



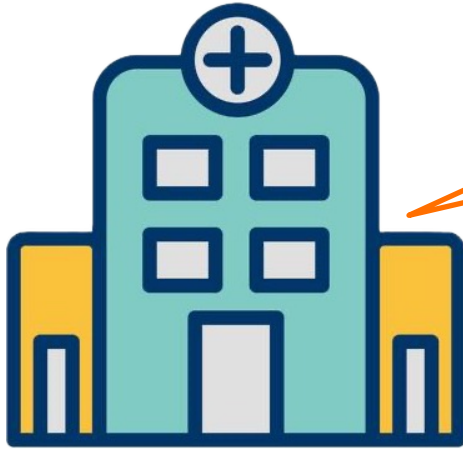
dsOMOP: bridging OMOP CDM and DataSHIELD for secure federated analysis of standardized clinical data

David Sarrat González
Juan R González

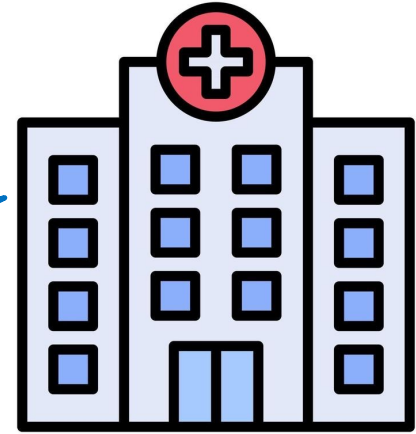


The problems of medical data sharing

The problems of data sharing



Let's share
our data!



OK!



⊘ Legal and regulatory compliance



⊘ **Legal and regulatory compliance**

⊘ **Skepticism from data owners / control**





⊘ **Legal and regulatory compliance**

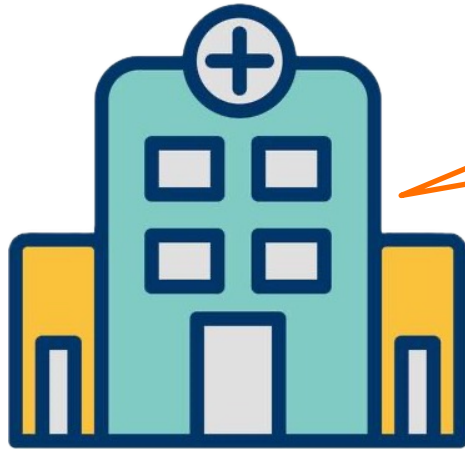
⊘ **Skepticism from data owners / control**



⊘ **Bureaucracy and administrative complexity / Disagreements**

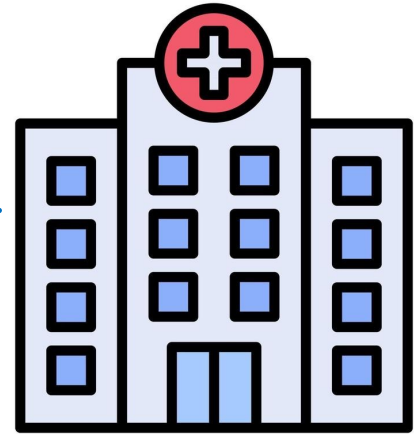
The traditional solution to share data

Traditional data sharing

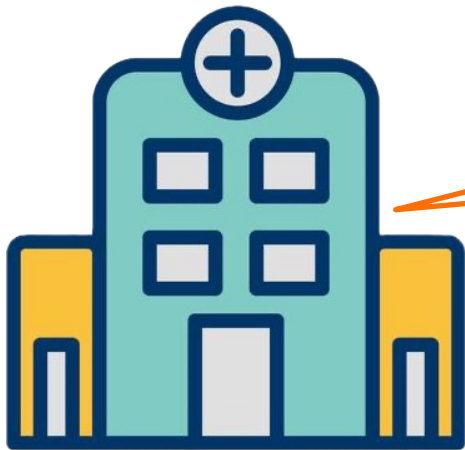


Can you
perform these
operations on
your data?

