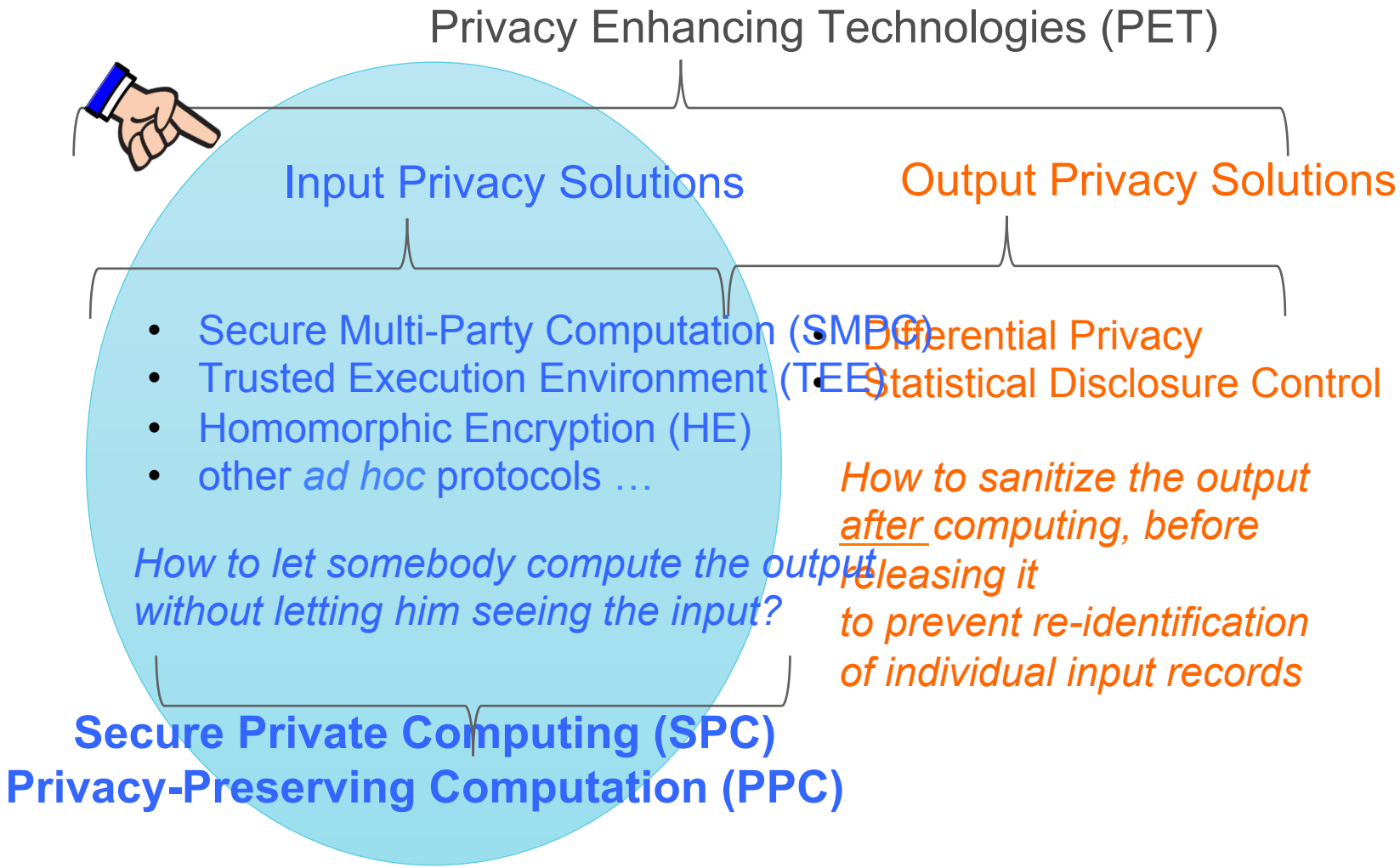
Secure Private Computing-as-a-service

Proposal for a technical public consultation by the UNECE HLG-MOS Project on Input Privacy Preservation (IPP)

