



Beyond the lab: How Multi-Party Secure Private Computing-as-a-service (MPSPCaaS) can foster the adoption of Input Privacy technologies in the ESS.

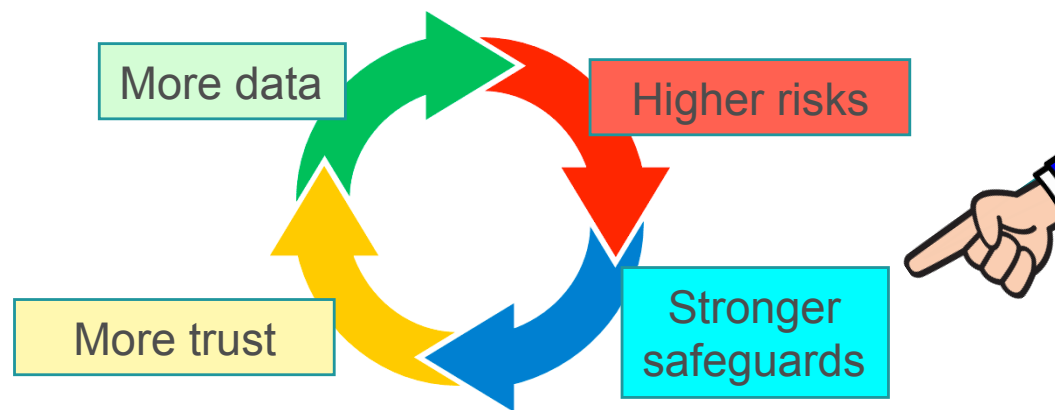
Fabio Ricciato

Unit A5 'Methodology; Innovation in Official Statistics'
Eurostat

WG Methodology meeting
Luxembourg, 23 March 2023

Why? Context and drivers

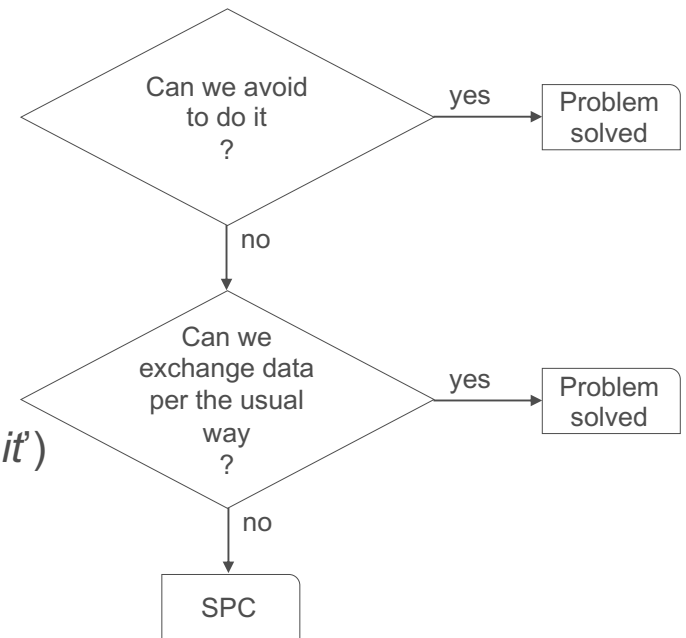
- Several trends Official Statistics innovation concur to increase the appetite for **cross-organisational data processing** in the context of
 - Data held by NSI in different Member States concerning cross-border phenomena (e.g., int'l trade, migration, ...)
 - Statistics based on data held by other public bodies (e.g., admin. records)
 - New statistics based on privately held data requiring integration across different providers (often competitors in the same business sector) and with data held by NSI
- Increasing awareness of the importance of **personal data** protection by the general public



Options

- Abstain

- Don't compute the statistics at all (*sorry, we can't do it*)
- (in some cases) process the data independently and combine the aggregate output data, accepting the lower quality of final statistics



- Exchange input data (e.g., microdata) between the involved entities or with a Trusted Third Party (TTP), with simple protection

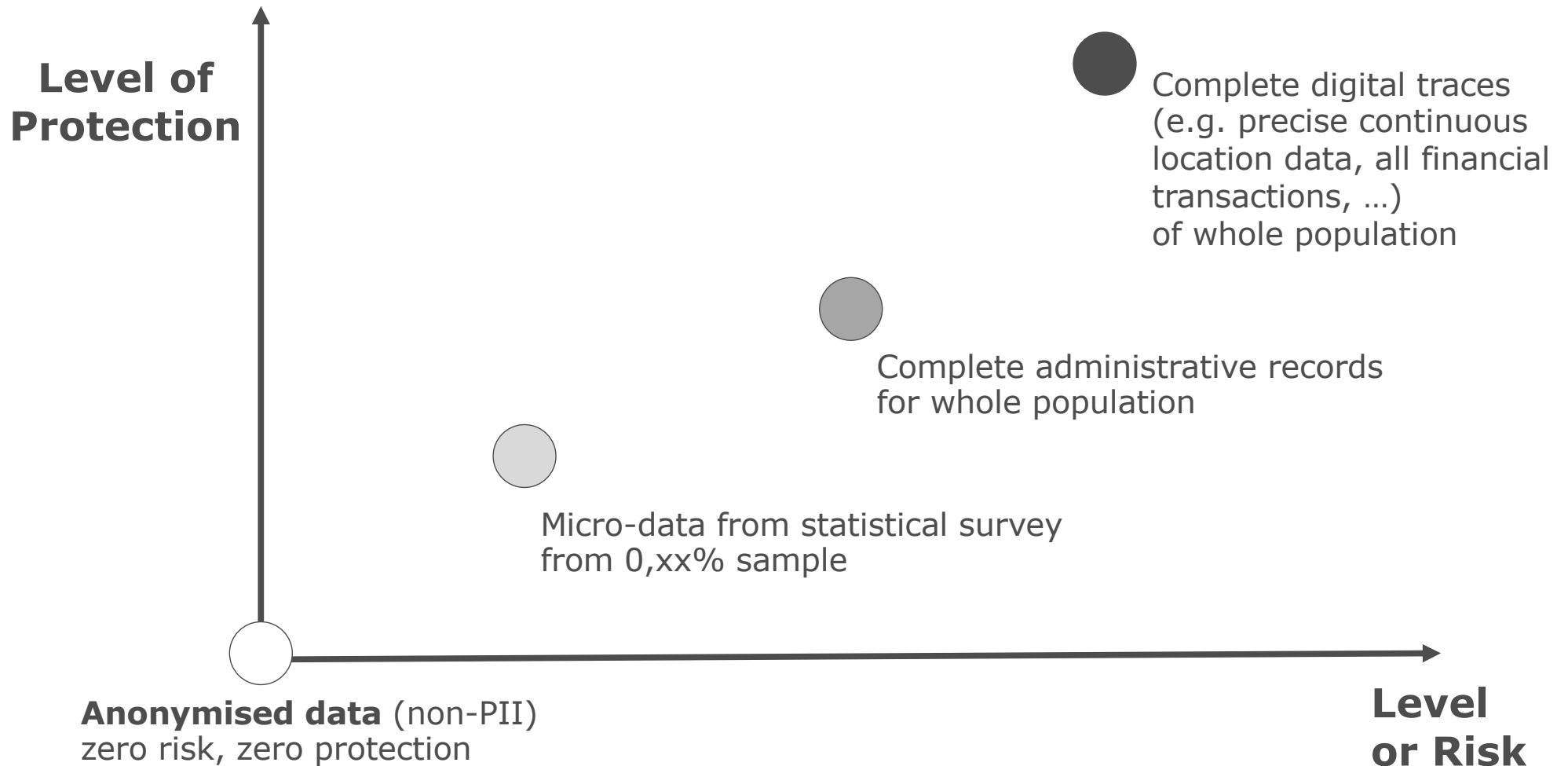
- Move your data outside to somebody else
- Delegate control over your data to "somebody else" → **TRUST**
- Multiplying the copies of the data = multiplying the **RISKS**

