On the (re)use of Mobile Network Operator (MNO) data for official statistics: view and activities by Eurostat

Fabio Ricciato
Eurostat - Unit A5 'Methodology; Innovation in Official Statistics'

EASA Workshop on population density
6. October 2023

Eurostat and the ESS

- *Eurostat is …*
  - **the statistical office of the EU**
  - **a DG of the European Commission**
  - **the coordinator of the ESS**
- *The European Statistical System (ESS) is the partnership between*
  - Eurostat (coordinator)
  - National Statistical Institutes (NSIs) in each EU country
  - Other National Authorities (ONAs) in each EU country

- *Eurostat (i) produces European statistics and (ii) contributes to harmonise methodologies, definitions, criteria, etc. within the ESS*

European Commission

Official Statistics

- *"Official statistics" vs. "experimental statistics"*
  - Regular production vs. one-off/short series
  - Complete vs. partial fulfillment of quality criteria

- *Data sources for statistical production*
  - Census, surveys – designed and collected by NSI
  - Administrative data
  - "Big Data" (future)

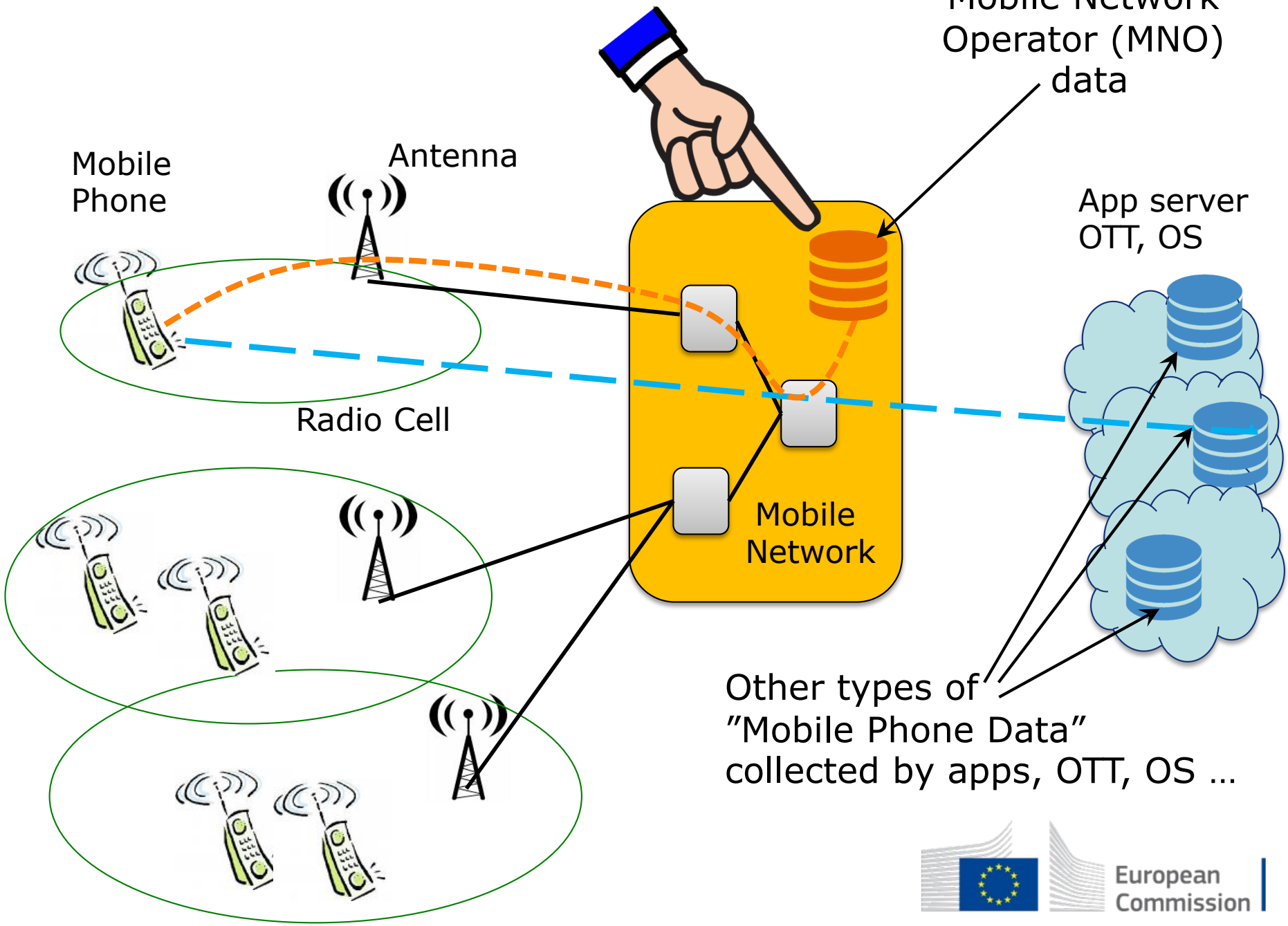
    Mobile Network Operator (MNO) data

    …

- *Goal: regular production of official statistics based on MNO data*

European Commission

What are MNO data?



European Commission

MNO data or MPD data?

- *'MNO data' are location data collected by the telecom operators*
  - Obtained by the "interaction" between the mobile device and the mobile network.
  - Collected primarily for billing (CDR) and network maintenance (**signalling**) purposes

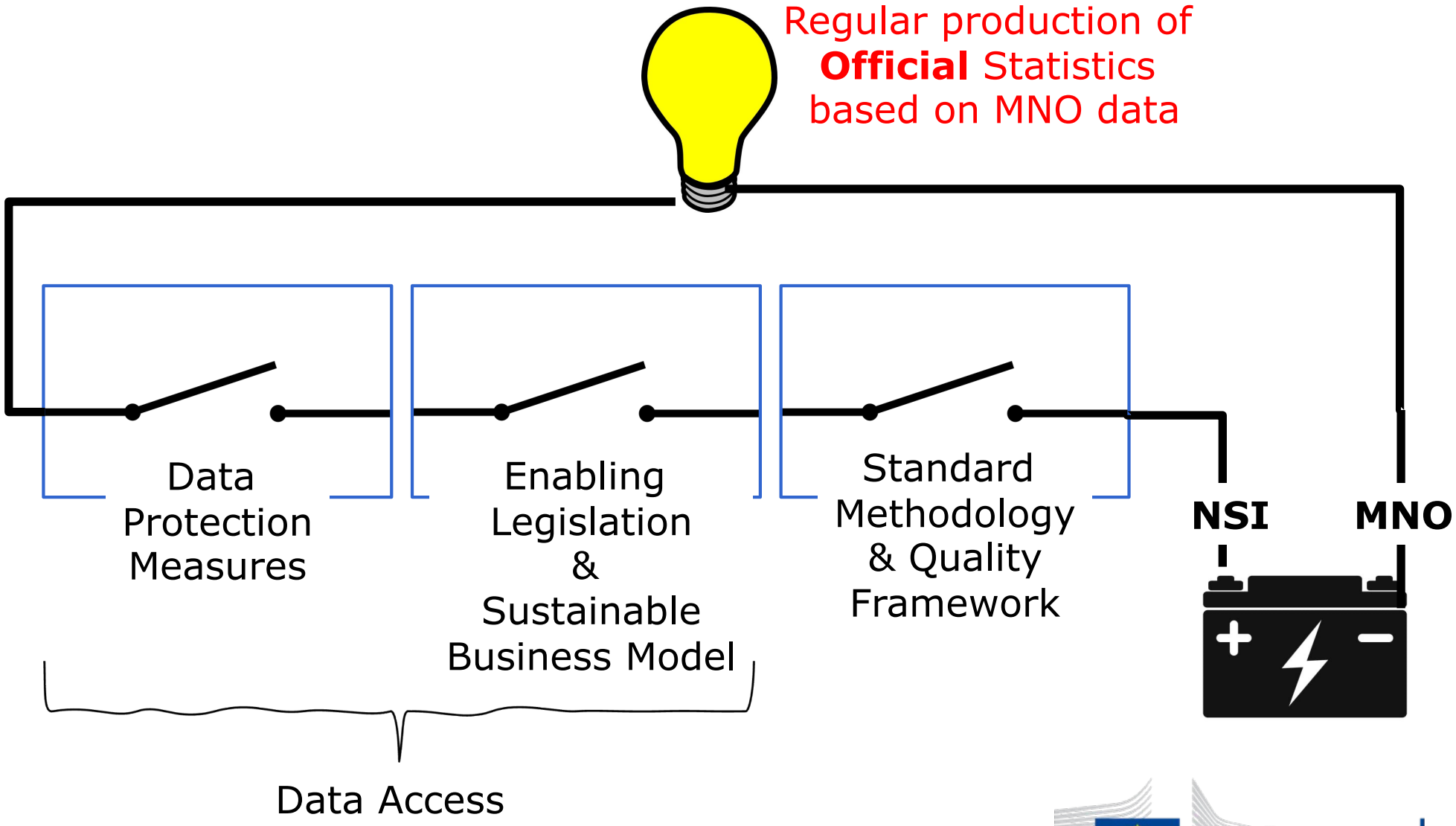
- *MNO or MPD?*
  - Location data produced by the mobile device and collected by apps, OS, Over-The-Top (OTT) companies are also called 'Mobile Phone Data" (MPD) --- but they are not "MNO data" and are not in the scope of this presentation

