"When GDPR met COED …"
a tale of two made one for the other

Fabio Ricciato
Unit A5 'Methodology; Innovation in Official Statistics'
Eurostat

COED workshop
KU Leuven, 10 October 2023

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.

Summary of this talk

- GDPR and COED are more than friends to each other
  - Like the characters in the movie, it took a while to realise that …
  - *NB: I'll use the equivalent term Secure Private Computing (SPC) instead of COED*
- Multi-Party SPC systems are game-changers
  - From individual game to team game
- What Eurostat is doing in the field
  - MPSPCaaS concept and project

A NEW COMEDY BY ROB REINER
BILLY CRYSTAL MEG RYAN
Can two friends sleep together and still love each other in the morning?
When Harry Met Sally...

Eurostat and the ESS

- Eurostat is …
  - the statistical office of the EU
  - a DG of the European Commission
  - the coordinator of the ESS
- The European Statistical System (ESS) is the **partnership** between
  - Eurostat (coordinator)
  - National Statistical Institutes (NSIs) in each EU country
  - Other National Authorities (ONAs) in each EU country
- Eurostat (i) produces European statistics and (ii) contributes to harmonise methodologies, definitions, criteria, etc. in the ESS

About myself

- Not a statistician, not a lawyer, not a cryptographer…
- Education in Electrical Engineering
- Previous academic life in technology research
    - telecommunication systems, computer networks, mobile networks, signal processing, software-defined radio, radio localisation…
    - First encounter with secret sharing back in 2006 for collaborative network monitoring…
- Since 2018 with Eurostat
    - Dealing with “innovation in official statistics”

There are I-PETs and O-PETs

- **Input Privacy (Enhancing) Technologies** (I-PET for short) allow (i) computing the exact predefined output y and delivering it to the predefined output party/ies while (ii) preventing anybody from learning anything about the input data x other than what can be inferred directly from y (including of course "seeing" the input data themselves) … all the above is of course valid under certain conditions (→ scenario assumptions, attack model)
  - HE, SMPC (secret sharing), TEE…
- **Output Privacy (Enhancing) Technologies** (O-PET for short) aim at producing a quasi-result $y* \approx y$ that fulfils two conflicting conditions, namely (i) it is sufficiently close to the exact result y to be still useful for the intended purpose, but at the same time (ii) it does not allow to infer back individual identities or characteristics of the data subjects represented in the input data x.
  - DP, SD, FL, …

There are I-PETs and O-PETs

- **Input Privacy (Enhancing) Technologies** (I-P
  (i) computing the exact predefi
  predefined outpu dy from l
  anythi what can be inferre
  ng the input data themselves)
  nd under certain conditions
  ns assumptions, attack model)
  - HE, SMPC (secret sharing), TEE…

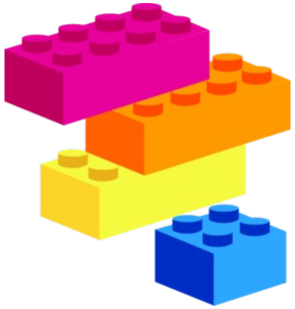
a.k.a. Secure Private Computing (SPC)

In scope of this talk ✓

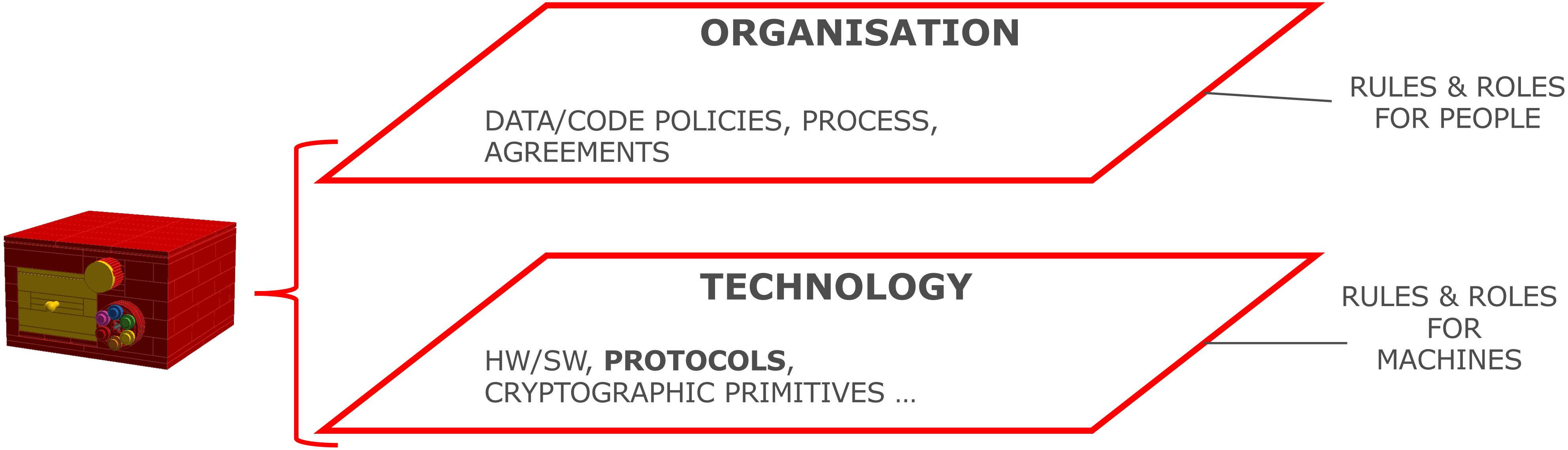
- **Output Privacy (Enhancing) Technologies** (O-PET for sho at
  pr quasi-result $y* \approx y$ that fulfils tw ons,
  na is sufficiently close t be still useful for
  th d purpose (ii) it does not allow to infer
  ba ual aracteristics of the data subjects
  represented in t nput data x.
  - DP, SD, FL, …

✗ NOT in scope of this talk

SPC system



- SPC technologies are the **bricks, not a magic stick –** one needs to engineer a whole system solution (hardware, software and … *humanware*)
- SPC ***enforces technologically*** governance policies for **data & code**
  - stipulating *ex-ante* what output information is computed on the data, with what code and who will see it
  - adopting technological solutions that prevent any other entity seeing any other information (including the input data themselves) if certain conditions are met (attack model, trust model) – and verify *ex-post*



ORGANISATION

DATA/CODE POLICIES, PROCESS, AGREEMENTS

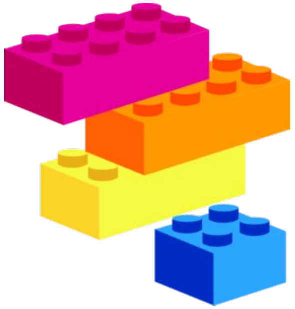
RULES & ROLES FOR PEOPLE

TECHNOLOGY

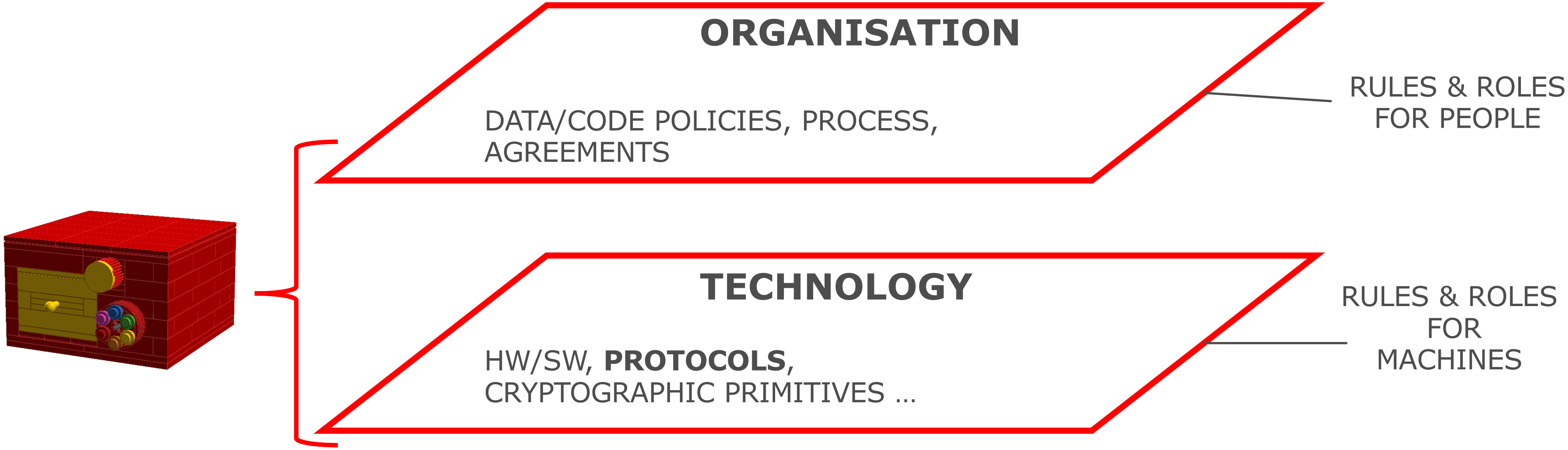
HW/SW, **PROTOCOLS**, CRYPTOGRAPHIC PRIMITIVES …

RULES & ROLES FOR MACHINES

SPC system



- A SPC solution is a *system of safeguards* comprising
  - **Technological** components (e.g., SMPC + TEE + …)
  - **Organisational** components: policies, processes, agreements…



