



8th International  
Conference on  
**BIG DATA**  
& Data Science for Official Statistics

**BILBAO 2024**

Informing Climate Change and  
Sustainable Development Policies  
with Integrated Data

**BILBAO. SPAIN** | **10-14 JUNE 2024** | **#UNBigData2024**

# The Role of Privacy Enhancing Technologies in Safeguarding Mobile Phone Data

**Matjaž Jug, Chair of the UN Privacy Enhancing Technologies Task Team**

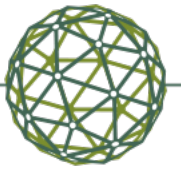
June 2024





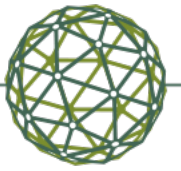
## Overview

- Why is Privacy data protection important?
- What are Privacy Enhancing Technologies?
- How can PETs help in the context of MPD?
- New approach to International Collaboration on PETs
- Examples of Case Studies



# Introduction

**Why is Privacy data protection important?**



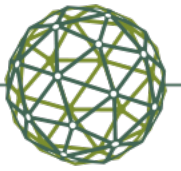
# Smartphone's location data is a “magnet” for Privacy issues

[Strava Data Heat Maps Expose Military Base Locations Around the World | WIRED](#)

[Your Apps Know Where You Were Last Night, and They're Not Keeping It Secret - The New York Times \(nytimes.com\)](#)

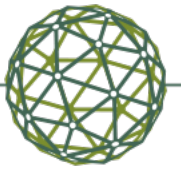
[Federal Agencies Use Cellphone Location Data for Immigration Enforcement - WSJ](#)

[Opinion | Twelve Million Phones, One Dataset, Zero Privacy - The New York Times \(nytimes.com\)](#)



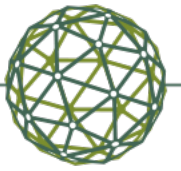
## Problem Definition

- Today, high-value data is siloed or locked down, significantly **hampering research and innovation**;
- This is often due to a combination of privacy, security, IP, cost, or legal reasons;
- When data is shared, things too often go wrong => data breaches;
- PETs, whilst not a silver bullet, have the potential to overcome this tradeoff between privacy and usability of data.



# Privacy-Enhancing Technologies

**What are PETs?**



## What are PETs?

“Privacy Enhancing Technologies (PETs) are a suite of tools that can help maximise the use of data by reducing risks inherent to **data use.**”

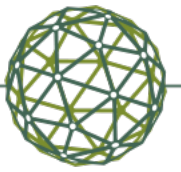
[Privacy Enhancing Technologies | Royal Society](#)

### From privacy to partnership

The role of privacy enhancing technologies in data governance and collaborative analysis



[From privacy to partnership | The Royal Society](#)



# What are PETs?

## Input Privacy

A guarantee that one or more parties can participate in a computation, in such a way that neither party learns anything about the other party's inputs to the computation.

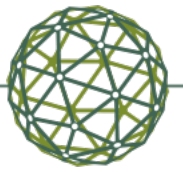
- ◆ SECURE ENCLAVES
- ◆ FUNCTIONAL ENCRYPTION
- ◆ HOMOMORPHIC ENCRYPTION
- ◆ SECURE MULTI-PARTY COMPUTATION
- ◆ FEDERATED LEARNING
- ◆ ZERO-KNOWLEDGE PROOFS

## Output Privacy

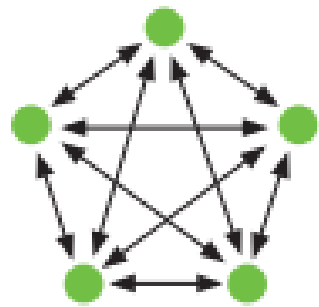
A guarantee that sensitive information in the data cannot be reverse engineered from the outputs of the computation.