Why we are interested in MNO data?

- *'MNO data' contain spatio-temporal information for all mobile devices*
  - Record in the form **<phone ID, timestamp, cell ID>**
  - They can be used to extract statistics about human presence (where people are) and mobility (where they move to/from and when)
  - … taking "phones" as proxyies for "humans"

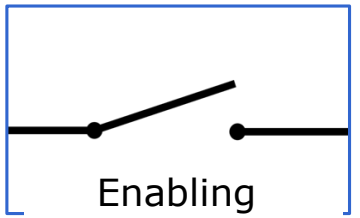
- *Appealing aspects of statistics based on MNO data*
  - Timeliness, temporal granularity, temporal continuity
  - Spatial coverage
  - Population coverage
- *Limitations and challenges*
  - Many …

Series of Challenges



European Commission

Data Access –ongoing actions



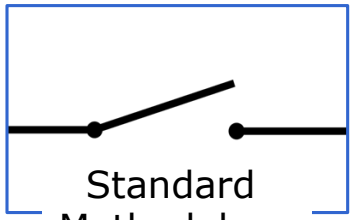
Enabling Legislation & Sustainable Business Model

- *Expert Group on facilitating the use of new data sources for official statistics (B2G4S)*
  - Final report with a set of recommendations, June 2022 https://ec.europa.eu/eurostat/en/web/products-statistical-reports/-/ks-ft-22-004
  - Gives an overview of the critical aspects of PHD, applicable also to MNO data

- *Legislation: EC proposal for amending Regulation N. 223/2009 on European statistics*
  - Relevant for all Privately-Held Data (PHD)
  - Target: adoption by 2024
  - Secondary legislation may be still needed afterwards for operational aspects

European Commission

Methodology



- ***ESS Task Force*** *on MNO data for Official Statistics (TF MNO)*
  - Established in 2021 with the official mandate to steer methodological development in the field
  - NSI representatives from 19 countries

- *Position paper by the TF MNO clarifying the methodological approach (2023)*
  - Reusing Mobile Network Operator data for Official Statistics: the case for a common methodological framework for the European Statistical System
  - https://ec.europa.eu/eurostat/en/web/products-statistical-reports/w/ks-ft-23-001

