Towards a Shared Infrastructure for Multi-Party Secure Private Computing in the ESS: the JOCONDE project

Fabio Ricciato
Eurostat, Unit A.5 Methodology; Innovation in official statistics

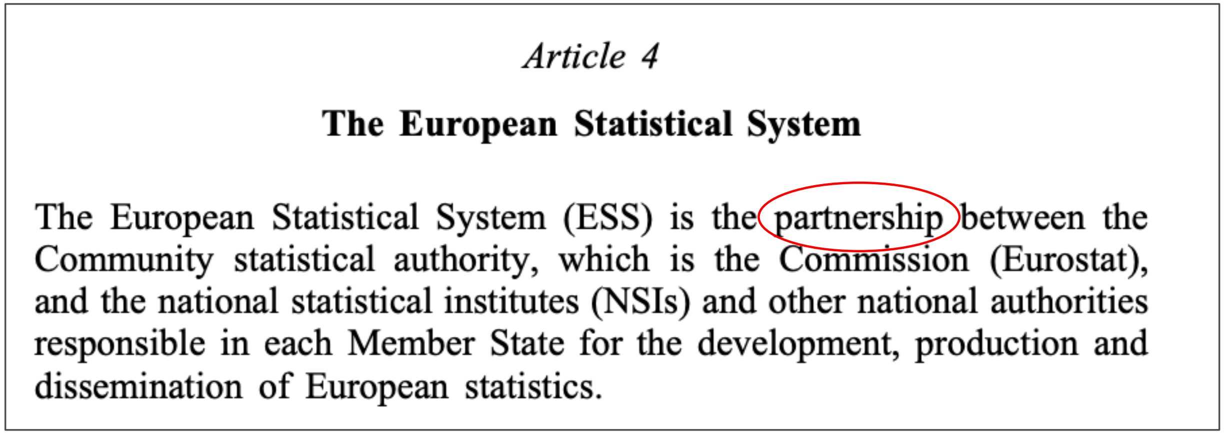
UN PET Lab meeting (online)

22 November 2024

Caveat: The information and views set out in this presentation are those of the author and do not necessarily reflect the official opinion of the European Union. Neither the European Union institutions and bodies nor any person acting on their behalf may be held responsible for the use which may be made of the information contained therein.

European Statistical System (ESS)

- The ESS is a partnership between statistical organisations
  - Independency
  - Collaboration
  - Governance



Article 4

The European Statistical System

The European Statistical System (ESS) is the partnership between the Community statistical authority, which is the Commission (Eurostat), and the national statistical institutes (NSIs) and other national authorities responsible in each Member State for the development, production and dissemination of European statistics.

Regulation (EC) No 223/2009 on European statistics (link)

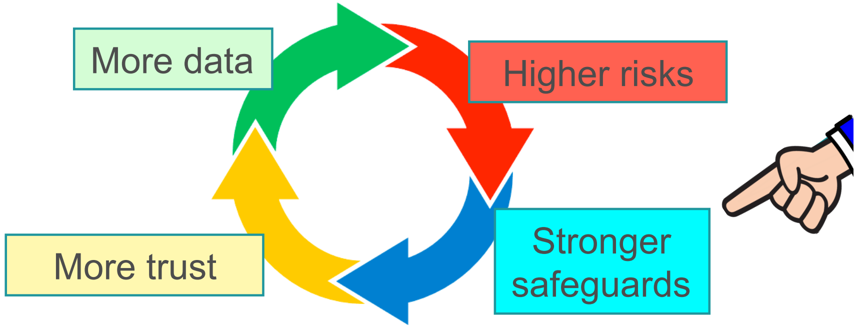
PET = ~~Privacy~~ Partnership(*) Enhancing Technologies

