Input Privacy-Preservation Techniques Project

Presented at 2022 Workshop on the Modernisation of Official Statistics

Dennis Ramondt - UNECE Project Manager

The work of the following people:

Fabio Ricciato	Eurostat	Ralph Schreijen	Statistics Netherlands
Konstantinos Giannakouris	Eurostat	Eric Deeben	ONS, UK
Luis Clemente	INEGI, Mexico	Dang Trung Nam	GSO, Vietnam
Fabrizio De Fausti	Istat, Italy	Benjamin Santos	Statistics Canada
Massimo De Cubellis	Istat, Italy	Taeke Gjaltema	UNECE
Mauro Bruno	Istat, Italy	Wai Kit	UNECE
Monica Scannapieco	Istat, Italy	Priscila Marentes	INEGI, Mexico
Abel Dasylva	Statistics Canada	Daniel Owen	ONS, UK
Robert McLellan	Statistics Canada	Mat Weldon	ONS, UK
Philippe Gagné	Statistics Canada	Julian Templeton	Statistics Canada
Saeid Molladavoudi	Statistics Canada	Dennis Ramondt	UNECE
Matjaz Jug	Statistics Netherlands		

in collaboration with members of the UN-petlab

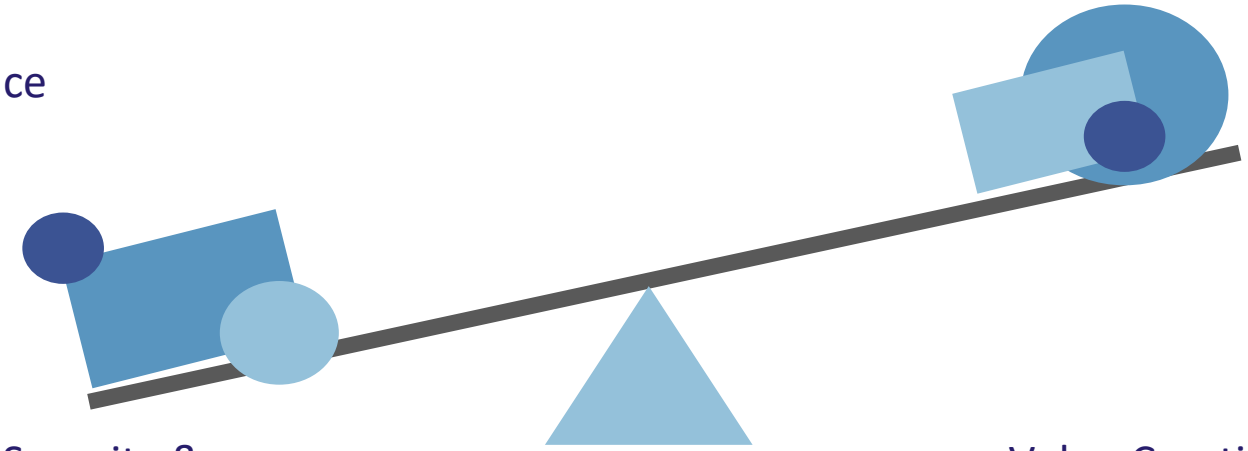
Outline

Input privacy
Recap results 2021
Track 1, Private set intersection
Track 2, Private machine learning
Track 3, organize public consultation
Closing words

The Balancing Act

We continually play the game of balancing data usability and security:

PPTs *do not solve* the balancing act of security, privacy and data use in and of themselves, but offer risk mitigation that may be the difference between a project no-go and go.



Security & Privacy

Value Creation from Statistical Inference

Input & Output Privacy

A function takes in some **input** (from one or more parties) and produces some **outputs** (given to one or more parties).

Input Privacy Approaches

Input privacy focuses on how to **ensure privacy of inputs of one or more parties entering a joint function**.