- Adopting a **Secure Private Computing (SPC)** solution

- Let the (predefined, aggregated) statistics being computed without letting the data being seen
- Remain in **CONTROL** (*Trust is good but control is better*)
- Lower the **RISKS**



Proportionality – a key GDPR concept

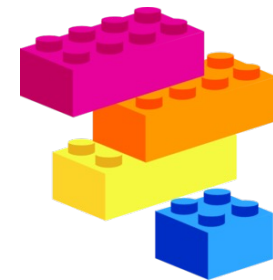


Secure Private Computing (SPC)

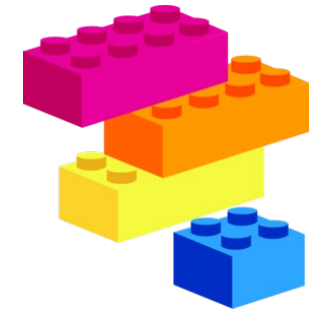
- Privacy Enhancing Technologies (PET) is an umbrella term comprising two distinct groups of methods/approaches that address distinct (often complementary) problems:



- **Input Privacy** (aka **Secure Private Computing**, aka Privacy-Preserving Computation)
 - Compute the output without exposing the input
 - Multi-Party Computation (MPC), Trusted Execution Environment (TEE), Homomorphic Encryption (HE) ...
- **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), Synthetic Data (SD) ...



SPC system

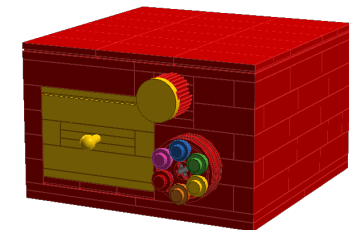


- A SPC solution is a *system of safeguards* comprising
 - **Technological** components (e.g., MPC + TEE)
 - **Organisational** components: policies, processes, agreements...
 - GDPR Art. 89 “**Technical and Organisational Measures**”

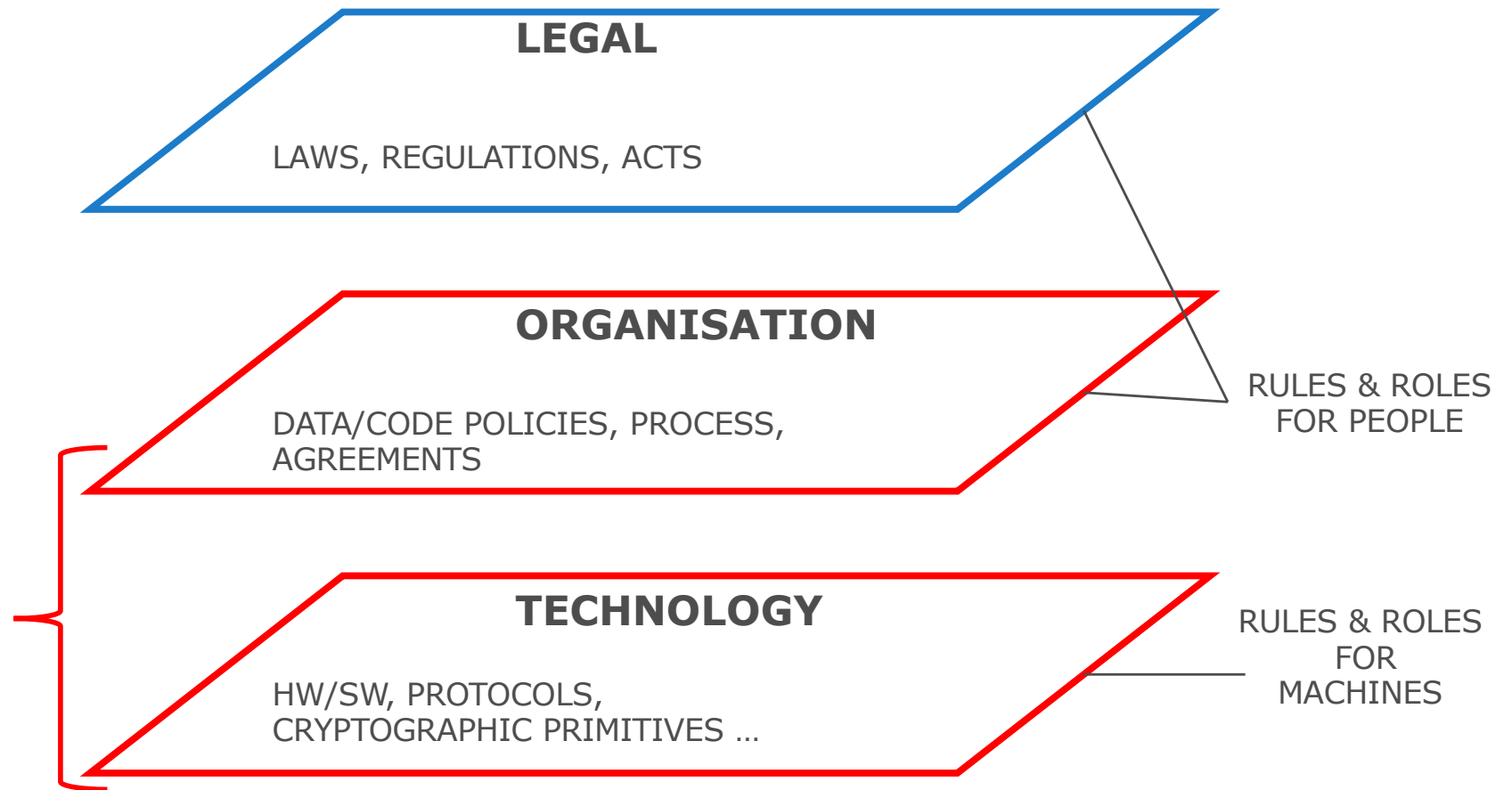
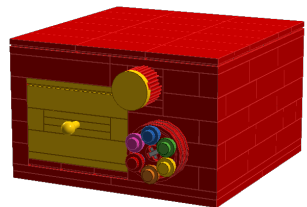
Article 89

