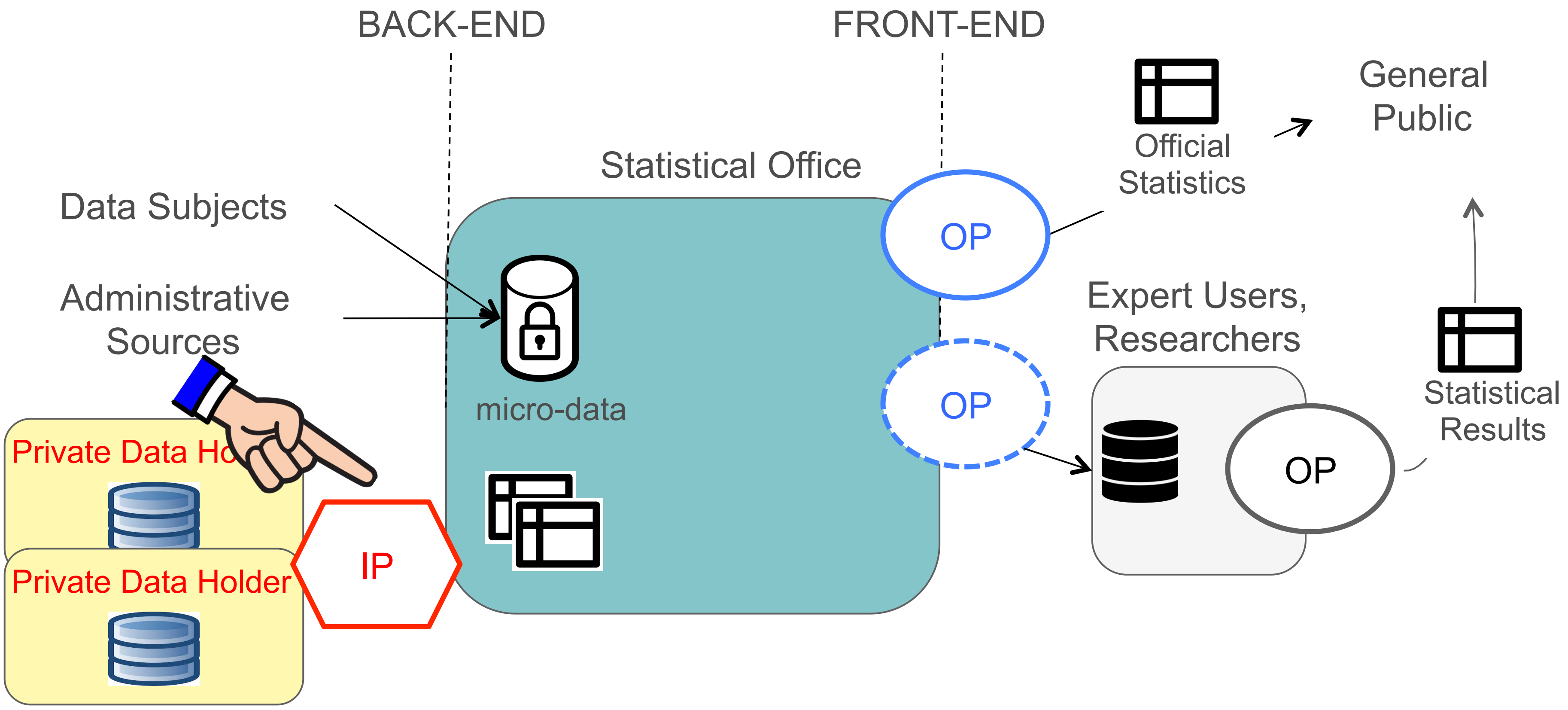
A proof-of-concept solution for secure processing of mobile network operator data for official statistics

Fabio Ricciato,
Eurostat, Unit A.5 Methodology; Innovation in official statistics

Joint UNECE/Eurostat Expert Meeting on Statistical Data Confidentiality
1-3 December 2021

Input Privacy (IP) & Output Privacy (OP)



F. Ricciato, A. Bujnowska
A reflection on privacy and data confidentiality in Official Statistics, ISI 2019
https://ec.europa.eu/eurostat/cros/system/files/isi_paper_ricciato_bujnowska_final.pdf

