Mobile Big Data Processing Without Compromising Privacy

Margus Tiru / Positium

Dan Bogdanov / Cybernetica

positium

CYBERNETICA

The Basics

- Mobile Big Data is protected by personal data protection legislation
- EU: 2018 GDPR will increase the protection and make it harder to access and process directly or indirectly identifiable data for MNOs and governments
- Anonymous Mobile Big Data (without personal ID, IMSI or pseudonymised code), loses its value: ~~longevity~~, ~~movements~~, ~~anchor points~~, ~~long- and short-term migration, usual environment, tourism trip duration~~, etc., = very limited possibilities

positium

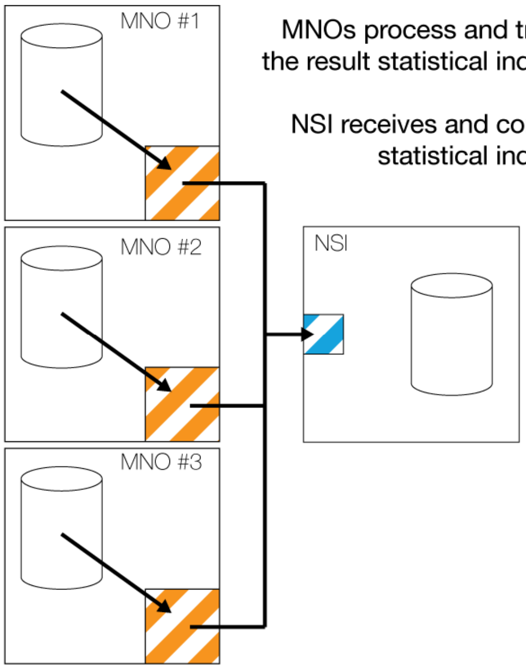
CYBERNETICA

Current Approach

- Get microdata from MNOs using National Statistics Act, but legislation not in place and probably will not be in near future
- Process data in MNOs' infrastructure could be a solution (OPAL), must validate methodology, high cost for MNOs
- Get anonymous data = aggregated data = less valuable

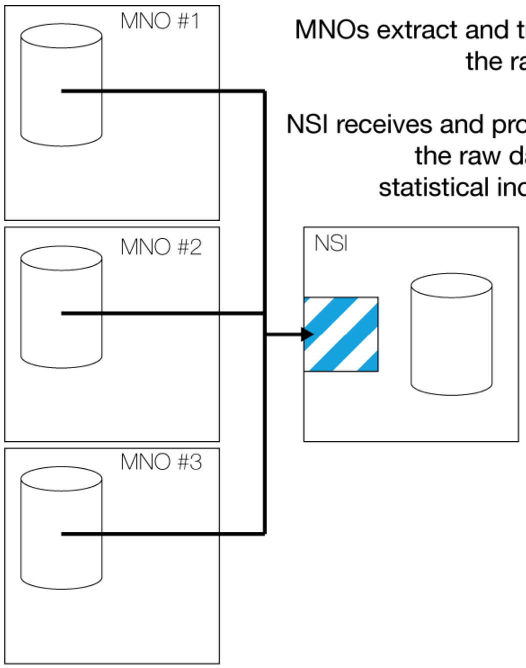
positium

Distributed vs Centralised Data Processing



MNOs process and transmit the result statistical indicators

NSI receives and combines statistical indicators



MNOs extract and transmit the raw data

NSI receives and processes the raw data into statistical indicators positium

CYBERNETICA

Processing personal microdata without direct intrusion to the personal data of the subscribers

Sharemind-Enabled Positium Data Mediator

sharemind

New Solution?

- Sharemind is a secure computing platform developed by Cybernetica
- Sharemind HI (Hardware Isolation) is based on Intel® Software Guard Extensions (SGX) technology
- Microdata encrypted by data controller (MNO), transmitted to NSO
- Raw data access impossible for the NSO
- Methodology and algorithms validated and approved by NSO, MNOs and third party (DPA) run on Sharemind HI and produce the expected validated, aggregated results
- MNO and third parties (DPA) can validate that data was processed exactly according to the pre-agreed algorithms and no personal data was extracted

