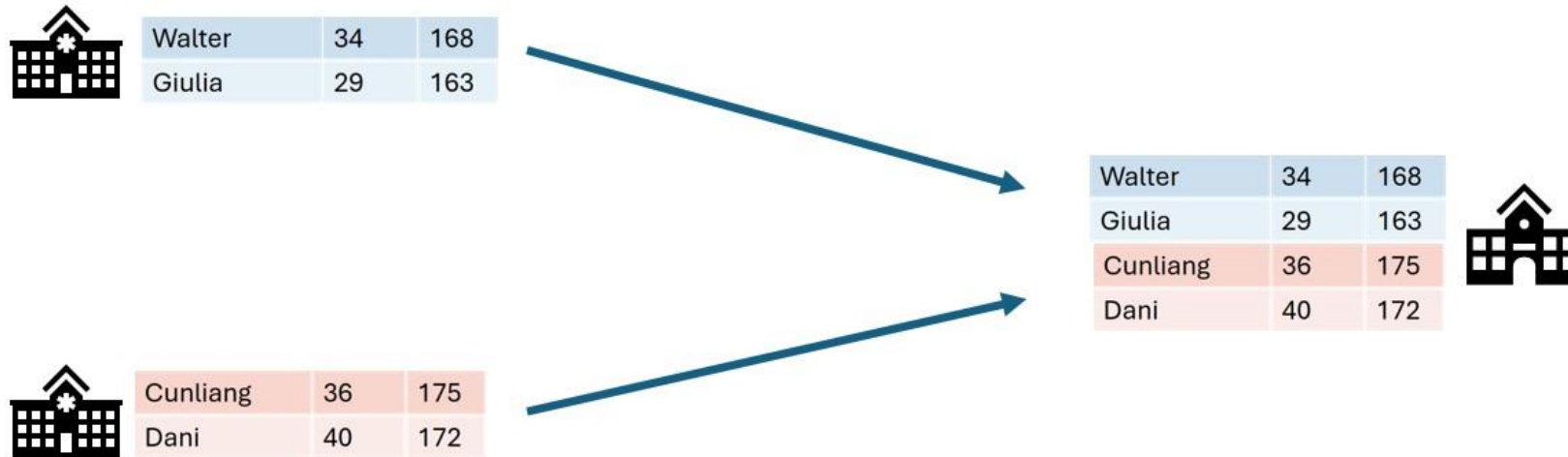
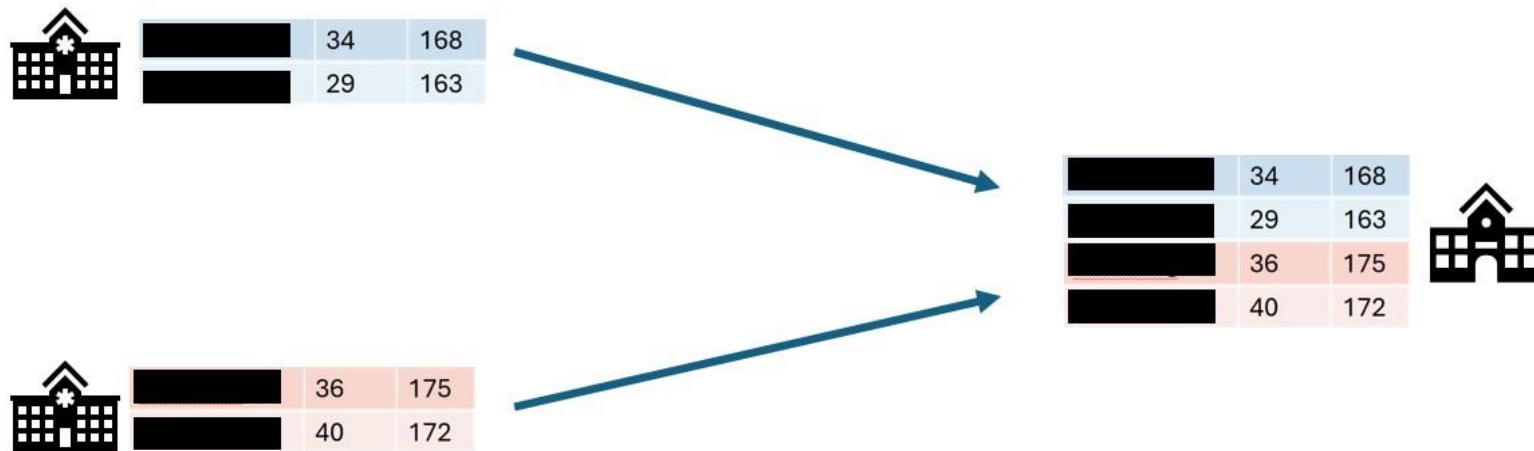