ORGANISATION

DATA/CODE POLICIES, PROCESS, AGREEMENTS

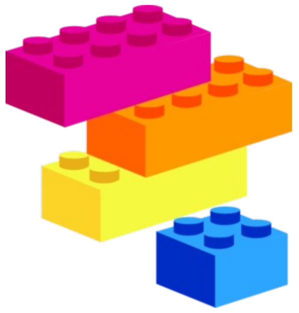
RULES & ROLES FOR PEOPLE

TECHNOLOGY

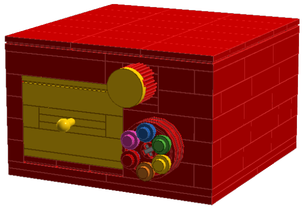
HW/SW, **PROTOCOLS**, CRYPTOGRAPHIC PRIMITIVES …

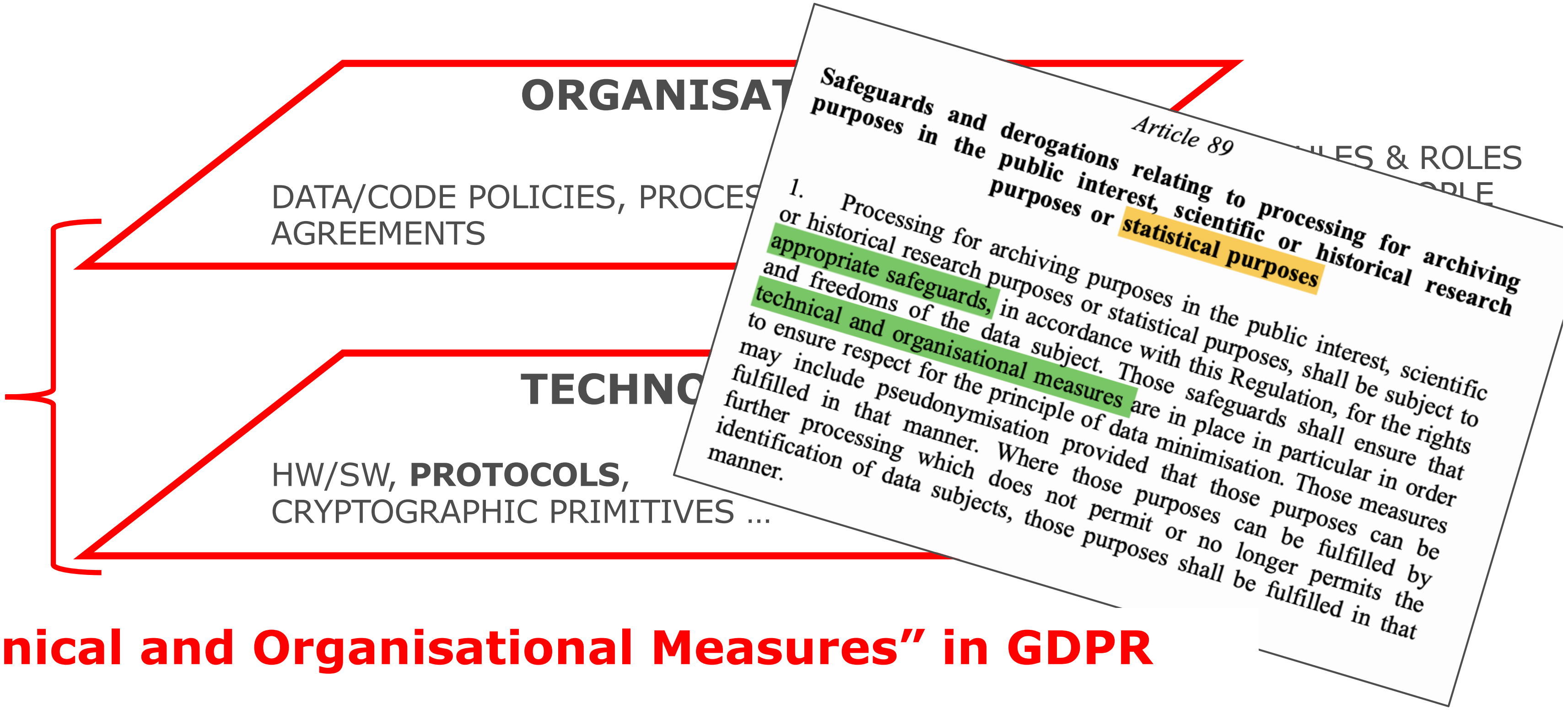
RULES & ROLES FOR MACHINES

SPC system



- A SPC solution is a *system of safeguards* comprising
  - **Technological** components (e.g., SMPC + TEE + …)
  - **Organisational** components: policies, processes, agreements…





ORGANISAT

DATA/CODE POLICIES, PROCES
AGREEMENTS

LES & ROLES
OPLE

TECHNO

HW/SW, **PROTOCOLS**,
CRYPTOGRAPHIC PRIMITIVES …

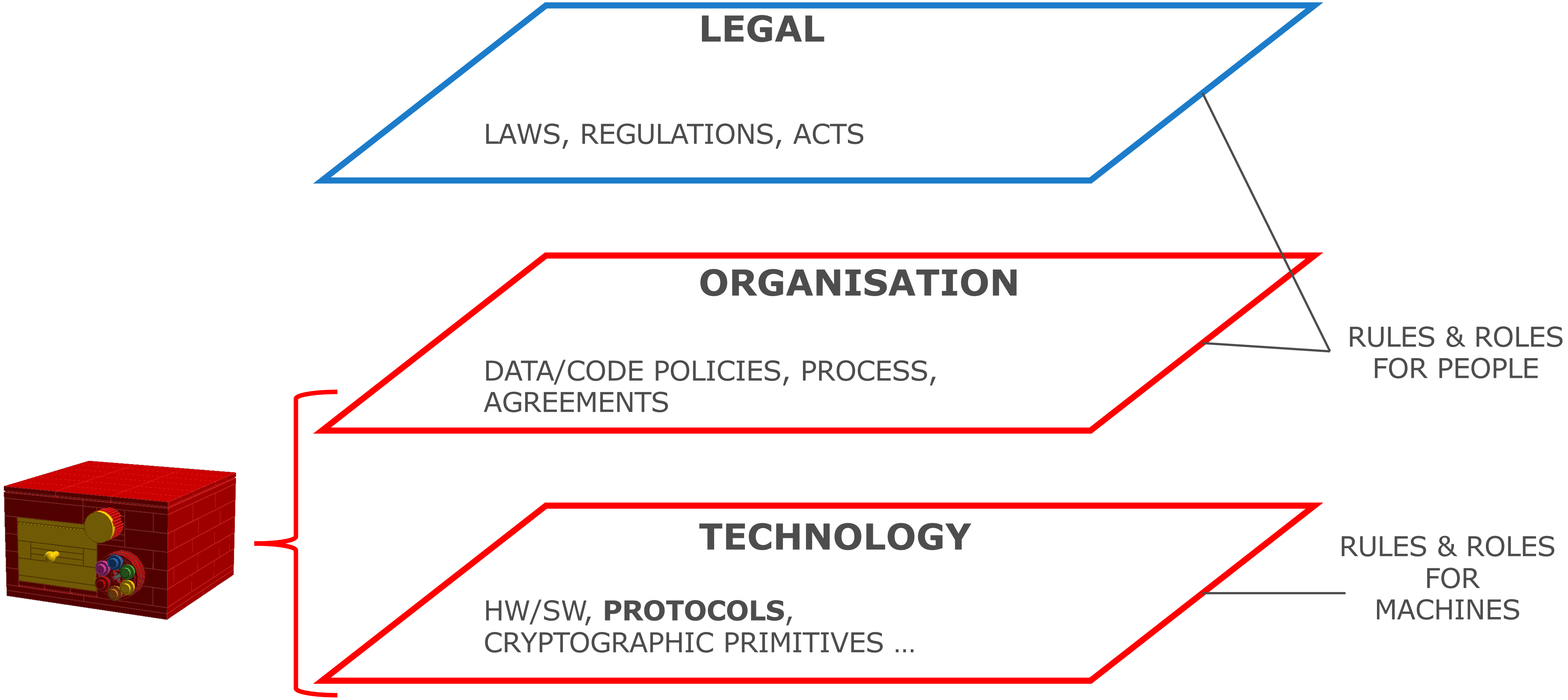
Article 89

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.

