: 8515 = Current, 8516 = Former, 8517 = Never...

Measurement table:

person_id	measurement_concept_id	value_as_number
1	3023540	1.04
1	3025315	5.89
...

measurement_concept_id: 3023540 for FEV1, 3025315 for FVC...

Complexity

I will join "Person" and "Observation" tables on "person_id", then select "year_of_birth" from "person" for "Age", select "value_as_concept_id" from "Observation" for "Smoking_Status", then filter rows by "observation_concept_id" 4005823 for "Smoking Status", then I will get a "value_as_concept_id" where 8515 is "Current", 8516 is "Former", 8517 is "Never", and then...



Researcher

Person table:

person_id	year_of_birth	gender_concept_id
1	1940	8507
2	1973	8532
...

gender_concept_id: 8507 for Male, 8532 for Female

Observation table:

person_id	observation_concept_id	value_as_concept_id
1	4005823	8515
2	4005823	8516
...

observation_concept_id: 4005823 for Smoking Status

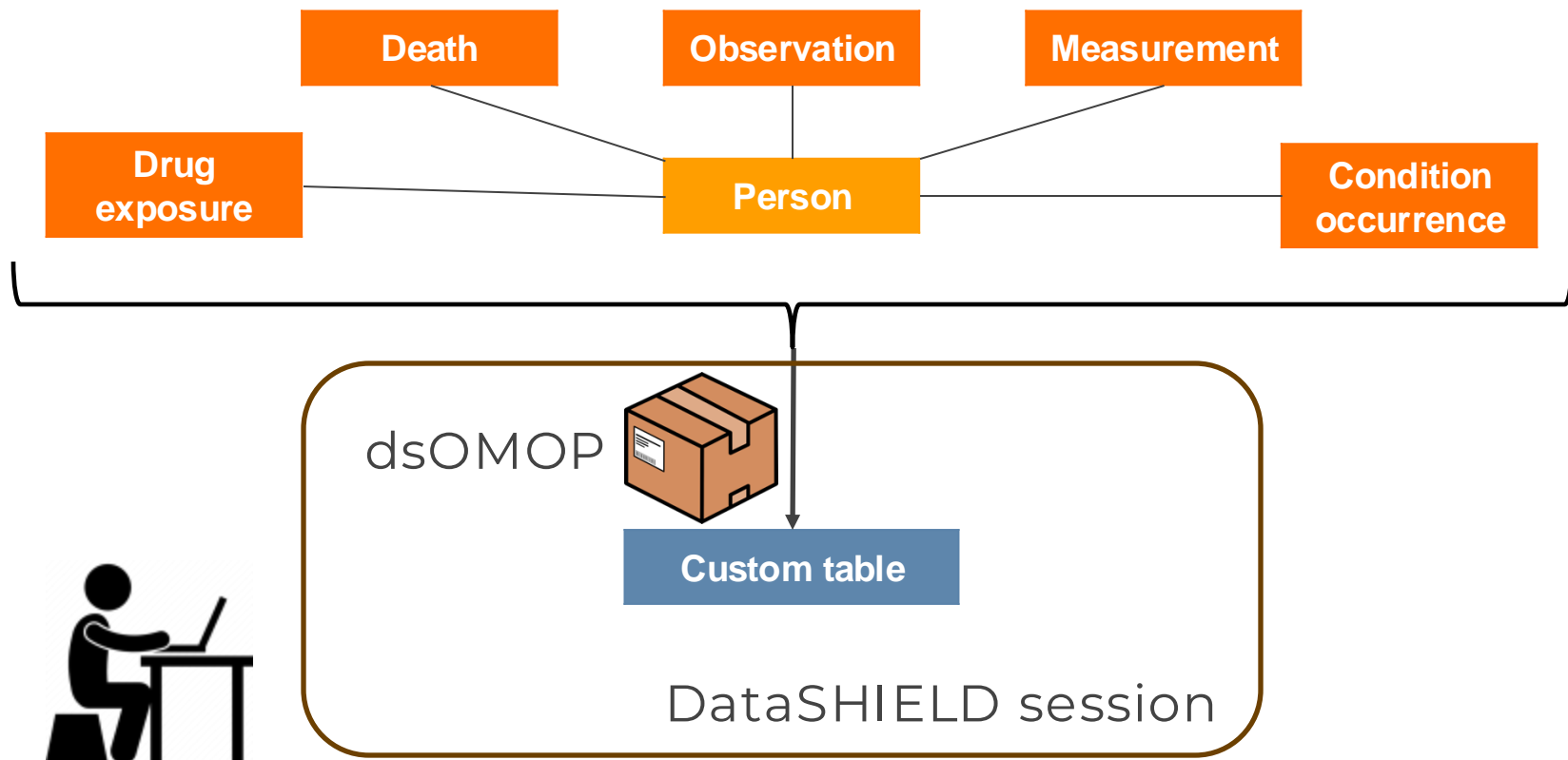
value_as_concept_id: 8515 = Current, 8516 = Former, 8517 = Never...

