UN Committee of Experts on Big Data and Data Science for Official Statistics

Privacy-Enhancing Technologies (PETs) Task Team

Enabling International Collaboration via UN PET Lab

2 September 2022

Raphaël de Fondeville (FSO Switzerland), Matjaž Jug (Statistics Netherlands),
Jack Fitzsimons (Oblivious.ai), Saeid Molladavoudi (Statistics Canada)

Data Protection & Privacy

Why Privacy-Enhancing Technologies?

Workstreams

The team's work spans multiple valuable workstreams:

1. **UN Handbooks**
   (first published in 2018)
2. **Exchange of Experience**
   (for example, application of PET in COVID-19 response activities in 2020)
3. **Training Courses**
   (in partnership with *openmined.org)*
4. **Experimentation**
   (UN Global Platform infrastructure + PET technologies)
5. **Promotion**
   (participation in events and other projects)
6. **Building PET Community of Practice in Official Statistics**
   (UN PET Lab initiative)

UN Handbooks – What's new?

This year, we have some welcome additions to the previous Handbook on Privacy Preserving Technologies, including:

- **Description of statistical case studies**
  (in collaboration with UNECE Input Privacy project and other initiatives)

- **New methods and technologies**
  (for example synthetic data, distributed learning)

- **Systematic inventory of all relevant international standards**
  (existing standards and standards in development)

- **Legal Guidelines for the use of PETs**

Estimated report delivery: **Q3 2022**

Building Active PET Community of Practice

- **Additional Statistical Use Cases**
  - ✓ Asymmetries in cross-border trade statistics
  - ✓ Private Machine Learning on Human Activity Recognition with Federated Learning

- **Experimentation on Statistical Use Cases in Safe Environment**
  (using no sensitive data!)

- **Technical Support**

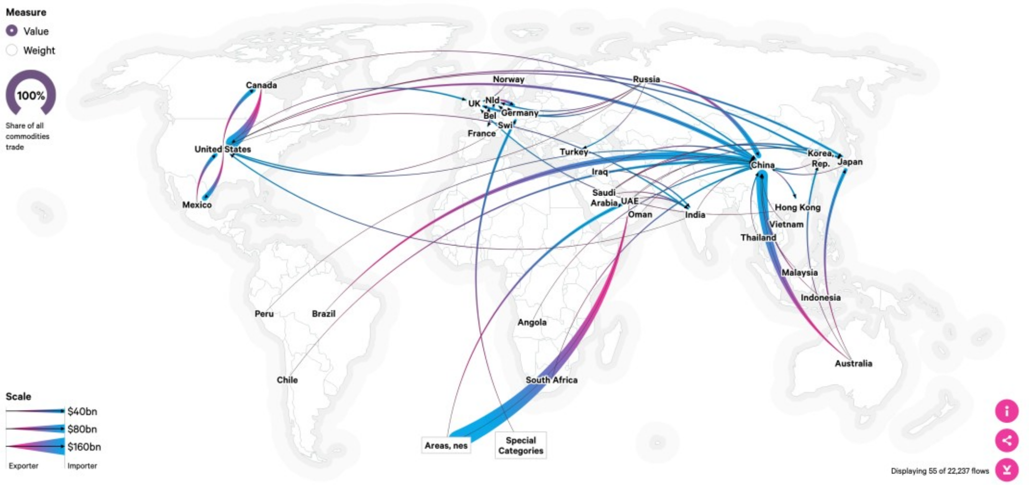
- **Demonstrations, PoCs..**

Check for more info here:

https://unstats.un.org/bigdata/task-teams/privacy

Case Study I: Cross Border Trade Statistics

- UN Platform Comtrade provides trade statistics.
- Amounts between pairs of countries should match.
- Ambiguities, disparities and errors are quite frequent.