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

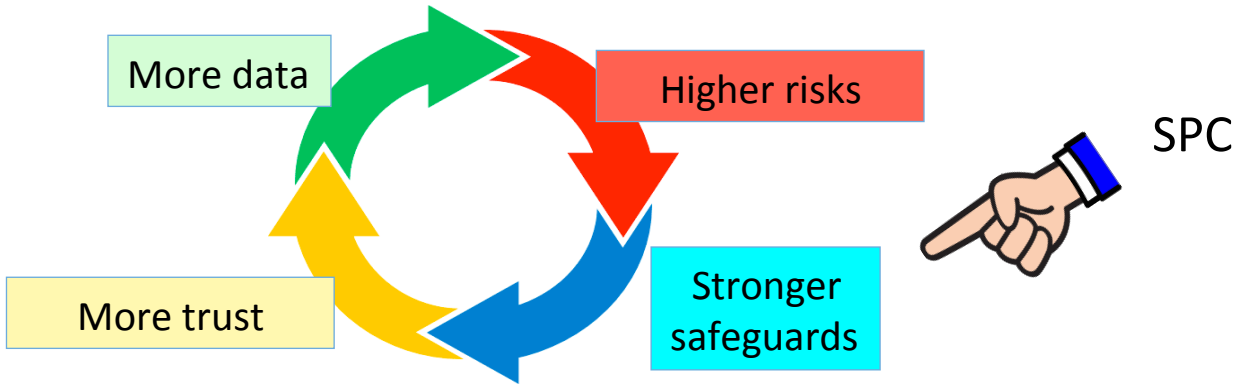
2021 Workshop on the Modernisation of Official Statistics - IPP Project Webminar
16 November 2021

Secure Private Computing (**SPC**)



Context: inter-organization data processing

- Increasing **appetite** for producing information (e.g., statistics, analyses) from the **combination of data held by different organizations** (private companies, public institutions)
  - Statistical authority/ies acting as output party, input party or both
- Increasing **pressure** to strengthen safeguards, "*technical and organisational measures*" for protecting the data
  - legal requirements by Data Protection Authorities
  - necessary condition to build public trust and public acceptance