How It Works? Step 1

Mobile Big Data

+

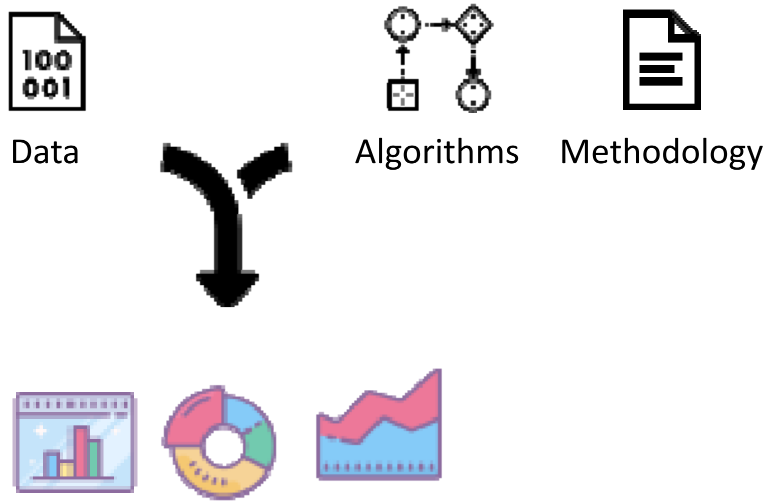
Methodology

+

Algorithms

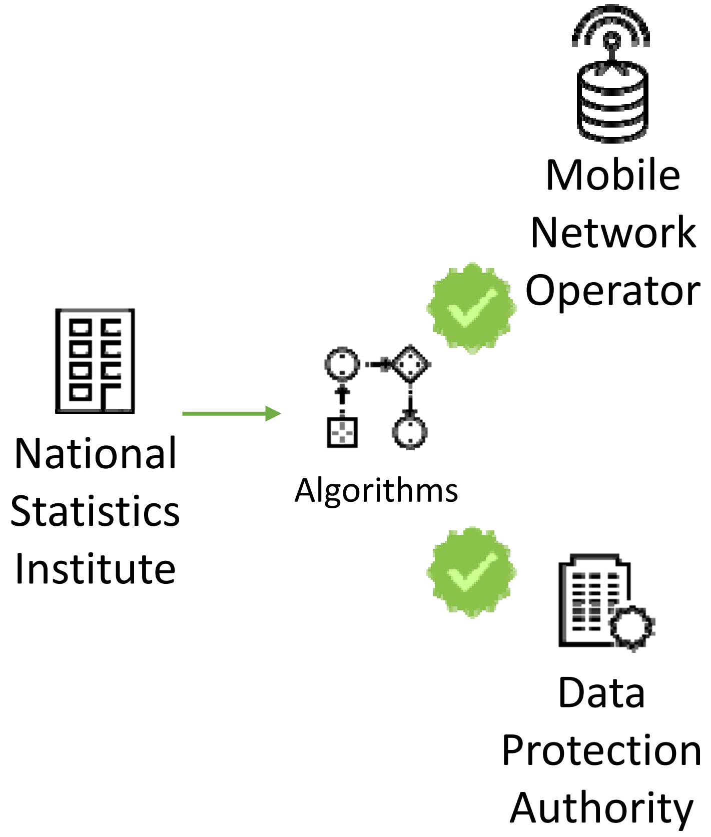
=

Valuable statistical indicators



How It Works? Step 2