(*) Credit to Andrew Trask https://docsend.com/view/db577xmkswv9ujap

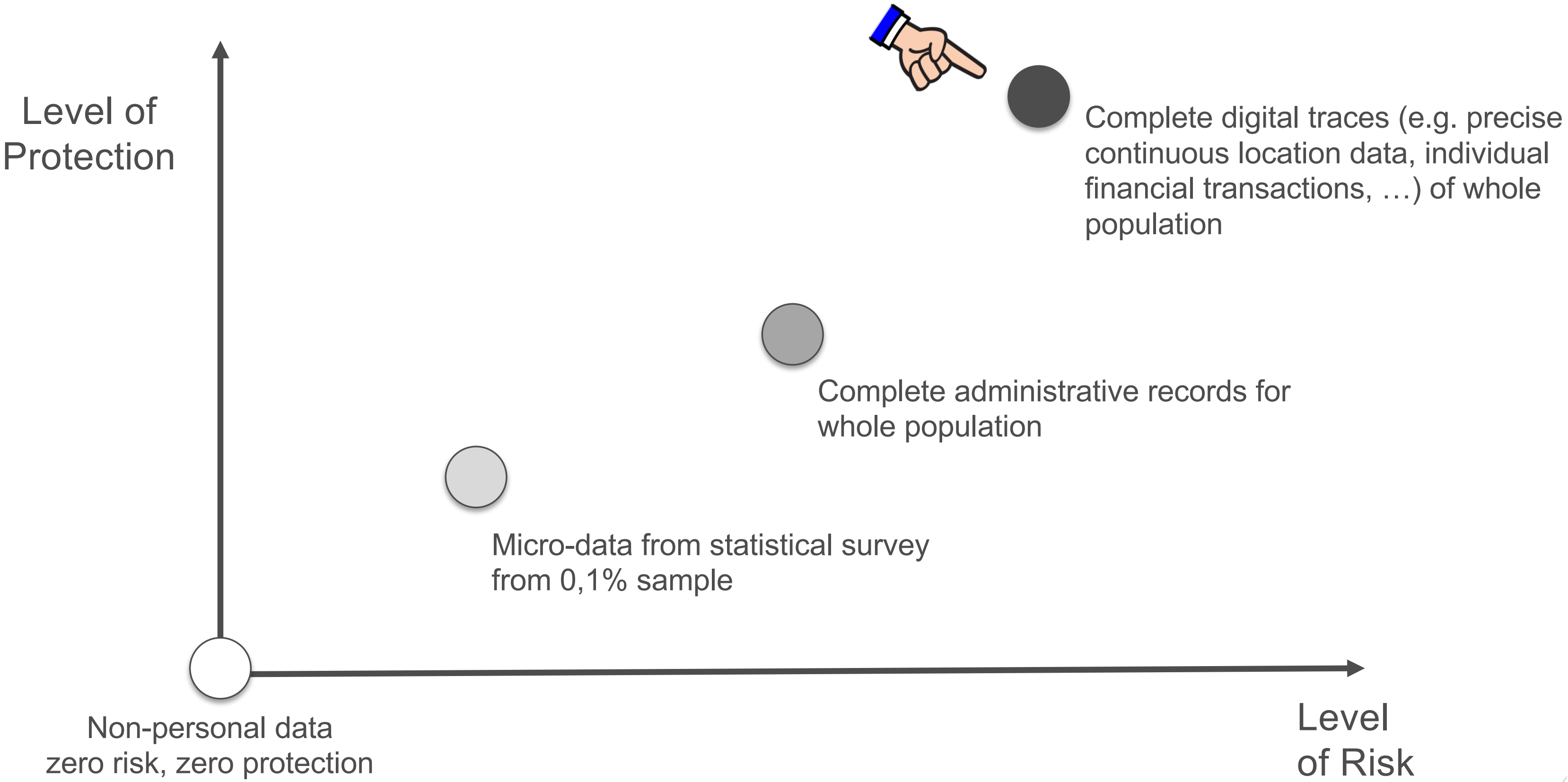
Why caring? Context and drivers

- Innovation trends concur to increase the demand for **cross-organisational data processing**
  - Data held by **NSIs in different Member States** concerning cross-border phenomena (e.g., international trade, migration, …)
  - Statistics based on data held by other **public bodies** (e.g., administrative records)
  - New statistics based on **privately held data**, requiring integration across different providers (e.g., Mobile Network Operators) and possibly with statistical data

- Increasing attention by the general public to **personal data protection**



Proportionality – a key GDPR concept



Level of Protection

Complete digital traces (e.g. precise continuous location data, individual financial transactions, …) of whole population

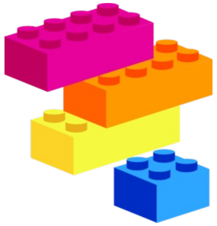
Complete administrative records for whole population

Micro-data from statistical survey from 0,1% sample

Non-personal data
zero risk, zero protection

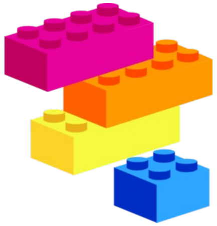
Level of Risk

Secure Private Computing (**SPC**)



- Privacy Enhancing Technologies (PET) is an umbrella term comprising two distinct groups of methods/approaches that address distinct (often complementary) problems:
- Input Privacy (aka **Secure Private Computing**, aka Privacy-Preserving Computation)
  - Compute the output without exposing the input (e.g., Computing over Encrypted Data)
  - Multi-Party Computation (MPC) based on secret-sharing, homomorphic encryption, garbled circuits; Trusted Execution Environment (TEE); …
- Output Privacy (not in the scope of this presentation)
  - Modify the output to avoid disclosing information about the input
  - **Statistical Disclosure Control (SDC)**, Differential Privacy (DP)

Secure Private Computing (**SPC**)



- Privacy Enhancing Technologies (PET) is an umbrella term comprising two distinct groups of methods/approaches that address distinct (often complementary) problems:

Input Privacy (aka **Secure Private Computing**, aka Privacy-Preserving Computation)

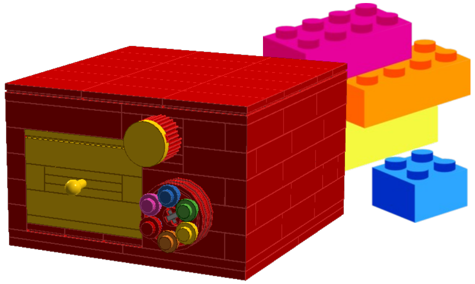
- Compute the output without exposing the input (e.g., Computing over Encrypted Data)
- Multi-Party Computation (MPC) based on secret-sharing, homomorphic encr[illegible] circuits; Trusted Execution Environment (TEE); …