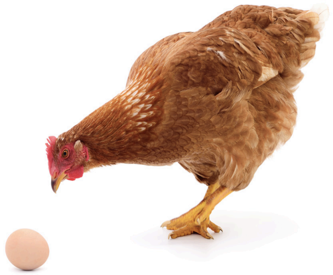
Background and motivations

- Project in collaboration between Eurostat and Cybernetica (1 year, closed mid 2021)

Main goals for Eurostat:

- Familiarize with the **technical** and **legal** aspects of **Input Privacy** technologies (a.k.a. *Secure Private Computing* technologies, a.k.a. *Privacy-Preserving Computation*)
- Understand specification / design / development processes and legal requirements in Europe (Data Protection Impact Assessment, DPIA)
- Assess feasibility of potential application in one specific application domain, namely (re)use of Mobile Network Operator (MNO) data of Official Statistics

Reference scenario 1/2



- Technology can provide a solution to a well specified problem. But the terms of the problem are defined also by what is (expected to be) feasible
  - → technological solution and application scenario a bit like chicken and egg …
- A **real-world application scenario** cannot be *precisely* specified yet
  - **Business** aspects still open (NSI – MNO relationship)
  - Privacy requirements depending on **national legislations** and orientations of Data Protection Authorities (DPA) → heterogeneity across member states
  - **Legal** framework in an evolving stage (forthcoming ePrivacy regulation, Data Act, revision of EU Statistics legislation 223/2009 …)
  - **Methodological aspects** still open – algorithms for transforming MNO data into officia statistiscs not yet consolidated

- Uncertainties in scenario/problem specification → take assumptions

European Commission

Reference scenario 2/2

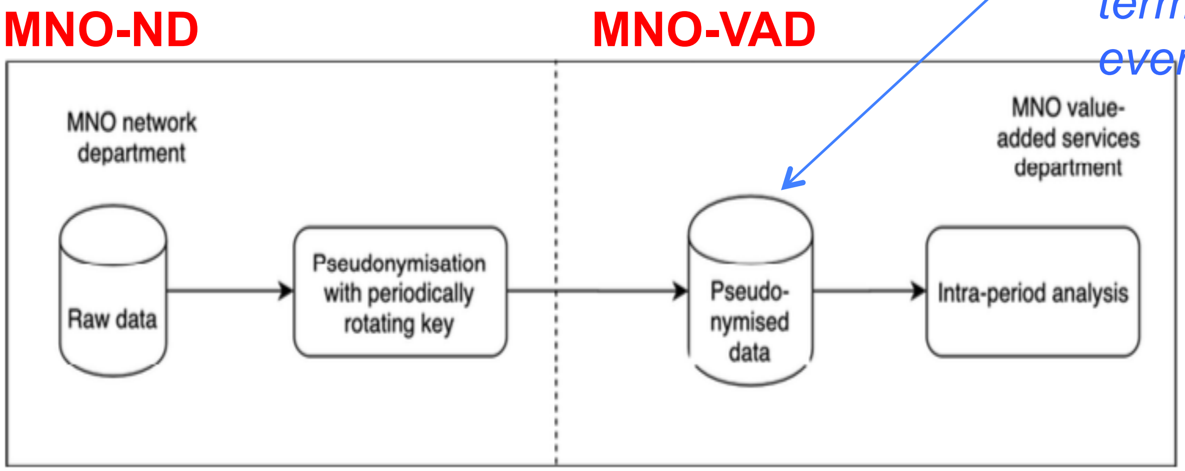
- Reference application scenario (= the problem)
  - Specified entirely by Eurostat, given as input to the project
  - Based on a set of assumptions - about national legal context, business relationships, statistical methodologies, etc.
  - Designed to “stress the technology” – harsher-than-real technical requirements

- Key ingredients
  - 1 NSI and 1 MNO (multi-MNO fusion NOT in scope, left for a separate future project)
  - Tension between short-term (re)pseudonymisation cycle at MNO (24 hours)
    and long-term analysis requirement by NSI (3 months or more)
  - Fusion of confidential MNO data (not visible to NSI) with confidential NSI data (not visible to MNO) for calibration of final statistics