Introduction to Privacy Enhancing Technology

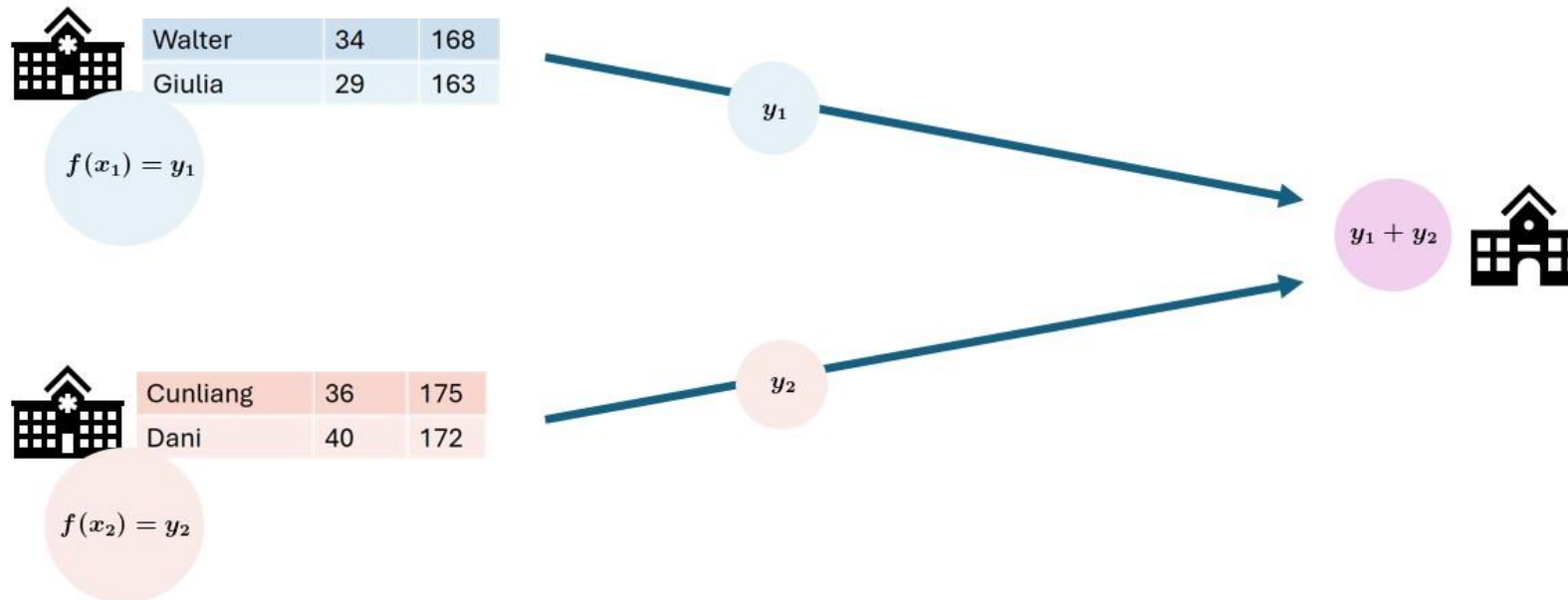
Classic analysis: Collect the data in one place