In the scope of this project

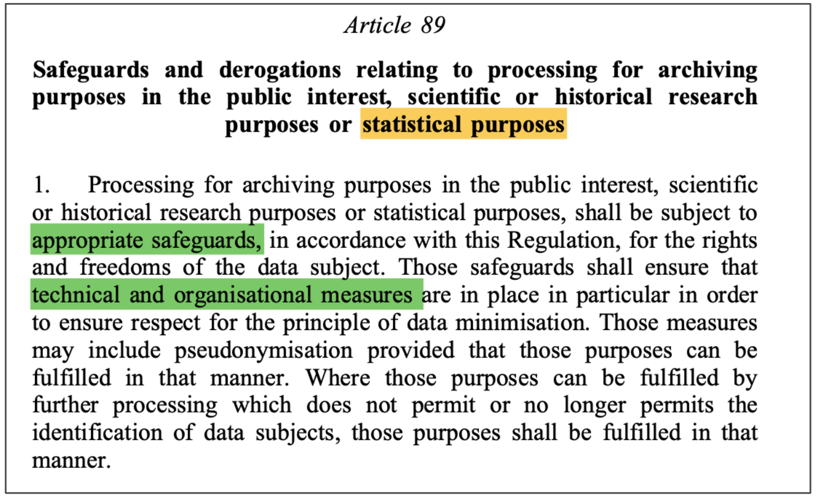
- Output Privacy (not in the scope o[illegible]s prese[illegible]tion)
  - Modify the output to avoid dis[illegible]ing [illegible] input
  - **Statistical Disclosure Con[illegible]l ([illegible]DC)**, Differential Privacy (DP)

NOT in the scope of this project

SPC as a “system” of safeguards



- SPC solution is a ***system*** *of safeguards* comprising
  - **Technological** components: MPC, TEE, authentication, encryption,…
  - **Organisational** components: policies, processes, agreements…
  - Fits well with “**Technical and Organisational Measures**” in GDPR, Art. 89



Article 89

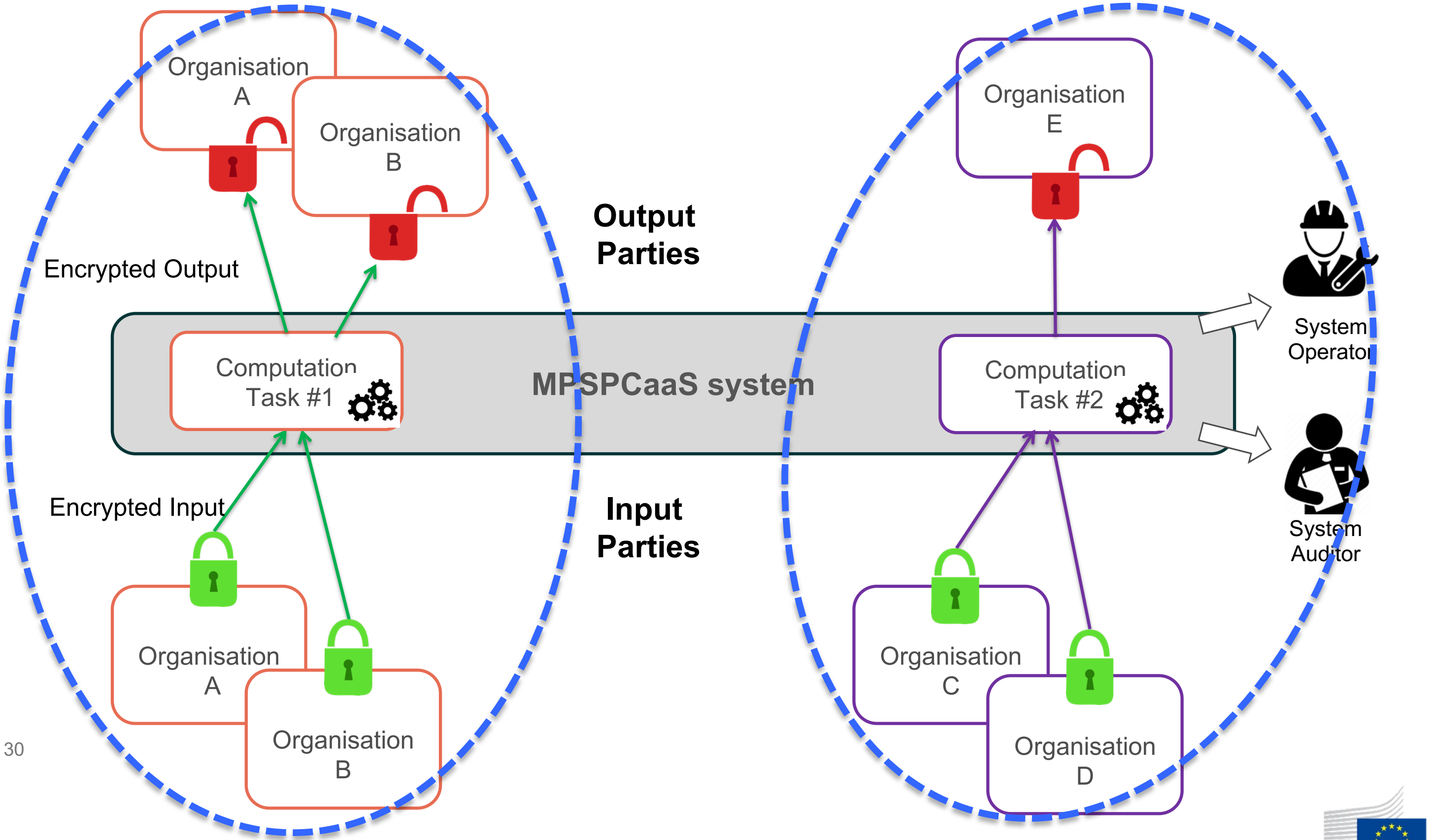
Safeguards and derogations relating to processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes

1. Processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes, shall be subject to appropriate safeguards, in accordance with this Regulation, for the rights and freedoms of the data subject. Those safeguards shall ensure that technical and organisational measures are in place in particular in order to ensure respect for the principle of data minimisation. Those measures may include pseudonymisation provided that those purposes can be fulfilled in that manner. Where those purposes can be fulfilled by further processing which does not permit or no longer permits the identification of data subjects, those purposes shall be fulfilled in that manner.

Risk vs cost

- Designing and building a **robust SPC system** is *costly*
  - Highly specialised **skills**: cryptography, HW/SW security, …
  - €€€ for HW/SW infrastructure building, deploying, maintenance
- **Saving on costs** by lowering robustness? NO!
  - Contradicts primary motivation for SPC: “lowering the risk”
- Alternative to saving: **Sharing!** Build a **Shared SPC solution**
  - Build one “shared SPC solution” to be used by multiple organisations

Multi-Party Secure Private Computing-as-a-service – **MPSPCaaS**



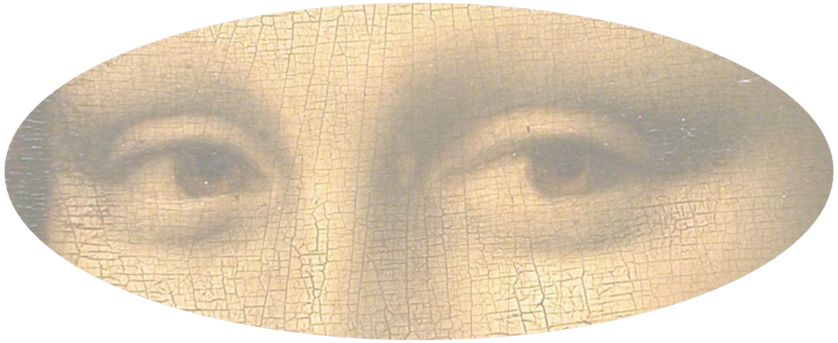
30

MPSPCaaS: from conceptualisation to specification

- 2021 – MPSPCaaS concept first proposed by Eurostat in UNECE HLG-MOS project on Input Privacy Preservation (IPP), 2021-2022
  - Open Technical Consultation organised as part of IPP project; presentations and exhange of ideas with data protection and privacy experts (ENISA workshop, MPC alliance, …)
- 2023 – Launch of open call for tenders by Eurostat and evaluation of tenders
  - ESTAT/2023/OP/0004 **Specification**, **feasibility analysis** and **prototype demonstration** of a multi-party secure private computing system for processing confidential sets of micro-data across organisations in support of statistical innovation (link)
- 2024 – Award of contract and launch of JOCONDE project

Reference: *Steps Toward a Shared Infrastructure for Multi-Party Secure Private Computing in Official Statistics*, JOS 03/2024 https://journals.sagepub.com/doi/10.1177/0282423X241235259

JOCONDE project



- **J**oint **O**n-demand **CO**mputation with **N**o **D**ata **E**xchange
  - Started in April 2023, duration 24 months, ending March 2026
  - In collaboration with Cybernetica – Estonian company specialised in security and privacy technologies (https://cyber.ee)
- Goals
  - **Define MPSPCaaS system specifications** at all levels – technology, organisational, legal – based on extensive analysis of state-of-the-art
  - **Demonstrate** based on **prototype implementation** feasibility, usability, scalability
  - *The results from JOCONDE will enable procurement and deployment of production system in follow-up projects 2026+*

6 Tasks for JOCONDE

Input for discussion with external experts

- Task 1 – Usage scenarios and system requirements
- Task 2 – Technology analysis

First deliverables D1.1 and D2.1 out by end of 2024

- Task 3 – Legal aspects
- Task 4 – System Specifications and Architecture

First deliverables D3.1 and D4.1 planned for Jan'25

- Task 5 – Demonstrator prototype and functional testing
- Task 6 – Trust building plan

Testing based on demonstrator implementation, possibly involving volunteering NSIs as beta-testers, in Q3/Q4 2025