European Commission

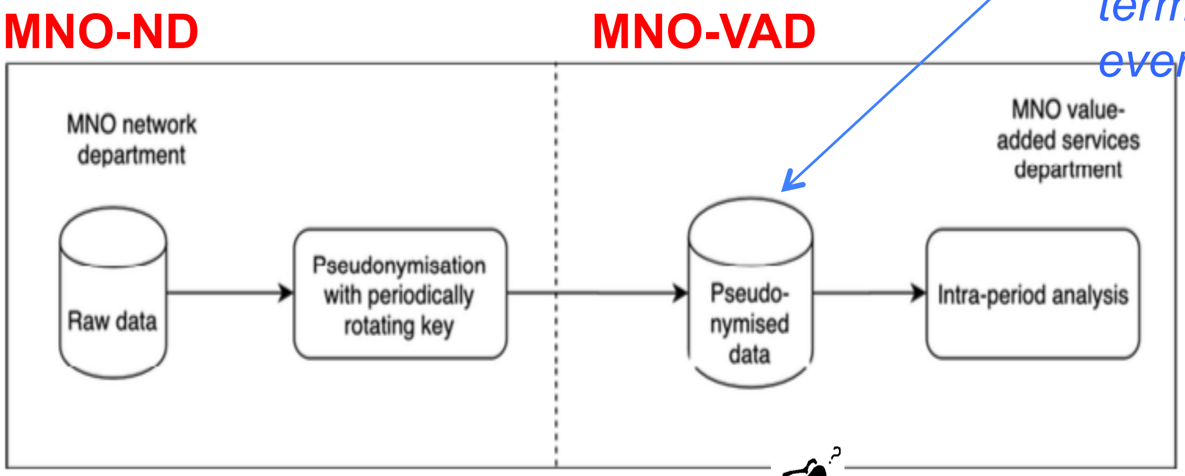
Reference scenario – legacy workflow

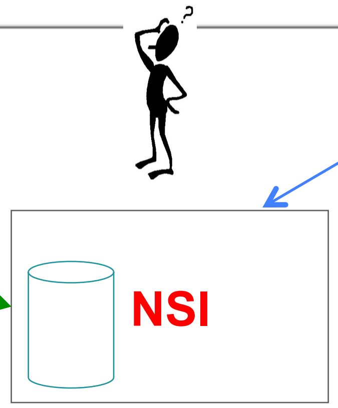
MNO data are (re)used for secondary purpose (e.g. commercial analytics) only after being pseudonymised with short-term pseudonyms that are changed every period T =24 hours



Reference scenario – NSI reqirements

MNO data are (re)used for secondary purpose (e.g. commercial analytics) only after being pseudonymised with short-term pseudonyms that are changed every period T =24 hours





*Furthermore, the methodology requires the **aggregate** data to be calibrated based on detailed census-grid data held by NSI that cannot be shared with the MNO*

The statistical methodology by NSI requires the same user to be tracked over long periods (3 months) to identify its “usual environment”.

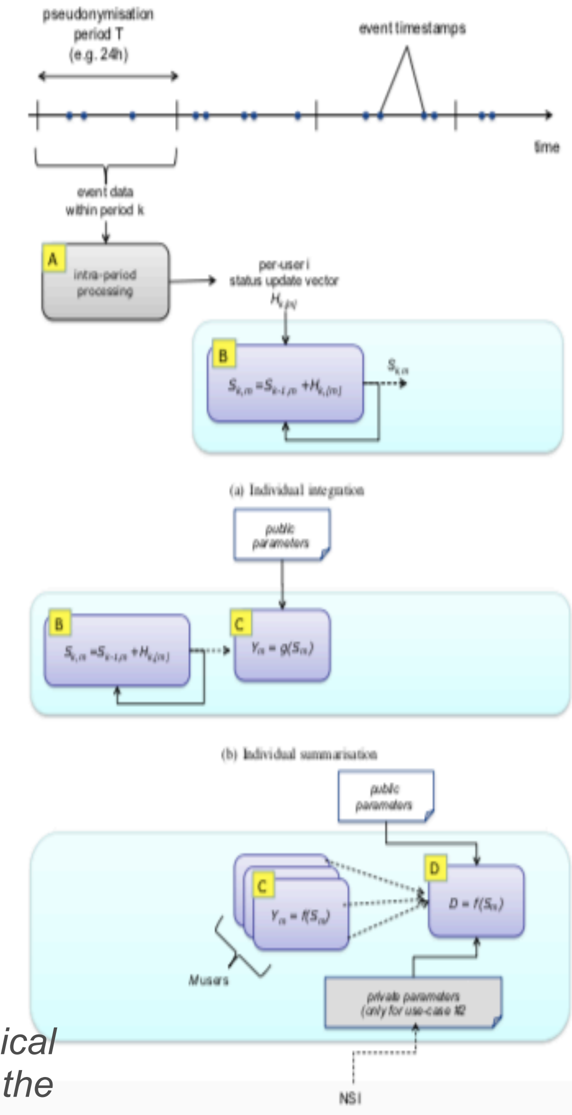
Statistical methodology

- A statistical “toy methodology” was invented (by Eurostat) specifically and exclusively for this project (*)
- Inspired by the statistical concepts of “Functional Urban Area” adapted to MNO data (→ Functional Urban Footprint)