“Technical and Organisational Measures” in GDPR

3 normative layers



LEGAL

LAWS, REGULATIONS, ACTS

ORGANISATION

DATA/CODE POLICIES, PROCESS, AGREEMENTS

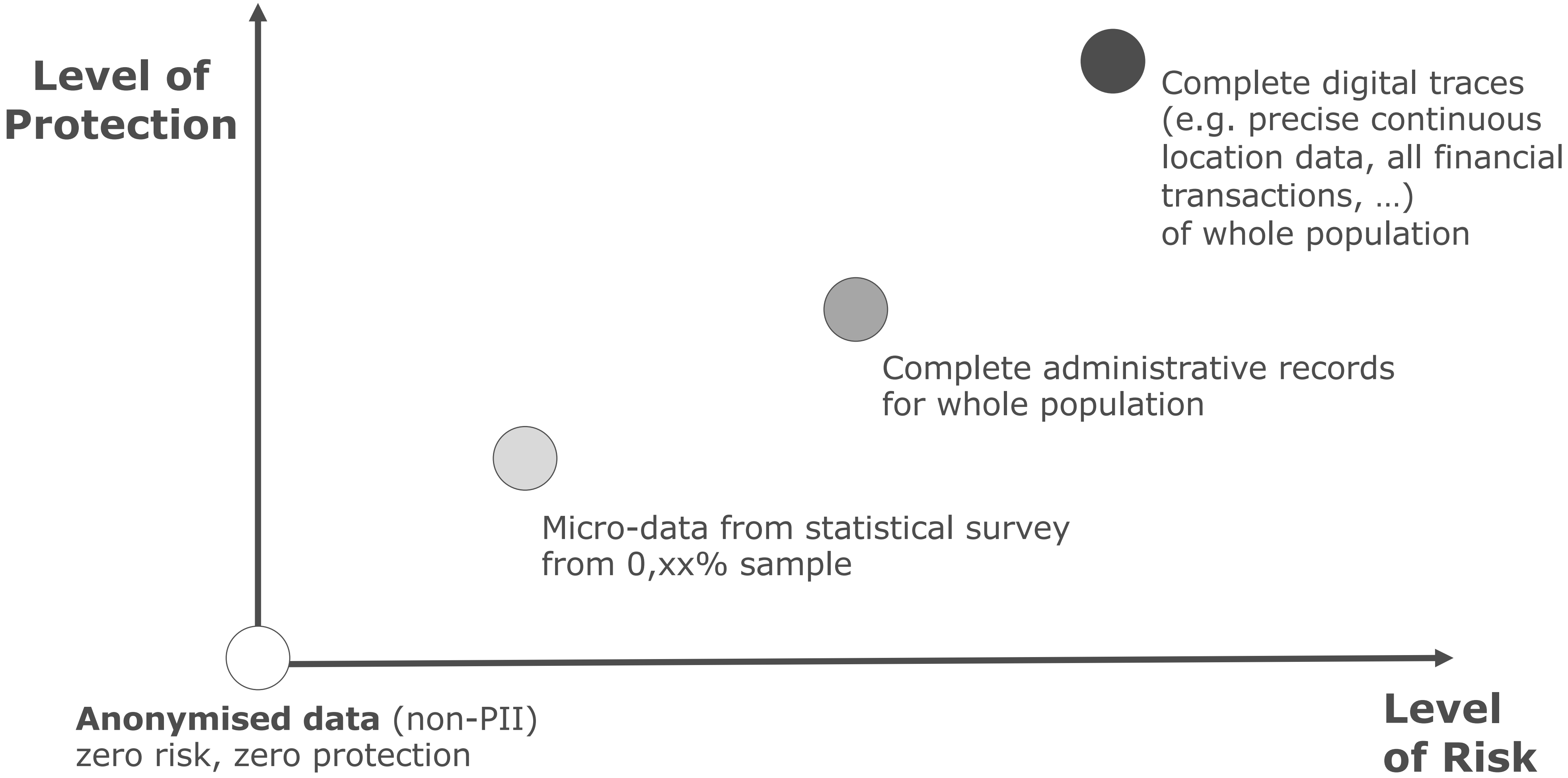
TECHNOLOGY

HW/SW, **PROTOCOLS**, CRYPTOGRAPHIC PRIMITIVES …

RULES & ROLES FOR PEOPLE

RULES & ROLES FOR MACHINES

Proportionality – a key GDPR concept



Level of Protection

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

Complete administrative records for whole population

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

Anonymised data (non-PII)
zero risk, zero protection

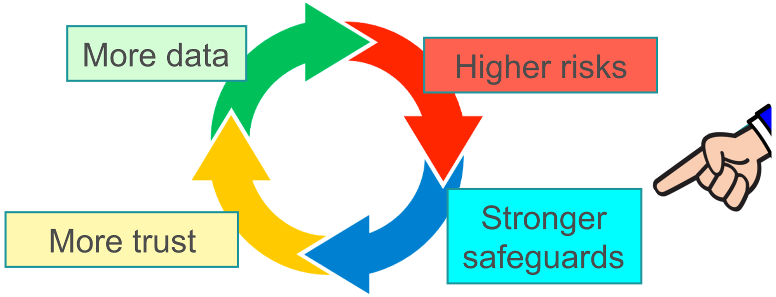
Level of Risk

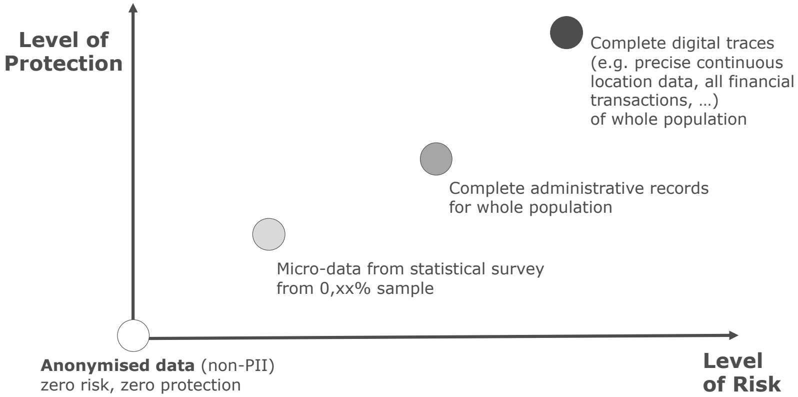
PII Personal Identifiable Information

Why SPC in Official Statistics?

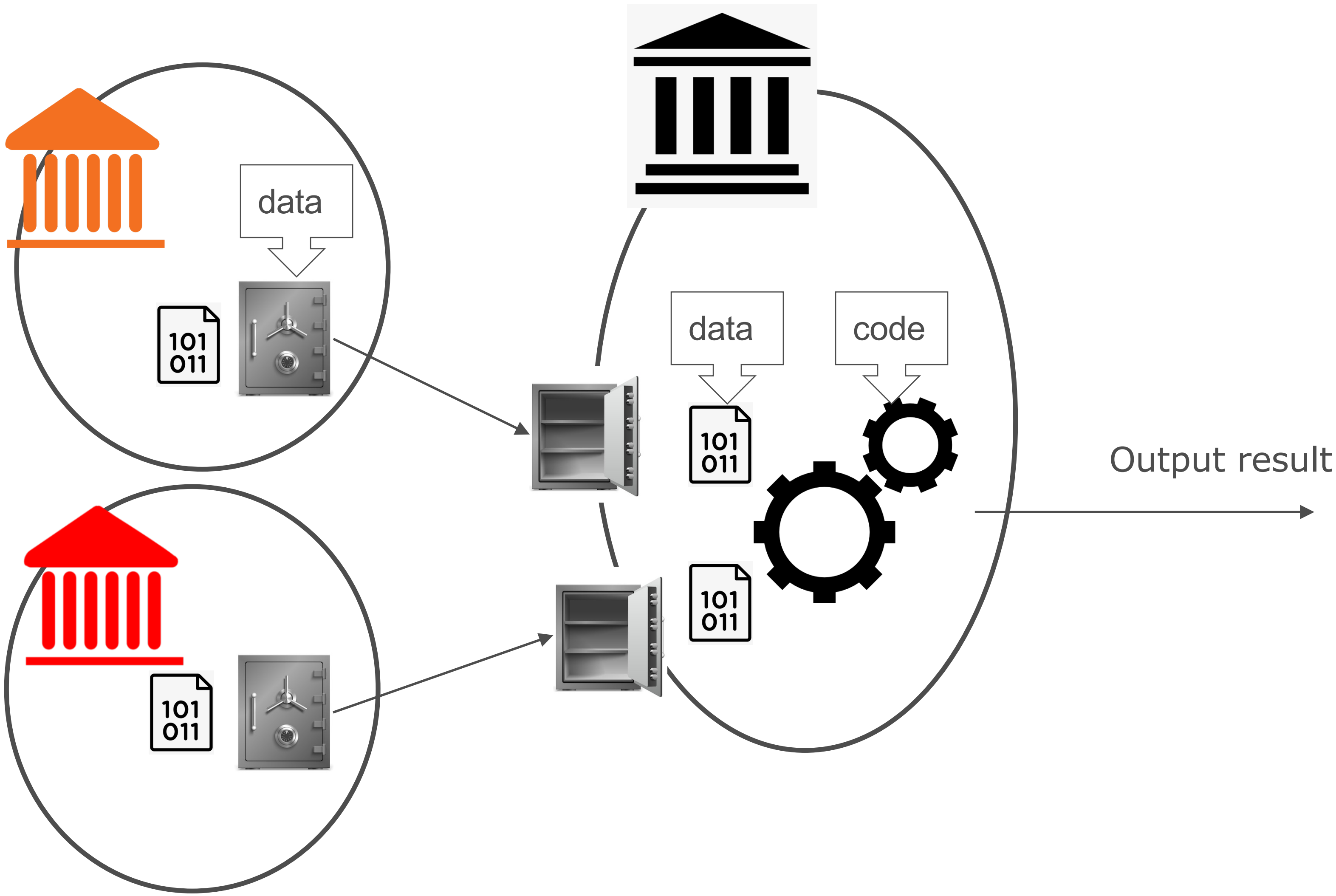
- Several trends Official Statistics innovation concur to increase the appetite for **cross-organisational data processing** in the context of
  - Data held by NSI in different Member States concerning cross-border phenomena (e.g., int'l trade, migration, …)
  - Statistics based on data held by other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