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.




Legal compliance



Risk vs cost

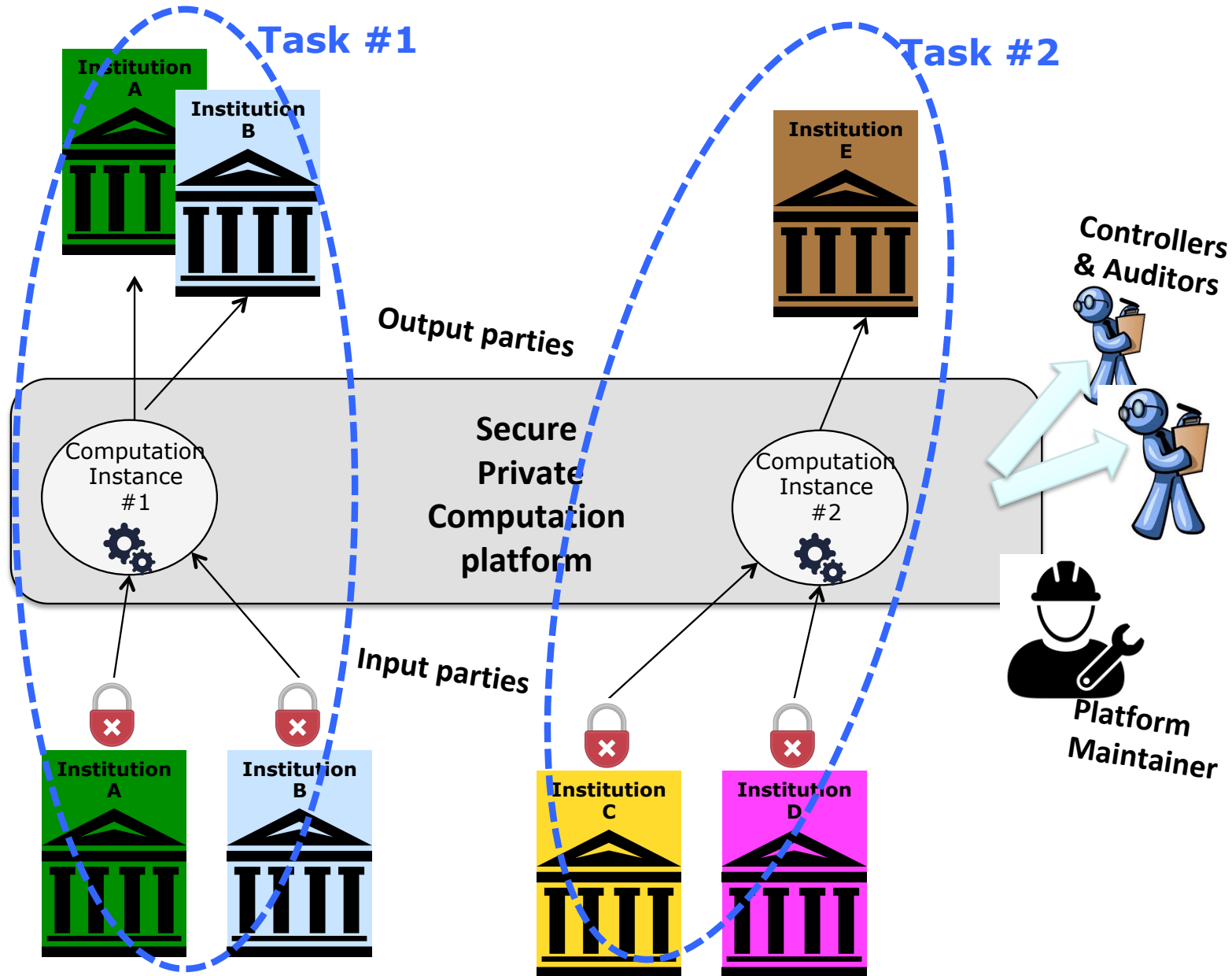


- Designing and building a **robust SPC system** is *costly*
 - Highly specialised skills: cryptography, HW/SW security, ...
 - €€€ for HW/SW infrastructure building, deploying, maintenance
 - May be too expensive for a single user and use-case
- Saving on costs → lower robustness → increase the risk 
 - This contradicts the primary motivation for SPC in the first place, i.e., “lowering the risk” and remaining in control
- Alternative: **shared SPC solution**
 - Build once, use many times (by multiple organisations, for multiple use-cases)
 - SPC-as-a-service (**SPCaaS**)