Main steps

- (i) summarization of main long-term individual locations;
- (ii) aggregation over mobile users (counting users in locations);
- (iii) simple calibration based on confidential NSI data;
- (iv) Statistical Disclosure Control (SDC) filters based on *k*-anonymity
  (NB: integration of Output Privacy mechanisms inside the methodology!).

- The aggregate data delivered to the NSI are non-personal



(a) Individual integration

(b) Individual summarisation

(*) Caveat: the toy-methodology is NOT meant to represent an official methodological proposal for MNO data processing. It serves exclusively the purpose of “stressing the solution” developed by this specific project with a methodology that is reasonably articulated, neither too trivial nor unnecessarily complicated, and anticipative of possible elements of future official methodologies.

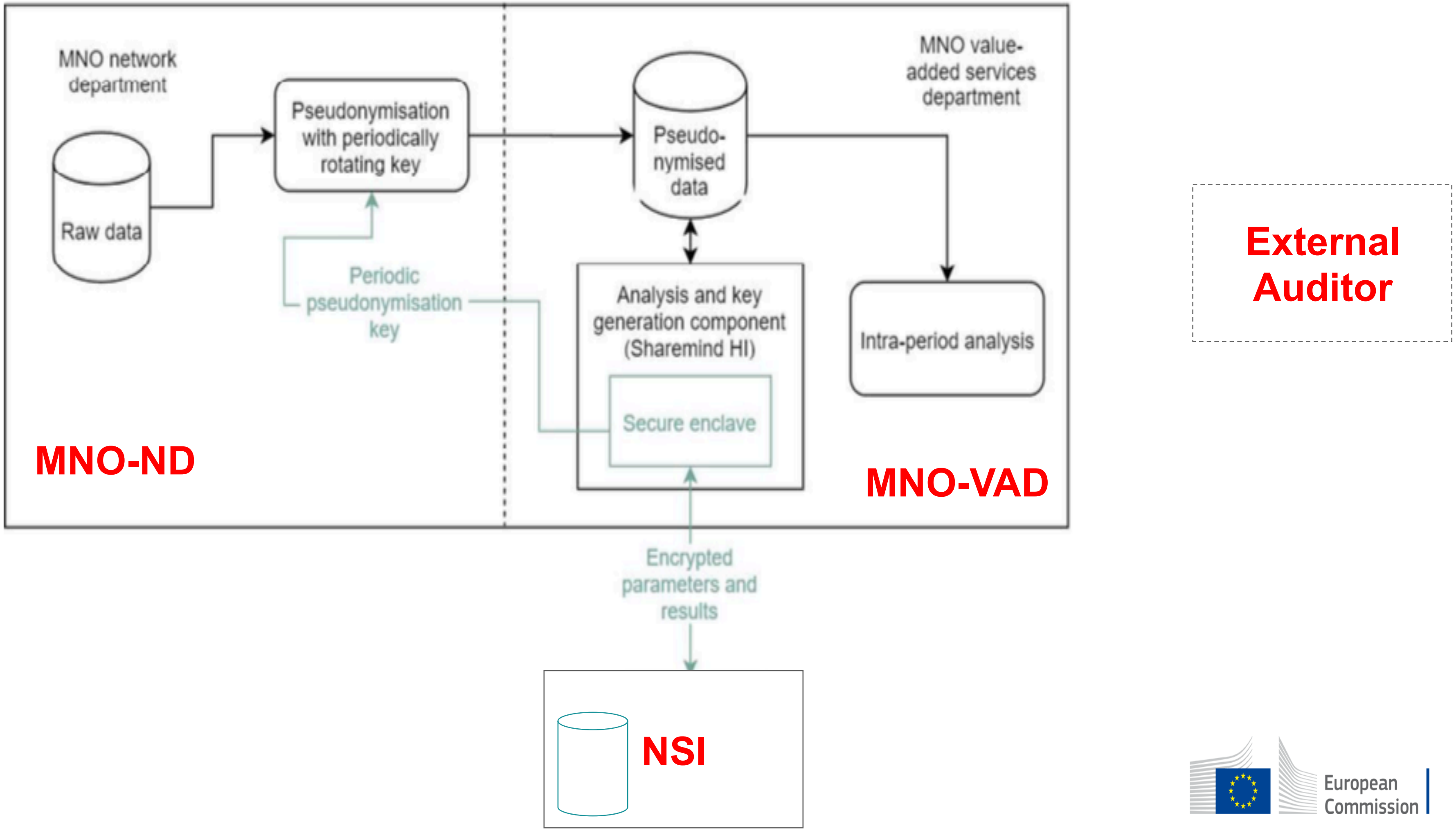
Technology

Technology of choice

- **Trusted Execution Environment (TEE)** with hardware isolation
- Based on **Intel SGX** technology
- **Sharemind HI** (proprietary platform developed by Cybernetica)

- Motivations
  - Technological **maturity** – ideally, a turn-key commercial solution that could be deployed today in production settings
  - **Scalability** – should be able to crunch data from 100 Mio users (large EU MNO) with commercial-off-the-shelf hardware
  - **Flexibility** – introducing changes to the processing methods should not be too costly/difficult

Solution architecture



MNO network department

Raw data

Pseudonymisation with periodically rotating key

Periodic pseudonymisation key

MNO-ND

MNO value-added services department

Pseudo-nymised data

Analysis and key generation component (Sharemind HI)