Stakeholders

Data Protection Authorities (DPA)
EDPS / EDPB

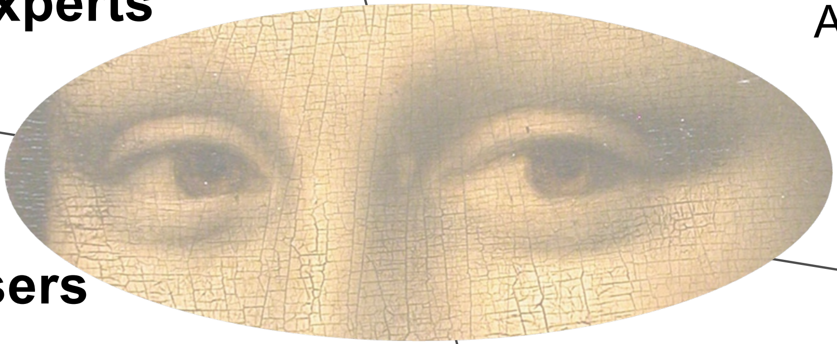
Legal Experts

Technology Experts

Academic Researchers

Research & Technology
Organisations (RTOs)

Industry R&D



System Users

ESS Members
National Statistical Institutes (NSI)
Other National Authorities (ONAs)

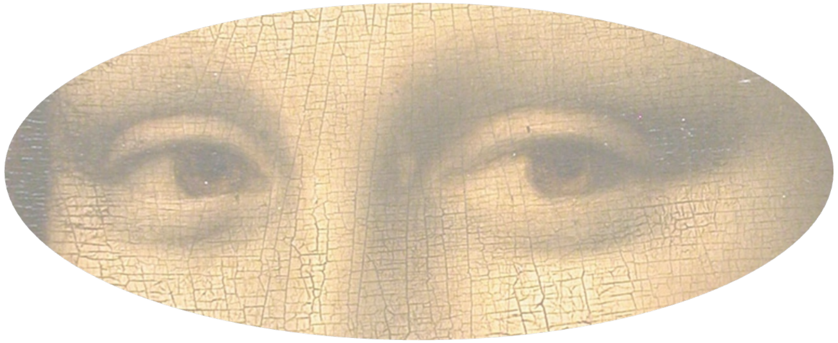
ESS partners
Government authorities (national, EU, extra-EU)

Private Data Holders

Civil Society

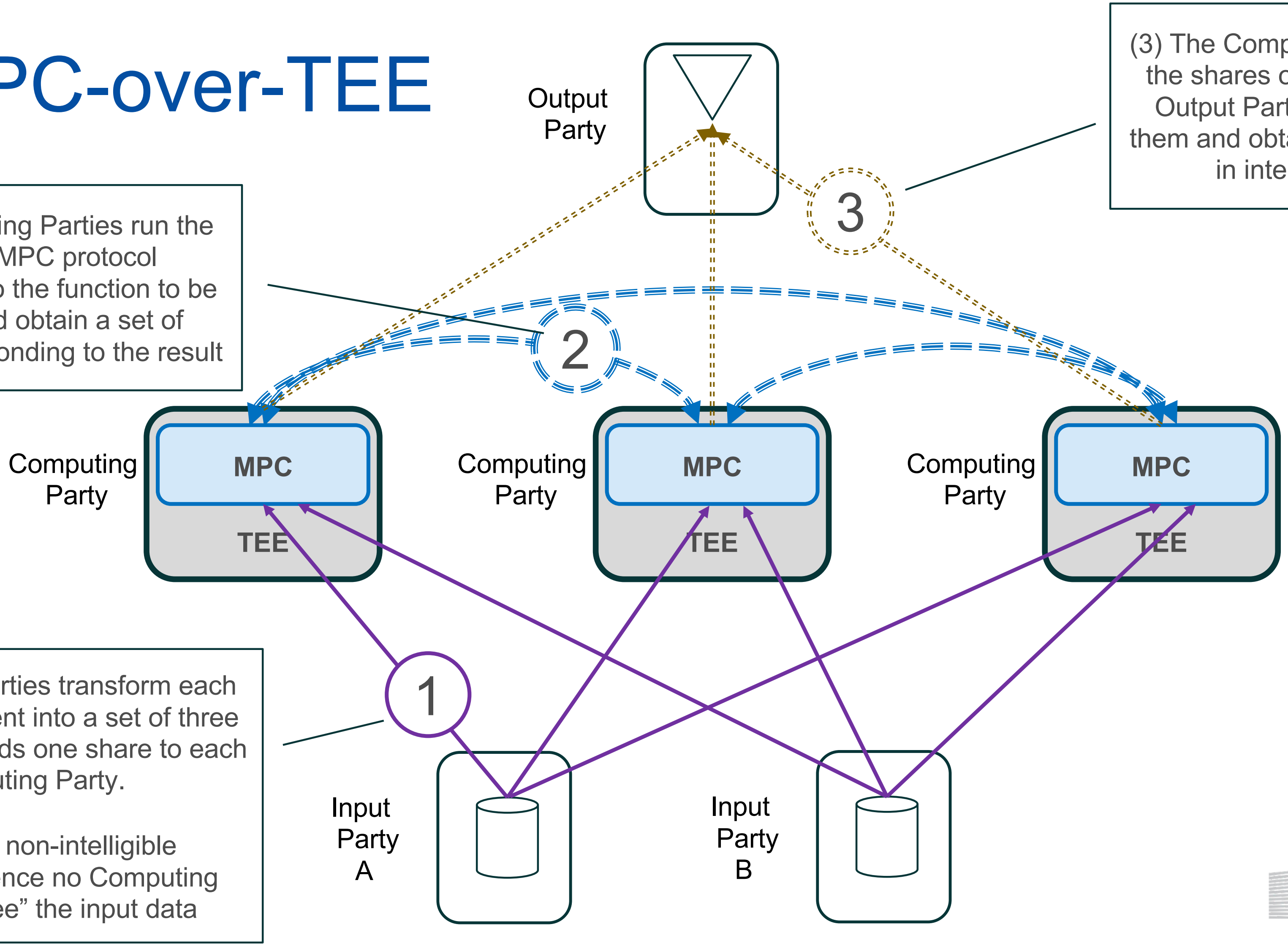
NGOs
Digital Rights Activists
Ethical Hackers