- Increasing awareness by the general public of **personal data** protection

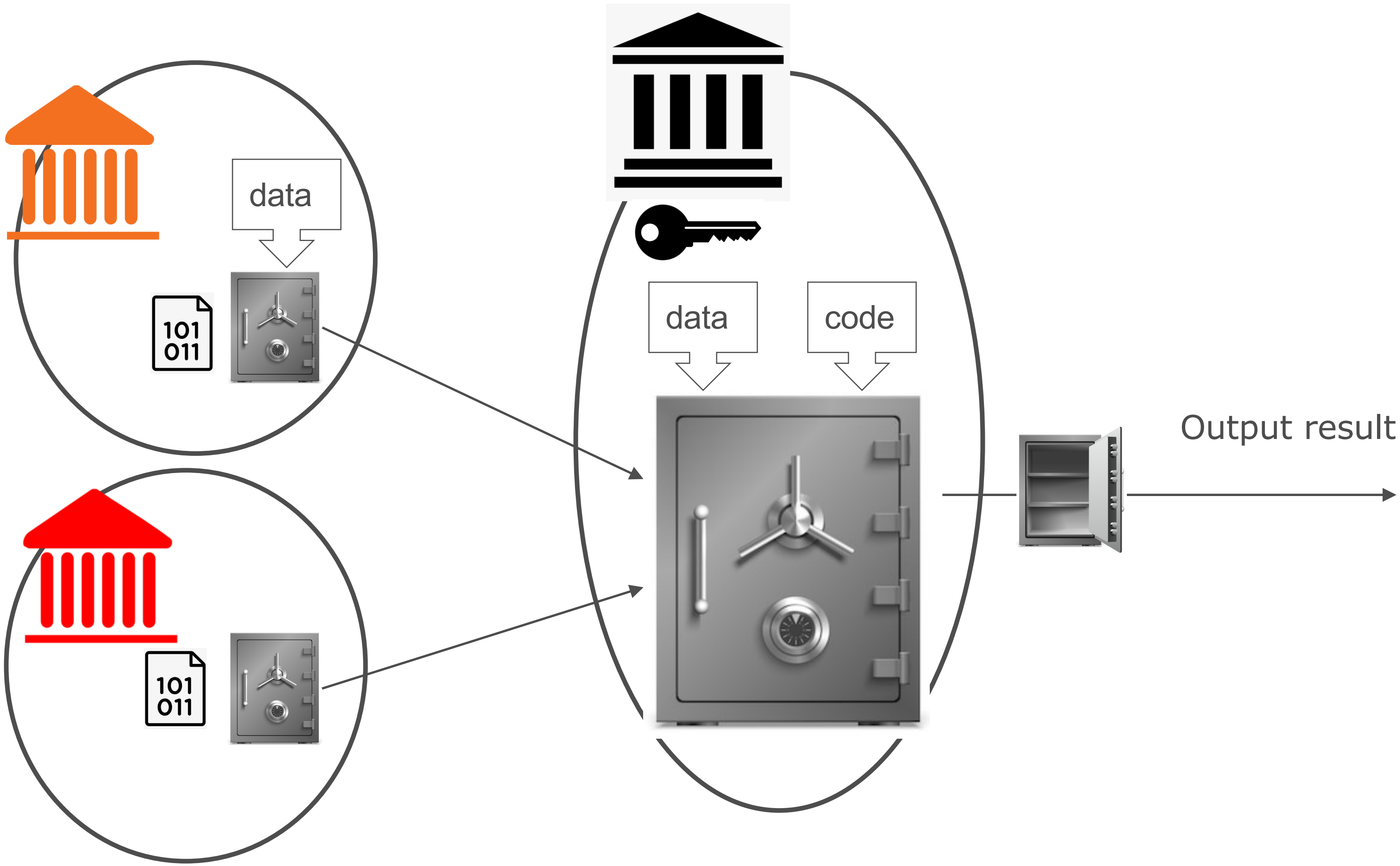




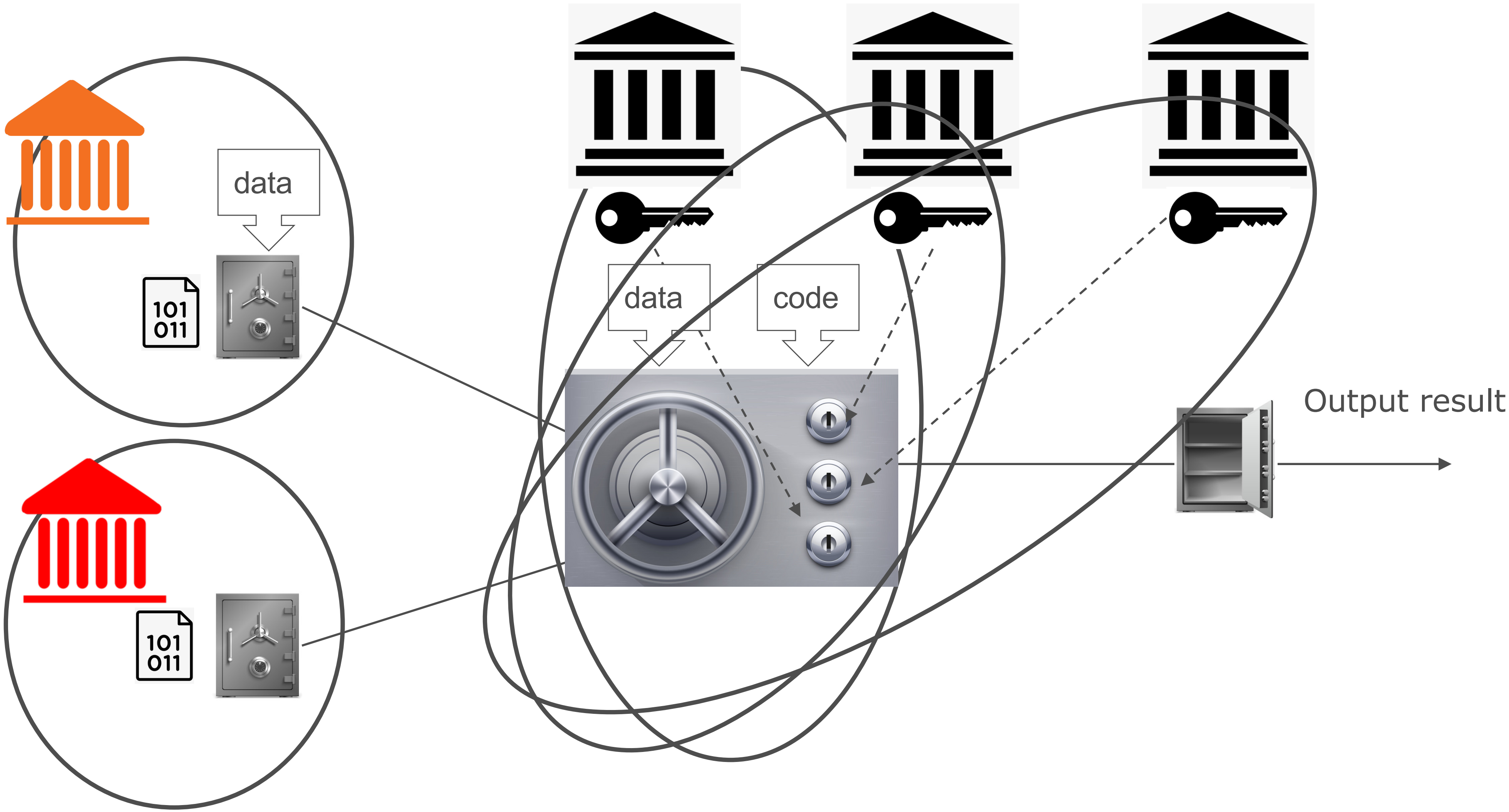
(Single) Trusted Third Party with computation in the clear



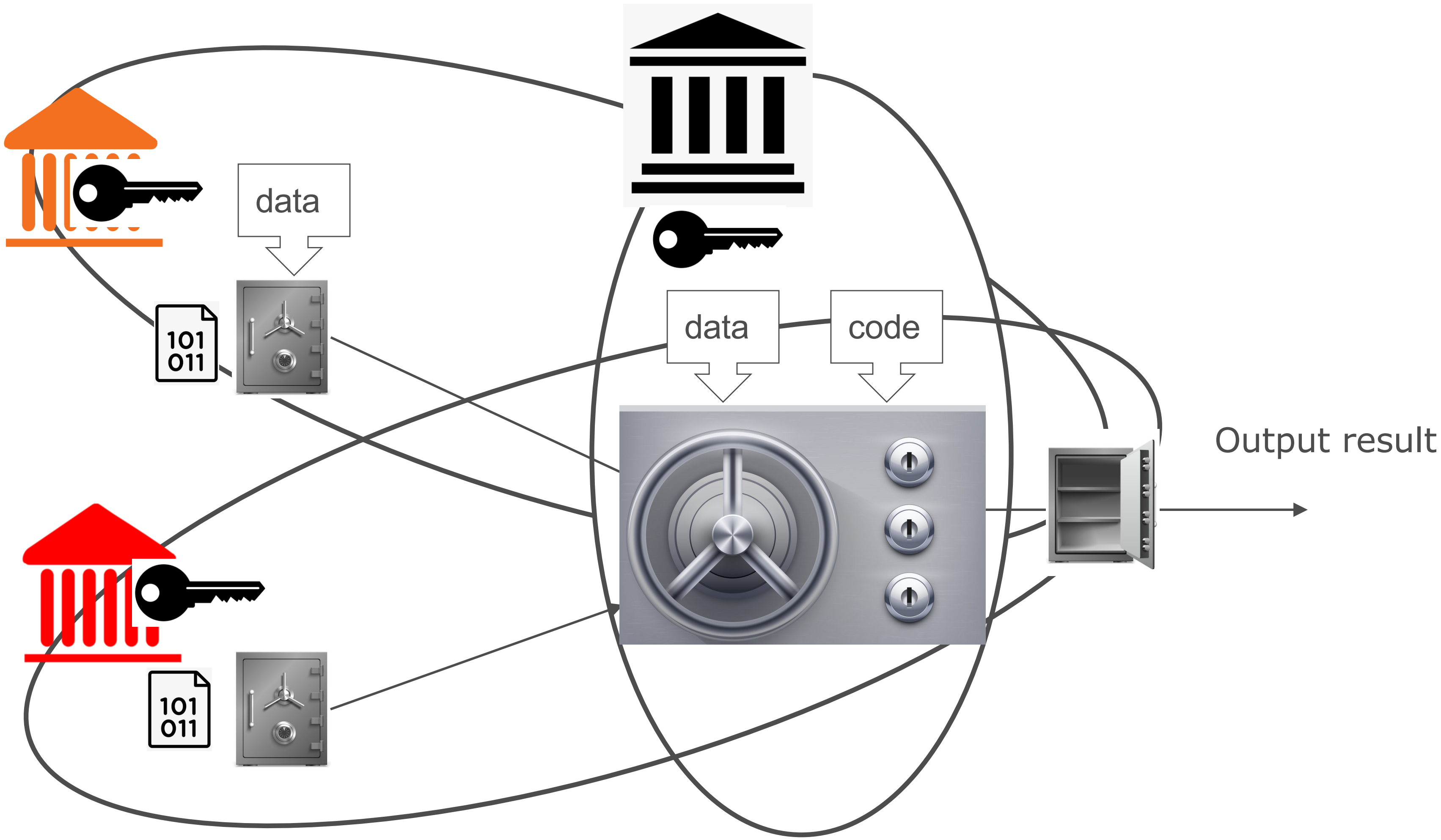
(Single) Trusted Third Party with protected computation (one key)



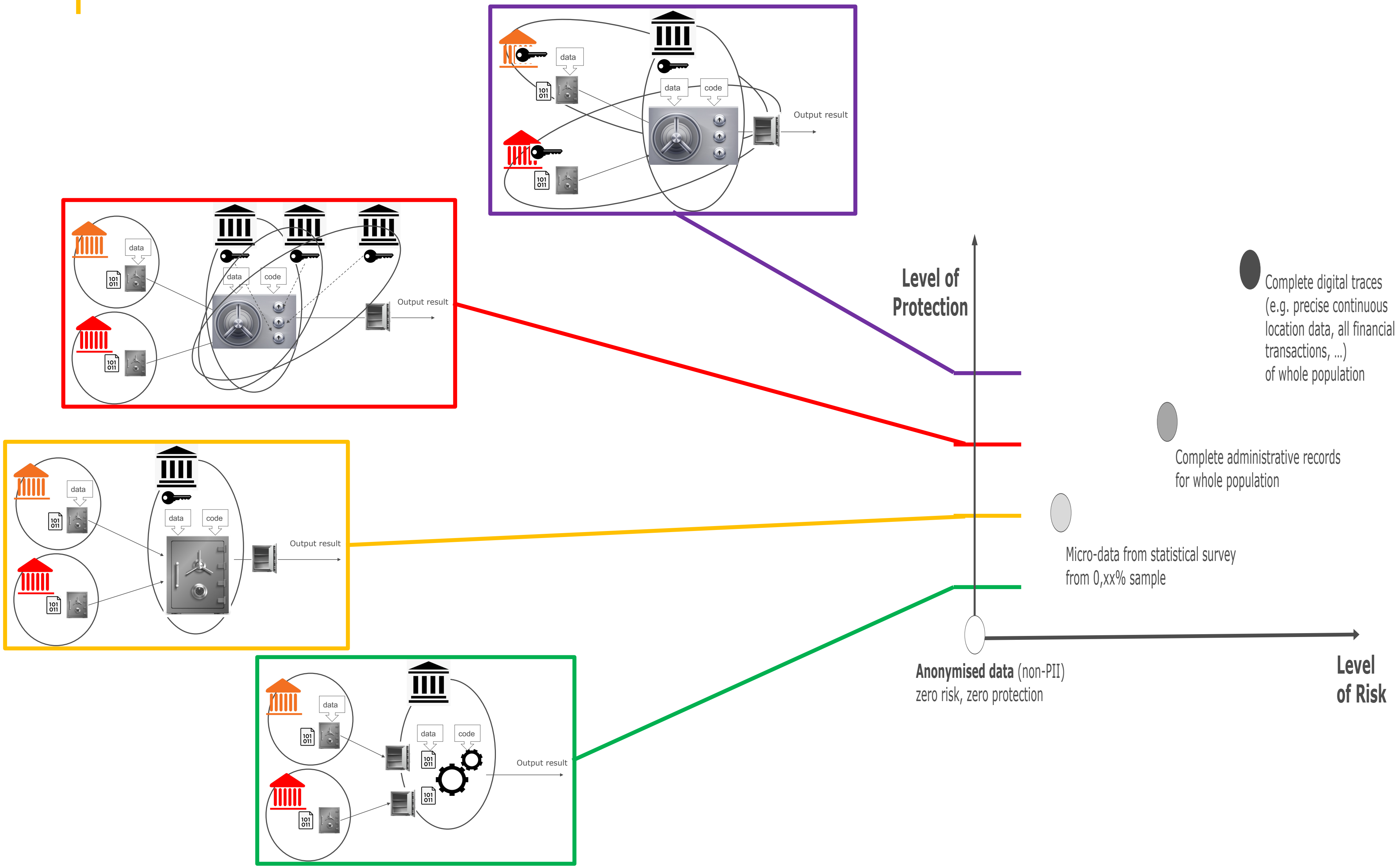
Multi-Party computation with external Processing Parties (multi-key)