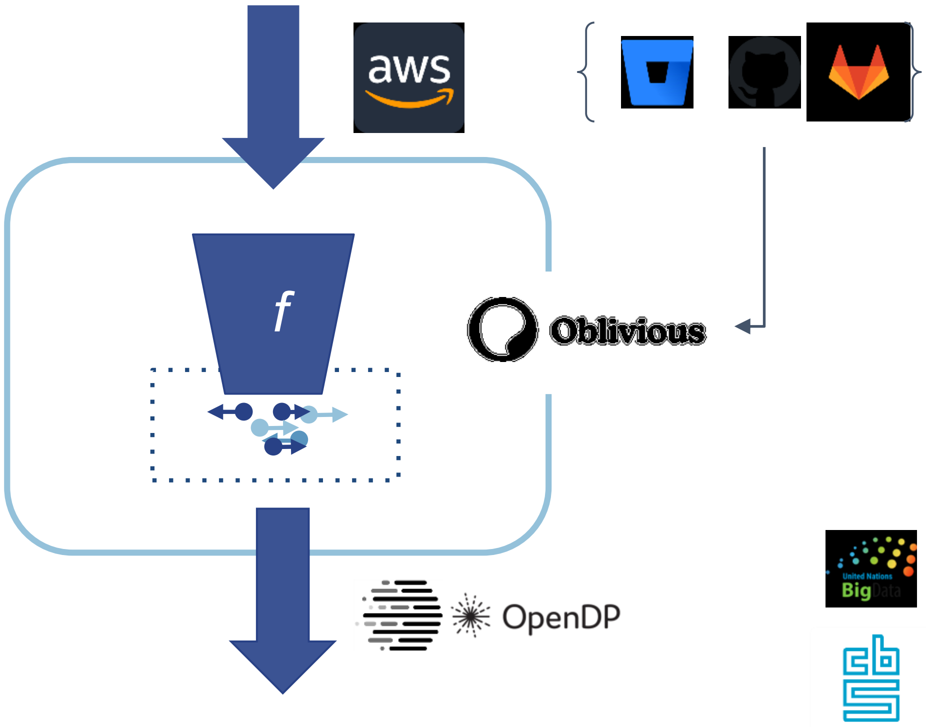
Goal: Using PETs to share **more granular information** between countries and to enable the linkage of additional heterogeneous data sources.

Case Study I: Cross Border Trade Statistics

Combining input and output privacy:

- ***Protect sensitive fine-grained data from each country***
  (*may contain data pertaining to a company*)

- ***Prevent statistics from revealing original data***
  (fixed epsilon for differential privacy)



Case Study I: Connecting to an Enclave

1. Request for Attestation

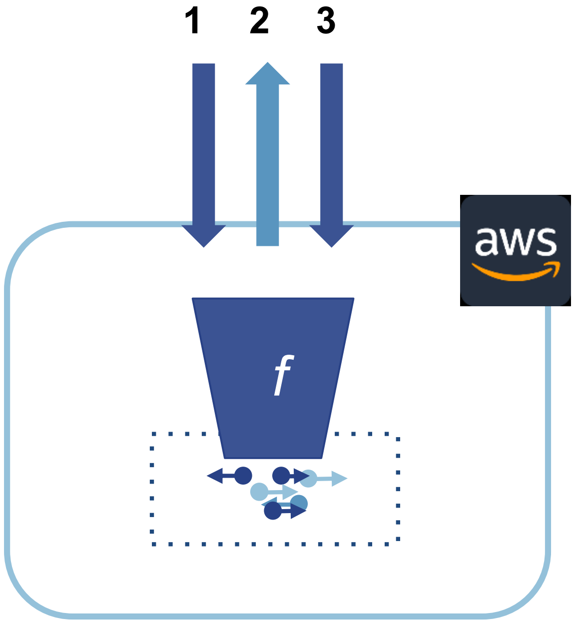
1. The client connecting to the enclave asking for proof of what is running.
2. HTTP packets signed with clients secret key.

2. Attestation & Public Key

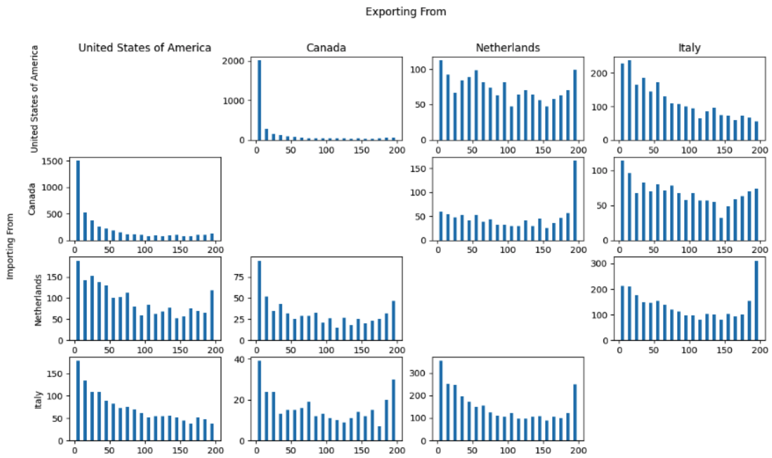
1. Enclave authenticates client and generated a new symmetric key pair.
2. Returns an attestation document with encrypted symmetric key embedded.