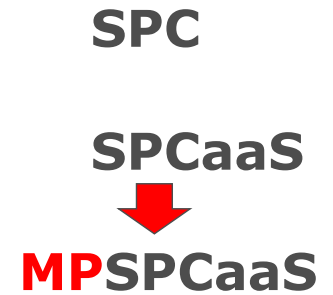


Secure Private Computing-as-a-service (SPCaaS)

SPC
↓
SPCaaS



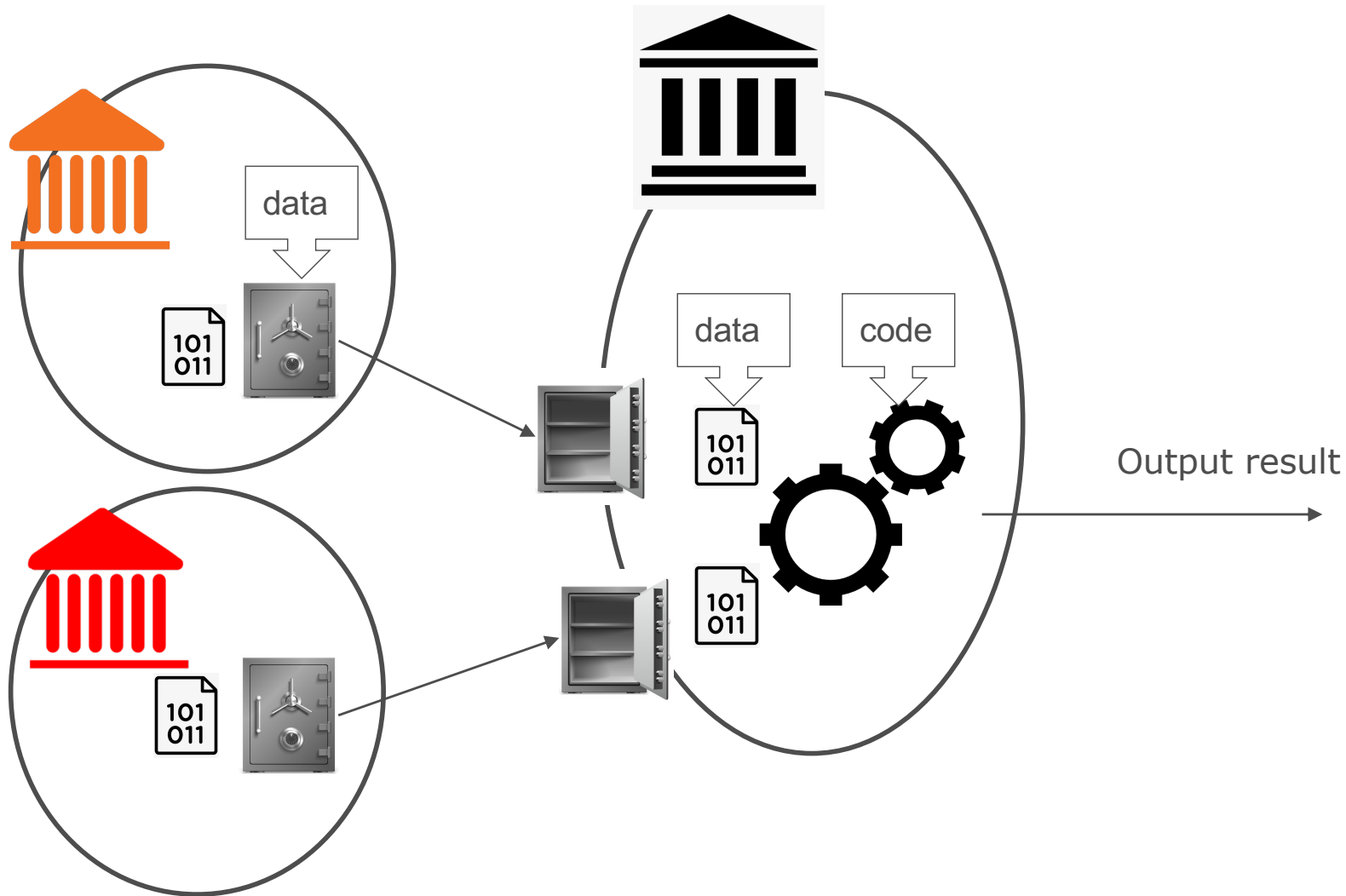
Why “Multi-Party”?