Technological approach



- Full overlay MPC over TEE
- Architecture based on 3 logical planes: M-plane, C-plane and D-plane
  - Inspired by telecom and computer network architectures
- D-plane supporting multiple MPC engines, open-source and proprietary
  - Configurable at Computation Task establishment
  - Evolvability – no vendor lock-in – coverage of design space
- Standard programming language with open-source compiler
  - To abstract cryptographic complexity from users

MPC-over-TEE



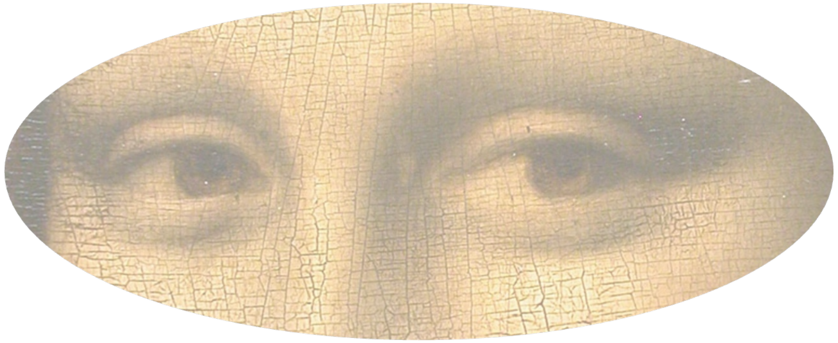
(3) The Computing Parties send the shares of the result to the Output Party who combines them and obtains the final result in intelligible form

(2) The Computing Parties run the interactive MPC protocol corresponding to the function to be evaluated and obtain a set of "shares" corresponding to the result

(1) The Input Parties transform each input data element into a set of three "shares" and sends one share to each Computing Party.

Shares are non-intelligible individually, hence no Computing Party can "see" the input data

Technological approach



- Full overlay MPC over TEE
- Architecture based on 3 logical planes: M-plane, C-plane and D-plane
  - Inspired by telecom and computer network architectures
- D-plane supporting multiple MPC engines, open-source and proprietary
  - Configurable at Computation Task establishment
  - Evolvability – no vendor lock-in – coverage of design space
- Standard programming language with open-source compiler
  - To abstract cryptographic complexity from users

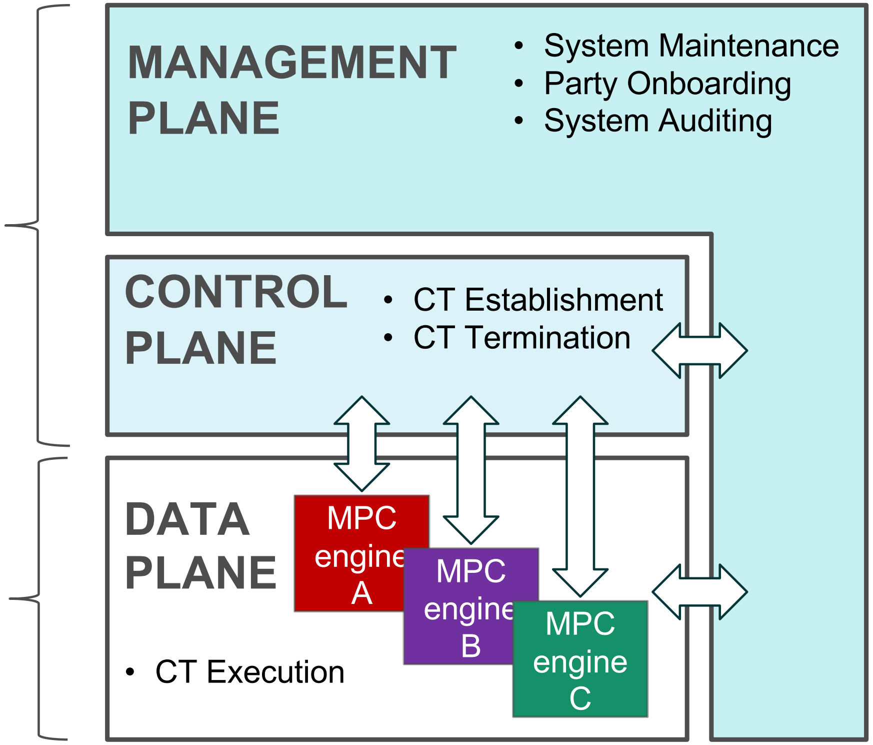
M-plane, C-plane and D-plane

M-plane and C-plane

- Developed by JOCONDE
- Fully open-source

D-plane

- Reusing existing components
- Supporting multiple TEE platforms
- Supporting multiple MPC engines, open-source and proprietary (selectable on per-CT basis)