3. Encrypted Payloads

1. Finally encrypted data can be sent back and forth safely.

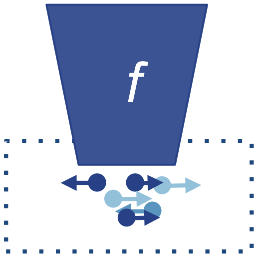


Case Study I: Differentially Private Statistics



Differentially private statistics are performed with a fixed epsilon, guaranteeing the privacy of the disseminated information.

We have used the **OpenDP/SmartNoise** framework to achieve this.



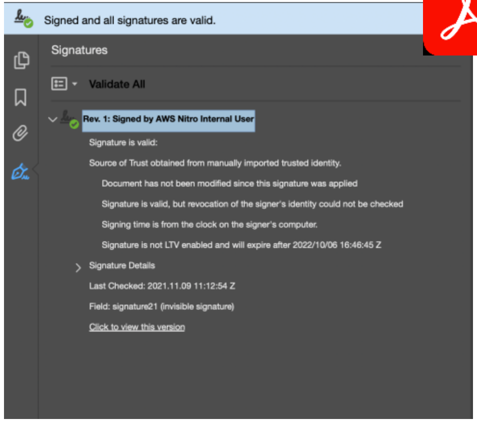
OpenDP

Case Study I: Trust to the Decision Makers

Embedded in the PDF is this image.

It contains the actual **Attestation Document** of the enclaved used for the multiparty computation.

By extracting it, we find the **public cert of a CA** created in the enclave. By adding it to the trusted authorities in our favourite PDF reader, we know that the document has not been modified since its creation.



Case Study II: Text Classification with HE

Description	ID1	ID2	Value
“mochi ice cream bon bons”	054	78	$5.31
“chipotle barbeque sauce”	201	34	$3.80

ML TC

NAPCS	ID1	ID2	Value
5611121	054	78	$5.31
5611132	201	34	$3.80

Statistics

Goal: automatic classification of retail goods into the North American Product Classification System (NAPCS).

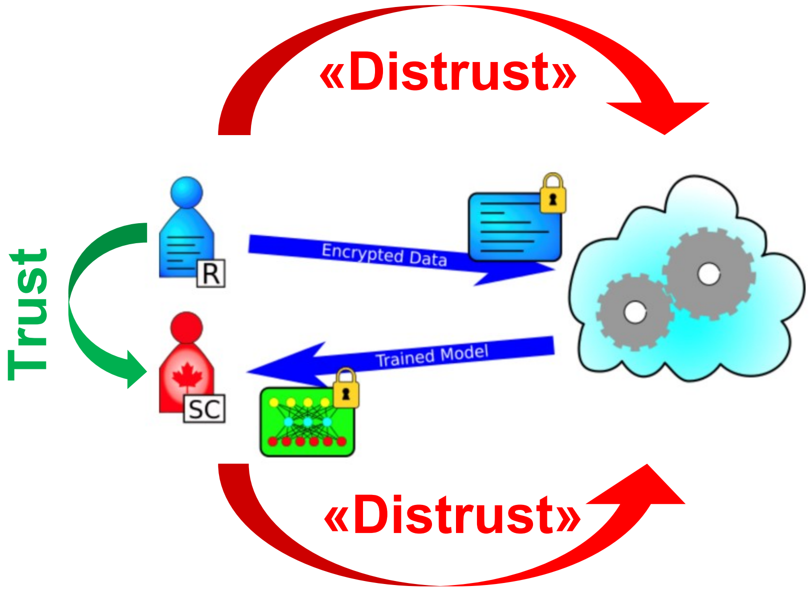
Zanussi, Z., Santos B., & Molladavoudi, S. (2021). Supervised Text Classification with Leveled Homomorphic Encryption . *Proceedings of the 63rd ISI World Statistics Congress*, 298–303.

Case Study II: Text Classification with HE

Purpose: PoC to migrate machine learning workloads to a cloud environment.

Solution: leveled HE scheme to train Neural Network classifiers.

Data: USDA FoodData 50,000 entries from 5 different NAPCS codes.



Zanussi, Z., Santos B., & Molladavoudi, S. (2021). Supervised Text Classification with Leveled Homomorphic Encryption . *Proceedings of the 63rd ISI World Statistics Congress*, 298–303.

Case Study II: Text Classification with HE

Model	Training Time	Accuracy (%)
Clear Text	15s	74.3
Single Layer NN	47h	67
Ensemble	7h	74.2

- It is today possible to use «off the shelf» implementations (Microsoft SEAL) to train ML alogrithms while preserving the privacy.
- Performance degradation introduced by HE approximations is manageable.

Zanussi, Z., Santos B., & Molladavoudi, S. (2021). Supervised Text Classification with Leveled Homomorphic Encryption . *Proceedings of the 63rd ISI World Statistics Congress*, 298–303.

Facts that matter