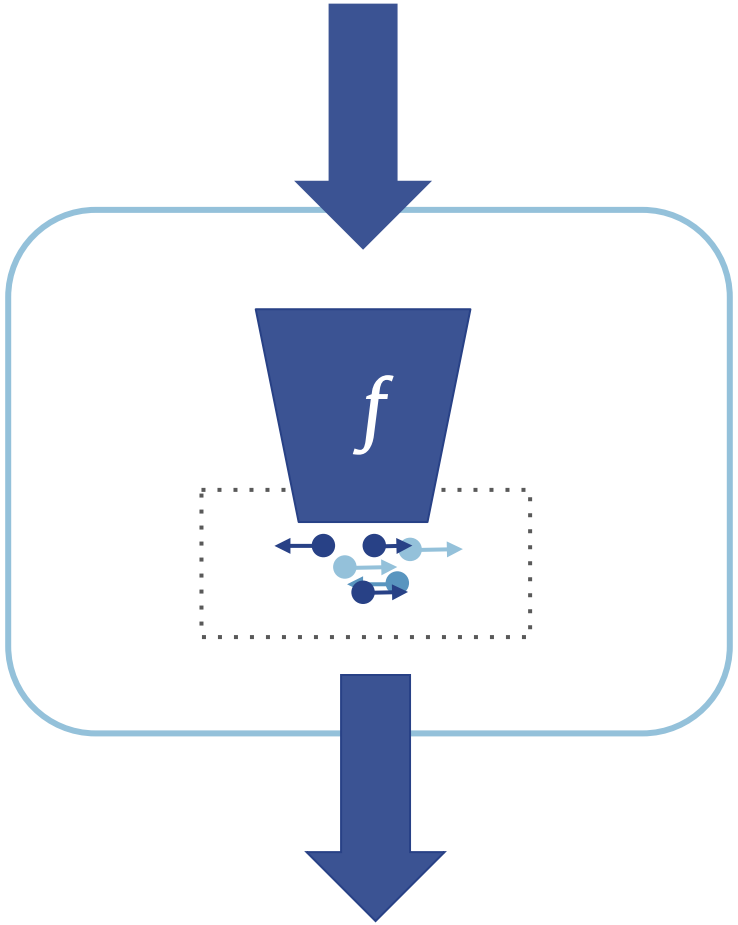
Output Privacy Approaches

Output privacy typically relies on:

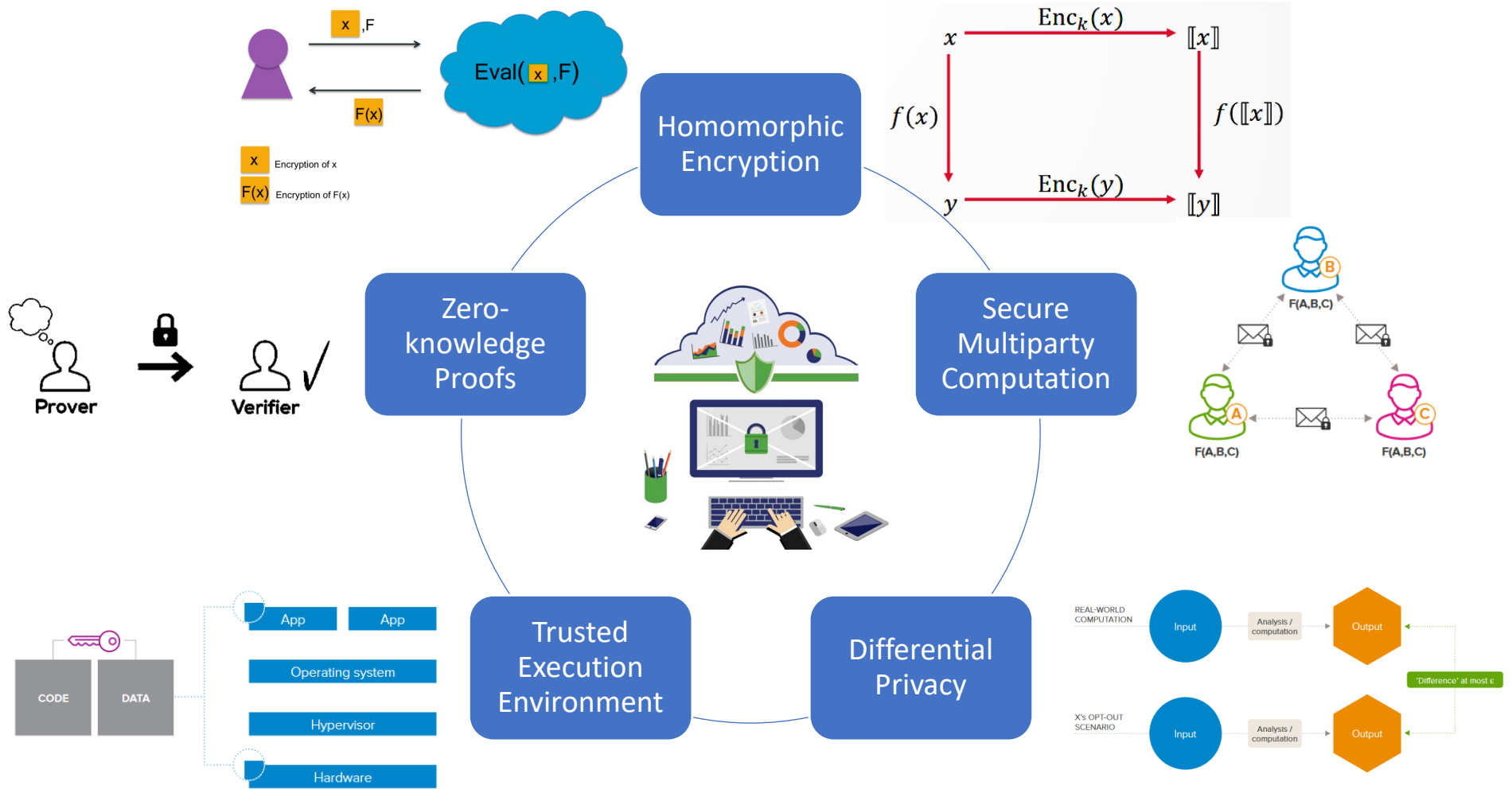
- ***Aggregation & sensitivity analysis***
  (*classic data disclosure controls*)

