RosemanLabs

PRIVACY BY DESIGN

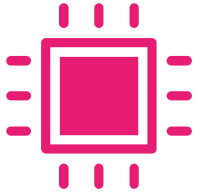
Multi-Party Computation for healthcare

Freya de Mink & Aldo Gunsing, Roseman Labs

Collaboration on sensitive data in challenging

Legal complexity

- GDPR
- Anti-trust laws



Technical complexity

- Security
- Confidentiality
- Standardization

Organizational complexity

- Governance, contracts
- Trusted third party
- Costly and time consuming

* https://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/collaborating-for-the-common-good

Roseman Labs helps organizations to collaborate on sensitive data

Combine and analyze data, without disclosing source data

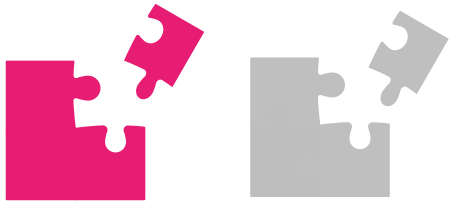
Strong security and legal compliance

Guaranteed by Multi-Party Computation (MPC)

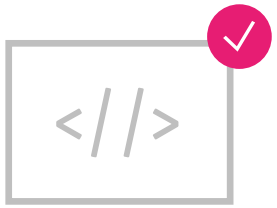
Easy to use software

Used by the National Cyber Security Center

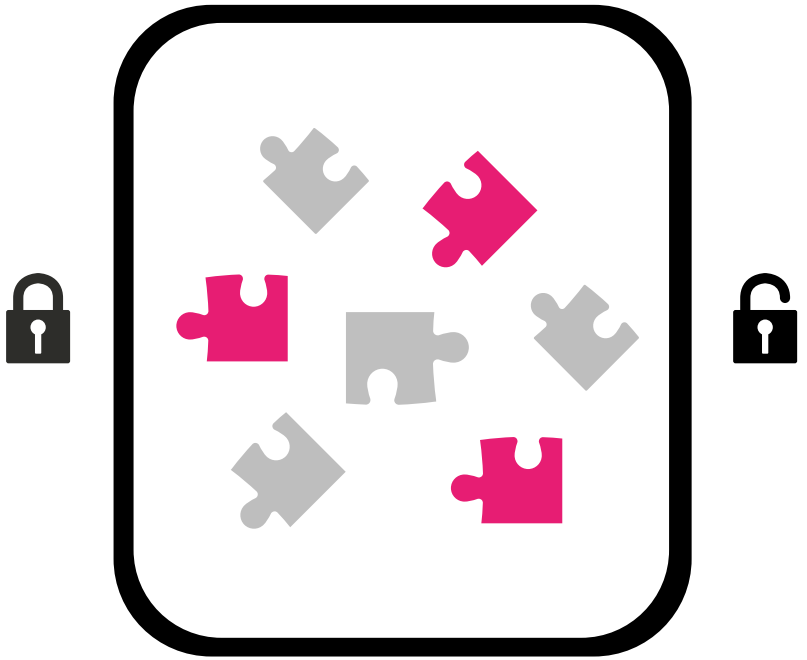
Safely combine data with Roseman Labs



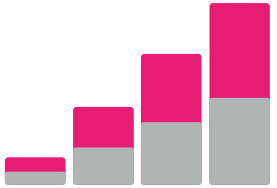
Data owners make data available



Approve analysis

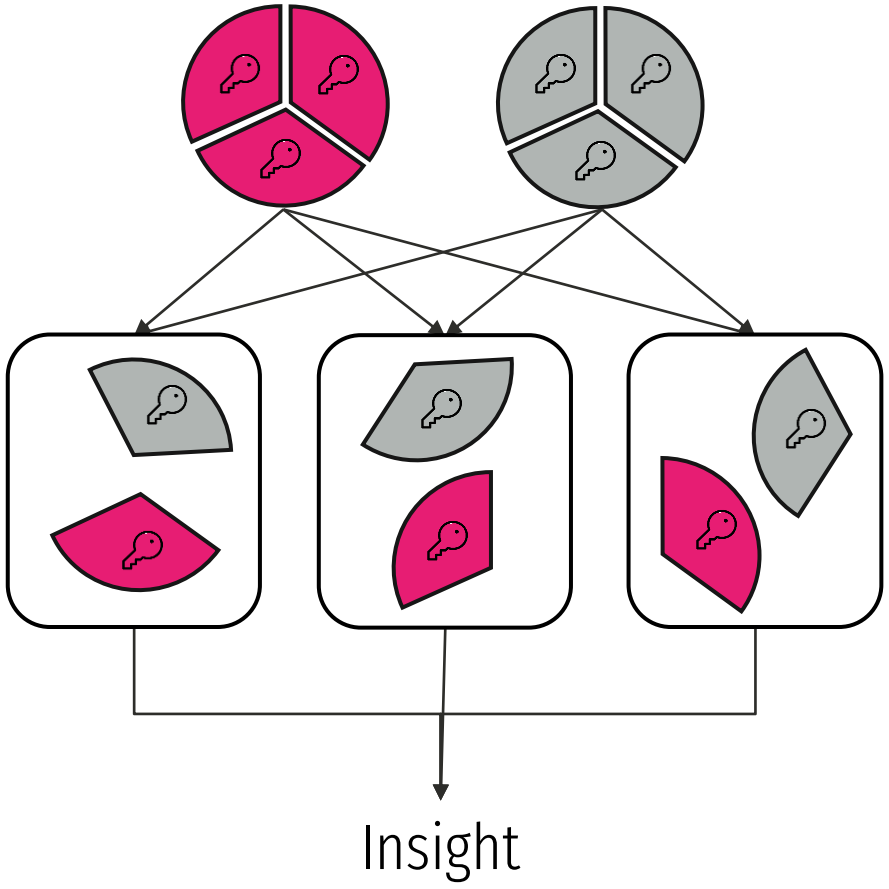


Combine tables and run analysis



Share insights

MPC in a nutshell



Data encrypted at the source into “secret shares[1]”

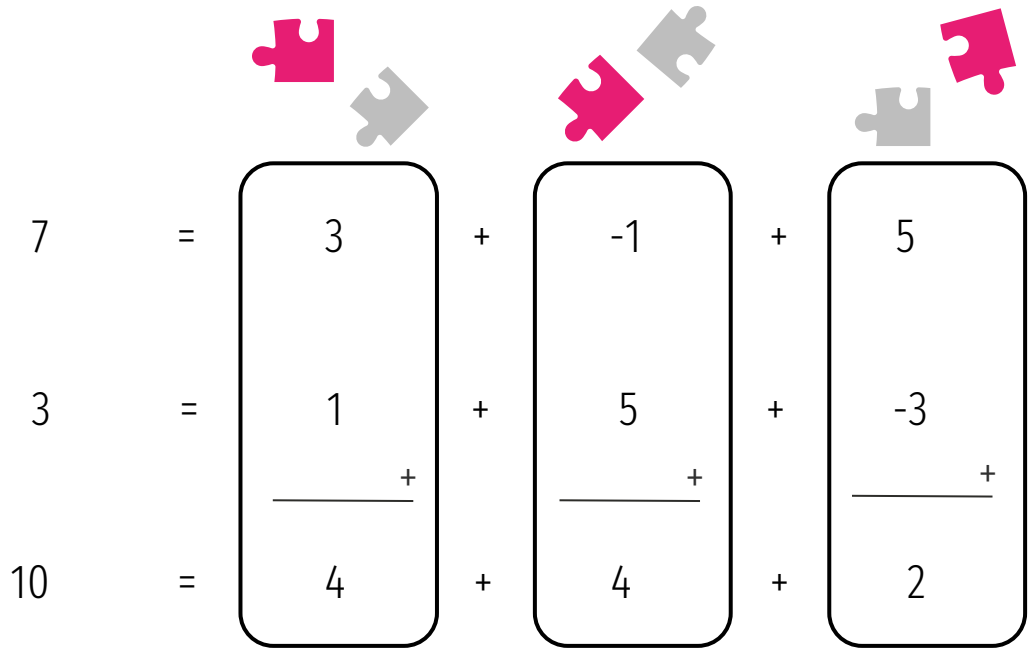
Secret shares distributed over multiple severs