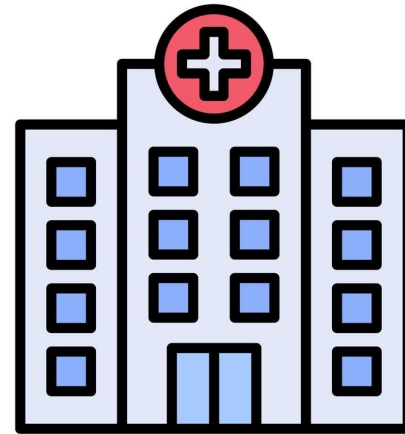
Sure!



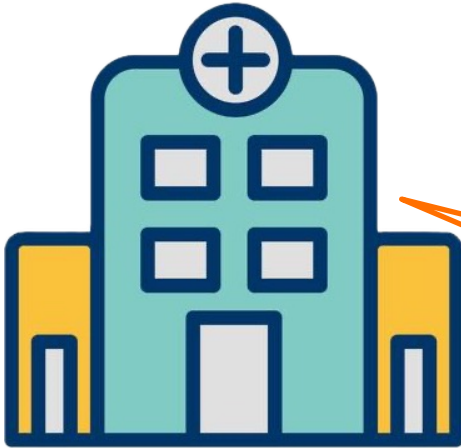
 **It's slow**



This is taking
too long...

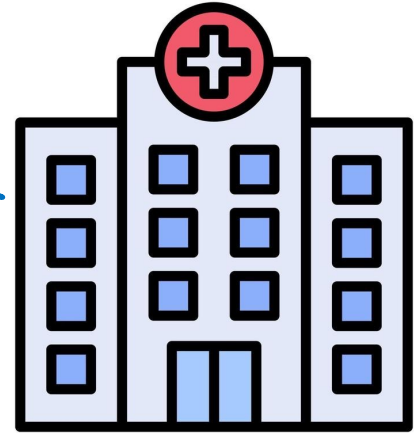


🚫 **It's prone to errors**



I finally have
your results!

I think these
are wrong...



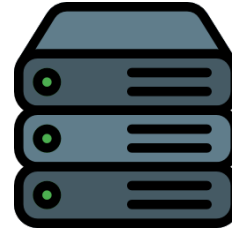
Federated data analysis with DataSHIELD



- ✓ ***Federated*** clinical data analysis platform
- ✓ Open source
- ✓ Widely recognized and used
- ✓ Focus on privacy and security



Medical data owner



DataSHIELD
server



Data
uploading

ID
1
2
...
N-1
N

Clinical data



DataSHIELD
server

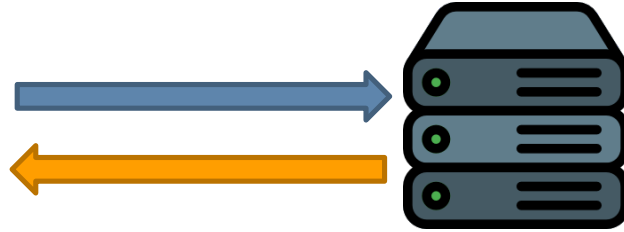


Medical data owner

You can now
connect and
interact with
my data!



DataSHIELD
client



DataSHIELD
server



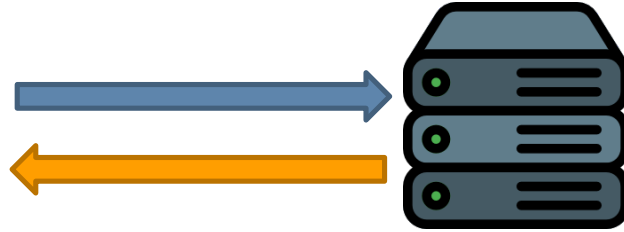
Server's R
session



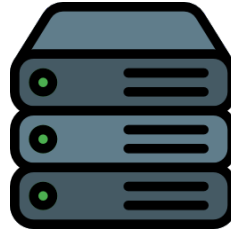
ID
1
2
...
N-1
...

