Update on the HLG MOS project on Input Privacy Preservation

Presented to Work Session on Statistical Data Confidentiality 2021

Dennis Ramondt - UNECE Project Manager

High Level Group for the Modernisation of Official Statistics

- Committed Chief Statisticians actively steering the modernization of statistical organisations
- Work collaboratively to identify trends, threats, and opportunities in modernising statistical organisations
- Executive Board is responsible for the strategic management of on going activities
- Working groups (ongoing) and projects (time-limited)

Introduction to the project

- Initiative ISTAT and Statistics Netherlands
- Started in summer 2020
- On-line project

Participants:

- UNECE
- ISTAT
- Eurostat
- Statistics Netherlands
- StaCan
- ONS
- GSO
- INEGI

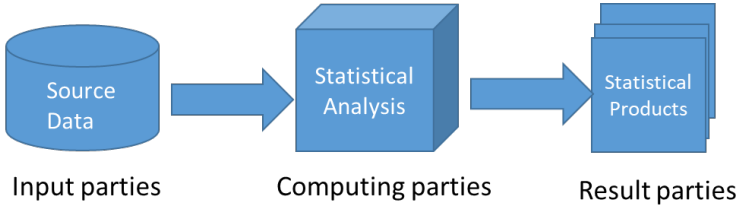
Why Privacy preserving techniques?

Modern statistical organizations:

- needs to become part of a data ecosystem
- acquire and integrate data from multiple sources
- provide richer statistical products

Risk for disclosing information violating individual privacy rights

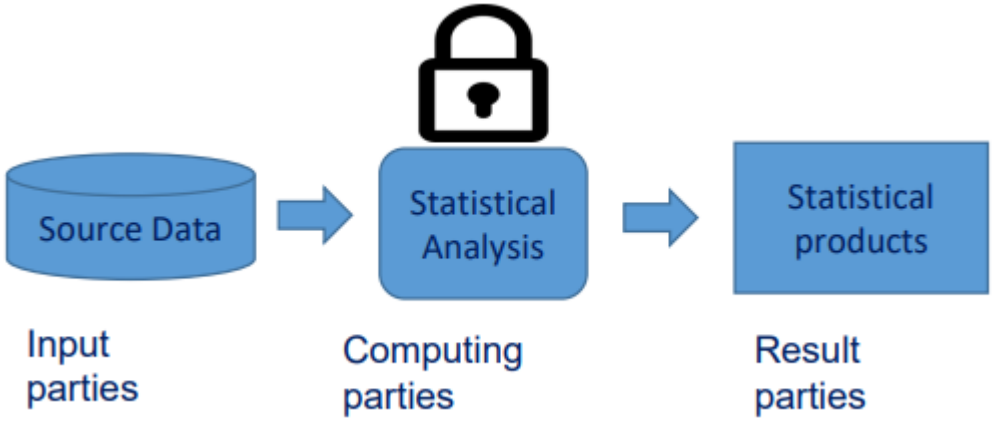
Overview Privacy protection



Output side: By design applied by NSO's

Input privacy *means that the Computing Party cannot access or derive any input value provided by Input Parties, nor access intermediate values or statistical results during processing of the data (unless the value has been specifically selected for disclosure).* [UN Handbook on Privacy-Preserving Computation Techniques, 2019]

Input Privacy Vs Output Privacy



Input privacy:

- Input privacy techniques are based on data «transformations» that preserve source data privacy
- Examples of input privacy techniques: Secure multi-party computation (SMC), homomorphic encryption, trusted execution environment, etc.

Output privacy:

- Output privacy aims at reducing the risk of privacy breaches in the phases of disseminating or exchanging statistical products
- Examples of output privacy techniques or Statistical Disclosure Control techniques: perturbation methods (e.g. differential privacy), non pertubation methods (e.g. local suppression)

WP1: Documenting Use cases

- A template how to document IPP use cases
- Documentation of IPP use cases
- Generalization for wider usability and use within project

WP2A: Private set intersection

- Investigate four different scenario’s
- Mini pilot
- Lessons learned

WP2B: Private machine learning

Pilot goal:

- Build a simulated environment to validate the concept of multi party privacy preserving Machine Learning (PPML) for both, training and inference

Scope:

- Investigate best practises and open source tools for disturbed and collaborative ML training among multiple organisations in a low trust environment whilst mutually benefitting from the outcomes (the final model) or allowing safe 3rd party access

Environment :

- Simulated multi organisational set-up with several NSO's gathering data from individuals (sensors) to predict their activities (time use and well-being surveys).

WP2B: Private machine learning, next steps

- Extend the scope to more complex models and other distributed data related to members of HLG-MOS
- Incorporate Secure Multi-party Computation for secure aggregation of weights during training, as well as inference
- Integrate Differential Privacy as part of the protocol to protect output privacy
- Collaborate with the Openmined community to use their software stack
- On Board the project to the UN PET-Lab infrastructure

WP2C: Organize public consultation

Context:

- Increasing **appetite** for producing information (e.g., statistics, analyses) from 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 requirement by Data Protection Authorities
    - Necessary condition to archive public trust, public acceptance

WP3C Organize public consultation

Goal:

- Secure Private Computing-as-a-service
  - Designed/produced/deployed/certified/advertised/etc by public institution (or consortium thereof) acting as SPC provider
  - Used on demand by SPC clients
  - NB: the marginal costs (per project) for clients is not zero, but should be anyway much lower than setting up a ad hoc infrastructure to a single use-case
- Public consultation to pull expert knowledge
  - Key challenge: how to build trust into the infrastructure?
  - Idea: ask the question to experts from various domains, via a public consultation (informal, technical)
  - Public consultation as a way to pull expert knowledge

Next steps

- Continuation of the project
- Apply collaboration more to complex problems
- More practical use cases
- Decision in January by HLG-MOS