- ***Perturbation***
  (for example differential privacy)

Both endeavour to prevent reverse engineerability of the original data.



Privacy-Preserving Technologies



Pictures: [Privacy Preserving Technologies Report, Royal Society, 2019](#).

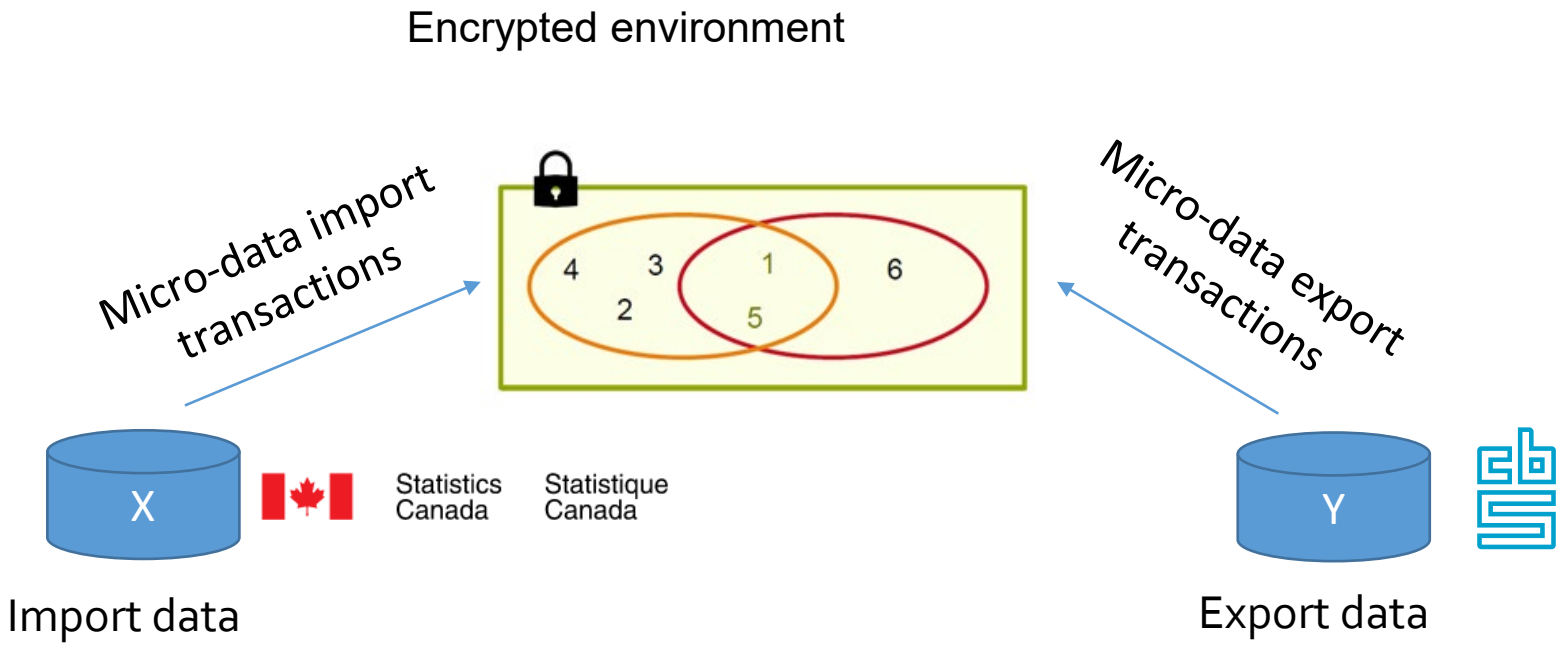
Highlights 2021

Framework for describing use cases for Input privacy-preservation

Documented 5 use cases

Described 2 generic scenario's

Private set intersection



International trade use case (PSI)

A PSI use case between two NSIs

Match international trade micro-data at the transaction level without a unique identifier

- Export data from a 1st NSI,
- More detailed import data, from a 2nd NSI.

