EUROPEAN CONFERENCE ON
QUALITY IN OFFICIAL STATISTICS
2024 ESTORIL - PORTUGAL

Multi-Party Secure Private Computing and Quality: a prospective alliance

Fabio Ricciato

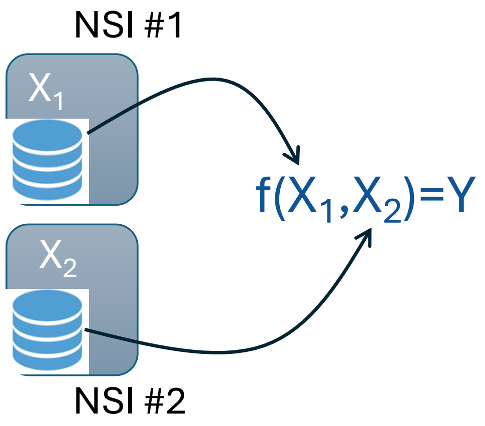
Eurostat, Luxembourg

INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

eurostat

The conference is partly
financed by the European Union

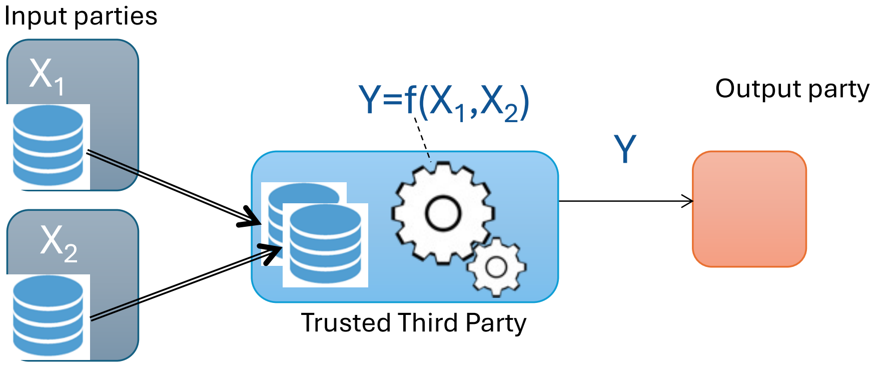
Statistics requiring combination of confidential micro-data sets held by different entities – Why?



- Two NSIs in different countries
  - Example: how many people are registered in both countries? (set intersection + counting)
- One NSI and one public administration in the same country
  - Example: linking micro-data from survey and administrative records
- One NSI and one or more private data holders
  - Example: combining pseudonymized micro-data from 3 Mobile Network Operators (MNO) for inbound roamers + survey data collected by NSI
- …
- The demand and use-cases for micro-data combination will increase following innovation trends in official statistics

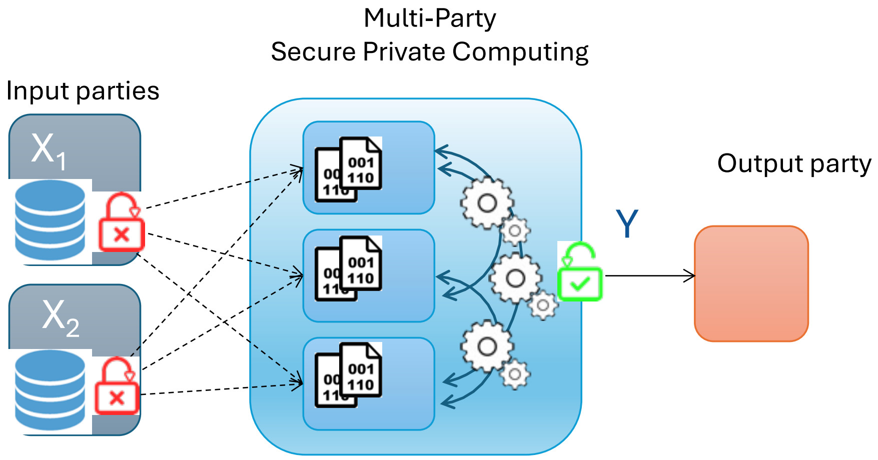
The traditional approach: *sharing data* (relies purely on *agreements*)

- Sharing the data X_1, X_2 *in intelligible form* ⇘⇘
  - One of the two parties, or an external Trusted Third Party, receives all the data (possibly with some simple pseudonymization, removal or replacement of direct identifiers)
  - The receiving party sees all data and runs the computation
  - The receiving party commits to: not use the data for anything else, keep the data secure, delete the data when not needed, not pass the data to other entities ...
- Pros: easy to implement
- Cons: risk that receiving party does not hold the commitments
  - Deliberately, or because infiltrated by a rogue attacker
- NB: you need to trust the *intentions and* the *capabilities* of a single receiving party