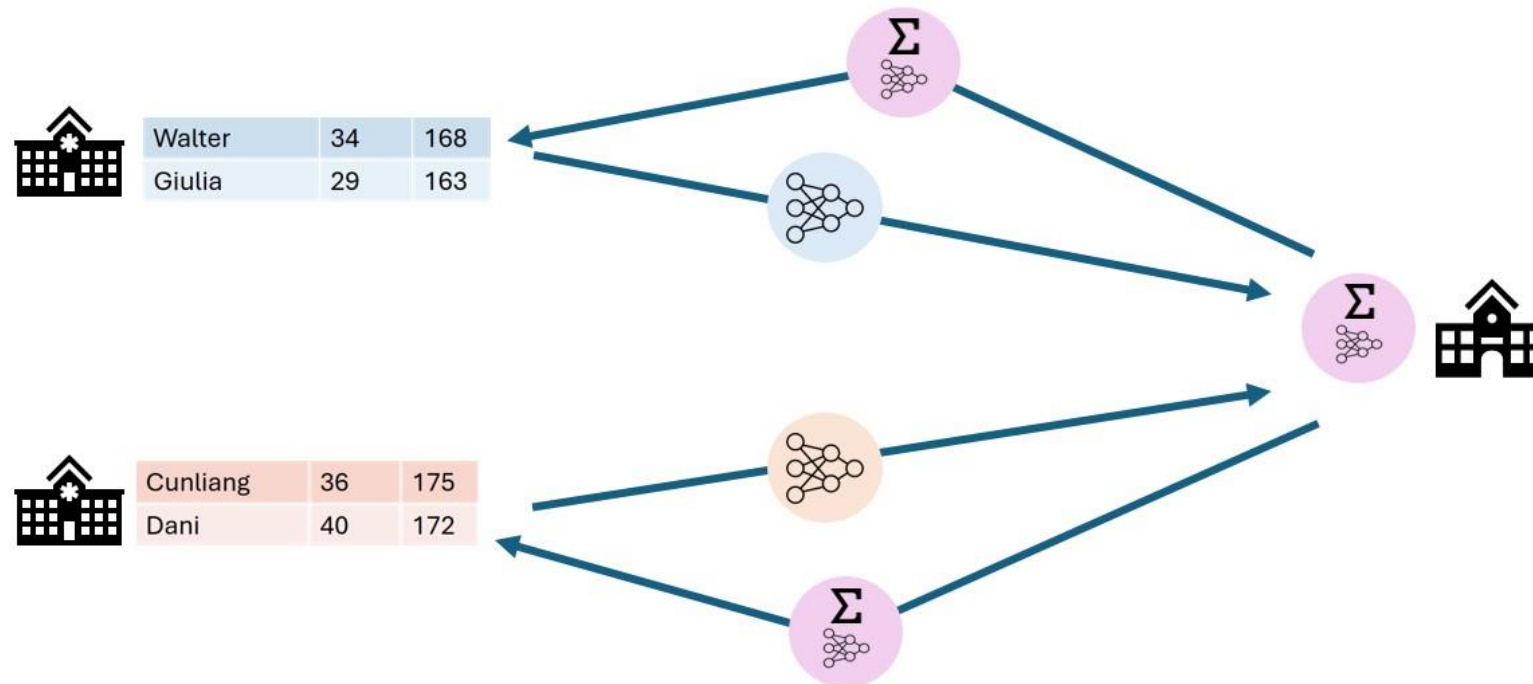
Data anonymization: Remove sensitive fields