- ◆ K-ANONYMIZATION
- ◆ DIFFERENTIAL PRIVACY
- ◆ SYNTHETIC DATA

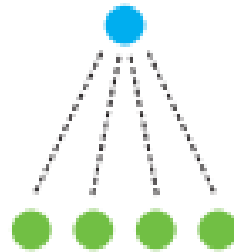




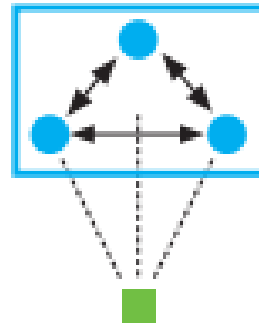
# Examples of some Input PETs



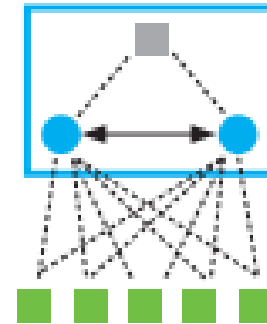
Standard Multiparty Computation



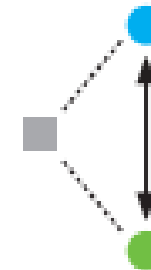
Federated Learning



Outsourced sMPC



Outsourced sMPC With Key Generator



Two Party Computation With Key Generator



Homomorphic Encryption

● Active Compute

■ Data Owner

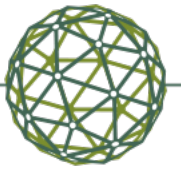
● Active Compute & Data Owner

■ Key Generator



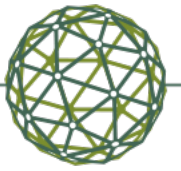
# The Role of PETs in the context of MPD

**How can PETs help in the context of MPD?**



## Different Considerations of PETs

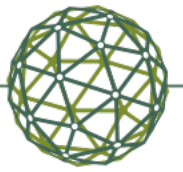
- **Functional** aspects: applicability of certain PETs for specific types of use cases, privacy guarantees, usability issues..
- **Technical** aspects: methodological limitations, technical maturity, scalability, energy consumption..
- **Legal & regulatory** aspects: purpose limitation, data minimisation, storage limitation, integrity and confidentiality..
- **Organizational** aspects: data management, skills, ethics..



## Application of PETs for MPD?

High-level use cases:

- Facilitating **secure access to highly sensitive privately-held datasets** in line with the Privacy-by-Design principle;
- Enhancing and streamlining existing **risk management** and accreditation processes for researchers;
- Enabling **collaborations between organizations**, both domestically and cross-border, via privacy-preserving joins;
- Testing MPD **methodology** on real-world like (e.g. synthetic) data;
- Minimising the **privacy disclosure risk** associated with outputs, e.g. the use of differential privacy for output protection.



# Example of Secure Enclave (Trusted Execution Environment) for Mobile Phone Data (used in Estonia & Indonesia MPD cases)

## Workflow:

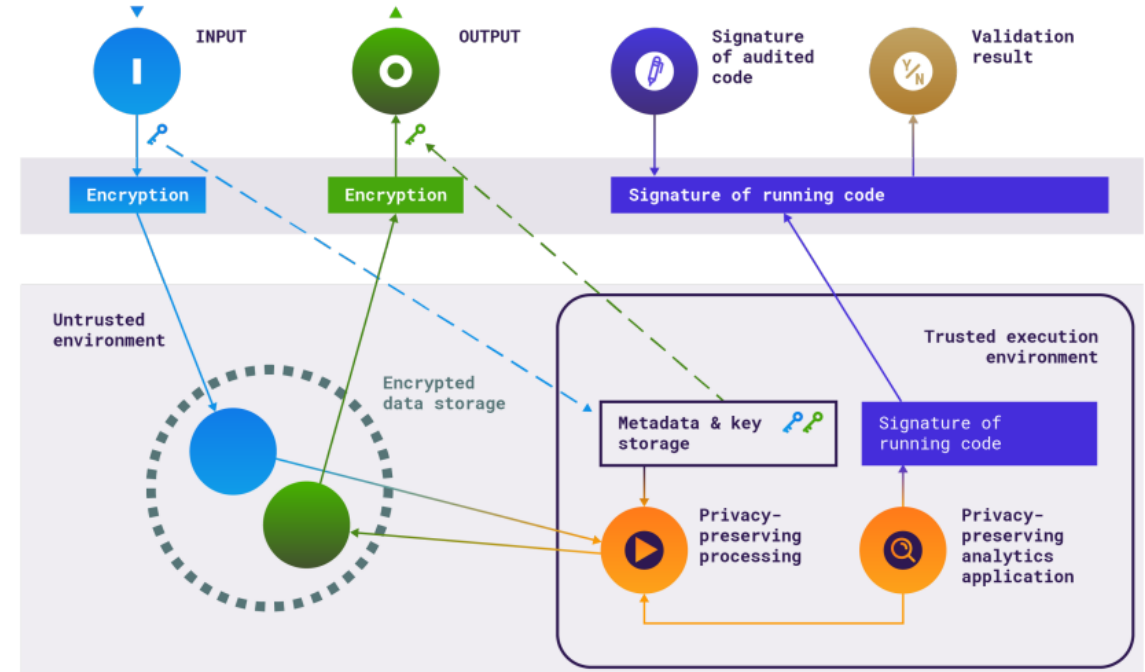
- Data owners encrypt data on site and upload to secure enclave;
- Platform runs queries without removing the protection;
- Authorized users receive results in an encrypted form;
- Platform can give proofs of its activities to third parties.