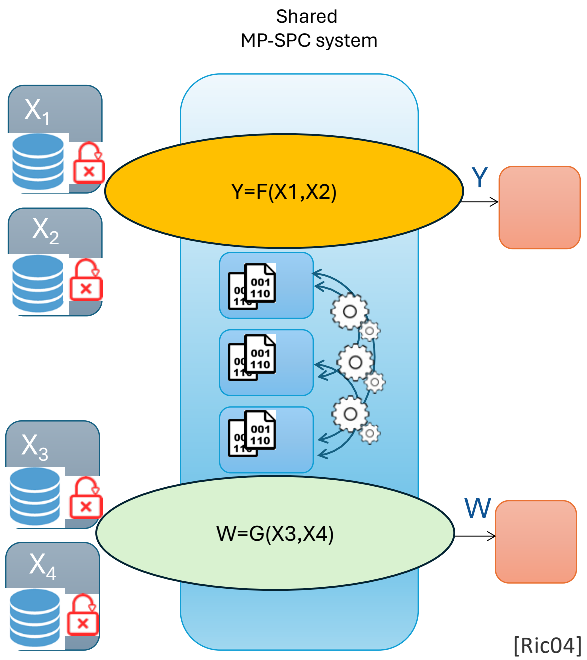


An alternative approach: *sharing computation* (relies on *technology*)

- **Multi-Party Secure Private Computing (MP-SPC) system**: multiple parties collaborate ("follow the protocol") to compute the exact output result Y
- Data is never exchanged in intelligible form
- The protocol requires transmission of something that "contains" the information needed to compute Y but does not reveal the input X_i to any party
- Think of some sort of
  - "encrypting" the input X_1, X_2 without ever "decrypting" them!
  - compute Y on the encrypted data
  - decrypting only the output Y
  - All exchanged information is provably deleted afterwards
- Pros: commitments are enforced by the technology; no single party holds all the data
- Cons: more costly to implement
- NB: you need to trust the intentions of *all* PPs *collectively, not individually* → stronger model



Shared MP-SPC system: built once, used by many



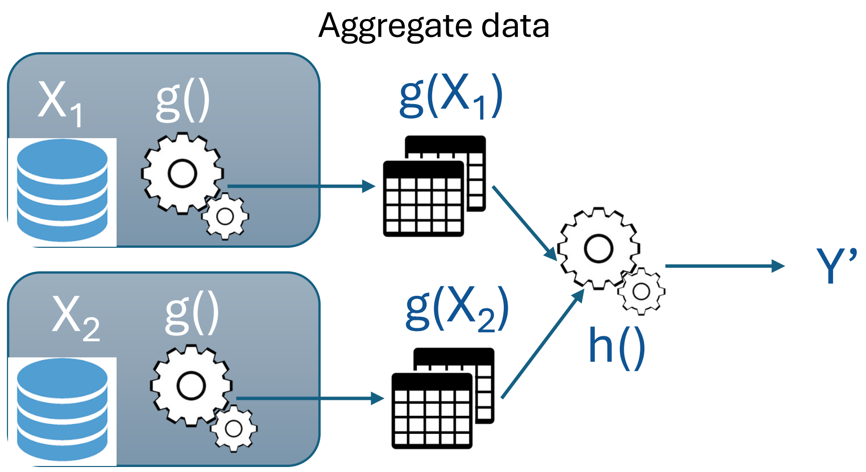
- Designing, developing, deploying and maintaining a MP-SPC system is costly and requires highly specialized skills
- Idea: build a common shared system for the whole ESS, to be used "on-demand" (as needed) by all ESS members and their partners
- *Servitisation* of secure computation: MP-SPC-as-a-Service (MPSPCaaS)
  - See [Ric04] for extensive description

A MP-SPC system is ultimately a "**data governance system**" where detailed policies, rules and roles are explicitly **defined at the organizational level** and **enforced at the technological level**

[Ric04] Ricciato, F. Steps Toward a Shared Infrastructure for Multi-Party Secure Private Computing in Official Statistic, JOS March 2024
https://journals.sagepub.com/doi/10.1177/0282423X241235259

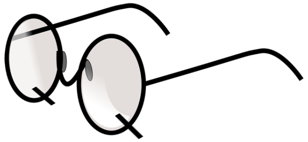
Are there alternatives to micro-data combination?

