

Input Privacy-Preservation Techniques Project

Dennis Ramondt

Statistics Netherlands and UNECE Project Manager

The project

Start July 2020 (delayed)

Participants:

- Eurostat
- Istat
- ONS
- Statistics Canada
- Statistics Netherlands
- UNECE

Objective

*“to investigate statistical use cases that require protection on the input side,
assess and determine applicability of selected classes of techniques for main scenarios,
identify opportunities for sharing across statistical community and
create community of practice across statistical organizations and external partners (academia, private sector).”*

Area and scope

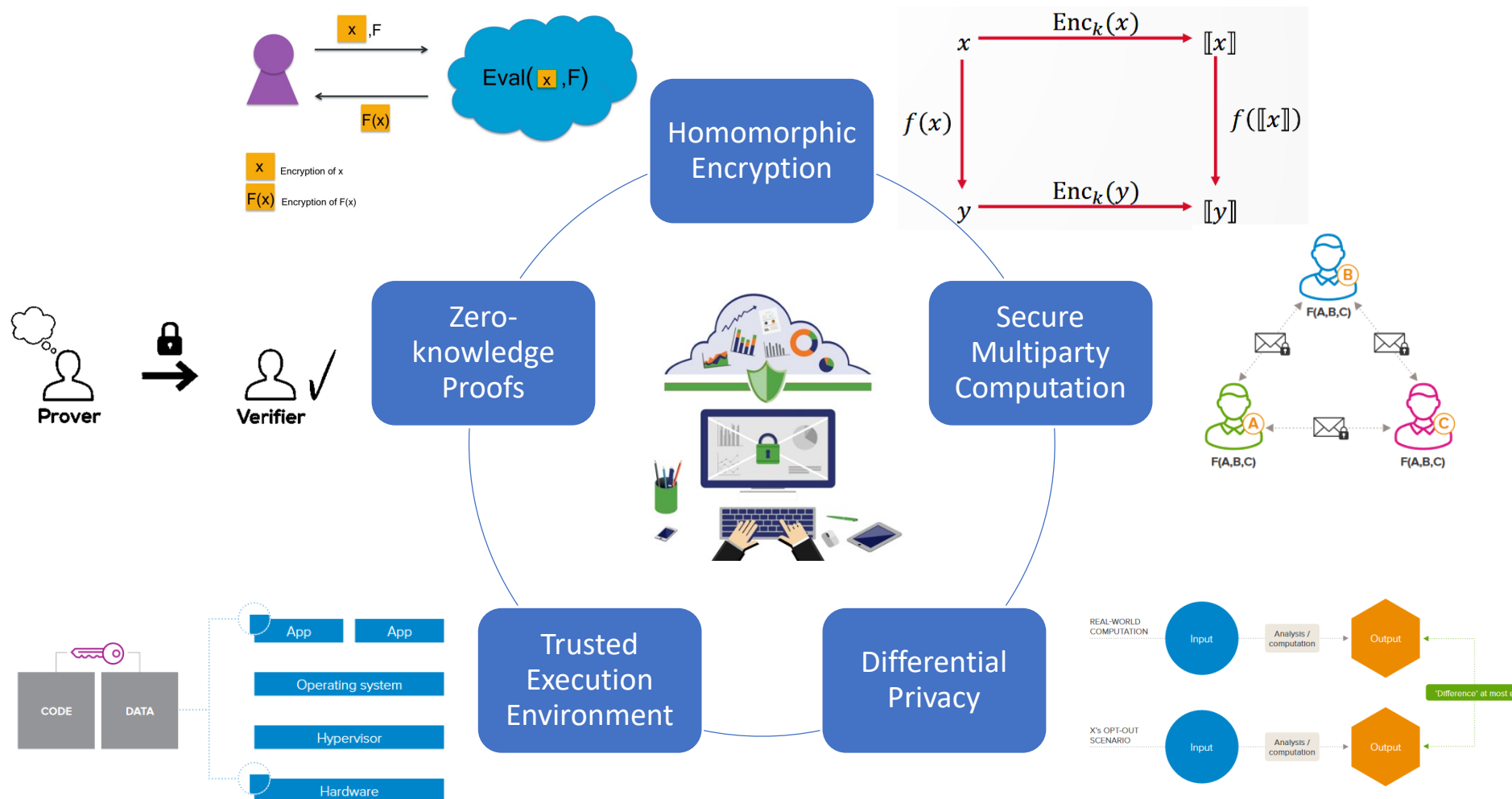
Privacy Preserving Technologies:

- **Input privacy (scope of the project)**
- output privacy

Based on:

- UN handbook on Privacy preserving techniques

Privacy-Preserving Technologies

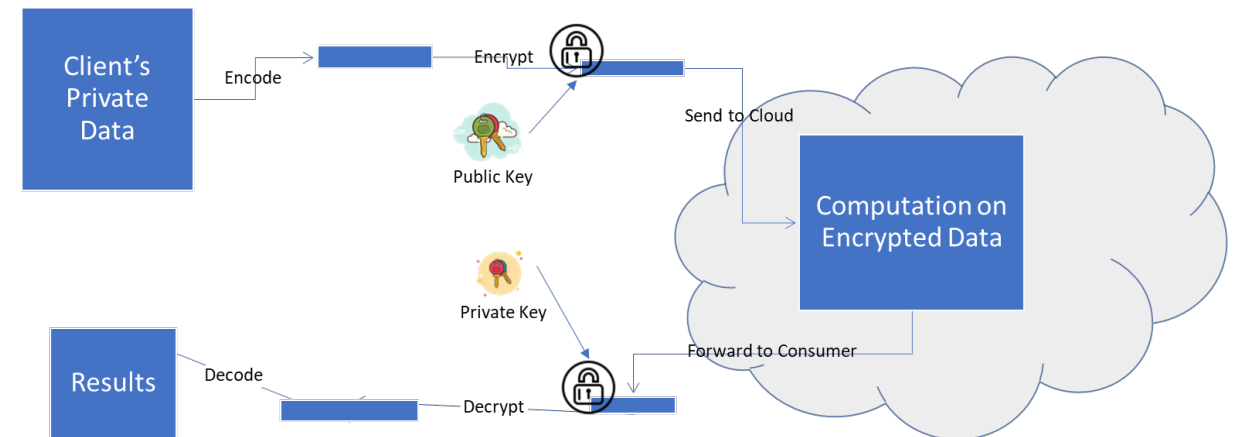


Use cases (first inventory)

Topic	Context
POC's : <ul style="list-style-type: none"> Machine Learning (ML) text classification algorithm. Post-ML aggregation and statistical analysis. 	Proof of concept with retailer scanner data through the use of homomorphic encryption in the Microsoft Azure Cloud infrastructure
Private Set Intersection with Analytics Case Study: Istat and Bank of Italy	The two parties do not exchange data directly, except those necessary to calculate the intersection on the $A \cap B$ keys, but use a neutral third party; Information enrichment takes place only in terms of aggregated data (counts); The third party can carry out checks on the counts returned and ensure that the result cannot be traced back to the individual elements of the population. Privacy preservation is not guaranteed in the event that one of the two parties agrees dishonestly with the third party
Use of PPT in statistics involving healthcare	Through the use of MPC and Block chain technology retrieve and combine datasets to gain insight into the impact of innovative healthcare treatment on patient recovery time and costs
Use of PPT in statistics involving healthcare	Use of PPT in statistics concerning the prevention and containment of disease by researching the correlation between socio-economic and environmental factors and health and biological factors.
COVID-19 containment	Gaining insight in the spread and development of COVID-19 by using aggregated and anonymized signaling data and reference data from telecom-provider antennas. The data concerns the number of people present and the number of visitors in a "gemeente" (county), these people being the subscribers and ad hoc users of telecom-provider. The entire process to gain insight should happen in near real-time but take no longer than one hour.
Use of PPT by 12 EU member states and EUROSTAT for the realization of ESS Smart data collection platform	Through the use of smartphone data collection combine smart sensors data with additional data sources to realize statistical purposes (European Social Surveys such as Household Budget Survey and Time Use Survey)

Homomorphic Encryption at Statistics Canada

- Completed two proof-of-concepts (experiments) on homomorphic encryption.
- **Scenario:** Retailer Scanner Data
- **Scope:**
 1. Machine Learning (ML) text classification algorithm.
 2. Post-ML aggregation and statistical analysis.
- **Data:** Synthetic product descriptions and price values.
- **Infrastructure:** Microsoft Azure cloud infrastructure.
- **Tools:** Used open source libraries, e.g. Microsoft SEAL.
- **Results:** HE-based solutions are *secure* and *practical* with the technology that is available today.



First results

WP0. Refine scope

- ❖ Project flyer was made

WP1. Documenting statistical use-cases relevant for application of privacy-preserving techniques

- ❖ Classifications how to document the use cases

Planning 2021

WP1. Finish the work (Documenting statistical use-cases relevant for application of privacy-preserving techniques and plan WP2 and WP3)

WP2. Secure Multiparty Computation (SMC) methods

WP3. Homomorphic Encryption (HE) methods

WP4. Identify opportunities for operationalization of methods and sharing of solutions

Project outcomes

- Set of best practices for the chosen techniques
- Insight into the applicability of the different Privacy-preserving techniques in the input domain of NSIs
- Community of statisticians in the area of PPT

Q & A