Federated data analysis: Send analysis to the source, then aggregate

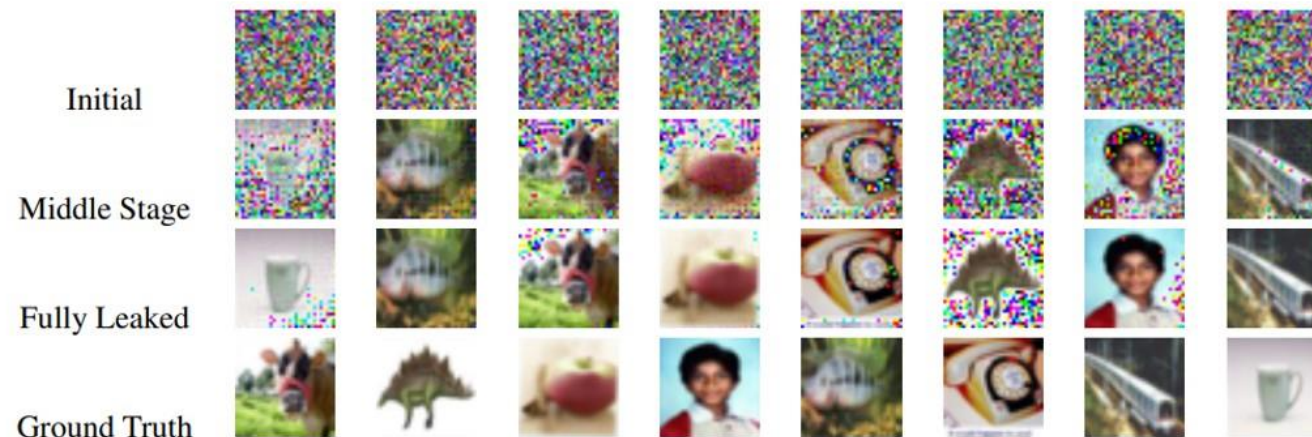


Federated learning: Train models locally, then aggregate



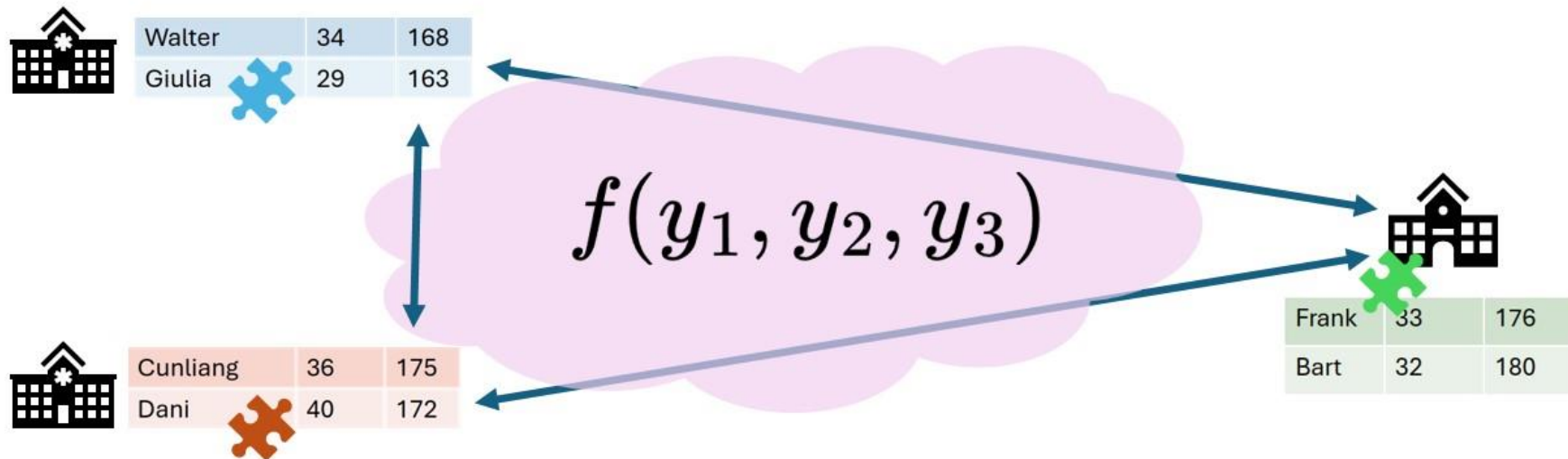
Federated learning and analysis can still leak data!

Gradient leakage



Zhu, Ligeng, Zhijian Liu, and Song Han. "Deep leakage from gradients."
Advances in neural information processing systems 32 (2019).

Secure multiparty computation: Send around encrypted puzzle pieces



Secret sharing: an example



SECRET SHARING, AN EXAMPLE

Mees, Sara and Noor want to know how much they weigh in total. Mees weighs 43 kg, Sara weighs 39, Noor weighs 45. They create secret shares for their weights that they give to their peers.