Servers jointly execute calculation on encrypted data

Only the result is revealed

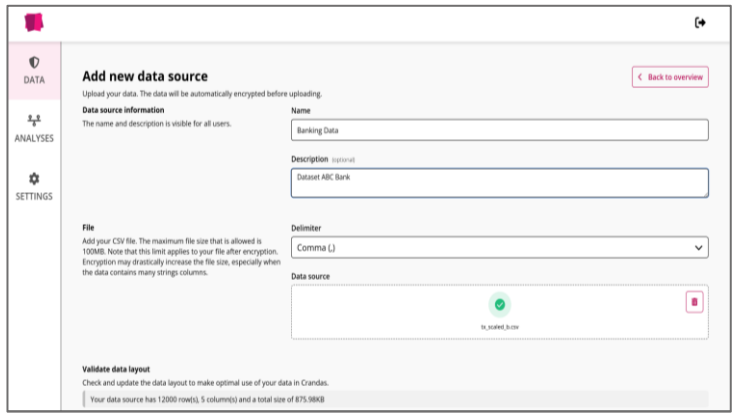
1. Secret shares are (for instance) created by breaking input numbers into a summation of three random numbers. For example, the number 9, could be converted into a 5, a 3 and a 1, because 5+3+1=9

Calculation example



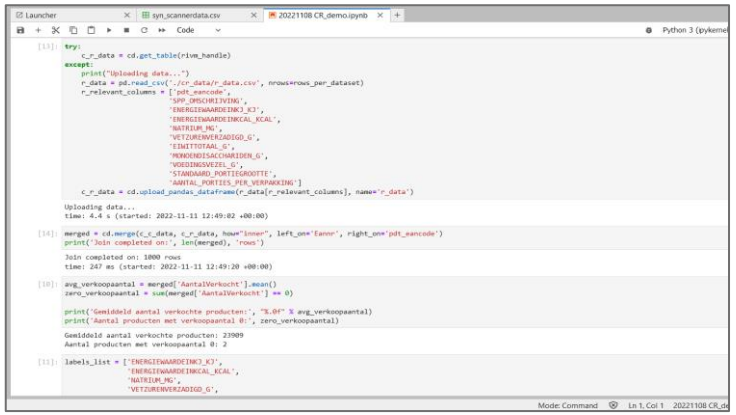
Sophisticated cryptography made easily accessible

Web portal



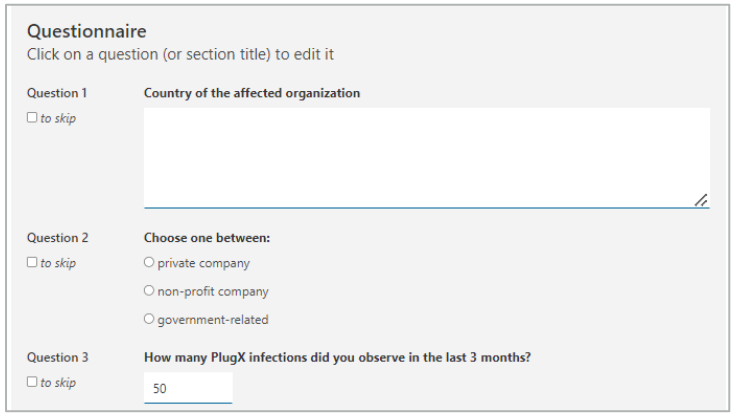
- Manage collaboration: users, rights, data sources
- Make data sources available
- Review and approve scripts

Python package



- Write python scripts with Roseman Labs Crandas library
- Deploy across different data sources
- Request approval and run your scripts

Data request forms



- Enable easy data collection among a large number of participating parties – structured and unstructured

Benefits

More data: Use sources that the data owner is not prepared / able to share with you

Safe: data is protected by highest security level at all time. Data is not shared. Strong GDPR compliance: purpose binding, control, data minimization

Faster: setting up a collaboration can be done on weeks, rather than months or years

(For reference) MPC performance for different operations

Milliseconds	Seconds	Minutes	Hours
• SVM (inference)	• Sums / Voting / Surveys (1M rows)	• SVM (training)	• Neural networks (training, LeNet)
• Private set intersection (1k rows)	• Private set intersection (1M rows)	• Decision trees (training, 10k rows)	• Random forests (training)
• Neural networks (inference, LeNet)	• Filtering (1M rows)	• k-means clustering (10k rows)	
• Decision trees (inference)	• Linear regr. training (1M rows)	• Logistic regr. training (100k rows)	
	• Neural networks (inference, VGG16)	• Regex (100k rows)	
	• Decision trees (training, 1k rows)		
	• Random forests (inference)		

This indication of computational time is based on a regular server (up to 32 cores). The exact compute time depends on a number of factors, such as server type (number of CPU cores), number precision, and network latency.

Contact: Freya.de.Mink@rosemanlabs.com

RosemanLabs
PRIVACY BY DESIGN