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



Level of protection proportional to risk



Level of Protection

Level of Risk

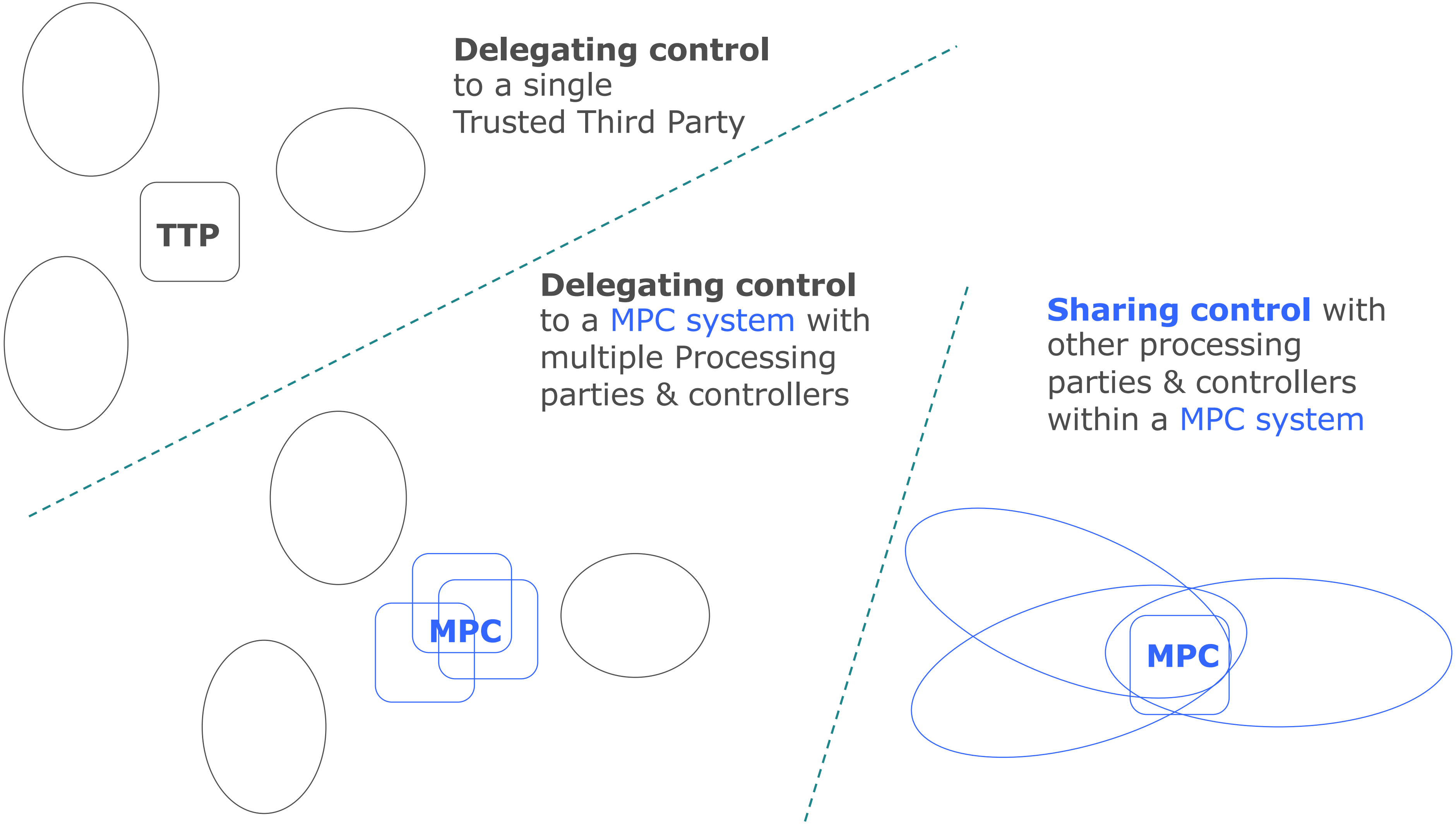
Complete digital traces
(e.g. precise continuous
location data, all financial
transactions, …)
of whole population

Complete administrative records
for whole population

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

Anonymised data (non-PII)
zero risk, zero protection

From delegation to sharing (of processing control)



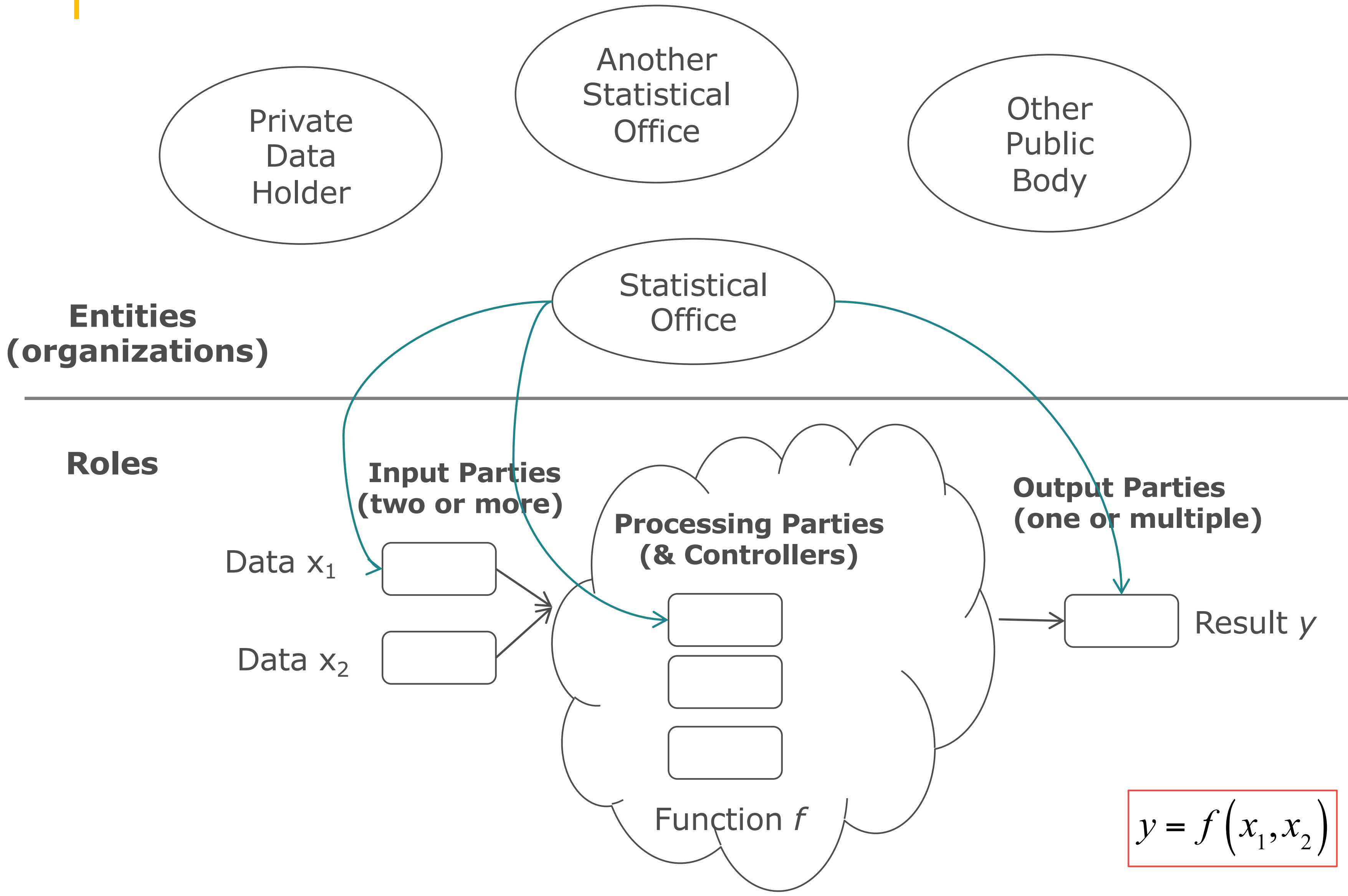
Delegating control to a single Trusted Third Party

Delegating control to a MPC system with multiple Processing parties & controllers