Protected
data

DataSHIELD
client



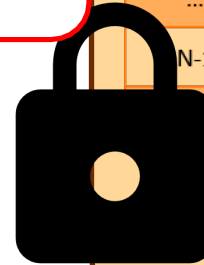
DataSHIELD
server



Server's R
session

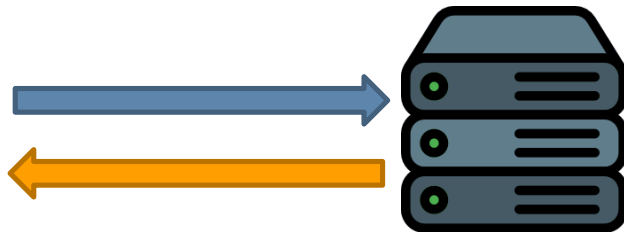


DataSHIELD **only** provides
the necessary statistical
information!



ID
1
2
...
N-1
...

Protected
data

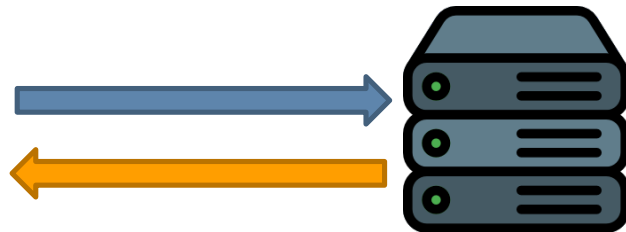
DataSHIELD
clientDataSHIELD
serverServer's R
session

Hey! You can't
have access to
individual patient
information!



ID
1
2
...
N-1
...

Protected
data

DataSHIELD
clientDataSHIELD
serverServer's R
session

I won't give you the
results because
these are related to
very few patients!



ID
1
2
...
N-1
...

Protected
data

DataSHIELD
clientDataSHIELD
serverServer's R
sessionLocal R
session

Query



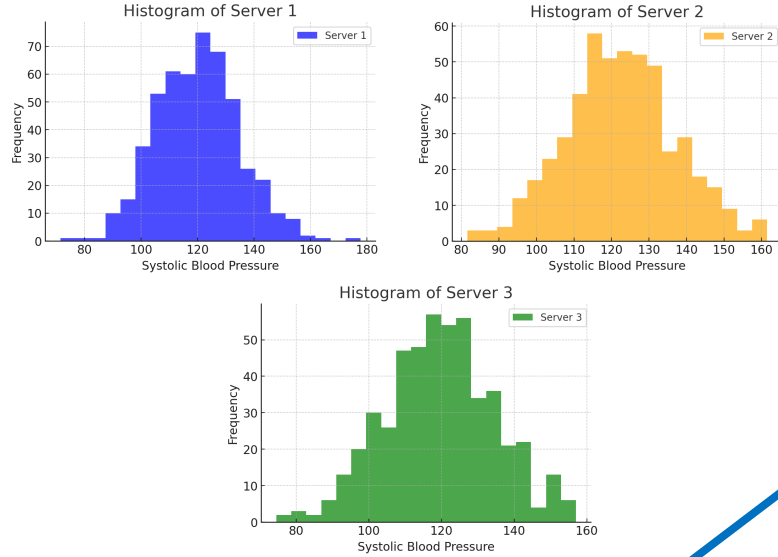
Researcher

OK, here are your
results, but I will
add **noise** so you
can't figure out the
original values!

ID
1
2
...
N-1
...