Secure enclave

Intra-period analysis

MNO-VAD

External Auditor

Encrypted parameters and results

NSI

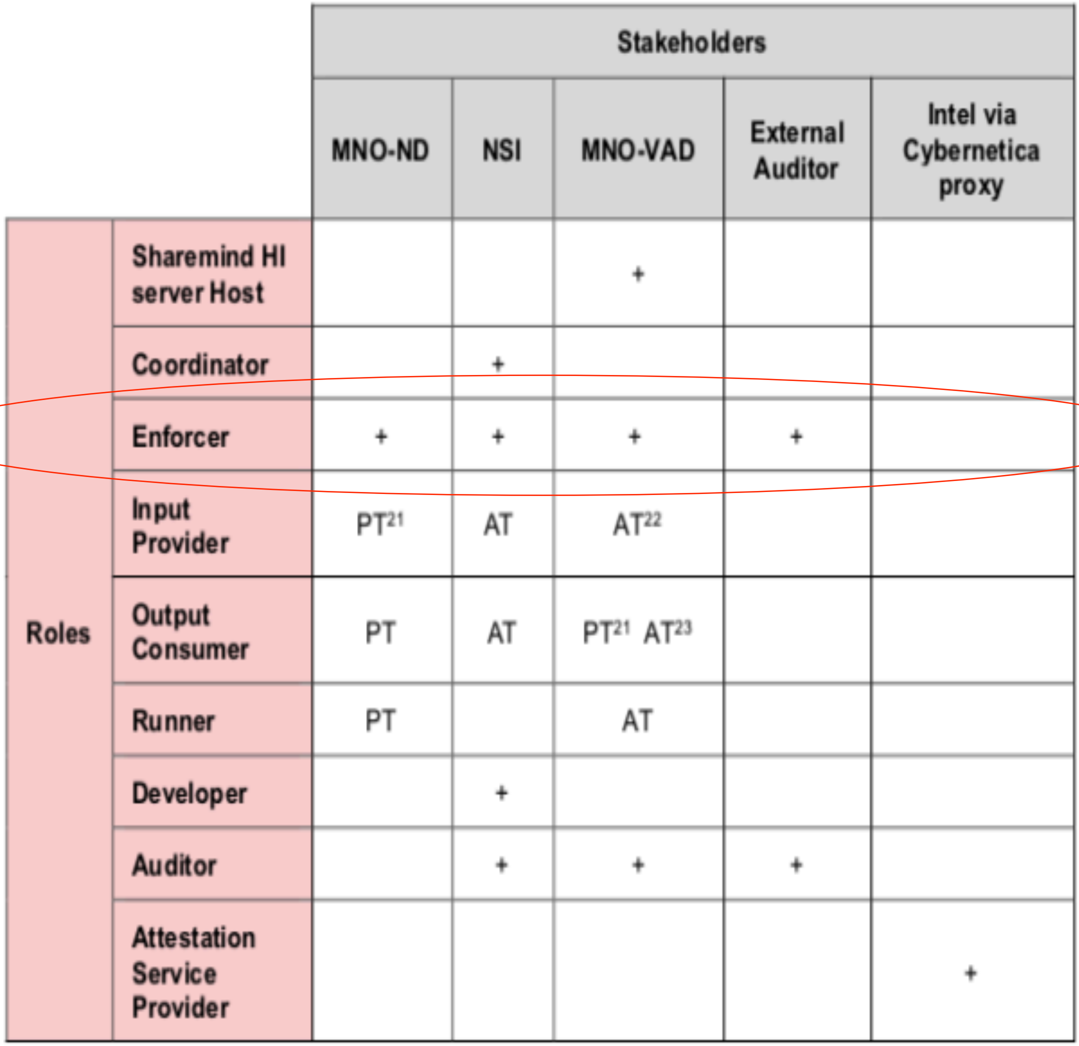
Table of Roles and Actor

The adopted technology (Sharemind HI) has a set of pre-defined "Roles".

Roles are assigned to Actors (stakeholders)

Enforcer role: only software code that is approved and signed by all Enforcers can be executed in the Secure Enclave (ex-ante control).

Assigning the Enforcer role to multiple actors is key to build trust into the solution without trusting any single actor (they are trusted collectively, not individually)



		Stakeholders				
		MNO-ND	NSI	MNO-VAD	External Auditor	Intel via Cybernetica proxy
Roles	Sharemind HI server Host			+		
	Coordinator		+			
	Enforcer	+	+	+	+	
	Input Provider	PT[21]	AT	AT[22]		
	Output Consumer	PT	AT	PT[21] AT[23]		
	Runner	PT		AT		
	Developer		+			
	Auditor		+	+	+	
	Attestation Service Provider					+

Table 1 Roles of Stakeholders in the Proposed Solution

Project results

- Design of the solution, including business processes, roles, architecture etc.

- Implementation of a proof-concept (PoC) version of the solution (based on reference scenario assumptions and implementing the toy methodology)
  - PoC implementation is ready to run in field pilots!

- Load testing on commercial-off-the-shelf hardware with synthetic data (emulating 100 Mio mobile users x 3 months).
  - Test result show computation load is not of concern (comp. time scaling linearly and anyway stay within acceptable range)

- Legal study + Data Protection Impact Assessment (DPIA) reference model
  - Will help NSI-MNO consortia that are interested in conducting a pilot testing of the solution with real data in their country to prepare a complete DPIA
  - Included a detailed assessment of security risks

Lessons learned 1/2

- As with all security technologies, Input Privacy technologies must encompass hardware, software and ***humanware*** levels (**business processes, roles**).