Options

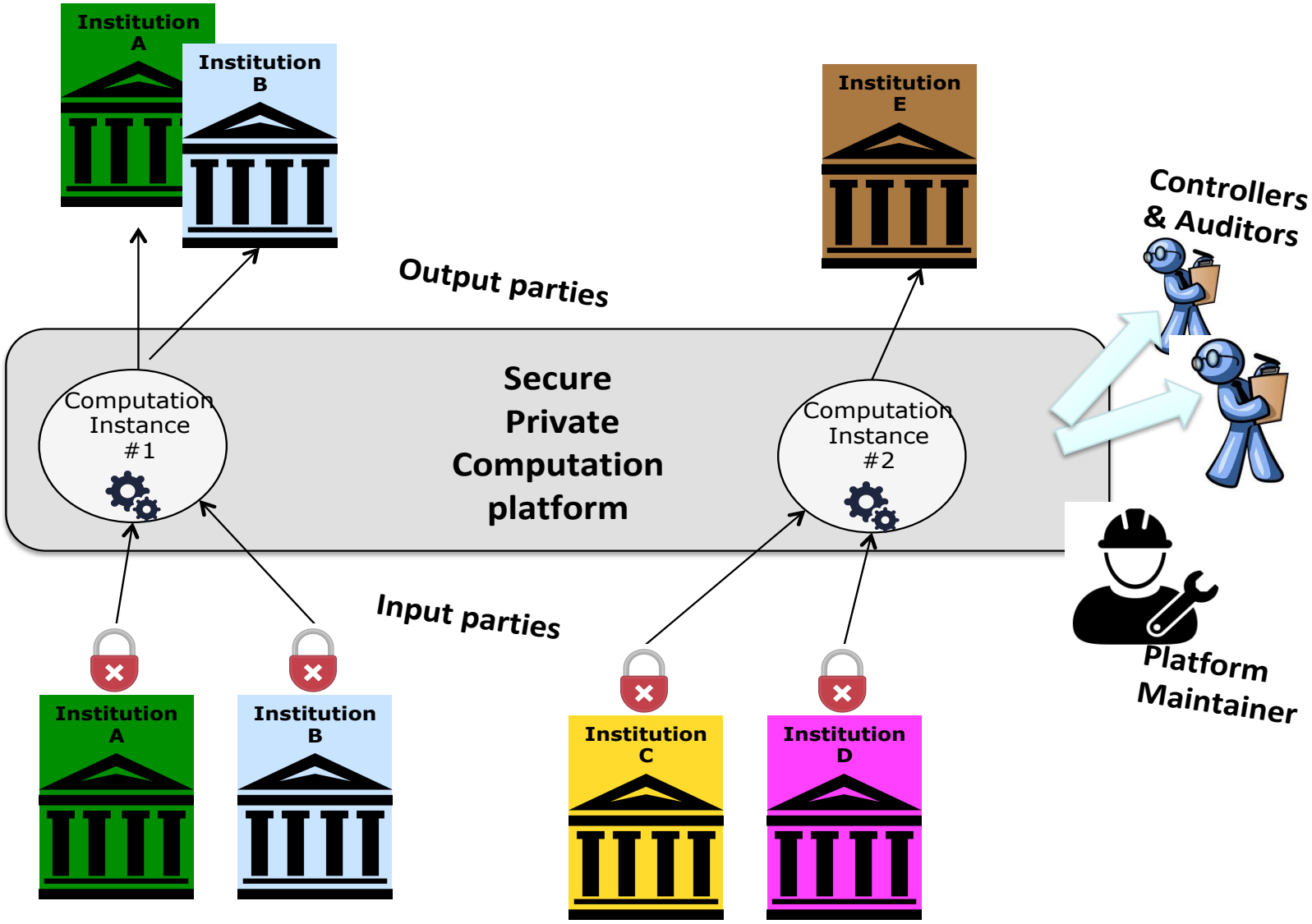
Consider two or more public organizations that have a need (or at least an interest) to produce information from the **joint processing** on their confidential data sets.

What options do they have?

1. Abstain from the project → Loss of public benefit
2. Execute the project via traditional data sharing, i.e. *move the data*
   → Increase of risks (for data mis-use, but also reputational)
3. Build an ad-hoc SPC infrastructure dedicated to the project
   → Often impractical due to **high costs** (time, staff resources, budget) and **lack of appropriate skills**
4. **SPC-as-a-service**: execute the project by using the SPC services made available *on demand* by a trusted SPC infrastructure that is …
   - designed/specified/procured/deployed/certified/etc. by a public institution (or consortium thereof) acting as **SPC provider** and made available *on demand* to SPC users

Note: the marginal costs (not priced) for SPC access is not zero, but should be anyway considerably lower than setting up (by internal development or procurement) ad-hoc infrastructure dedicated to a single use-case

Secure Private Computing-as-a-service



Key aspects

- What is a SPC infrastructure ?
  - The term *infrastructure* is meant here to refer to a combination of **technological and non-technological components**, including e.g. organisational measures, business processes, legal and contractual aspects, liabilities, etc.
  - **hardware + software + . . . *humanware***
- What is the role of SPC-as-a-service provider ?
  - Build the infrastructure & build **trust in** the infrastructure
  - NB: I'm not sure these two items are really distinct from each other, but it's anyway useful to spell them distinctively

Key design features of SPC-as-a-service model

- **No *single point of trust***
  - no single party should hold full control over the process and/or access to the data (not the SPC provider, not the technology provider, …)
  - distribute control across **multiple selected actors**
    - select actors that are **semi-trusted individually,** and can be **trusted collectively** → part of **trust engineering**
    - balance complexity (not too many) vs trustworthiness (not too few)
    - ensure **credibility** and **mutual independence** among selected actors

- **Close the data, open everything else**
  - For each computation instance, ensure **full transparency** as to (i) purpose of the processing; (ii) participating organisations; (ii) what input data are (re)used; (iii) detailed description of methods and desired output, including the kind of **output privacy** protections (if applicable)

Ok, let's build it … but first let's specify it!