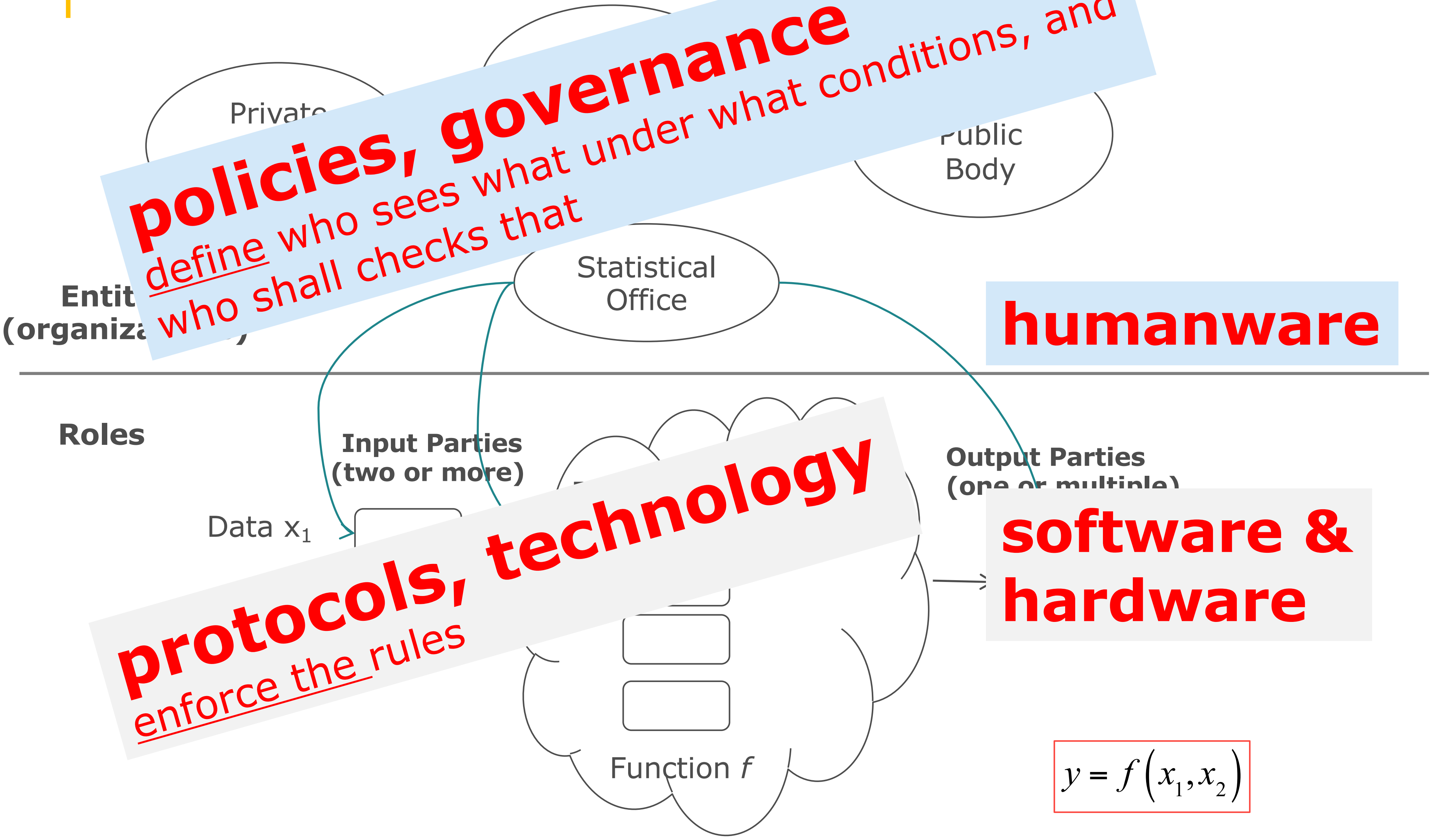
Sharing control with other processing parties & controllers within a MPC system

Explanation: ovals represent Input Parties and Output Parties.
Rectangles represent processing parties & controllers

Technical and organisational layers


Private Data Holder
Another Statistical Office
Other Public Body
Statistical Office
Entities (organizations)
Roles
Input Parties (two or more)
Data x_1
Data x_2
Processing Parties (& Controllers)
Function f
Output Parties (one or multiple)
Result y
$y = f\left(x_1, x_2\right)$

Technical and organisational layers



$y = f\left(x_1, x_2\right)$

policies, governance

define who sees what under what conditions, and who shall checks that

humanware

protocols, technology

enforce the rules

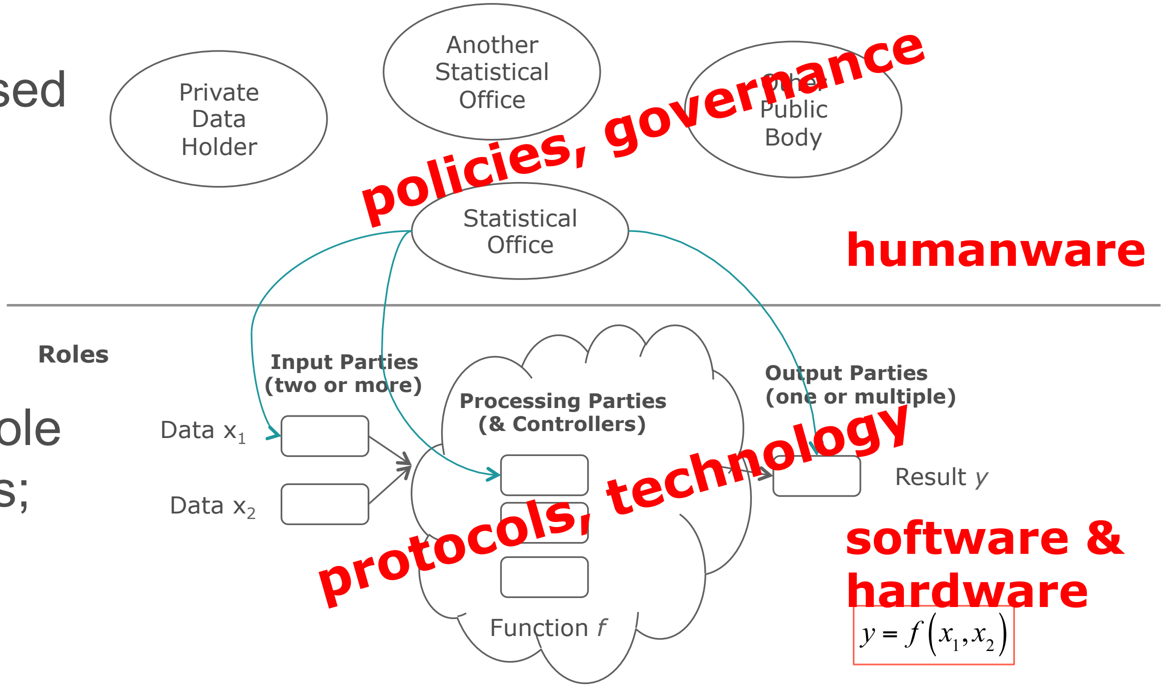
software & hardware

"…technical and organisational measures…"

Trust model

- The essential role of the is to enforce technologically the governance/policies (for data & code) defined among entities
- Truly ‘Multi-Party’ → avoid single-point-of-trust → the set of processing parties to be *trusted collectively, not individually*
- If you don’t trust the other processing parties, be a processing party yourself!

- The overall strength of MPC-based solution depends *jointly* on
- (i) robustness of policies/governance scheme;
- (ii) choice of entities taking the role of processig parties & controllers;
- (iii) strength of technology implementation



Engineering a strong SPC

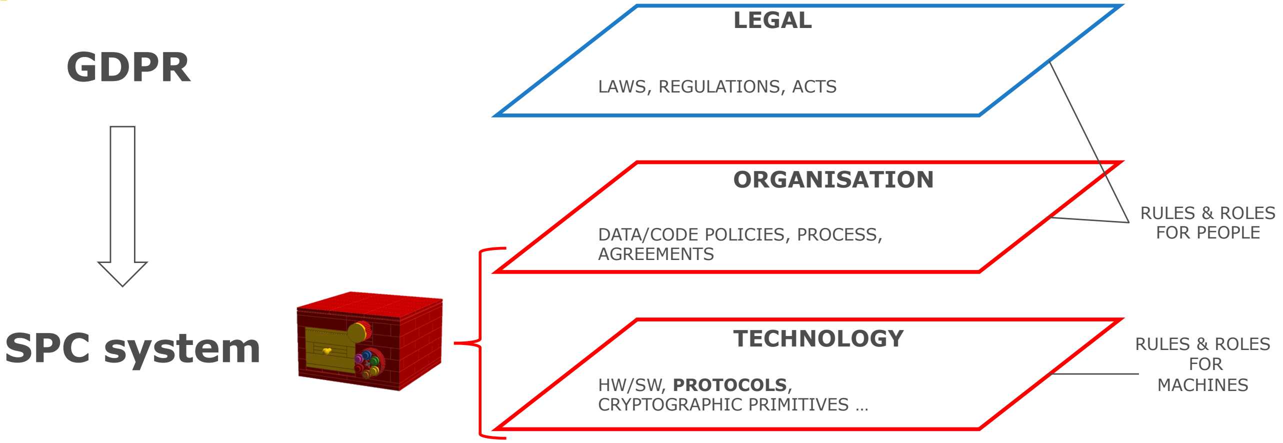
These are "*just*" system **design** aspects.
The design process starts from **requirements** …

- The overall strength of MPC-based solution depends *jointly* on
- (i) robustness of policies/governance scheme;
- (ii) choice of entities taking the role of processig parties & controllers;
- (iii) strength of technology implementation

e.g., mutual independence, (partly) antagonist goals,…

e.g., combine technologies with complementary guarantees, overlay multiple security layers

GDPR principles as design requirements (top-down approach)



GDPR

SPC system

LEGAL

LAWS, REGULATIONS, ACTS

ORGANISATION

DATA/CODE POLICIES, PROCESS, AGREEMENTS

TECHNOLOGY

HW/SW, **PROTOCOLS**, CRYPTOGRAPHIC PRIMITIVES …

RULES & ROLES FOR PEOPLE

RULES & ROLES FOR MACHINES

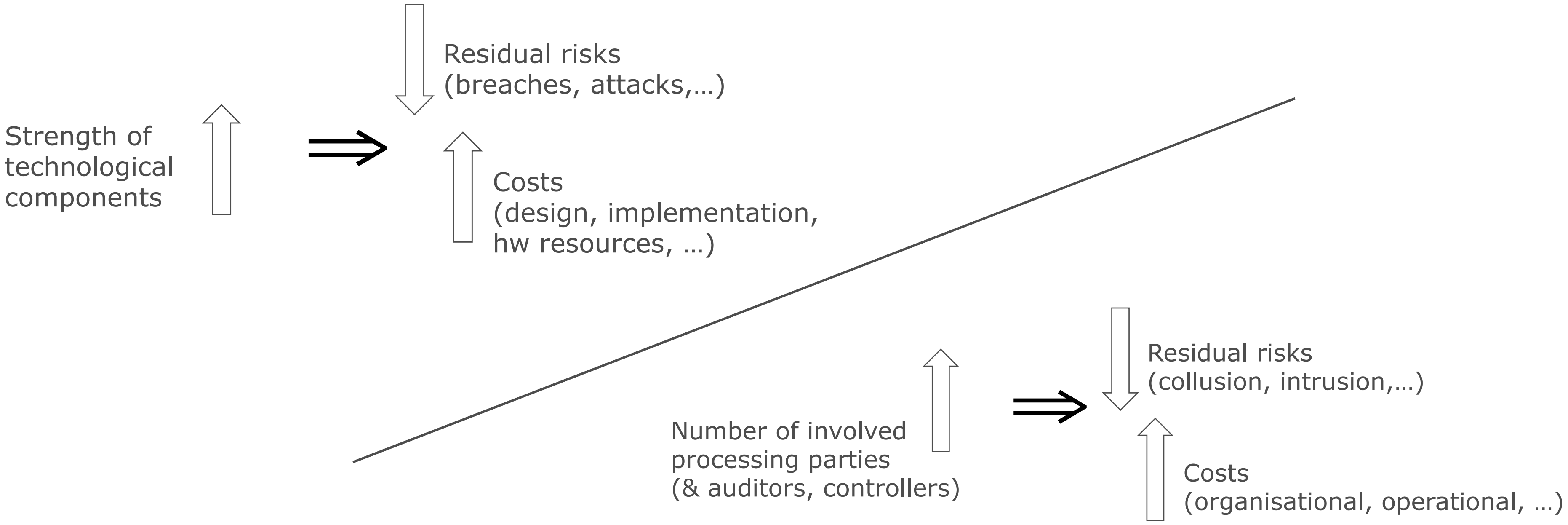
GDPR principle - requirement	System specifications at organisational level	System specifications at technical level
Lawfulness, fairness and transparency	…	…
Purpose limitation	…	…
Data minimisation	…	…
Accuracy	…	…
Storage limitation	…	…
Integrity and confidentiality	…	…
Accountability	…	…

Documenting the system → DPIA

- Good system design comes with good documentation
  - Document how requirements were addressed by functional specifications
  - Document and motivate scenario assumptions, including attack model, risks considered and countermeasures
  - …
- These are all elements of the Data Protection Impact Assessment (DPIA) in GDPR

What about the costs?