MANAGEMENT PLANE

- System Maintenance
- Party Onboarding
- System Auditing

CONTROL PLANE

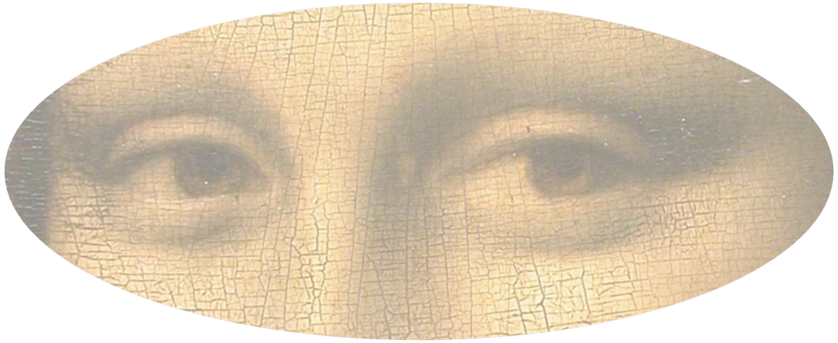
- CT Establishment
- CT Termination

DATA PLANE

- CT Execution

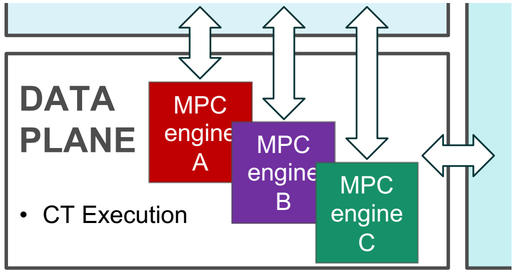
CT – Computation Task

Technological approach

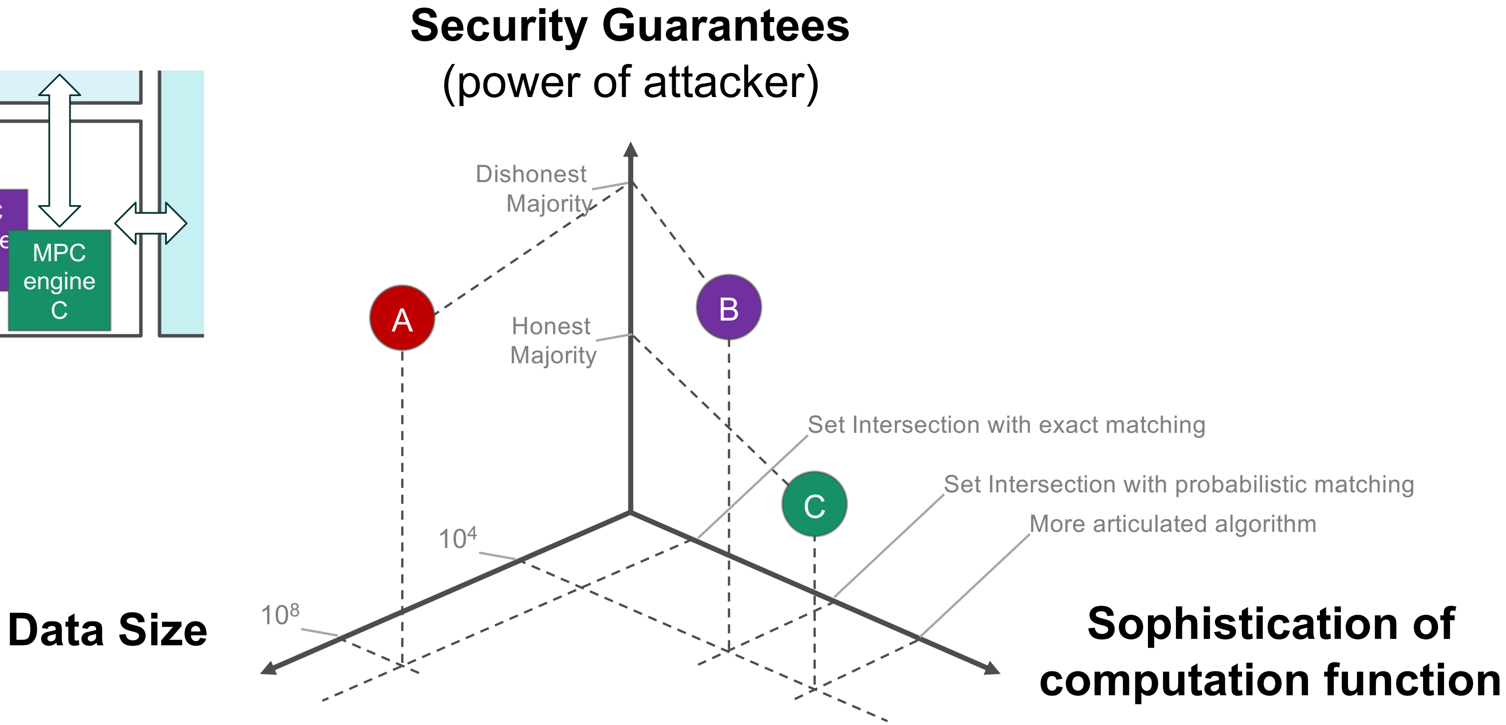


- Full overlay MPC over TEE
- Architecture based on 3 logical planes: M-plane, C-plane and D-plane
  - Inspired by telecom and computer network architectures
- D-plane supporting multiple MPC engines, open-source and proprietary
  - Configurable at Computation Task establishment
  - Evolvability – no vendor lock-in – coverage of design space
- Standard programming language with open-source compiler
  - To abstract cryptographic complexity from users

