Privacy-preserving data analytics via neural network models

Eleonora Ciceri (e.ciceri@mediaclinics.it), Marco Mosconi (m.mosconi@mediaclinics.it)

Problem statement

Data analytics can provide valuable knowledge to companies that leverage collected data to derive relevant information. However, processed data are often **highly sensitive** and thus their disclosure may harm **individuals' privacy**.
Nowadays, the current European **General Data Protection Regulation** (GDPR) represents a major challenge for companies, as they are required to follow a privacy-by-design approach to protect data and yet allow their processing.

Proposed solution

We developed dedicated **privacy-preserving data analytics modules** able to extract analytics on **protected** data, via the application of **artificial intelligence**. This approach ensures data subjects' privacy while still being **cost-effective** and **accurate**.

PAPAYA and the General Data Protection Regulation

Advances in technology and the capability of big data analytics and artificial intelligence have made it easier to create **profiles** and make **automated decisions**. Article 29 Data Protection Working Party, on its "Guidelines on Automated Individual decision-making and Profiling", reports that "profiling and automated decision-making can pose significant risks for individuals' rights and freedoms which require **appropriate safeguards**".

How PAPAYA addresses the challenge

Explicit consent
Data subjects have the ability to give or withdraw consent

Security measures
PETs are employed to extract analytics from subjects' data

Transparency
Data subjects can visualize the disclosed data and their rights

Auditability
Data controllers can visualize audit logs and handle DPIA

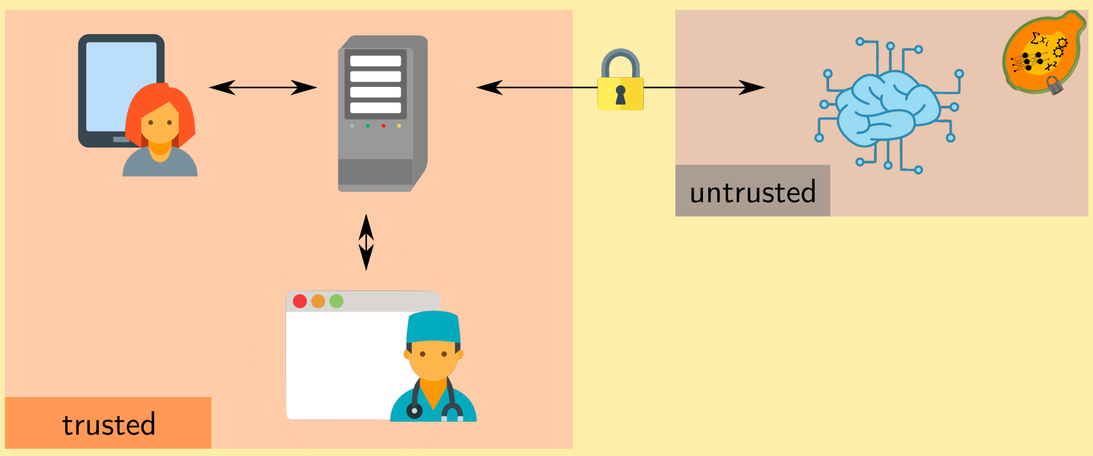
Applications in the healthcare field

Single data owner

The data controller needs to extract data analytics from his sensitive data in a privacy-preserving manner, thus he applies adequate **protection** (e.g., encryption) beforehand, and then outsources the task to a **third-party data processor** (e.g., the PAPAYA platform).

Advantages:
- data subjects' privacy is preserved
- the computational burden is on the data processors

Use case: ECG analysis

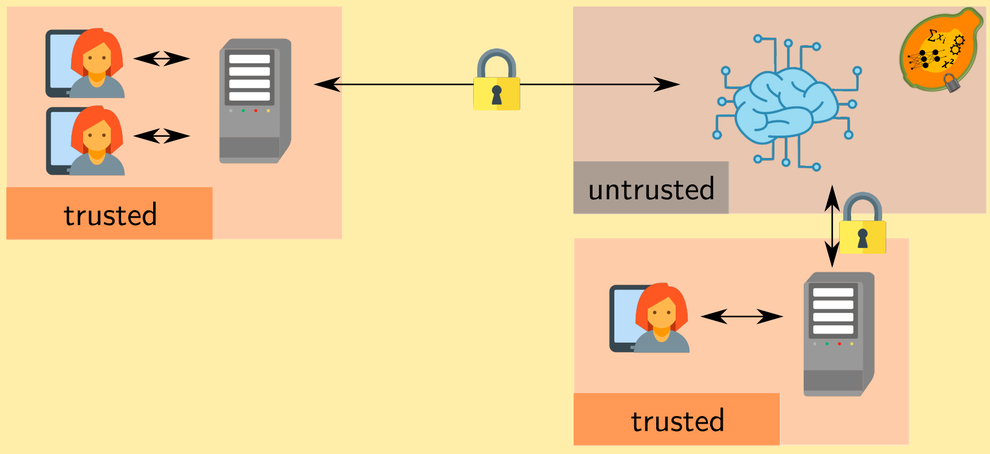


Multiple data owners

Multiple data controllers would like to perform analytical tasks over their **combined datasets**, disclosing the extracted analytics and keeping protected their original datasets. Thus, each controller **protects** his dataset (e.g., via encryption) before outsourcing it for analytics extraction.

Advantages:
- data subjects' privacy is preserved
- better models are built on larger datasets

Use case: Stress management



The research leading to these results has received funding from the European Union's Horizon 2020 Research and Innovation Programme, through the PAPAYA project, under Grant Agreement No. 786767. This poster reflects the view of the consortium only. The Research Executive Agency is not responsible for any use that may be made of the information it contains.