Protected
data

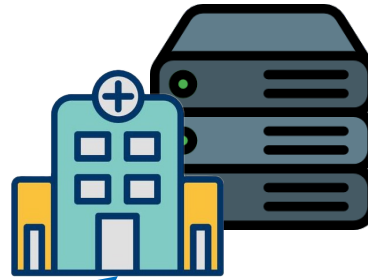
Distributed analysis



DataSHIELD
client



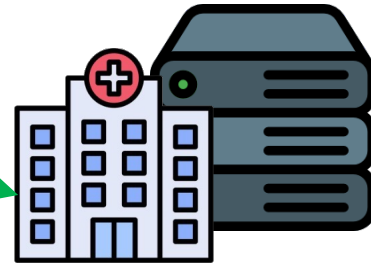
Researcher



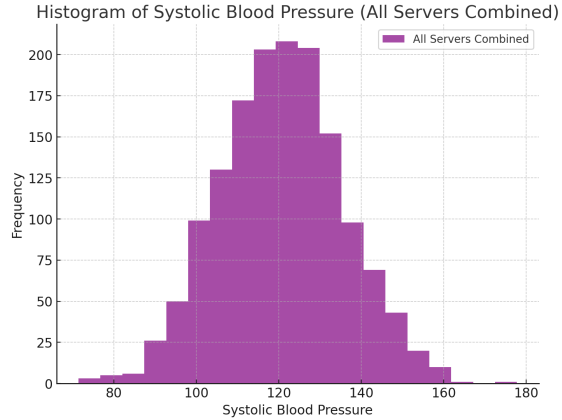
Server 1



Server 2



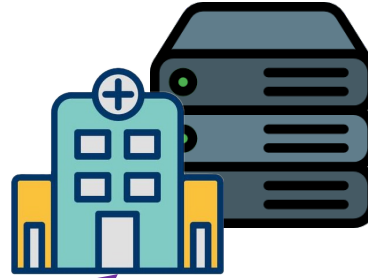
Server 3



DataSHIELD
client



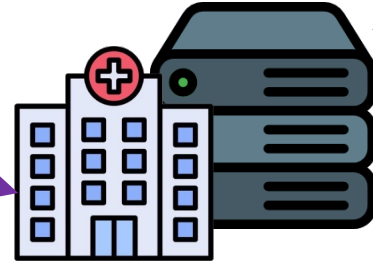
Researcher



Server 1



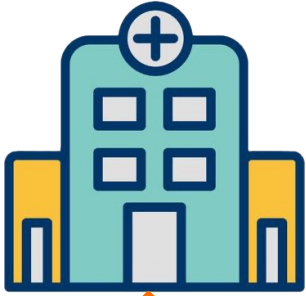
Server 2



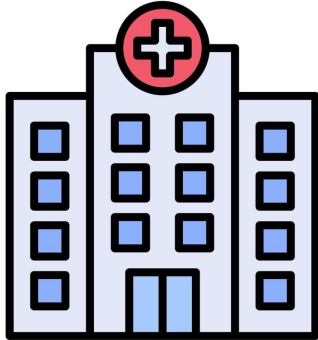
Server 3

Combination with OMOP CDM

Need for a CDM



We can now collaborate!

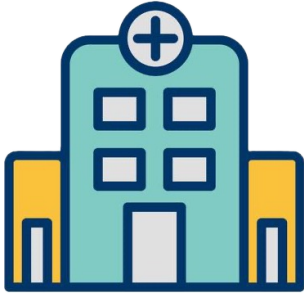


ID	Age	Sex	Current_Smoker	Conditions
1A	65	M	Y	Severe COPD
2A	45	F	N	Mild COPD
3A	55	M	Y	Moderate COPD, Diabetes Type 2

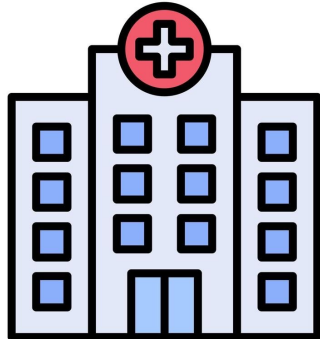
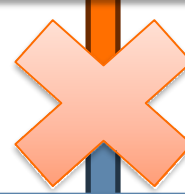


PatientCode	DateOfBirth	Gender	Observations	COPD_Level	Diabetes
B001	12/1/1945	Male	Smoker	High	
B002	5/15/1980	Female			Type 2
B003	11/23/1955	Male	Smoker	Medium	

Need for 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

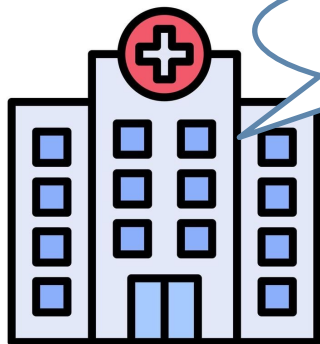


PatientCode	DateOfBirth	Gender	Observations	COPD_Level	Diabetes
B001	12/1/1945	Male	Smoker	High	
B002	5/15/1980	Female			Type 2
B003	11/23/1955	Male	Smoker	Medium	

Need for a CDM

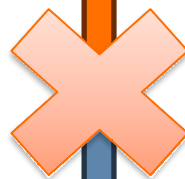


How do we
fix this?