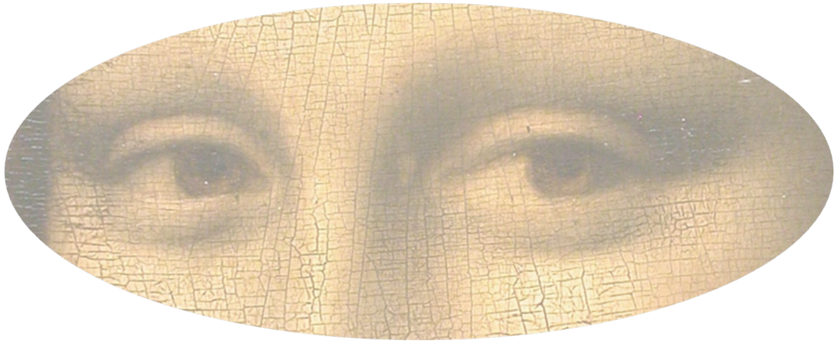
Design space for MPC protocols



- CT Execution

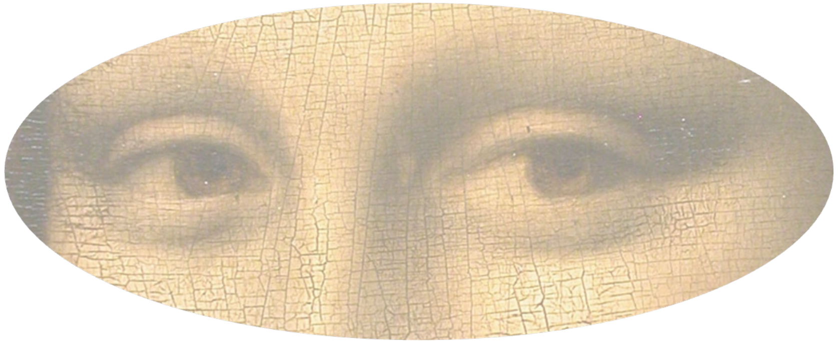

Security Guarantees
(power of attacker)
Dishonest Majority
Honest Majority
A
B
C
Set Intersection with exact matching
Set Intersection with probabilistic matching
More articulated algorithm
10^4
10^8
Data Size
Sophistication of computation function

Technological approach



- Full overlay MPC over TEE
- Architecture based on 3 logical planes: M-plane, C-plane and D-plane
  - Inspired by telecom and computer network architectures
- D-plane supporting multiple MPC engines, open-source and proprietary
  - Configurable at Computation Task establishment
  - Evolvability – no vendor lock-in – coverage of design space
- Standard programming language with open-source compiler
  - To abstract cryptographic complexity from users

Technological approach/2



- Secure deletion of all intermediate data (secret shares, keys,…) except audit logs upon Computation Task (CT) termination
  - Legal implications: liabilities limited to CT execution time
- No Single Point of Trust
  - Otherwise we would be building “just” a very complicated Trusted Third Party
- Not a mission-critical system
  - Availability not a primary design requirement

Task 3 – Legal aspects

- Task 3 shall perform a **legal analysis** to identify applicable legal requirements and identify open issues
- Task 3 will also prepare **reference legalware** (for CT-independent components) and provide **guidelines** (for CT-specific components)
  - **Reference models of agreements and contracts** between the parties (clarify liabilities, controller / processor roles, etc.)
  - Model of DPIA elements
- **The ultimate goal of Task 3 is to minimise as much as possible the legal burden for prospective users.**

Task 6 –Trust building plan

- Building a Trustworthy System is necessary, not sufficient for Bulding Trust.
- More is needed to ensure public trust and public acceptance
- How to convince key stakeholders and the general public of the genuinity and strength of the system?
  - Openness and transparency → how in practice?
  - Red team? → how in practice?

Demonstrator Prototype ≠ Production System

- The system prototype developed in JOCONDE serves multiple purposes
  - Provide a (first version of) **C-plane and M-plane open-source code** to be reused for the future production system (with further extension and consolidation)
  - Proof of concept: demonstrate its technical feasibility and usability as a production system (show it’s ready to move from lab to fab)
  - Testing – see if everything works as expected, find problems and fix them
  - Tasting – let some prospective users (beta testers) give it a try and see how they like it

Outlook

- Work in progress towards providing the ESS with a shared platform for on-demand multi-party computation on confidential data, offering alternative to traditional data sharing mechanisms
  - JOCONDE project https://cros.ec.europa.eu/joconde
  - Testing activities in late 2025 may possibly involve volunteering NSIs
  - JOCONDE results will enable procurement and deployment of production system in follow-up projects

- For further activities by Eurostat related to Privacy-Enhancing Technologies for Official Statistics refer to https://cros.ec.europa.eu/PET4OS

Thank you