## SHAREMIND HI

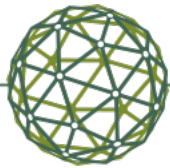
### CLIENT APPLICATION

Sharemind HI client library

### SHAREMIND HI APPLICATION SERVER



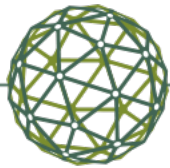
CONFIDENTIAL



## Legal & Regulatory aspects

- Involving legal experts early in any project is strongly advised;
- Specific PETs are typically not mandated by legislation, but PETs do enable compliance with legal requirements such as regarding “**data minimization**” or “**data protection by design and by default**”, and specific PETs might be recommended or required by a particular regulator for certain use cases;
- The use of PETs must be consistent with existing laws, policies, and ideally cultural norms, and PETs can open up new opportunities and affordances within this social structure;
- Any activity involving data from more than one jurisdiction will be more complicated;
- Different laws and jurisdictions may take different views on the adequacy of a PET for a given use case, so we encourage regulators to publish guidance about the use of PETs.

[2023\\_UN PET Guide.pdf](#)



## Regulation on European Statistics (amendment approved by the European Parliament on 13/3/2024)

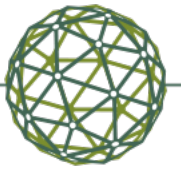
“..the particular safeguards, which should be applied when data sharing under Regulation (EC) No 223/2009 requires personal data to be processed, include **technical and organisational measures such as privacy-enhancing technologies** and the respect of the principles of **purpose limitation, data minimisation, storage limitation and integrity and confidentiality** as set out in Regulation (EU) 2016/679 and Regulation (EU) 2018/1725 and further elaborated in the European Statistics Code of Practice. ”

[TA \(europa.eu\)](#)

European Data Protection Supervisor Opinion 40/2023:

“Further, the EDPS recommends that such sharing of **personal data by private data holders** shall make use of **privacy-enhancing technologies** and shall take place using a **secure infrastructure.**”

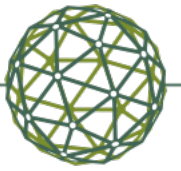
[2023-0753\\_d2563\\_opinion\\_en.pdf \(europa.eu\)](#)



# International Collaboration on PETs

**UN Task Team for Privacy Enhancing Technologies & UN PET Lab**





# UN Task Team for Privacy Enhancing Technologies

- **Methodology** (PET Task Team)
  - Research of applicability of PETs for use cases in Official Statistics.
- **Experimentation** (PET Lab)
  - Proofs-of-concept and pilot projects focused on the evaluation of PETs for real-world use cases in the official statistics community.
- **Outreach & Training** (events, webinars, Open Houses)
  - Sharing learnings and insights from the use of PETs with the wider statistical community through training, public events, and educational materials.
- **Support Services**
  - Advice to organizations utilizing PETs



## UN PET Guide


- Role of Privacy Enhancing Technologies in Official Statistics
- Methodologies & Approaches
- 18 Case Studies
- Standards
- Legal and Regulatory Issues