	Mees receives	Sara receives	Noor receives	Sum
Mees generates:	-11	50	4	43
Sara generates:	-12	17	34	39
Noor generates:	19	-38	64	45

They sum their shares:

Mees -4

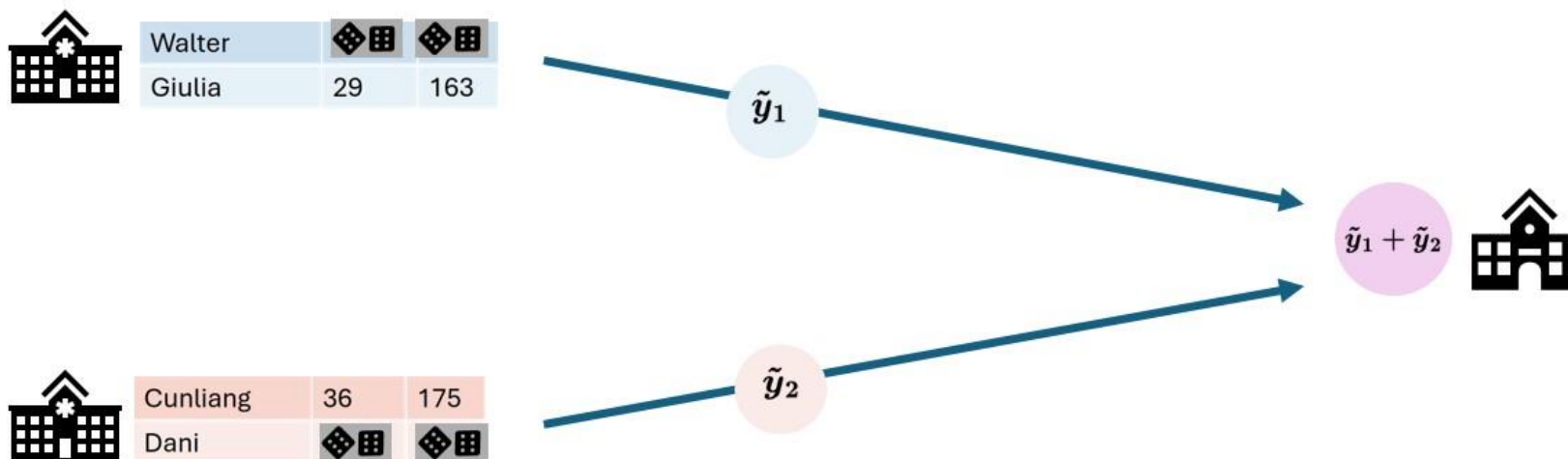
Sara 29

Noor 102

[↑
To Top](#)


They add their sums together: $-4 + 29 + 102 = 127$ In this way, they have aggregated their data without sharing their individual data with anyone else.

Differential privacy: Add noise



Data partitioning


Horizontal partitioning



Walter	34	168
Giulia	29	163
Cunliang	37	175
Dani	40	172

Vertical partitioning

Walter	34	Walter	168
Giulia	29	Giulia	163
Cunliang	37	Cunliang	175
Dani	40	Dani	172



Technology doesn't solve everything!

- Privacy enhancing technology is only a small part of the data sharing puzzle
- Some other factors
 - Trust
 - Regulations (either general or specific to the location)
 - Data harmonization

Summary

- With PETs you can derive insights from data without seeing individual records.
- PET analysis usually starts with the anonymization or pseudonymization of the data.
- In federated data analysis the analysis moves to the data, while in classic analysis the data moves around.
- In secure multiparty computation, computations are performed collaboratively without any one party being able to see all the raw data.
- Techniques from differential privacy add noise to the data to make it harder to reconstruct the original records from an aggregation.
- Privacy enhancing analyses usually stack multiple techniques on top of each other to provide multiple layers of protection.
- Horizontal partitioning means the records are split, while in vertical partitioning the features are split.
- In research on privacy sensitive data, technology is only one part of the story