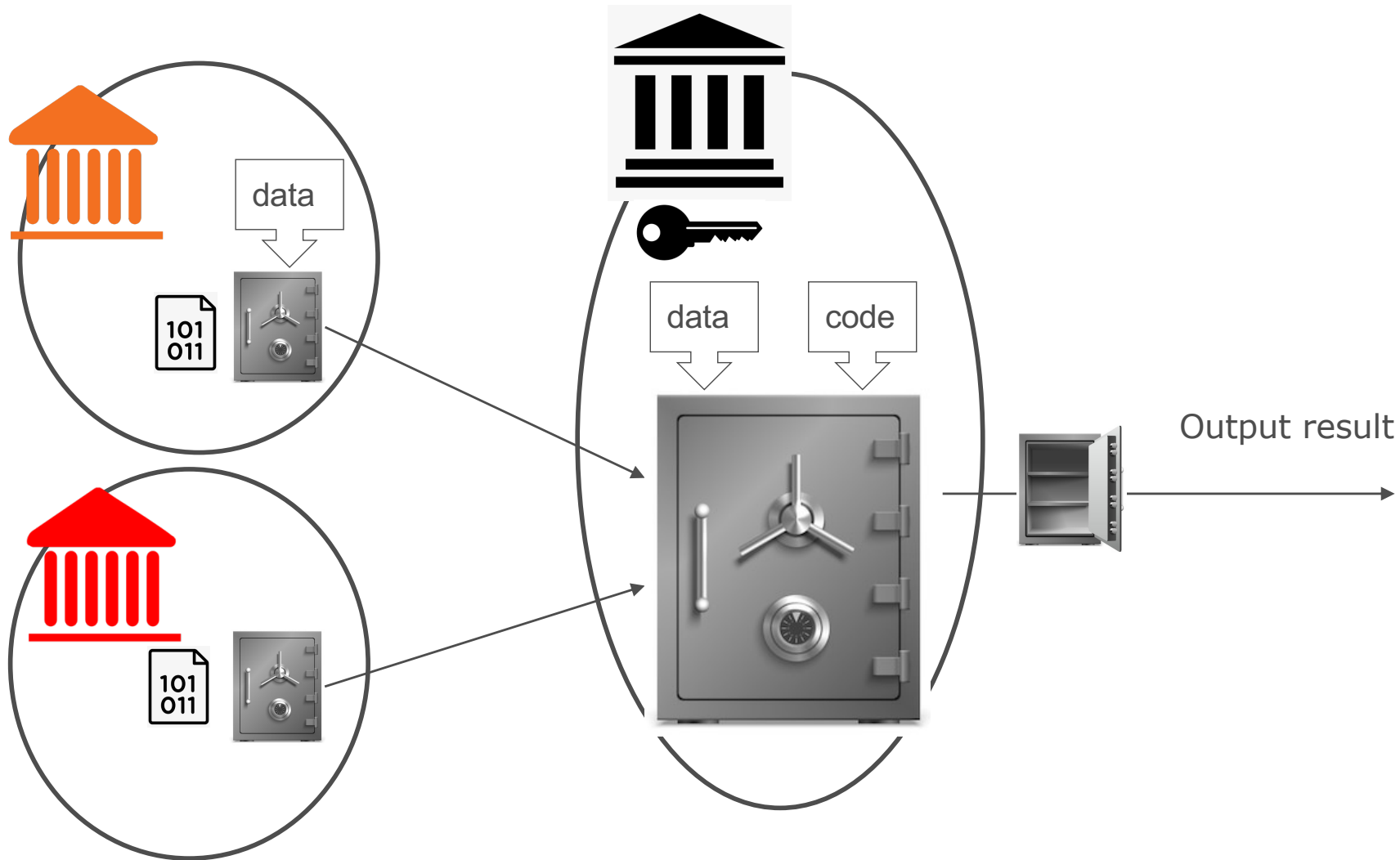
- The “MP” in MPSPCaaS ...
- And how is it possible to compute without seeing the data?
- What is the essence of Multi-Party Input Privacy Solutions?



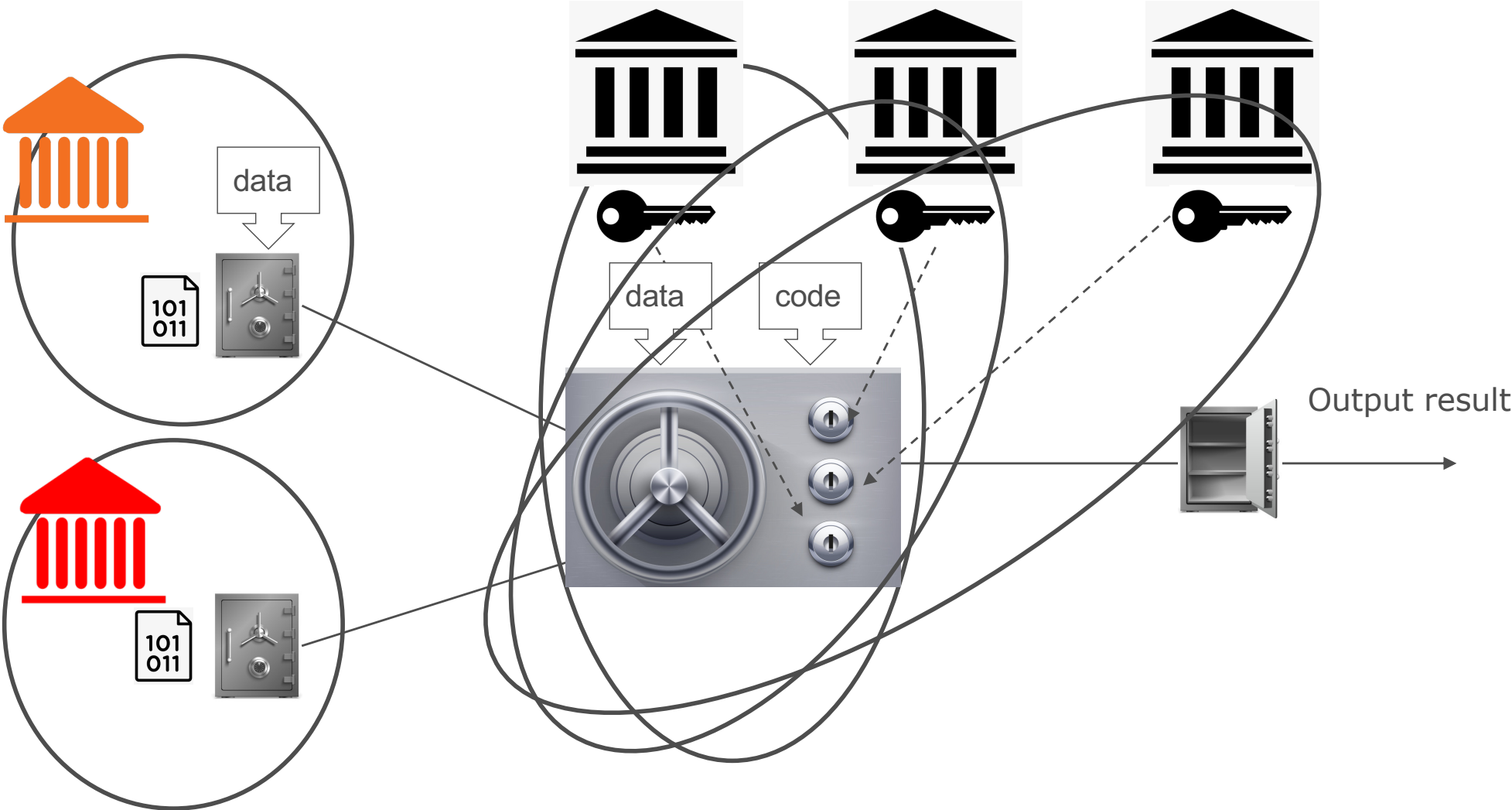
(Single) Trusted Third Party with computation in the clear