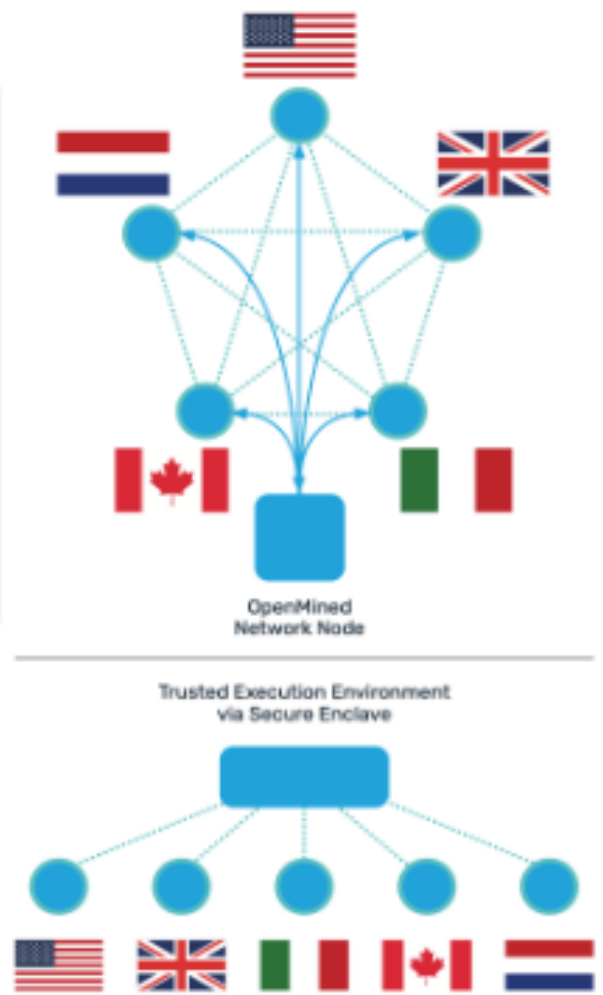


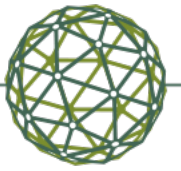
[2023\\_UN PET Guide.pdf](#)

# 16. United Nations PET Lab: International Trade

Created by David Buckley, last modified on Feb 09, 2023

Purpose	Enable multiple national statistical offices (NSOs) to perform reconciliation and joint analysis on independent but collected trade datasets.
Date	 Menu   Weekly edition   The world in brief   Search
PET	Science and technology   Data privacy
Details	<h2>The UN is testing technology that processes data confidentially</h2> <p>How to analyse data without revealing their secrets</p>
Partners	between parties.
Implementation status	Proof of concept (ongoing)
Resources	<a href="#">What is the UN PET Lab and Why is it Important?</a> <a href="#">The Economists write-up on the PET Lab</a>





## Case Studies Related to MPD

**Purpose, datasets, PETs used, implementation status, outcomes & lessons...**

## Case study repository

Created by David Buckley, last modified on Feb 08, 2023

A number of national statistical offices (NSOs) and government agencies are leveraging PETs to enable rich and innovative statistical analysis, whilst protecting the privacy and confidentiality of sensitive information contained within their datasets. This repository provides example case studies of where PETs have been deployed in the real world to facilitate the generation and dissemination of privacy-preserving statistics.

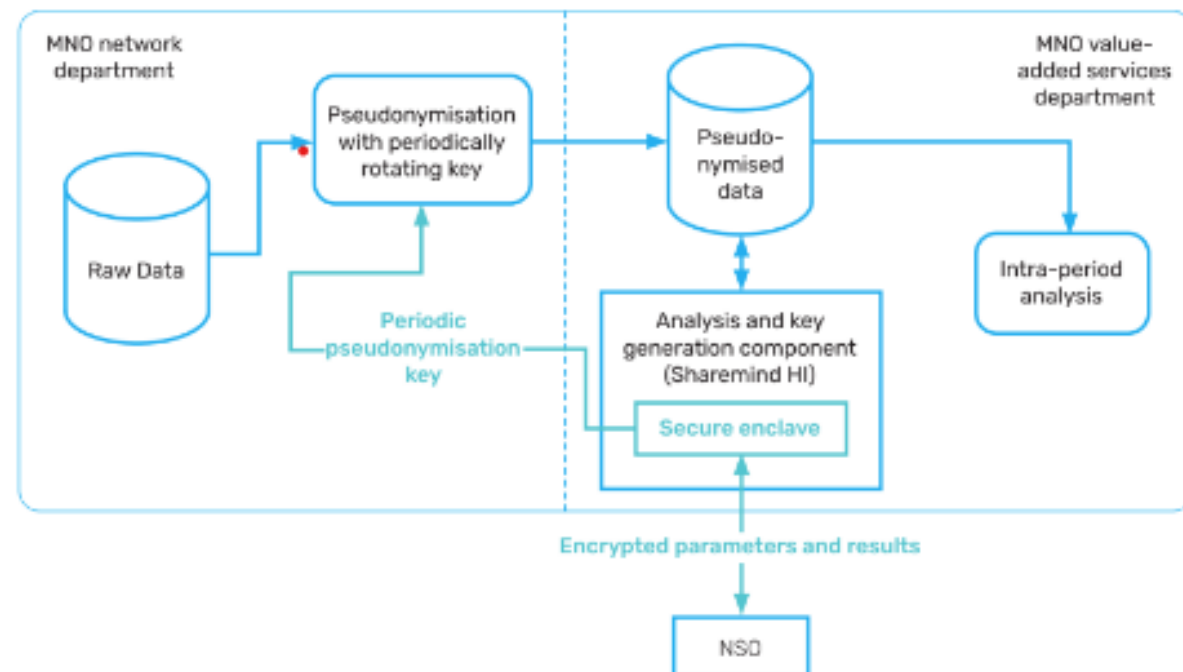
The repository includes the 18 case studies that are provided in Chapter 3 of the 2023 UN Guide on Privacy-Enhancing Technologies for Official Statistics (which can be found [here](#)). We intend to update and expand this repository moving forward,