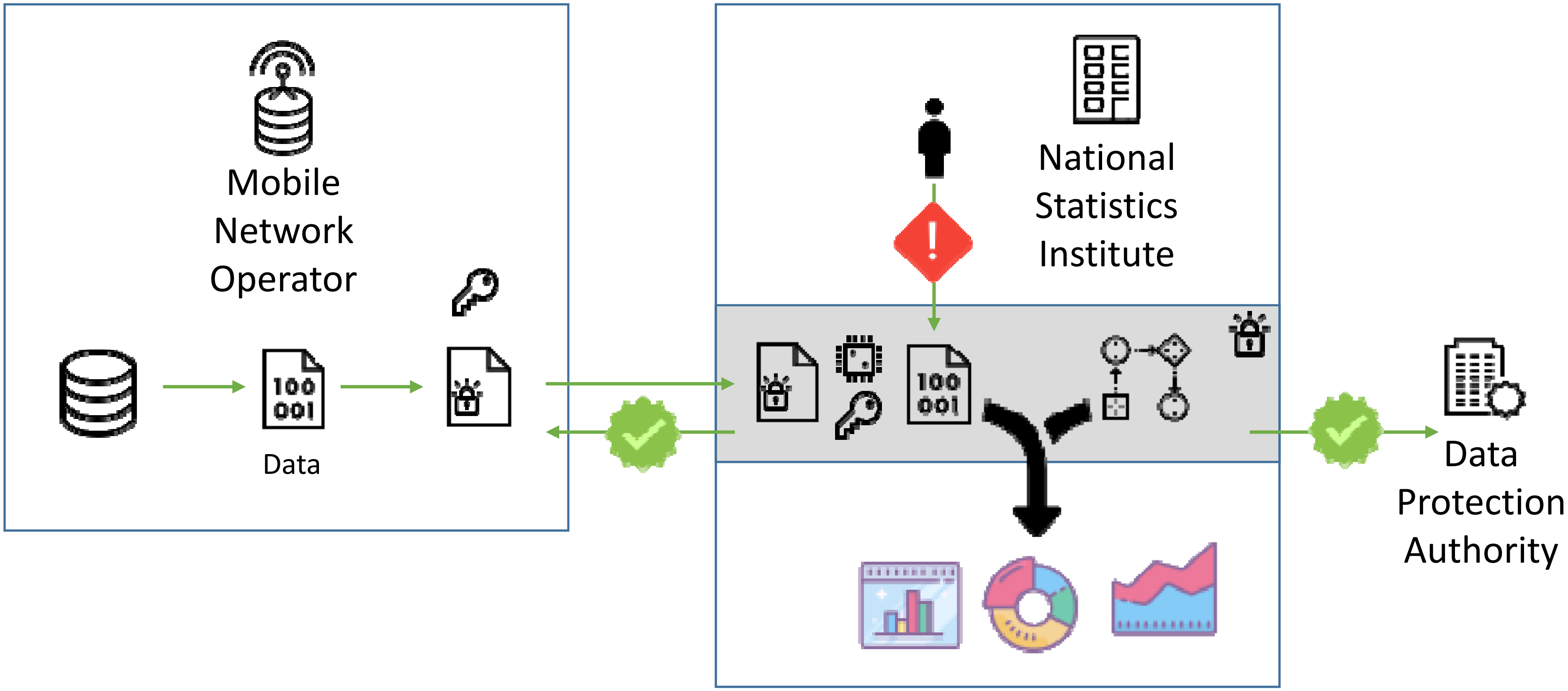
- NSI, MNOs, DPA agree on the algorithms and confirm that there are no privacy-intrusive algorithms there
- Output is only aggregated statistics ensured with statistical disclosure control methods