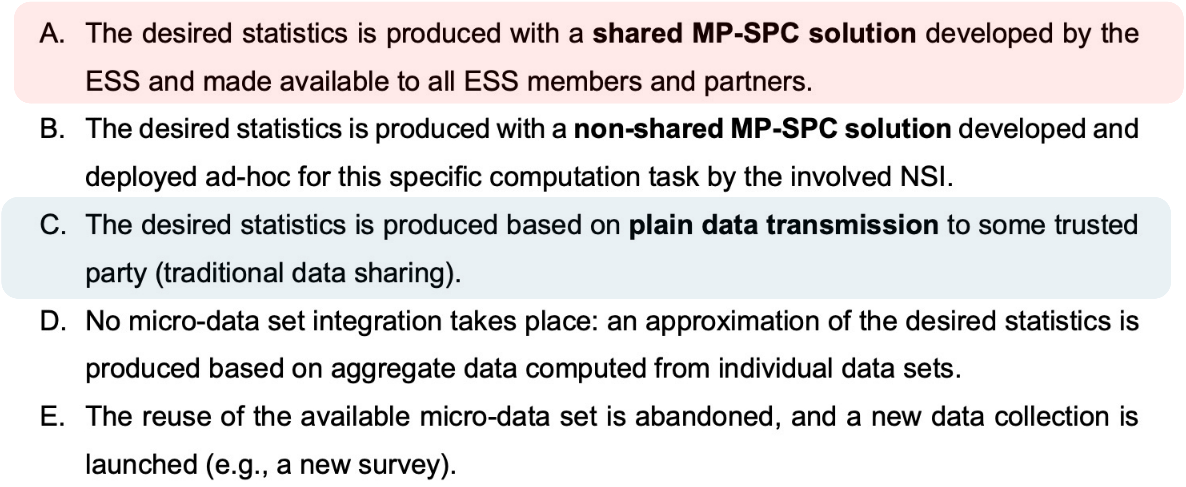
$$Y' = h(g(X_1),g(X_2)) \approx f(X_1,X_2) = Y$$



- Compute an approximation Y' of the target statistics Y that does not require combination of micro-data
  - Compute locally some aggregate intermediate data based on local function g(), then combine them with some composition function h()
  - Factorisation of f() into cascade g() and h() is not always feasible
  - When feasible, it may lead to a coarse approximation of the desired result
- Disregard these data and launch a new data collection

Looking through the quality glasses: quality considerations for *choosing* a shared MP-SPC system





A. The desired statistics is produced with a **shared MP-SPC solution** developed by the ESS and made available to all ESS members and partners.

B. The desired statistics is produced with a **non-shared MP-SPC solution** developed and deployed ad-hoc for this specific computation task by the involved NSI.

C. The desired statistics is produced based on **plain data transmission** to some trusted party (traditional data sharing).

D. No micro-data set integration takes place: an approximation of the desired statistics is produced based on aggregate data computed from individual data sets.

E. The reuse of the available micro-data set is abandoned, and a new data collection is launched (e.g., a new survey).

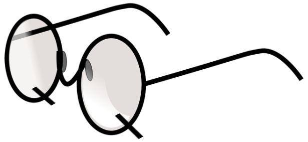
Trade-off between quality principles, specifically:

- Principle 5 . Statistical Confidentiality and Data Protection
- Principle 9. Non-excessive Burden on Respondents
- Principle 10. Cost effectiveness
- Principle 12. Accuracy and Reliability

Assessment to be done case-by-case.

Opting for a shared MP-SPC system can be regarded as "optimisation" across conflicting quality principles

Looking through the quality glasses: quality considerations for *designing* a shared MP-SPC system



Quality Assurance Framework of the European Statistical System

EUROPEAN STATISTICAL SYSTEM

Version 2.0

A MP-SPC system is ultimately a "**data governance system**" where detailed policies, rules and roles are explicitly **defined at the organizational level** and **enforced at the technological level**

In [Ric04] we have shown that

- *opting* for a shared MP-SPC system is a matter of compliance with **data protection** legislation…
- Hence its *design* should be guided by GDPR → **GDPR principles** may be interpreted as primary **requirements** for the specifications of a shared MP-SPC system

Along a parallel line of reasoning:

- *opting* for a shared MP-SPC system is (also) a matter of **quality optimisation**
- Hence its *design* should be guided also by QAF → selected **QAF elements** may provide guidance and inspiration for the specifications of a shared MP-SPC system
- Future QAF version 3.0 may consider making explicit reference to MP-SPC systems(?)