- [1. Boston Women's Workforce Council: Measuring salary disparity using secure multi-party computation](#)
- [2. European Statistical System: Developing Trusted Smart Surveys](#)
- [3. Eurostat: Processing of longitudinal mobile network operator data](#)
- [4. Indonesia Ministry of Tourism: Confidentially sharing datasets between two mobile network operators via a trusted execution environment](#)
- [5. Italian National Institute of Statistics and Bank of Italy: Enriching data analysis using privacy-preserving record linkage](#)
- [6. Office for National Statistics: Trialling the use of synthetic data at the United Kingdom's national statistics institute](#)
- [7. Samsung SDS \(Korea\): Data aggregation system](#)
- [8. Statistics Canada: Measuring the coverage of a data source using a private set intersection](#)
- [9. Statistics Canada: Training a machine learning model for private text classification using leveled homomorphic encryption](#)
- [10. Statistics Canada: Trialling the use of synthetic data](#)
- [11. Statistics Korea: Developing a privacy-preserving Statistical Data Hub Platform](#)
- [12. Statistics Netherlands: Developing privacy-preserving cardiovascular risk prediction models from distributed clinical and socioeconomic data](#)
- [13. Statistics Netherlands: Measuring effectiveness of an eHealth solution using private set intersection](#)
- [14. Twitter and OpenMined: Enabling Third-party Audits and Research Reproducibility over Unreleased Digital Assets](#)
- [15. United Nations Economic Commission for Europe: Trialling approaches to privacy-preserving federated machine learning](#)
- [16. United Nations PET Lab: International Trade](#)
- [17. United States Census Bureau: Deploying a differentially private Disclosure Avoidance System for the 2020 US Census](#)
- [18. United States Department of Education: Analysing student financial aid data using privacy-preserving record linkage](#)

