



# Multi-Party Computation for healthcare

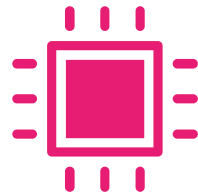
Freya de Mink & Aldo Gusing, Roseman Labs

# 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

# Roseman Labs helps organizations to collaborate on sensitive data



Combine and analyze data, without disclosing source data

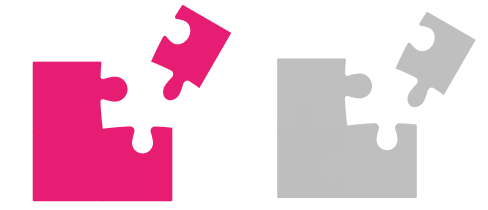
Strong security and legal compliance

Guaranteed by Multi-Party Computation (MPC)

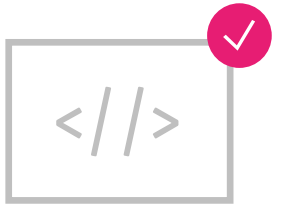
Easy to use software

Used by the National Cyber Security Center

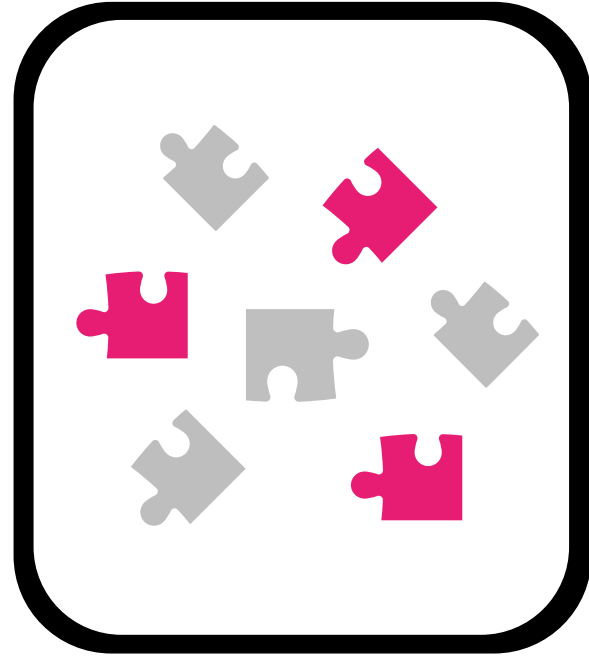
# Safely combine data with Roseman Labs



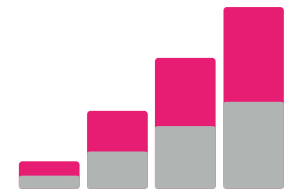
Data owners make data available



Approve analysis

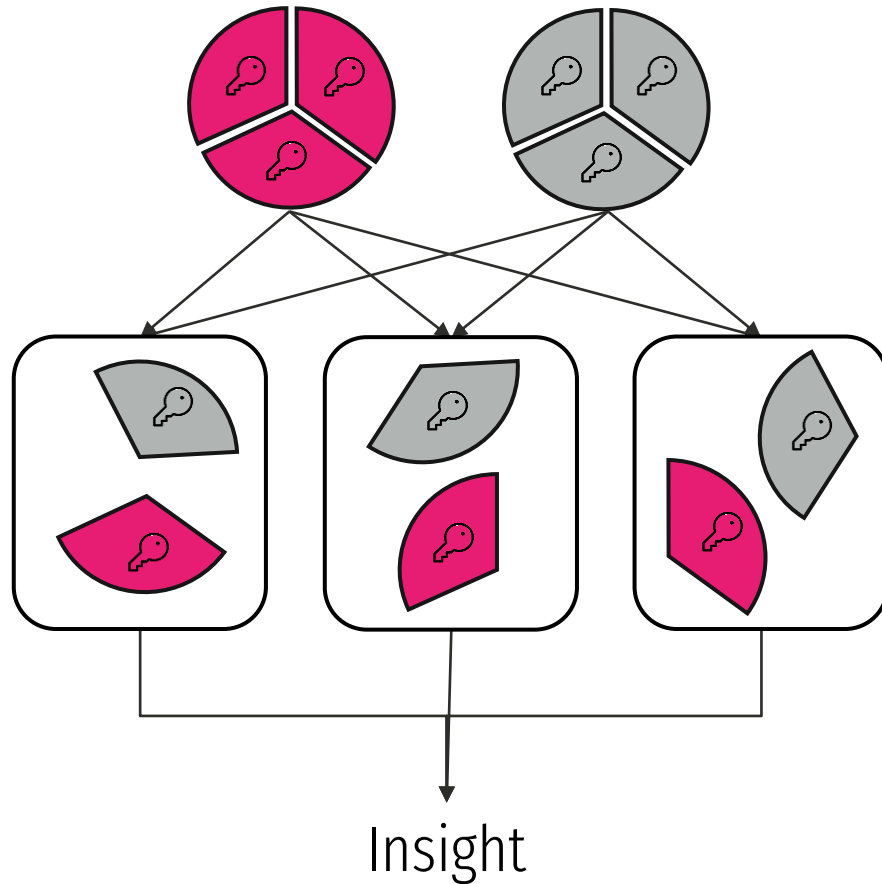


Combine tables and run analysis



Share insights

# MPC in a nutshell



Data encrypted at the source into “secret shares<sup>1</sup>”

Secret shares distributed over multiple servers

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  $5+3+1=9$

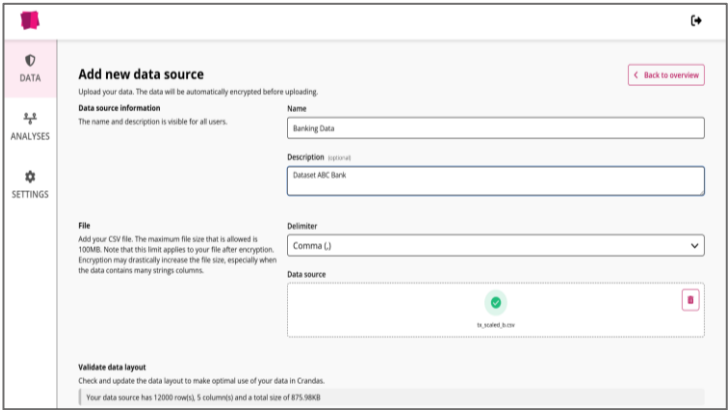
# Calculation example

$$\begin{array}{r} 7 \\ 3 \\ 10 \end{array} = \begin{array}{|c|} \hline 3 \\ \hline 1 \\ \hline 4 \\ \hline \end{array} + \begin{array}{|c|} \hline -1 \\ \hline 5 \\ \hline 4 \\ \hline \end{array} + \begin{array}{|c|} \hline 5 \\ \hline -3 \\ \hline 2 \\ \hline \end{array}$$

The diagram illustrates a calculation example using three vertical columns, each enclosed in a rounded rectangle. The columns are separated by plus signs. Above each column are two puzzle pieces: a red one and a grey one. The first column has a red puzzle piece above the top number and a grey one above the bottom number. The second column has a red puzzle piece above the top number and a grey one above the bottom number. The third column has a grey puzzle piece above the top number and a red one above the bottom number. The numbers in the columns are: Column 1: 3, 1, 4; Column 2: -1, 5, 4; Column 3: 5, -3, 2. Horizontal lines are drawn under the middle numbers (1, 5, -3) and the bottom numbers (4, 4, 2). Plus signs are placed to the right of each horizontal line. To the left of each row, there is an equals sign and a number: 7 for the top row, 3 for the middle row, and 10 for the bottom row.