[Ric04] Ricciato, F. Steps Toward a Shared Infrastructure for Multi-Party Secure Private Computing in Official Statistic, JOS March 2024 https://journals.sagepub.com/doi/10.1177/0282423X241235259

Examples from QAF version 2.0

Quality Assurance Framework of the European Statistical System

EUROPEAN STATISTICAL SYSTEM

Version 2.0

Indicator 4.2 (Procedures are in place to plan, monitor and improve the quality of the statistical processes, including the integration of data from multiple data sources)

Indicator 8.3 (Statistical processes are routinely monitored and revised as required)

→ *Setting in place a formal system for qualified external auditing and issue reporting for MP-SPC system operations*

Indicator 6.4 (Information on data sources, methods and procedures used is publicly available)

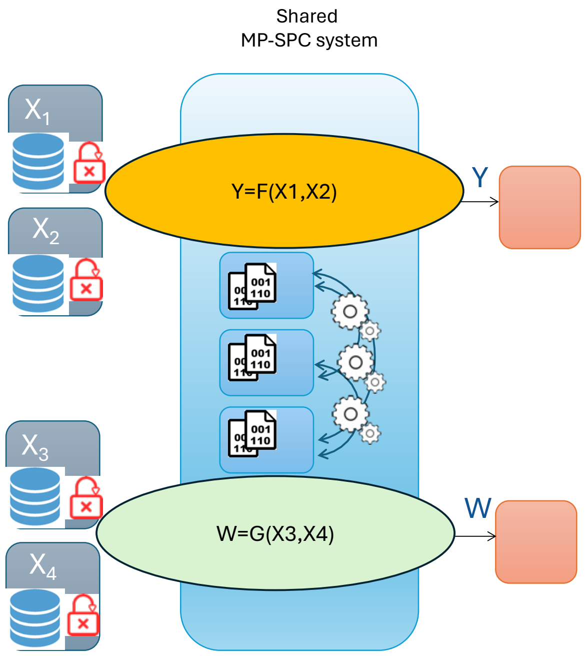
→ *The scripts source code that defines each computation task should be made public and open to scrutiny.*

→ *Logs of computation tasks should be auditable*

Indicator 5.4 (Guidelines and instructions are provided to staff on the protection of statistical confidentiality throughout the statistical processes. The confidentiality policy is made known to the public).

→ *Full documentation on MP-SPC system should be public*

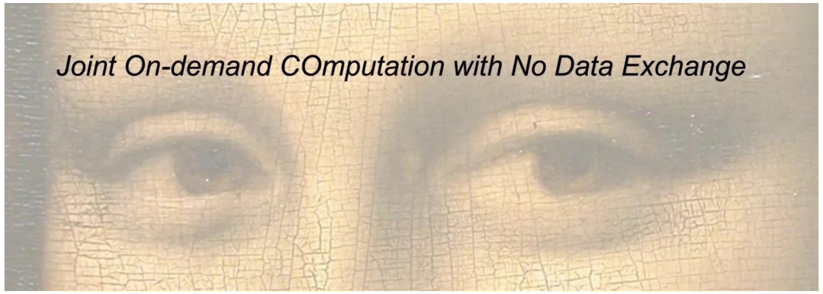
Take-home messages



- A MP-SPC system enables computation without exchange of data in intelligible form. No single entity has full control over the data or computation.
- Adopting a shared MP-SPC system for the ESS may be interpreted (also) as a matter of “quality optimisation”
- The design of a shared MP-SPC system could benefit from taking a close look at the QAF 2.0.

- Future QAF version 3.0 may consider making explicit reference to MP-SPC systems(?)

Towards implementation of a shared MP-SPC system for the ESS



- **J**oint **O**n-demand **CO**mputation with **N**o **D**ata **E**xchange **JOCONDE**
- A new project by Eurostat aimed at:

  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

- Started in April'24, will terminate in March'26 (24 months)
- In collaboration with cybernetica, an Estonian company specialised in security and privacy technologies (selected based on an open call for tenders)
- Multi-faceted project. List of Tasks:
  - Task 1 – Usage scenarios and system requirements
  - Task 2 – Technology analysis
  - Task 3 - Legal aspects
  - Task 4 – System specifications and architecture
  - Task 5 – Demonstrator prototype and functional testing
  - Task 6 – Trust building plan

https://cros.ec.europa.eu/joconde

Thank you!

INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

eurostat

The conference is partly
financed by the European Union