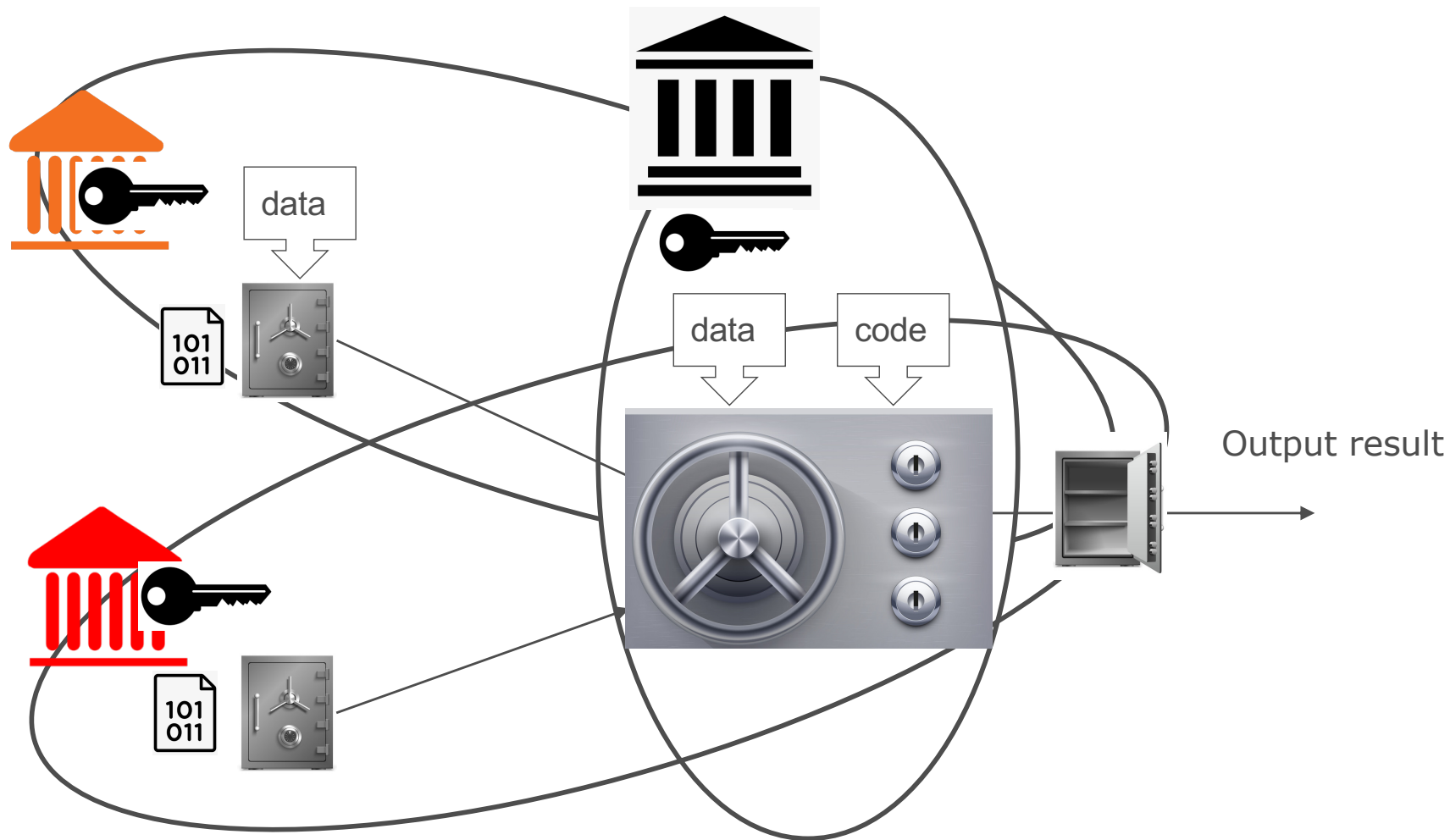
(Single) Trusted Third Party with protected computation (one key)