### 3. Eurostat: Processing of longitudinal mobile network operator data

Created by David Buckley, last modified on Feb 09, 2023

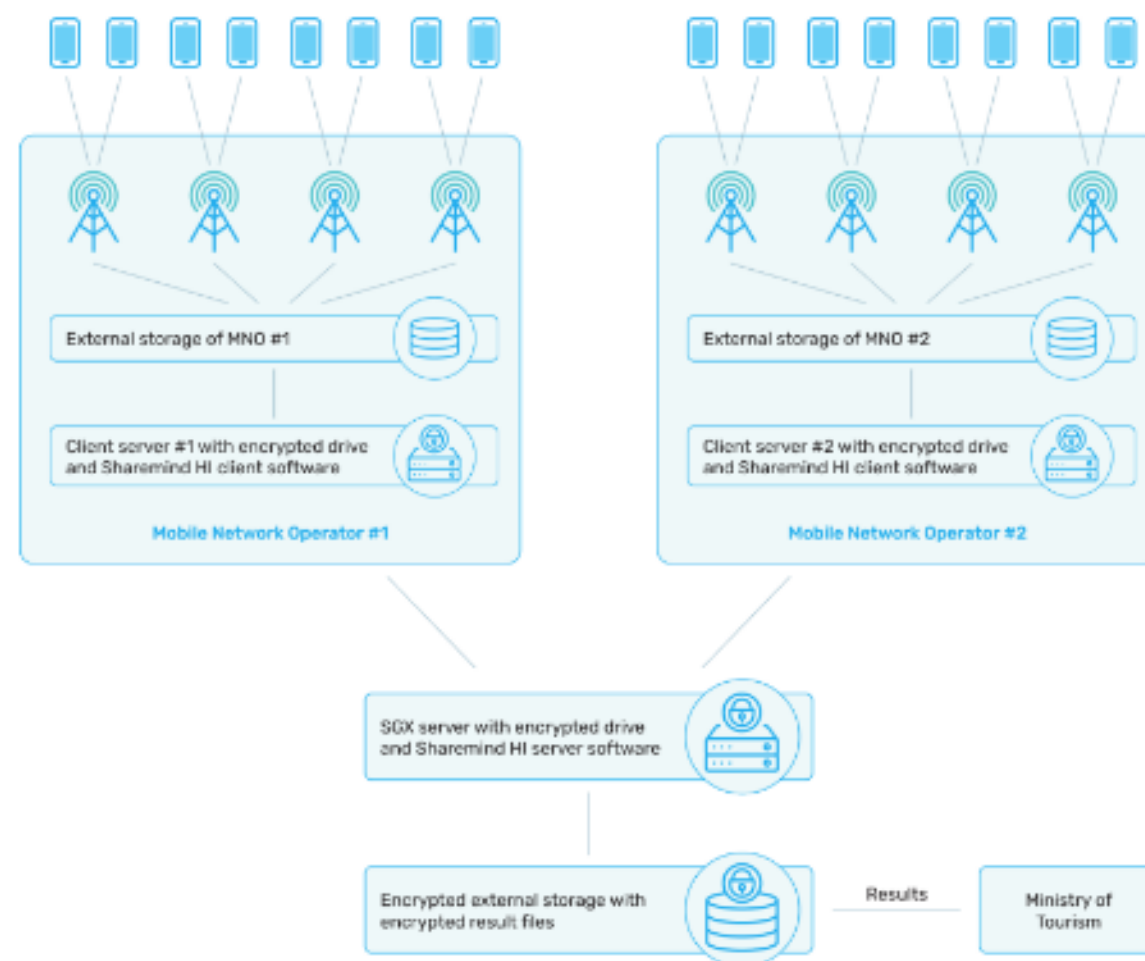
Purpose	To enable a NSO to safely and confidently conduct analysis on longitudinal Mobile Network Operator (MNO) mobility data.
Datasets	Summary of daily visited locations by individual (pseudonymised) subscribers extracted from (call data record) CDR or signalling data, for 100 million mobile subscribers.
PETs used	Trusted Execution Environment
Details of computation	Articulated workflow consisting of a chain of simple operations performed regularly. Longitudinal MNO analysis and integration of MNO and NSO data takes place within a secure enclave/trusted execution environment using a predefined set of algorithms that deliver aggregate (non-personal) data in output.
Parties and trust relationship	MNO and NSO act as both input and output parties, and their relationship is assumed honest-but-curious.
Implementation status	Proof of concept
Resources	<a href="https://ec.europa.eu/eurostat/cros/content/eurostat-cybemetica-project_en">https://ec.europa.eu/eurostat/cros/content/eurostat-cybemetica-project_en</a>

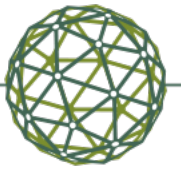


## 4. Indonesia Ministry of Tourism: Confidentially sharing datasets between two mobile network operators via a trusted execution environment

Created by David Buckley, last modified on Feb 09, 2023

Purpose	To generate tourism statistics from the combined data of two Confidential sharing of datasets of two mobile network operators (MNOs).
Datasets	A list of IMSIs from the two MNOs in border areas for the same time period. The IMSIs were uniformly hashed from the 7th digit onwards.
PETs used	Trusted Execution Environment
Details of computation	Statistics are generated from the input data in a trusted execution environment (Intel SGX), through the Sharemind HI platform
Parties and trust relationship	Two MNOs act as input parties; Sharemind serve in a compute role; the Ministry of Tourism acts as output party.
Implementation status	Production
Resources	<a href="https://sharemind.cyber.ee/sharemind-hi/">https://sharemind.cyber.ee/sharemind-hi/</a> <a href="https://netmob.org/assets/netmob19_withFCC.pdf">https://netmob.org/assets/netmob19_withFCC.pdf</a>





## Resources

- [Task Team on Privacy-Enhancing Technologies — UN-CEBD](#)
- [Emerging privacy-enhancing technologies: Current regulatory and policy approaches | en | OECD](#)
- [Privacy-enhancing technologies \(PETs\) | ICO](#)
- [Privacy Enhancing Technologies for Official Statistics \(PET4OS\) | Eurostat CROS \(europa.eu\)](#)
- [OpenMined Courses](#)



# Questions?



#UNBigData2024