- Designing and building a **robust SPC system** is *costly*
  - Highly specialised skills: cryptography, HW/SW security, …
  - €€€ for HW/SW infrastructure building, deploying, maintenance
  - Several **cost-vs.risk trade-offs**



Strength of technological components ↑ ⇒

↓ Residual risks (breaches, attacks,…)

↑ Costs (design, implementation, hw resources, …)

Number of involved processing parties (& auditors, controllers) ↑ ⇒

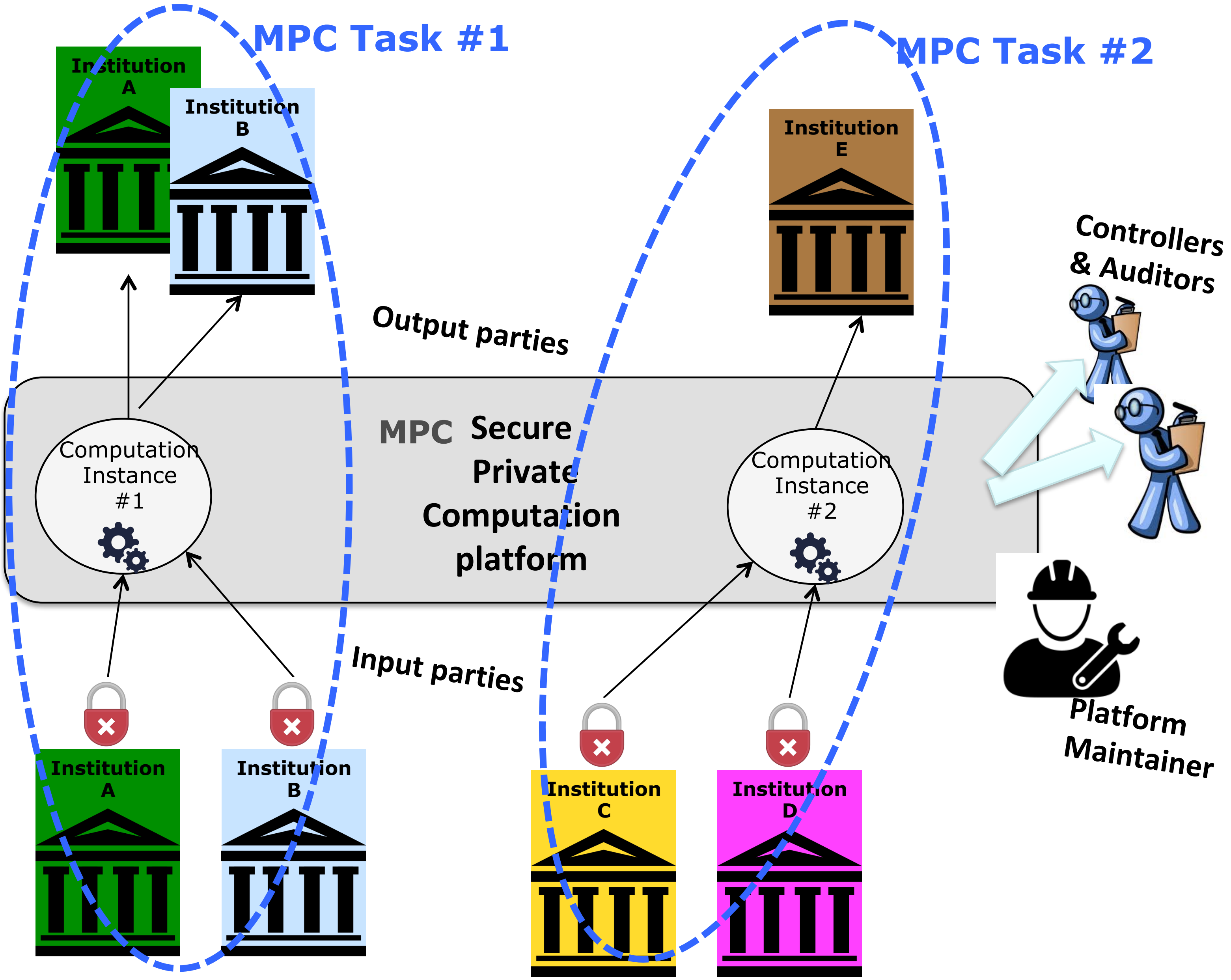
↓ Residual risks (collusion, intrusion,…)

↑ Costs (organisational, operational, …)

Lowering the risks *and* the costs

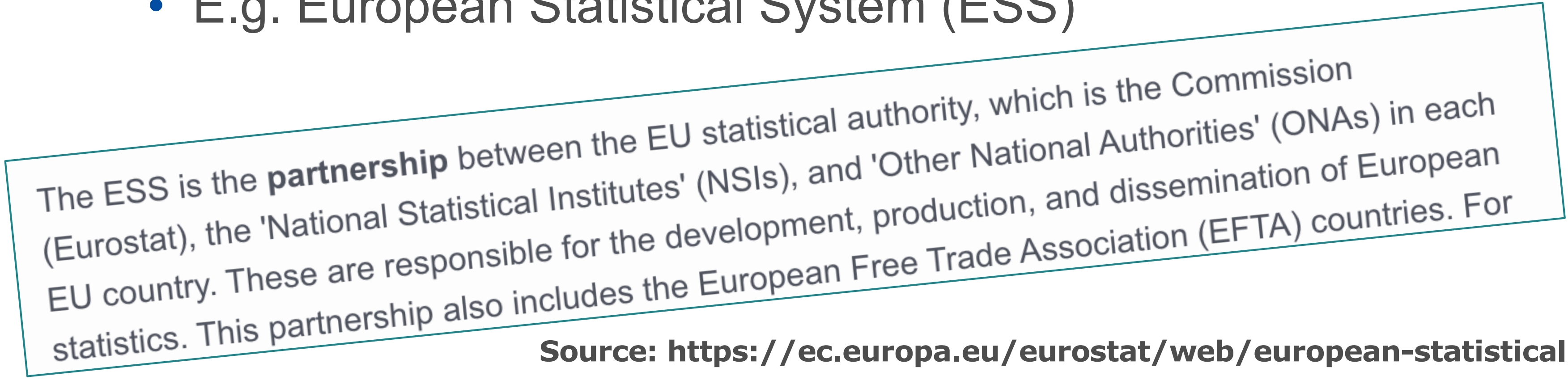
- Q. How to make the strongest possible Multi-Party Secure Private Computing (MPSPC) solution affordable for adopters?
  - Lowest risk at low cost
- Saving on costs → lower robustness → increase the risk
  - This contradicts the primary motivation for SPC in the first place, i.e., “lowering the risk” 🚫
- Alternative: **shared solution**
  - Joining forces, pooling resources, building once, use many times – by multiple organisations, for multiple use-cases
  - MPSPC-as-a-service (**MPSPCaaS**)

Multi-Party SPC-as-a-service (MPSPCaaS)



Multi-Party SPC-as-a-service (MPSPCaaS)

- Built and operated by a consortium/network/**partnership** *of* public institutions *for* public institutions and their private partners
  - E.g. European Statistical System (ESS)



The ESS is the **partnership** between the EU statistical authority, which is the Commission (Eurostat), the 'National Statistical Institutes' (NSIs), and 'Other National Authorities' (ONAs) in each EU country. These are responsible for the development, production, and dissemination of European statistics. This partnership also includes the European Free Trade Association (EFTA) countries. For

Source: https://ec.europa.eu/eurostat/web/european-statistical-system

  - PET as *Partnership* Enhancing Technology *(*)*

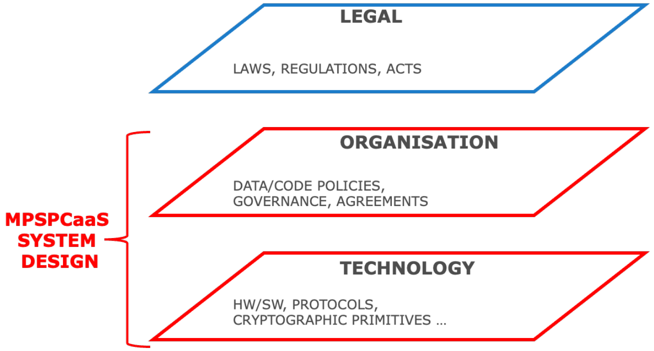
(*) Credit to Andrew Trask for inventing the term. Source: The Coming Age of Collaborative Computing https://medium.com/lunar-ventures/the-age-of-collaborative-computing-e73374b7aedc

MPSPCaaS concept

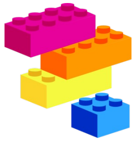
- First proposed by Eurostat in the context of UNECE HLG-MOS project on Input Privacy Preservation (IPP, 2021-2022)
  - (2021) Discussed internally to IPP project team
  - (2022) Open Technical Consultation organised within the IPP project
  - Presentations and exhange of views with data protection and privacy experts (ENISA workshop, MPC alliance, …)

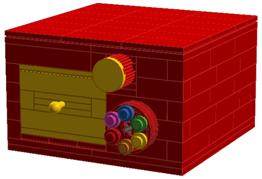
- 2023 Eurostat Call for Tender
  - *Specification, feasibility analysis and prototype demonstration of a multi-party secure private computing system for processing confidential sets of micro-data across organisations in support of statistical innovation (TSS-PET)*
  - Published on 7/4/2023 with submission deadline 31/7/2023 (now closed)
    - https://etendering.ted.europa.eu/cft/cft-display.html?cftId=12503
  - Currently in evaluation phase.
  - Planned duration 2 years