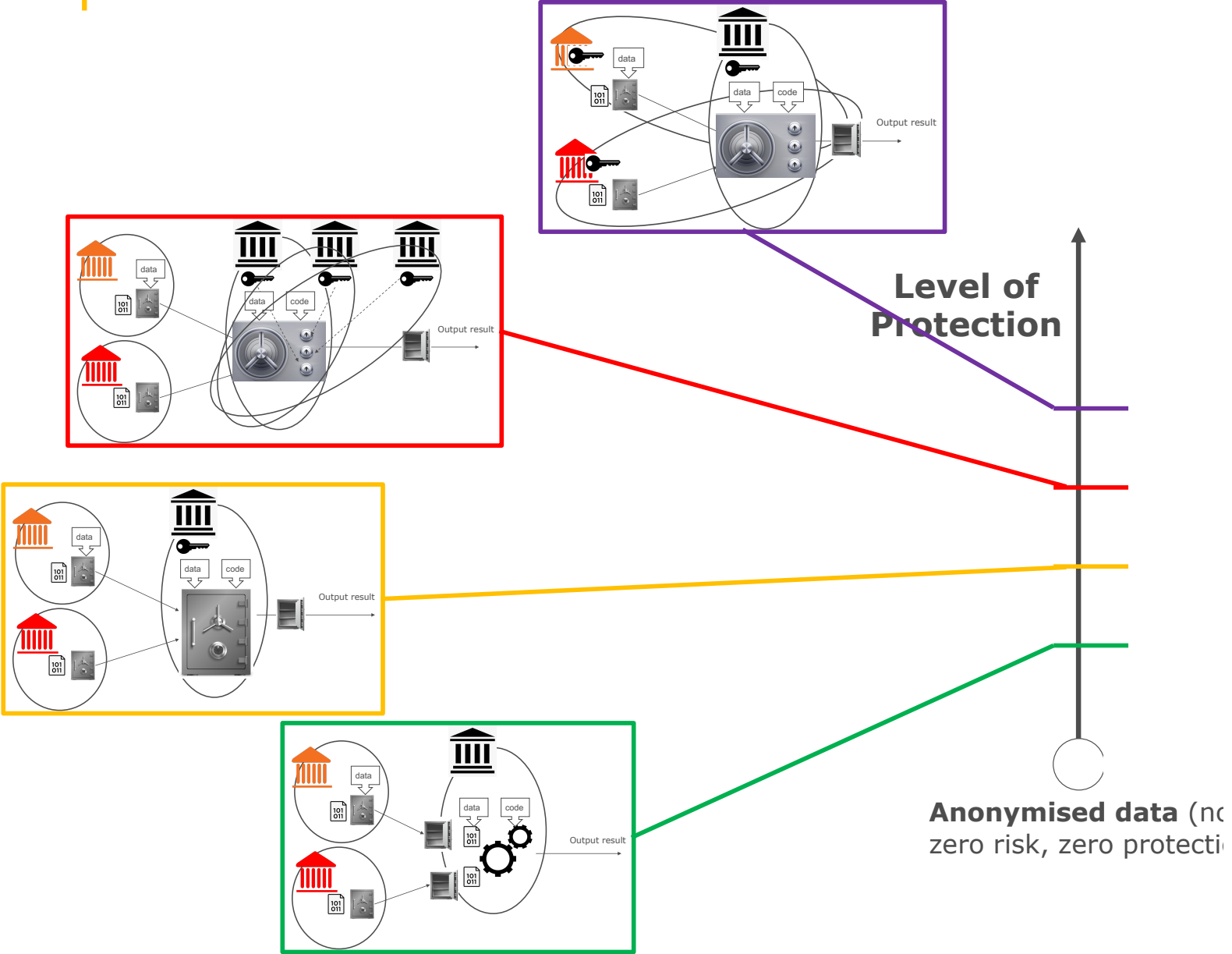
Multi-Party computation with external Processing Parties (multi-key)



Multi-Party computation with data holders acting also as Processing Parties



Level of protection proportional to risk



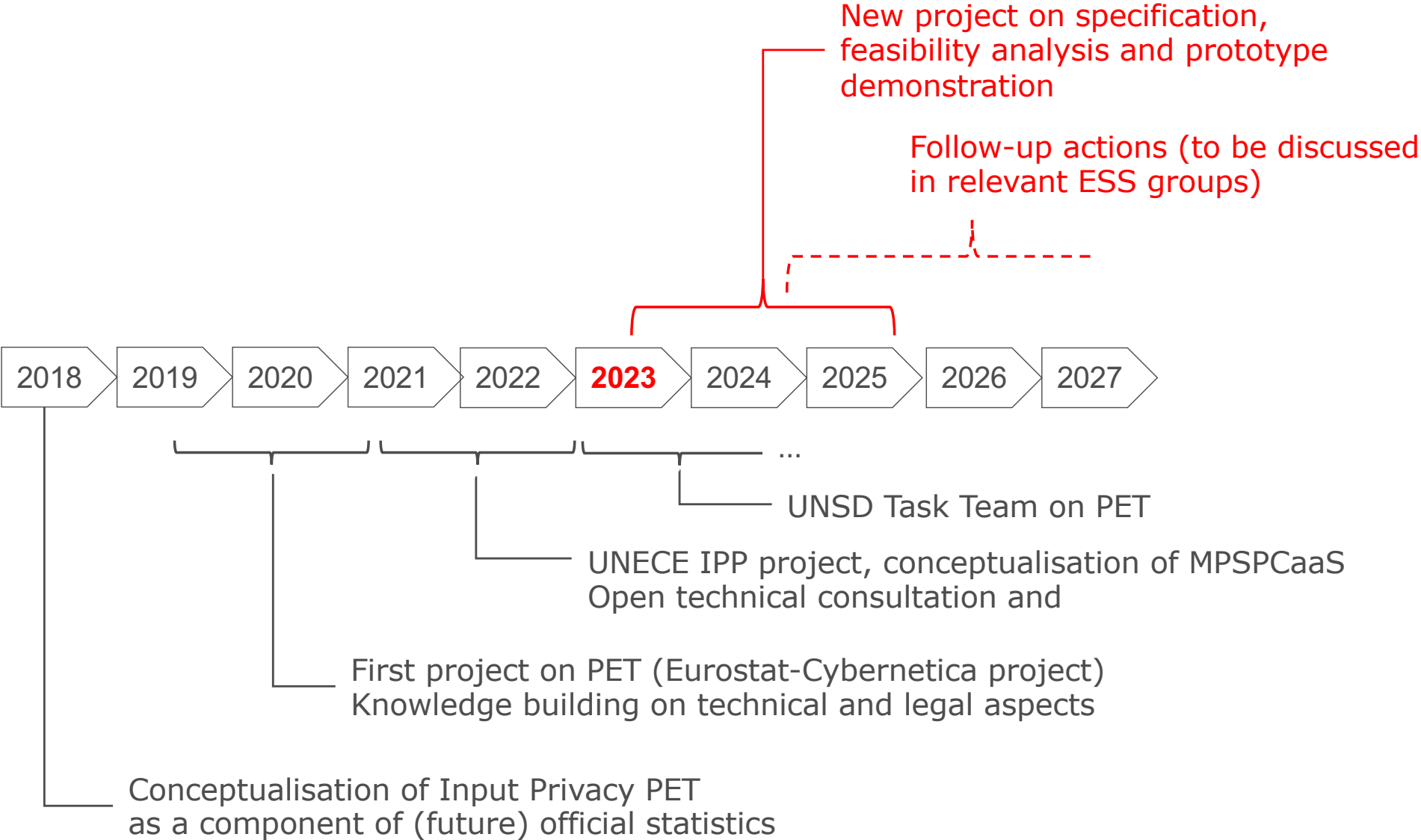
MPSPCaaS concept

- First proposed by Eurostat in the context of the UNECE HLG-MOS project on Input Privacy Preservation (IPP project, 2021-2022)
 - (2021) Discussed internally to IPP project team
 - (2022) Open Technical Consultation organised within the IPP project
 - Presentations and exchange of idea with data protection and privacy experts (ENISA workshop, MPC alliance, ...)
- Current state: clear understanding of technical and non-technical aspects and challenges (*“all unknowns seem to be known”*)
 - Technologies are there, but need to be tailored to adopters needs
 - Collaboration between technology providers and adopters is key
- Next step: **specification**, feasibility **analysis** and **demonstration** of a **prototype** - new project to be launched in 2023 (duration 2 years)