positium

CYBERNETICA

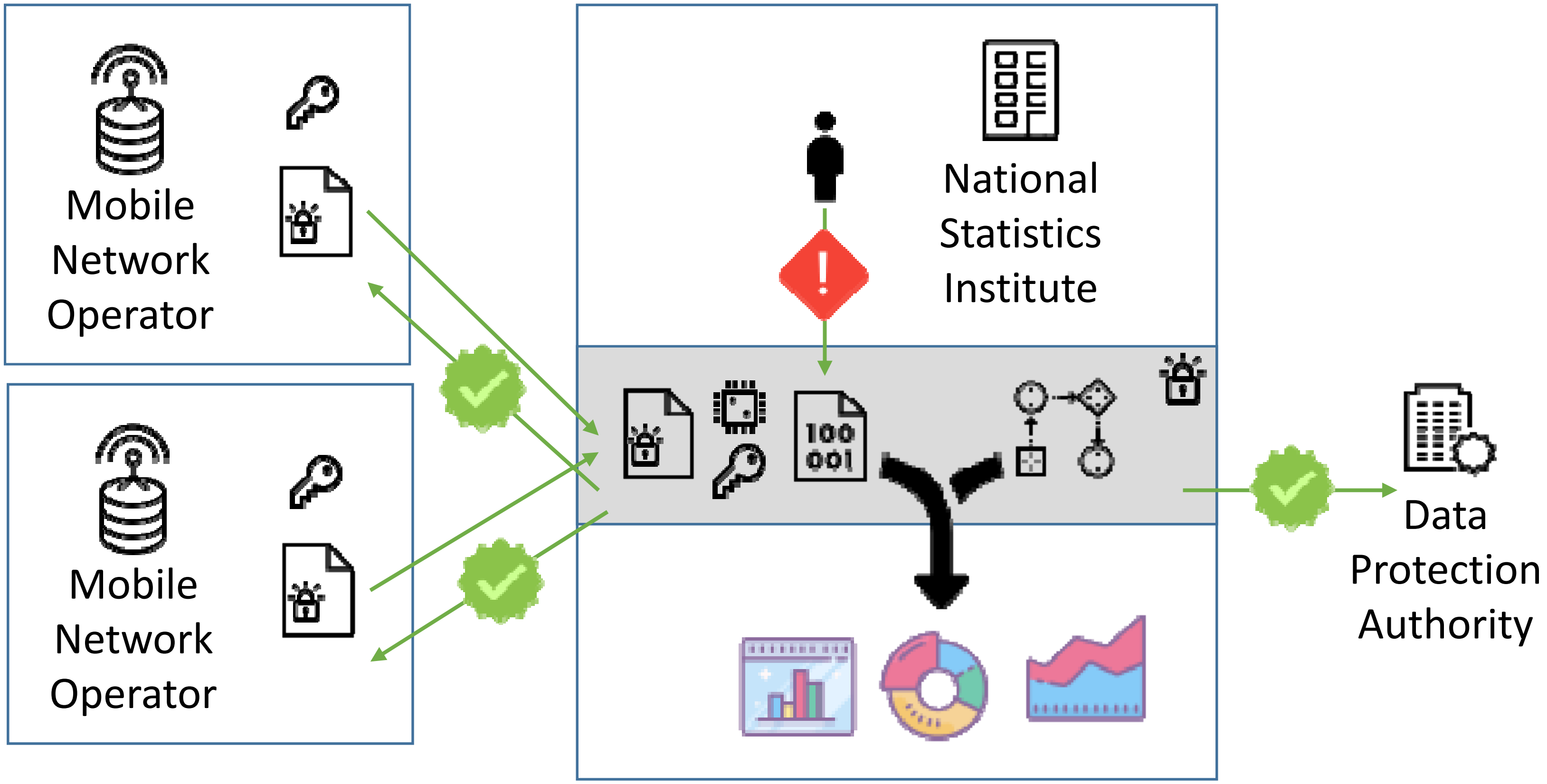
How It Works? Step 3 – One MNO



positium

CYBERNETICA

How It Works? Step 3 – Many MNOs

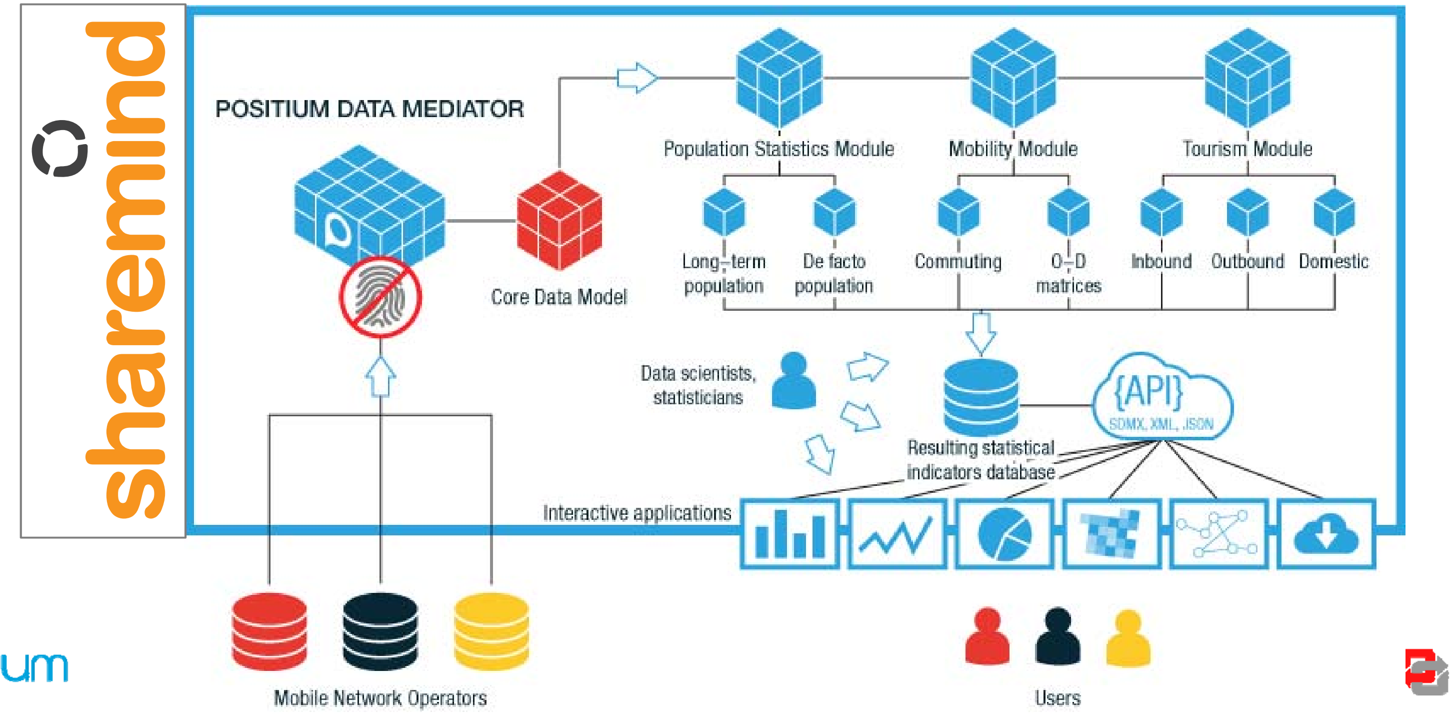


positium

CYBERNETICA

Sharemind HI Enabled Positium Data Mediator

- No loss of value of the data – results same as if processed with identifiable data
- Solution to data protection challenge



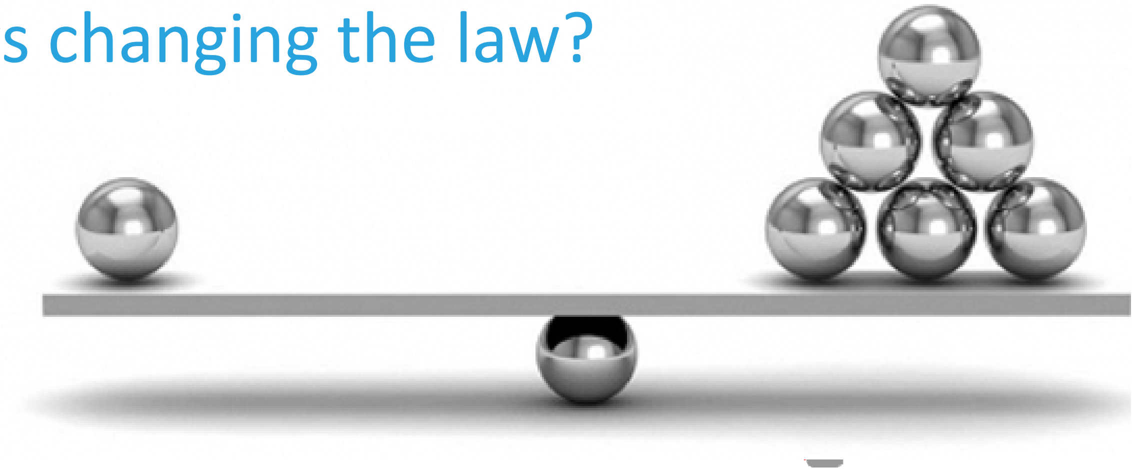
Challenges

- Validate the solution with Data Protection Authorities (where applicable)
- There are usually three issues that MNOs give as an excuse:
  - ~~Regulatory restrictions~~
  - ~~Too expensive to process in their infrastructure~~
- Test the solution on cloud infrastructure based on SGX (trials planned in early 2018)

Challenges

- NSI can’t “play” with microdata
  - But it is possible to request sample data from MNOs and play with that – no need to get access to all data
- Settle on new processes
  - E.g., are data files deleted after processing?

- Technical solution vs changing the law?



Thank You!

Margus Tiru

positium

margus.tiru@positium.com

Dan Bogdanov

CYBERNETICA

dan@cyber.ee