A total or a mean is computed based on the matching result.

International trade use case (cont'd)

Benefits

- Study the use of trade agreements, e.g. Canada-EU Trade Agreement (CETA).
- Resolve bilateral trade-asymmetries, e.g. imports into A from B reported by A differing from exports from B to A reported by B.

Private machine learning

- Experiment with 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
- analyze the impact of PETs when used to protect the machine learning model

Pilot:

- NSO offer PET based remote analytics service

Organize public consultation

Open technical consultation on:

Towards a trustworthy Multi-Party Secure Private Computing-as-a-service infrastructure for official statistics

The consultation is mainly targeted at:

- Privacy and security experts from both the technical and legal sides.
- Potential users of the envisioned MPSPC infrastructure, including but not limited to statistical authorities, public bodies and private companies.
- Digital activists and representative of civil society (e.g., citizen associations).
- Researchers and developers in relevant fields.

Organize public consultation

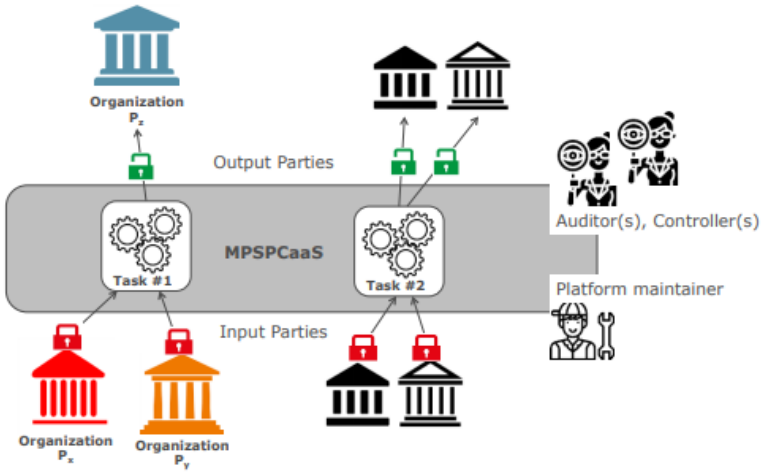


Figure 1 - Shared MPSPC-as-a-service infrastructure

Multi-party = no single point of trust

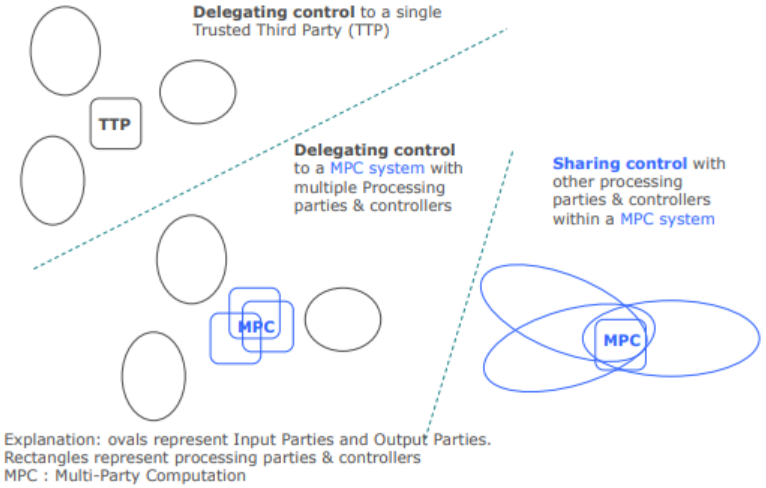


Figure 2 – Abstract representation of processing control distribution in the different paradigms. In the Trusted Third Party model a single entity is delegated full processing control (left). Multi-Party Secure Private Computation

Possible application areas for NSOs

1. **Advanced Data Collection**: access to (private) data sources with privacy-related barriers
2. **Datahubs:** providing data analytics environment for data with privacy-related barriers (NSO provides service and data)
3. **Data ecosystems**: enabling data collaboration in privacy-preserving data networks and alliances (NSO provides data but not service)

Next steps (in and after the project)

1. Final report with reports of the different tracks
2. Public wiki with the reports
3. Community of practitioners, continued *through the UN petlab*
4. Results of the technical consultation

Workshop Input privacy-preservation project
24 November 2022

Time	
14:00-14:15	Introduction to the project
14:15-15:00	Private Set Intersection
15:00-15:45	Private machine learning
15:45-16:30	Multi party private computing-as-a-service
16:30-17:00	Project conclusions
17:00	Closing Seminar

Q & A