# Sophisticated cryptography made easily accessible

## Web portal



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

## Python package

```
Launcher | syn_scannedata.csv | 20221108 CR_demo:zyhb | Python 3 (pykern)
[11] try:
    c_r_data = cd.get_table(rim_handle)
    except:
        print("Uploading data...")
        r_data = pd.read_csv("./r_data/r_data.csv", nrows=r_data.nrows)
        r_relevant_columns = ["SPM_DESCRIBING",
                               "ENERGIEWAARDERING_K3",
                               "ENERGIEWAARDERING_K4",
                               "METSILP_M",
                               "VETZUURVERZADIGD_G",
                               "EIJNTTOFAAL_G",
                               "HOUWENSCHEIDEN_G",
                               "WOLDSHOEVEN_G",
                               "STANNAAR_PUNTIGHOETJE",
                               "AMHNL_PORTELS_PER_VERPAKKING"]
        c_r_data = cd.upload_pandas_dataFrame(r_data[r_relevant_columns], name="r_data")
    Uploading data...
    Time: 4.4 s (started: 2022-11-11 12:49:02 +00:00)
[14] merged = cd.merge(c_r_data, c_r_data, how="inner", left_on="name", right_on="pid_mrcode")
    print("Join completed on:", len(merged), "rows")
    Join completed on: 1800 rows
    Time: 247 ms (started: 2022-11-11 12:49:20 +00:00)
[18] ang_verkoop aantal = merged["aantalVerkocht"].max()
    zero_verkoop aantal = size(merged["aantalVerkocht"]) == 0
    print("Gemiddeld aantal verkochte producten:", "E-04" % ang_verkoop aantal)
    print("aantal producten met verkoop aantal 0:", zero_verkoop aantal)
    Gemiddeld aantal verkochte producten: 23999
    aantal producten met verkoop aantal 0: 2
[11] label_list = ["ENERGIEWAARDERING_K3",
                  "ENERGIEWAARDERING_K4",
                  "METSILP_M",
                  "VETZUURVERZADIGD_G"]
```

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

## Data request forms

A screenshot of a questionnaire form titled 'Questionnaire' with the instruction 'Click on a question (or section title) to edit it'. It contains three questions: 'Question 1: Country of the affected organization' with a text input field and a 'to skip' checkbox; 'Question 2: Choose one between:' with three radio button options: 'private company', 'non-profit company', and 'government-related', and a 'to skip' checkbox; 'Question 3: How many PlugX infections did you observe in the last 3 months?' with a numeric input field containing '50' and a 'to skip' checkbox.

- 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
<ul style="list-style-type: none"><li>• SVM (inference)</li></ul>	<ul style="list-style-type: none"><li>• Sums / Voting / Surveys (1M rows)</li></ul>	<ul style="list-style-type: none"><li>• SVM (training)</li></ul>	<ul style="list-style-type: none"><li>• Neural networks (training, LeNet)</li></ul>
<ul style="list-style-type: none"><li>• Private set intersection (1k rows)</li></ul>	<ul style="list-style-type: none"><li>• Private set intersection (1M rows)</li></ul>	<ul style="list-style-type: none"><li>• Decision trees (training, 10k rows)</li></ul>	<ul style="list-style-type: none"><li>• Random forests (training)</li></ul>
<ul style="list-style-type: none"><li>• Neural networks (inference, LeNet)</li></ul>	<ul style="list-style-type: none"><li>• Filtering (1M rows)</li></ul>	<ul style="list-style-type: none"><li>• k-means clustering (10k rows)</li></ul>	
<ul style="list-style-type: none"><li>• Decision trees (inference)</li></ul>	<ul style="list-style-type: none"><li>• Linear regr. training (1M rows)</li><li>• Neural networks (inference, VGG16)</li><li>• Decision trees (training, 1k rows)</li><li>• Random forests (inference)</li></ul>	<ul style="list-style-type: none"><li>• Logistic regr. training (100k rows)</li><li>• Regex (100k rows)</li></ul>	

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](mailto:Freya.de.Mink@rosemanlabs.com)