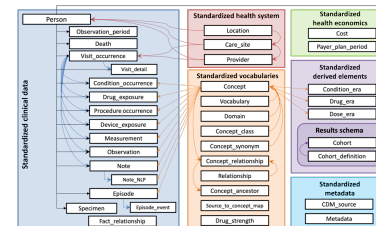
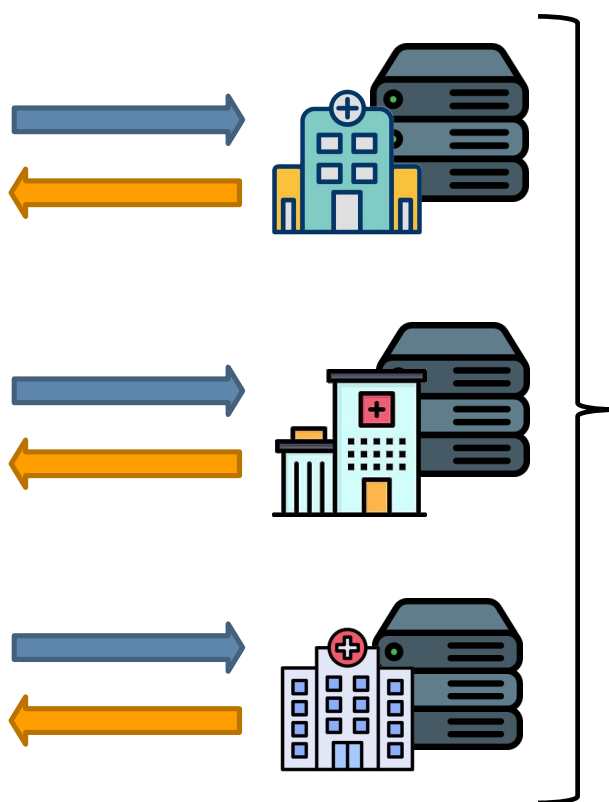


Let's use
OMOP CDM!

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



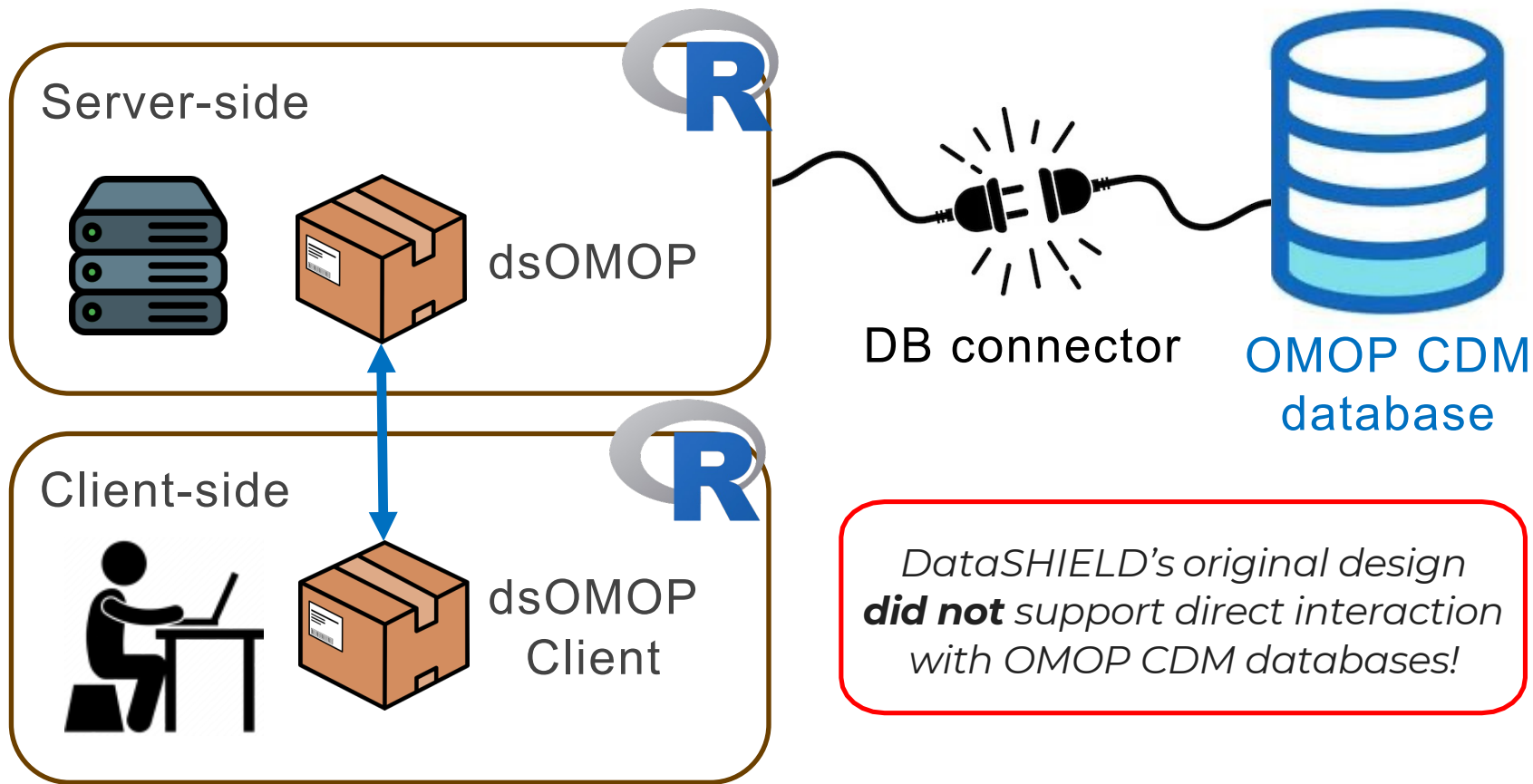
PatientCode	DateOfBirth	Gender	Observations	COPD_Level	Diabetes
B001	12/1/1945	Male	Smoker	High	
B002	5/15/1980	Female			Type 2
B003	11/23/1955	Male	Smoker	Medium	