- The design of roles and business processes (i.e., distribution of powers between the stakeholders!) at *humanware* level is key (hardware and software are enablers).
  - Successful technologies must allow for **flexible and easy (re)configuration of roles** and business processes.
- **Legal feasibility** proved to be challenging based on current legislation.
  - Complex interplay between three legislative domains: data protection (GDPR), statistical legislation and telecom legislation (ePrivacy Directive)
  - Country-specific situations due to national legislations
  - Adoption of Input Privacy solution does not eliminate the need to have a clear legal basis for the reuse of MNO data by NSI.

Lessons learned 2/2

- Technical scalability was not a big issue for the considered scenario and should not be a major worry in production scenarios
  - The secure enclave has limited memory but can use external memory hardware-accelerated encryption/decryption.

- Absolute security does not exist. The question is whether the solution is “secure enough” in the specific context
  - The goal is not to reduce the attack surface of MNO infrastructure, but to avoid increasing the attack surface when NSI comes into play
  - For the considered technology (Intel SGX), new exploits/weaknesses were discovered even during the project, but new patches and countermeasures were developed and deployed readily (secure IT infrastructure requires continuous updates, as any other IT)

Future work

- Project deliverables available publicly
  - from https://europa.eu/!RDkywK
  - including detailed scenario description, architecture, DPIA study, etc.
  - all code developed within the project available open-source from https://github.com/eurostat/mnodata-tee-poc

- Next step: in-field testing with real MNO and real NSI???
  - Eurostat is open to support NSI-MNO consortia interested in testing the developed solution in a real environment with real-world data

  - For folllow-up contact Fabio.Ricciato@ec.europa.eu

Thank you

For folllow-up contact Fabio.Ricciato@ec.europa.eu