Before building a (first version of) an SPC infrastructure, we need to formulate a list of specifications, i.e. answer questions like …

- What SPC services to offer ?
  - Initial focus on a **Private Set Operations with analytics.** Scenario: two or more input parties have lists of structured records (micro-data) and need to execute some simple analytic primitive (e.g. counting) on the *intersection or union* of their sets
- To which users?
  - Any combination of public/private organizations with the constraint that at least one input or output party is a statistical authority.
- …

How to build *trust into* the infrastructure?

- How to build *trust into* the infrastructure?
  - This is the main challenge - Trust Engineering https://doi.org/10.1017/dap.2020.7
  - Complex answer, as it intermingles technological and non-technological aspects.

- Idea: ask the question to those that will be eventually concerned via a public consultation (informal, technical):

- Public consultation as a way to pull expert knowledge
  - to identify possible solutions to known challenges but also to identify additional challenges and critical points
  - Side benefit: probe general interest for SPC-as-a-service model

Scope and targets of public consultation

- Which “experts” to address?
  Wide and diverse range of expertise are relevant, including
  - Technology experts (computer science, cryptography, IT security…) and legal experts
  - Privacy advocates, civil right activists
  - Researchers and scholars in relevant disciplines, e.g. Critical Data Studies, politics, e-government …
  - Potential SPC *users* : statistical authorities, other public bodies, private data holders

- Mind that **technology is a means, not the goal!**
  - Ask primarily “**what should be achieved [by the technology]”**
  - The question **“what technology can [help to] achieve that”** comes later

Examples of (initial) questions 1/2

1. **Certifications and technical standards**: what kinds of certifications and by which certification bodies do your think should be required? Which technical standards should the envisioned SPC infrastructure comply with?
2. **Open-source**: considering the current stage of technological maturity for SPC technologies, do you think there should be an explicit requirement that the SPC infrastructure to be based purely on open-source software and hardware components? What could be the benefits and the potential risks of imposing a stringent requirement in this sense?
3. **Independent audits, penetration tests**: should independent audits, penetration tests or other similar actions be required? If so, how should they be organised and by whom?
4. **Inter-operability**: in which ways the SPC provider may ensure interoperability of the SPC infrastructure and prevent vendor locked-in effects, considering that the most mature SPC solutions tend to be proprietary nowadays?

Examples of (initial) questions 2/2

6. **Distributed control**: How important (or not important) is to ensure that control over the computation process is shared among multiple actors, so as to avoid any single point of trust? If this requirement is important, how should these actors be selected (e.g. based on what criteria, whether government or non-governmental organizations, etc.)? And what would be the role, duty and commitment of the selected actors?
7. **Infrastructure governance model**: what are the key elements that a governance model for the SPC infrastructure should incorporate in order to strengthen public trustworthiness in the infrastructure? What entities should be called to "share control" ?
8. **Procedures**. What are the key ingredient of the procedure that should be put in place in order to ensure trustworthiness of each individual SPC transaction and of the SPC infrastructure as a whole? (e.g. preventive authorisation, ex-post controls, regular audits,…)
9. …

Proposed roadmap (tentative)

- Finalize formulation of questions and launch of consultation via EUsurvey - December'21
- Closing date for replies – end of March'22
- Analyse response and draft a summary report – end April'22

Thanks for your attention

Fabio.Ricciato@ec.europa.eu