Same data
model
=
1 script 4 all!

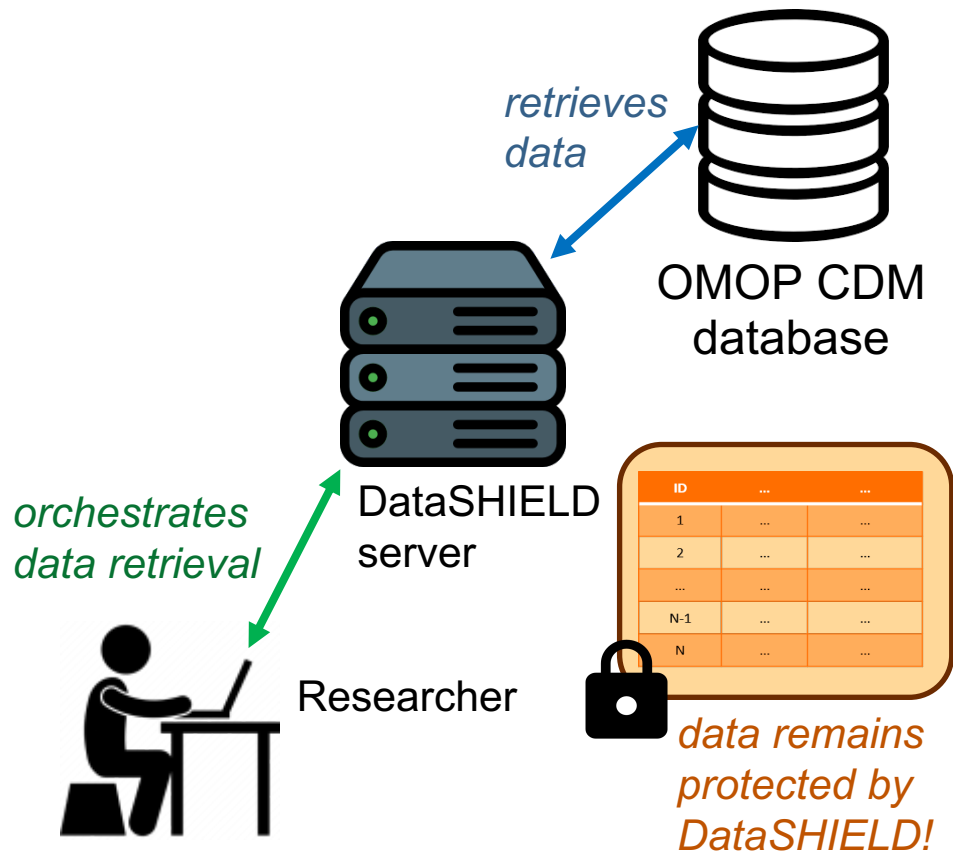


Development of dsOMOP



Generated data catalog:

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





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



David Sarrat
González



Xavier Escribà
Montagut



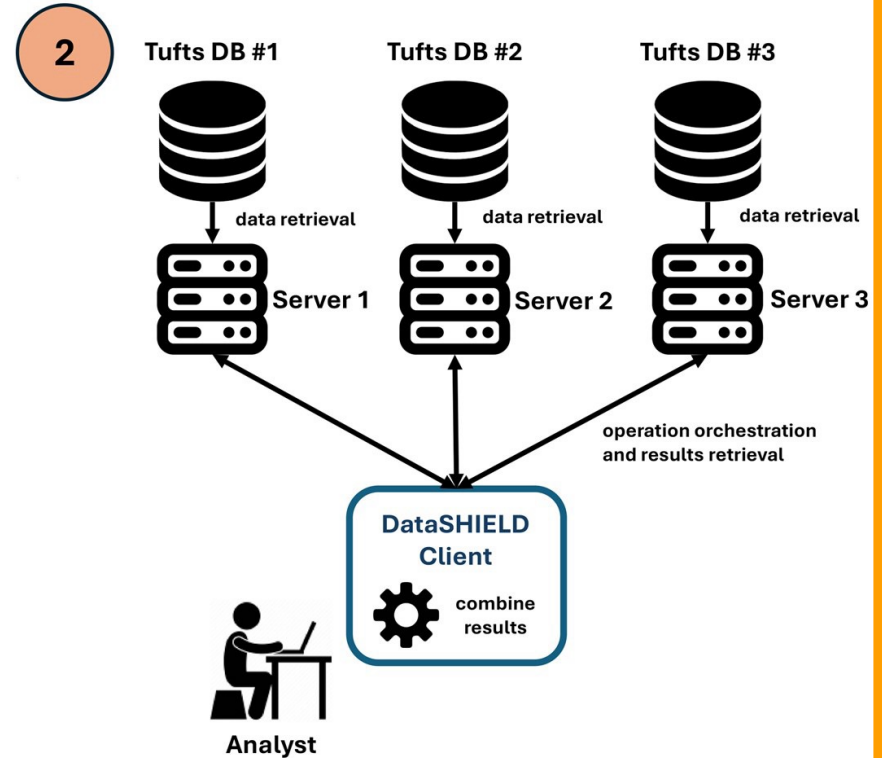
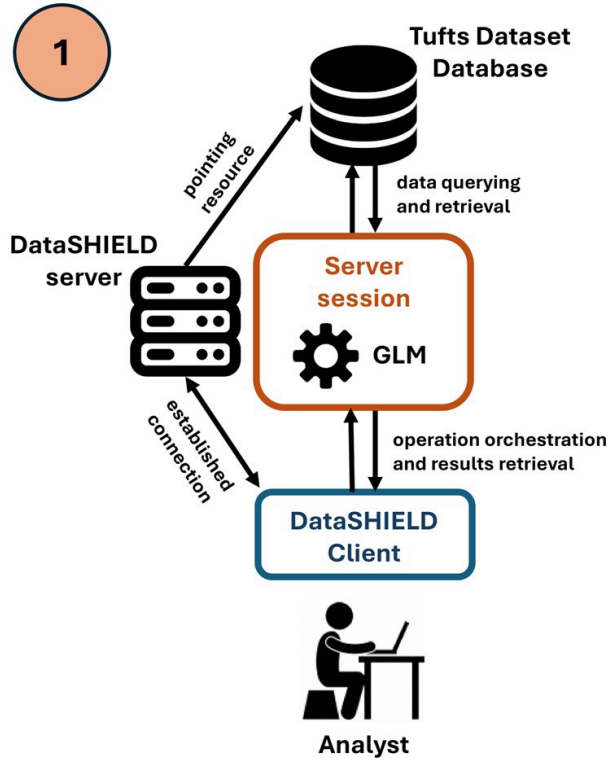
Juan R.
González