Measurement table:

person_id	measurement_concept_id	value as number
1	3023540	1.04
1	3025315	5.89
...

measurement_concept_id: 3023540 for FEV1, 3025315 for FVC...

dsOMOP's workflow



On-demand custom dataset creation!

Concept table



concept_id

concept_name

person_id	gender	4221442	4214963
1	8507	45877984	8517008
2	8532	4221688	4214964
3	8507	4221334	4214965
4	8532	45877984	4214964
5	8507	4221688	8517008

Patient_ID	Gender	COPD_Severity	Smoker
1	Male	Mild	Current
2	Female	Moderate	Never
3	Male	Severe	Former
4	Female	Mild	Never
5	Male	Moderate	Current

Introducing the dsOMOP framework



David Sarrat
González



Xavier Escribà
Montagut



Juan R.
González



- ✓ Enables integration with OMOP CDM databases
- ✓ Open source
- ✓ OMOP CDM-specific security checks
- ✓ User-friendly design
- ✓ Supports development of extensions



<https://isglobal-brge.github.io/dsOMOP>



dsOMOP extensions

dsOMOP

Server-Side DataSHIELD Integration for OMOP CDM Databases

This package facilitates interaction with remote databases in the OMOP CDM format from a DataSHIELD environment. It is responsible for fetching and transforming data from databases into a user-intelligible table format, integrated into the DataSHIELD workflow to ensure compliance with the DataSHIELD security model.

 [GitHub](#)

dsOMOPClient

Client-Side DataSHIELD Integration for OMOP CDM Databases

This package facilitates interaction with remote databases in the OMOP CDM format from a DataSHIELD environment. It enables users to invoke server-side functions that perform fetching and transforming of data from OMOP CDM databases, integrating these operations into the DataSHIELD workflow to maintain adherence to the DataSHIELD security model.

 [GitHub](#)

 [User guide](#)

dsOMOPHelper

dsOMOP Helper Functions

This package provides a set of functions to help the user to work with the dsOMOPClient package in DataSHIELD. It provides plug-and-play functionalities for data selection and fetching, streamlining interactions with most simple use cases.

 [GitHub](#)

 [User guide](#)

dsOMOP.oracle

dsOMOP Oracle Extension

Extends the functionality of the dsOMOP package to support OMOP CDM databases in Oracle. Requires the oracle.resourcer package.

 [GitHub](#)



Setting up a connection to a DB:

Add Resource

Category
OMOP CDM

The resource is in OMOP CDM format.

Type
OMOP CDM Database

dsOMOP - OMOP CDM Database Resources
Connection to an OMOP CDM database. The connection will be established using DBI.