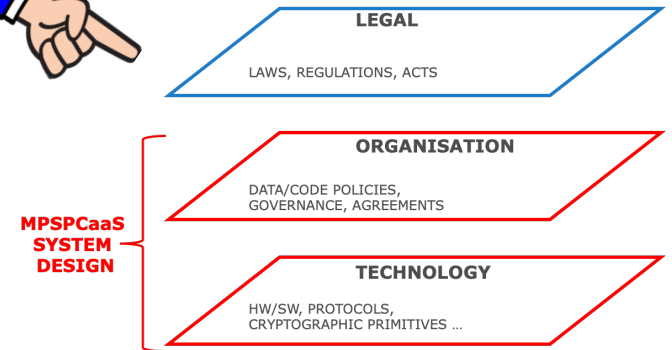
Questions to be answered at design phase

- **Technical Robustness** – How to engineer the system so as to achieve the highest level of security and robustness? What technical components do we need and how to combine them?
- **Usability** – How to make it easy and affordable for the users? How to lower the initial barrier for adoption?
- **Governance, liabilities and business model** – Who decides what? Who is liable for what? Who pays for what?
- **Public acceptance** – How to communicate to the public? How to prove publicly that we are not building yet another “big brother”?
- **Legal compliance:** What is need to ensure the SPC system is compliant with data protection laws?
 - **GDPR principles serving as “design requirements” for SPC system (!)**
 - Dialogue with data protection experts

Timeline - System Design



PET and European legislation



- Data Governance Act mentions 'secure processing environments'
- On 20/1/2023 the European Commission has adopted the proposal for a new regulation on European Statistics on Population and Housing (ESOP) making explicit reference in Recital (30), Art. 13 and Art. 14. ([link](#))
- EDPS opinion on ESOP published in 16/3/2023 ([link](#))

(30) When data sharing entails processing of personal data according to Regulation (EU) 2016/679 of the European Parliament and of the Council³⁷ or Regulation (EU) 2018/1725, the principles of purpose limitation, data minimisation, storage limitation and integrity and confidentiality should be fully applied. In particular, data sharing mechanisms based on privacy enhancing technologies that are specifically designed to implement these principles should be preferred over direct data transmission.

Article 13 Data sharing

1. Data shall be shared between the competent national authorities of different Member States, and between these competent national authorities and the Commission (Eurostat), exclusively for the purpose of developing and producing European statistics governed by this Regulation and of improving their quality.
2. In the interest of secure data sharing within the ESS, all necessary safeguards with regard to the physical and logical protection of data shall be taken. The Commission (Eurostat) shall set up a secure infrastructure to facilitate data sharing referred to in paragraph 1. Competent national authorities for statistics under this Regulation may use this secure data sharing infrastructure for the purpose specified in paragraph 1.
3. When the data concerned are confidential data within the meaning of Article 3, point 7, of Regulation (EC) No 223/2009 or personal data according to Regulations (EU) 2016/679 and (EU) 2018/1725, the sharing of such data shall be allowed and may take place on a voluntary basis provided it is:
 - (a) based on a request justifying the necessity to share the data in each individual case, in particular with regard to the quality issues to be specifically addressed;
 - (b) based preferably on privacy enhancing technologies that are specifically designed to implement the principles of Regulations (EU) 2016/679 and (EU) 2018/1725, with particular regard to purpose limitation, data minimisation, storage limitation, integrity and confidentiality;
 - (c) without prejudice to Chapter V of Regulation (EC) No 223/2009.
4. The Commission (Eurostat) and the Member States shall test and assess by means of pilot studies the fitness of relevant privacy enhancing technologies for data sharing.
5. Where the pilot studies under paragraph 4 of this Article identify effective and secure data sharing solutions for the purposes referred to in paragraph 1, the Commission may adopt implementing acts laying down technical specifications for the data sharing and measures for the confidentiality and security of information. These implementing acts shall be adopted in accordance with the examination procedure referred to in Article 18(2).

Article 14 Pilot and feasibility studies

1. The Commission (Eurostat) shall, where necessary and appropriate for the purposes of this Regulation, launch pilot and feasibility studies that aim at:
 - (a) assessing the availability of data sources and their quality, including of publicly and privately held data in Member States and at Union level;
 - (b) developing and assessing the feasibility of implementing new topics, detailed topics, statistical units, variables and their breakdowns;
 - (c) developing new methodologies and statistical techniques to reinforce quality;
 - (d) reducing asymmetries of migration flows;
 - (e) testing and assessing the fitness of relevant privacy enhancing technologies for secure data sharing within the ESS in accordance with Article 13(4);
2. Member States may participate in those studies but shall, together with the Commission (Eurostat), ensure the representativeness of those studies at Union level.
3. The results of those studies shall be evaluated by the Commission (Eurostat) in cooperation with Member States. The Commission (Eurostat) shall prepare in cooperation with the Member States reports on the findings of those studies.

Outlook



- Work is in progress ...
- ... advancing step-by-step, from initial **concept** through **specification** to future **deployment** of shared PET infrastructure for the ESS, based on the MPSPCaaS concept
 - Pilot testing (based on prototype demonstrator) could possibly start in second half of 2025
- Continuous dialogue with **technology specialists** and **data protection legal experts** to define certain open issues (known unknowns) at the edge between legal and technology

References

- PET4OS page on CROS portal
https://ec.europa.eu/eurostat/cros/content/privacy-enhancing-technologies-official-statistics-pet4os_en
 - With links to dedicated pages on Eurostat-Cybernetica project, various presentations and papers
- Final Report of UNECE HLG-MOS IPP project
<https://statswiki.unece.org/x/mQCQFw>

Questions? Comments?