MPSPCaaS project

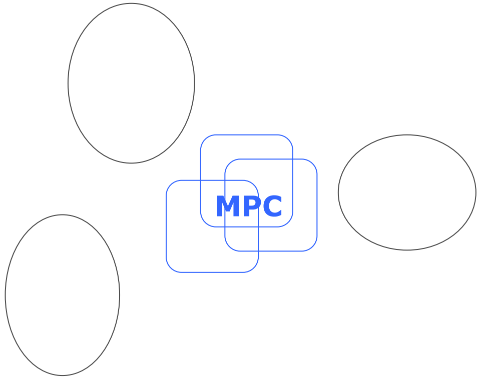

LEGAL
LAWS, REGULATIONS, ACTS
ORGANISATION
DATA/CODE POLICIES, GOVERNANCE, AGREEMENTS
TECHNOLOGY
HW/SW, PROTOCOLS, CRYPTOGRAPHIC PRIMITIVES …
MPSPCaaS SYSTEM DESIGN

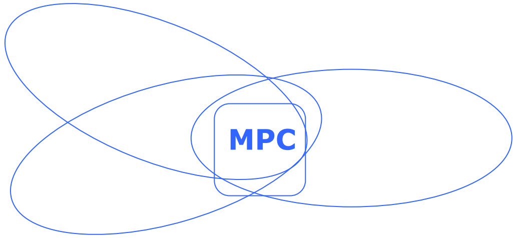
- Tasks to be performed by the contractor *in close consultation with Eurostat*
    - Task 1 – Usage scenarios and system requirements
    - Task 2 – Technology analysis
    - Task 3 – Legal aspects
    - Task 4 – System specifications and architecture
    - Task 5 – Demonstrator prototype and functional testing
    - Task 6 – Trust building plan (public acceptance)



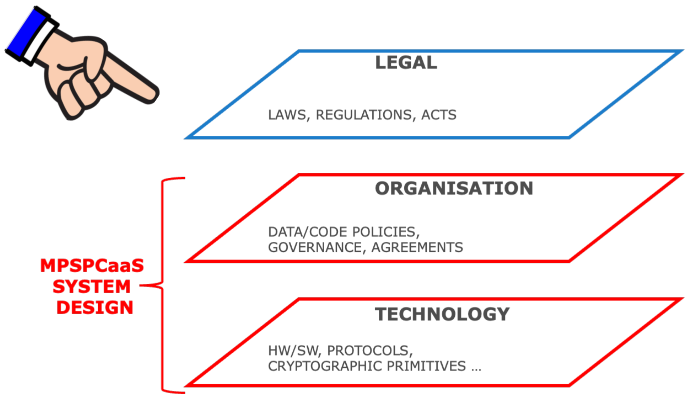


- Key requirements:
    - Multi-Party with at least 3 processing parties (fixed or custom)
    - Robust to collusion of / intrusion at 2 out of 3 processing parties
    - Embedding security measures at SW and HW level
    - Rich logging and auditing features for ex-post controls
    - Provable deletion of intermediate data (storage limitation)
    - No single point of trust
    - Scalability (for a relatively simple set of core operations)





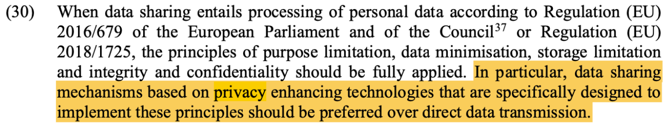
PET and European legislation – not just GDPR



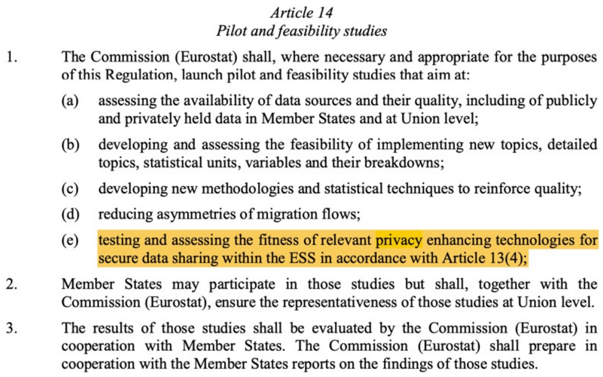
- Data Governance Act mentions ‘secure processing environments’

- Jan’23 adoption by European Commission (EC) of proposal for 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)

- July’23 adoption by EC of proposal for revising Regulation 223/2009 on European statistics (link).
- EDPS opinion published in Sept 2023 (link)

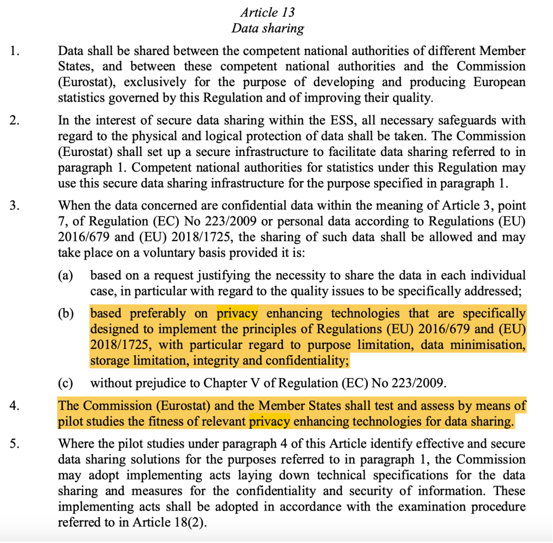


(30) When data sharing entails processing of personal data according to Regulation (EU) 2016/679 of the European Parliament and of the Council[37] or Regulation (EU) 2018/1725, the principles of purpose limitation, data minimisation, storage limitation and integrity and confidentiality should be fully 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 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.



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).

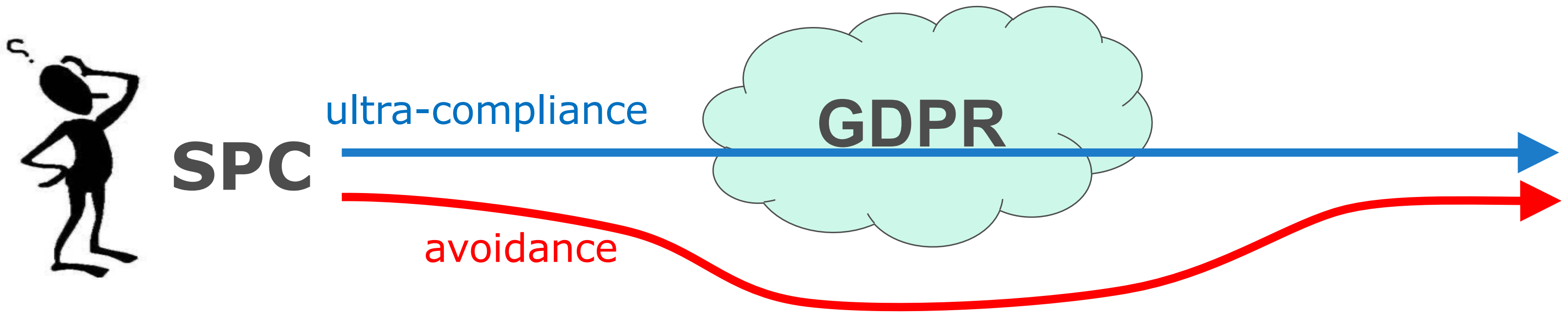
Official Statistics as a favourable incubator for MPSPCaaS

- A “**partnership**” of multiple organisations with common mandates and a culture of coordination, cooperation and sharing is already in place – it’s the ESS (!)
- **Legal enablers** for Official Statistics enshrined in GDPR - Art 89(1) statistics purposes non-incompatible with primary purpose
- **Methodological transparency**: methods are not secret! Methods are (should be) publicly available, anyway not subject to IPR
- For many use-cases, relatively **simple statistical methods** suffice (e.g., set intersection, low-dimensional regression) which helps scalability
- …

IPR : Intellectual Property Rights

SPC and GDPR: legal orientations

- SPC as very advanced (state-of-the-art) form of **pseudonymisation:** encrypted data are PII, hence in scope of GDP
- → coherent with ”absolute” interpretation of “anonymisation”
- SPC as a means to embody **data minimisation**, **purpose limitation** and other GDPR principles (ultra-compliance)



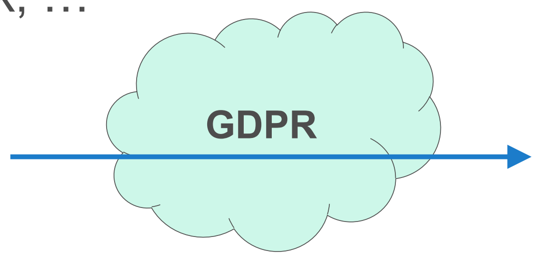
- SPC as “anonymisation” :
  encrypted data are non-PII, hence out of GDPR scope
- → based on the “relative” definition of “anonymisation”

PII : Personal Identifiable Information

SPC and GDPR: legal interpretations

- The final word is up to the competent authorities …
- In our current understanding, SPC-based solutions qualify as *processing of personal data,* remain subject to GDPR
  - SPC solutions as *supplementary "technical and organisational measures"* in the sense of GDPR Art. 89 (*,**)
  - Implications: need for legal basis, DPIA, assessing strength of SPC solutions vs level of risk, …

A NEW COMEDY BY ROB REINER
BILLY CRYSTAL MEG RYAN
Can two friends sleep together and still love each other in the morning?
When Harry Met Sally...



(*) In line with EDPB Recommendations 01/2020 on measures that supplement transfer tools to ensure compliance with the EU level of protection of personal data (Use Case 5: Split or multi-party processing)

(**) In line with ENISA view, see report on "Data Pseudonymisation: Advanced Techniques and Use Cases", January 2021

Take home messages

- In our view, well-designed SPC solutions represent today the strongest possible way to embody the GDPR principles (data minimisation, purpose limitation, storage limitation, integrity and confidentiality, etc.)
  - Embracing GDPR principles as design requirements
- Continuous dialogue (co-design) with **technology specialists** and **data protection legal experts** is needed to design robust (technically and legaly)and usable solutions
  - Consultation with Data Protection Authorities

- Work is in progress: Eurostat and the ESS advancing step-by-step, from initial **concept** through **specification** towards future **deployment** of shared PET infrastructure for the ESS, based on the MPSPCaaS concept
- This work by Eurostat in the ESS may serve as a lighthouse and inspiration for other public sectors (and maybe also private sector?) as to how data protection and data usage can be **reconciled**, rather than confronted or compromised.

Thank you for your attention

More about the work done at Eurostat on Privacy Enhancing Technologies for Official Statistics (PET4OS):

https://ec.europa.eu/eurostat/cros/content/privacy-enhancing-technologies-official-statistics-pet4os_en

(with links to all references in the presentation)