Parameters Credentials

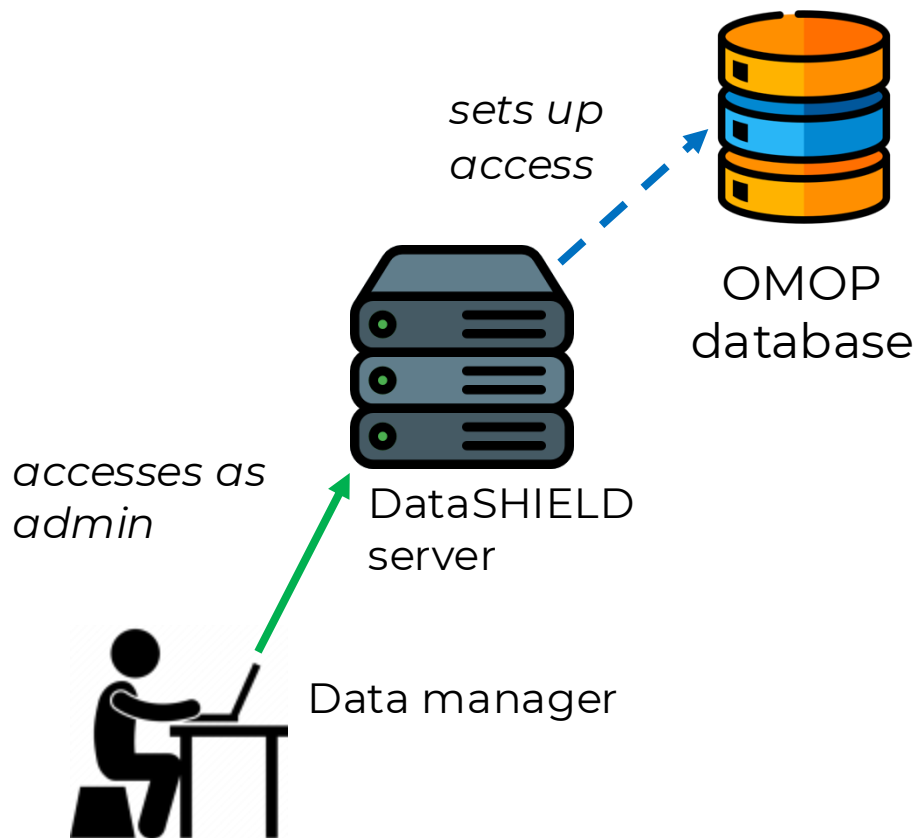
Database engine
PostgreSQL

Host name or IP address
localhost

Port number
5432

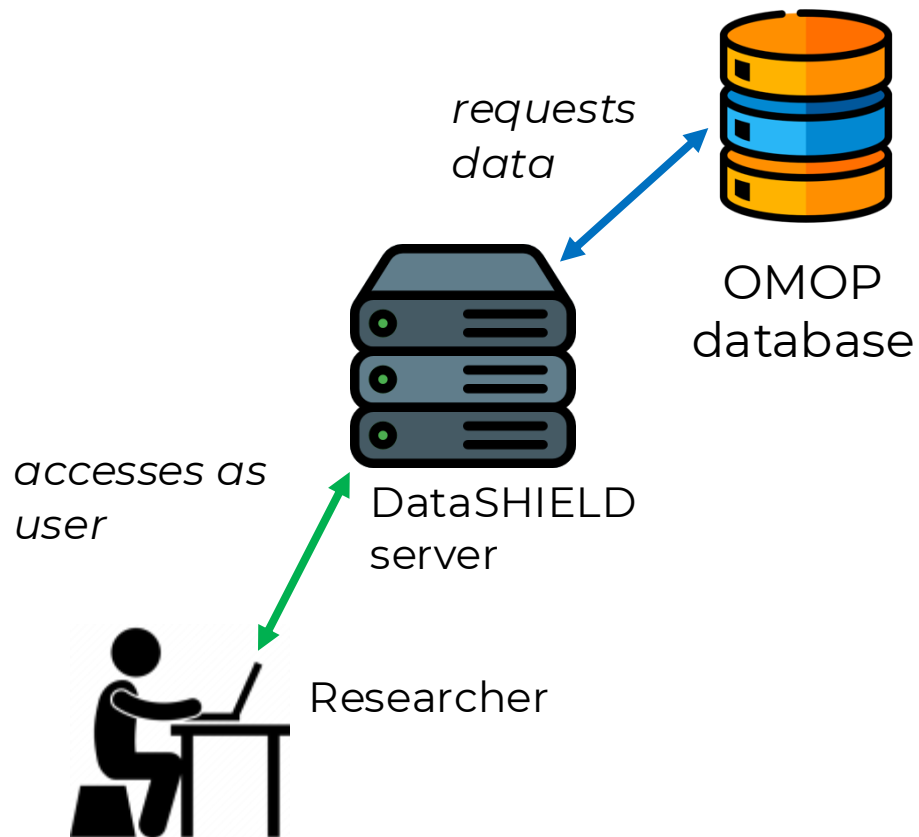
Database name
my_database

Save Cancel



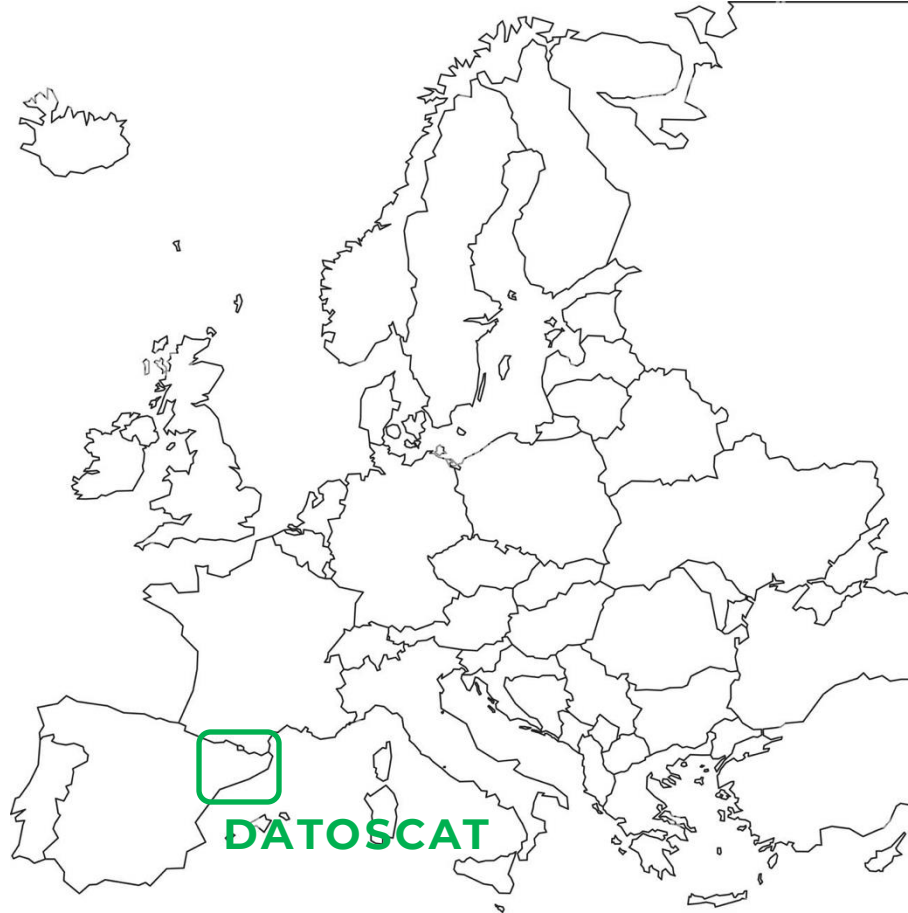
Available data catalog:

Code	Concept
4237017	Genetic test
4245261	Prothrombin time
4246053	Blood test
4261836	Thyroid panel
4326419	HIV-1 ELISA assay



dsOMOP's adoption status

Projects using dsOMOP





P4COPD

CADSET



**Let's proceed with a practical
example!**