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

Validation and study

Tufts Medical Center



OXFORD
ACADEMIC

Bioinformatics



JOURNAL ARTICLE

dsOMOP: bridging OMOP CDM and DataSHIELD for secure federated analysis of standardized clinical data

David Sarrat-González , Xavier Escribà-Montagut , Jared Houghtaling , Juan R González 

Bioinformatics, Volume 41, Issue 6, June 2025, btaf286, <https://doi.org/10.1093/bioinformatics/btaf286>

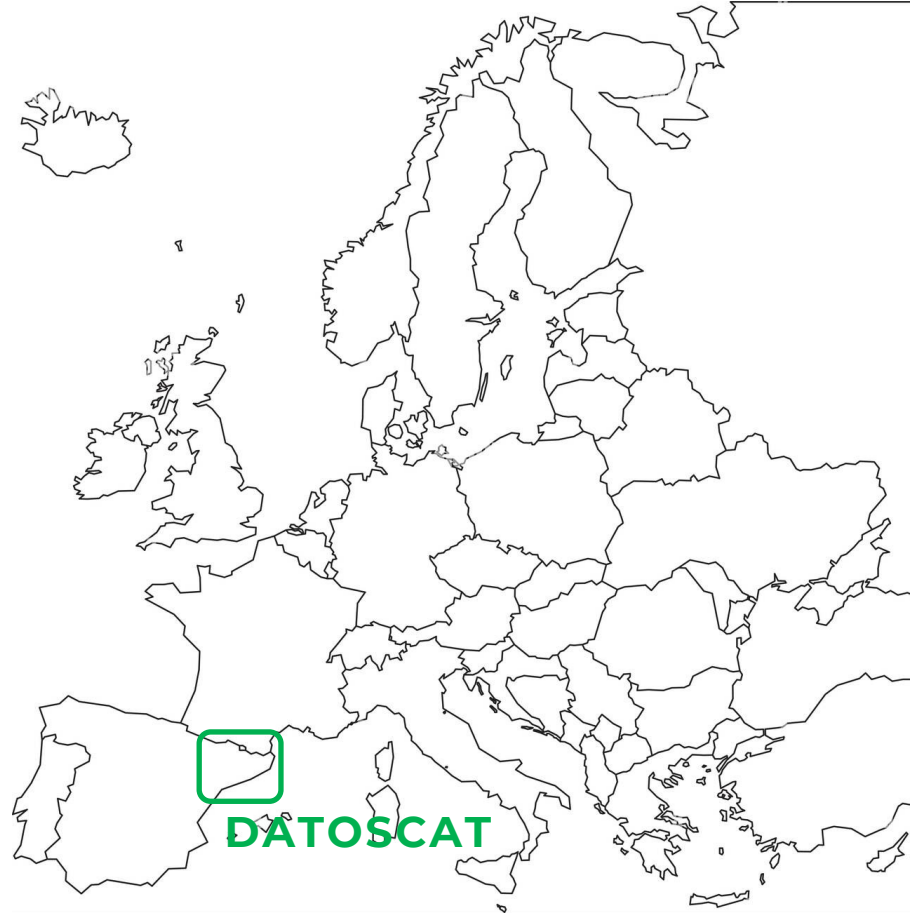
Published: 06 May 2025



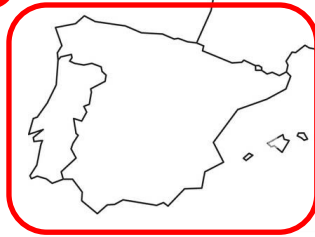
<https://doi.org/10.1093/bioinformatics/btaf286>

Adoption of dsOMOP

Adoption of dsOMOP



P4COPD





CADSET



**African Population and
Health Research Center**



- Kenya
- Senegal
- Cameroon
- Ethiopia

**Thank you very much
for your attention!**



Contact us:

david.sarrat@isglobal.org

juanr.gonzalez@isglobal.org