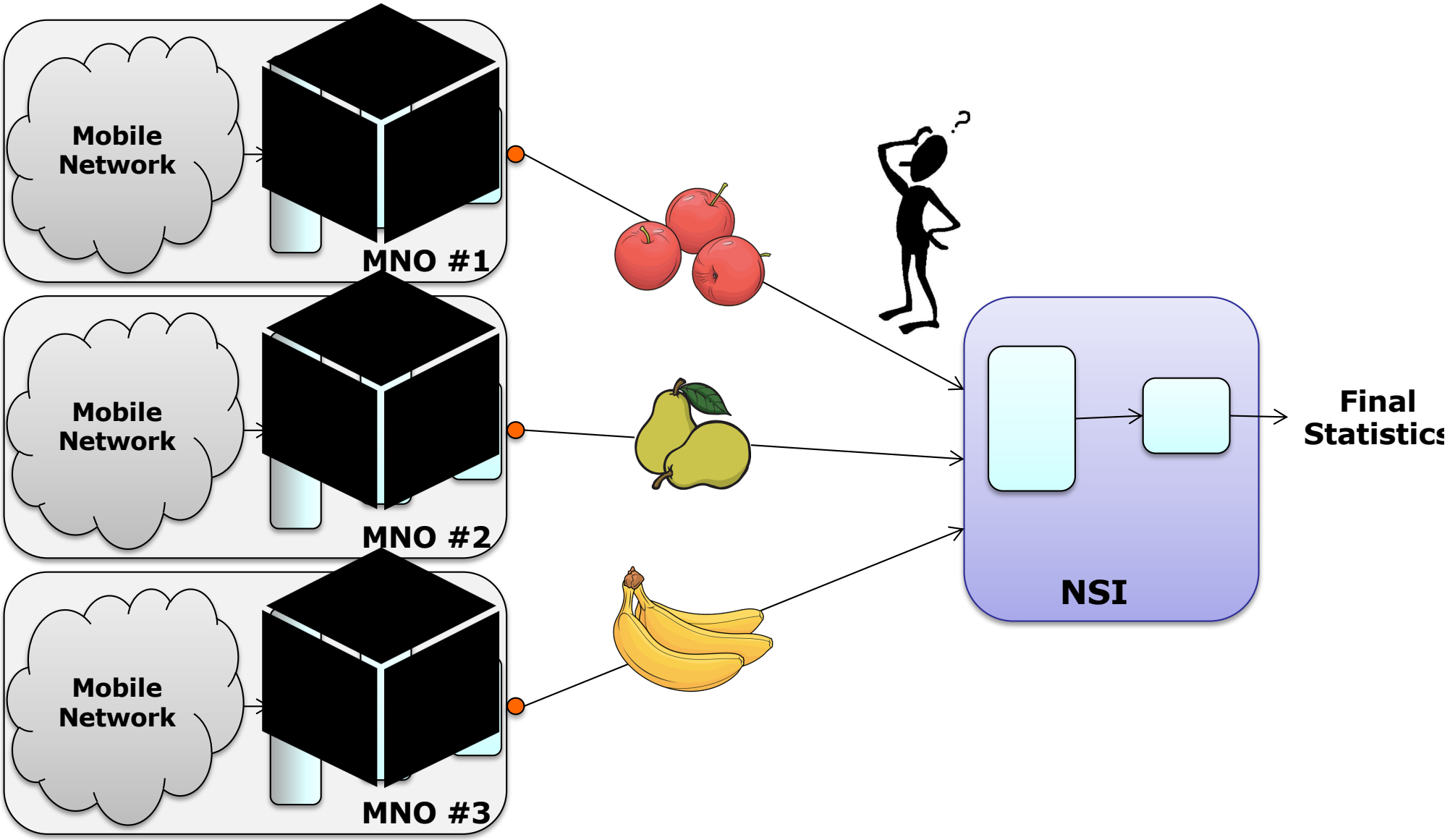
European Commission

Key points of the ESS view
(as expressed in the TF MNO position paper)

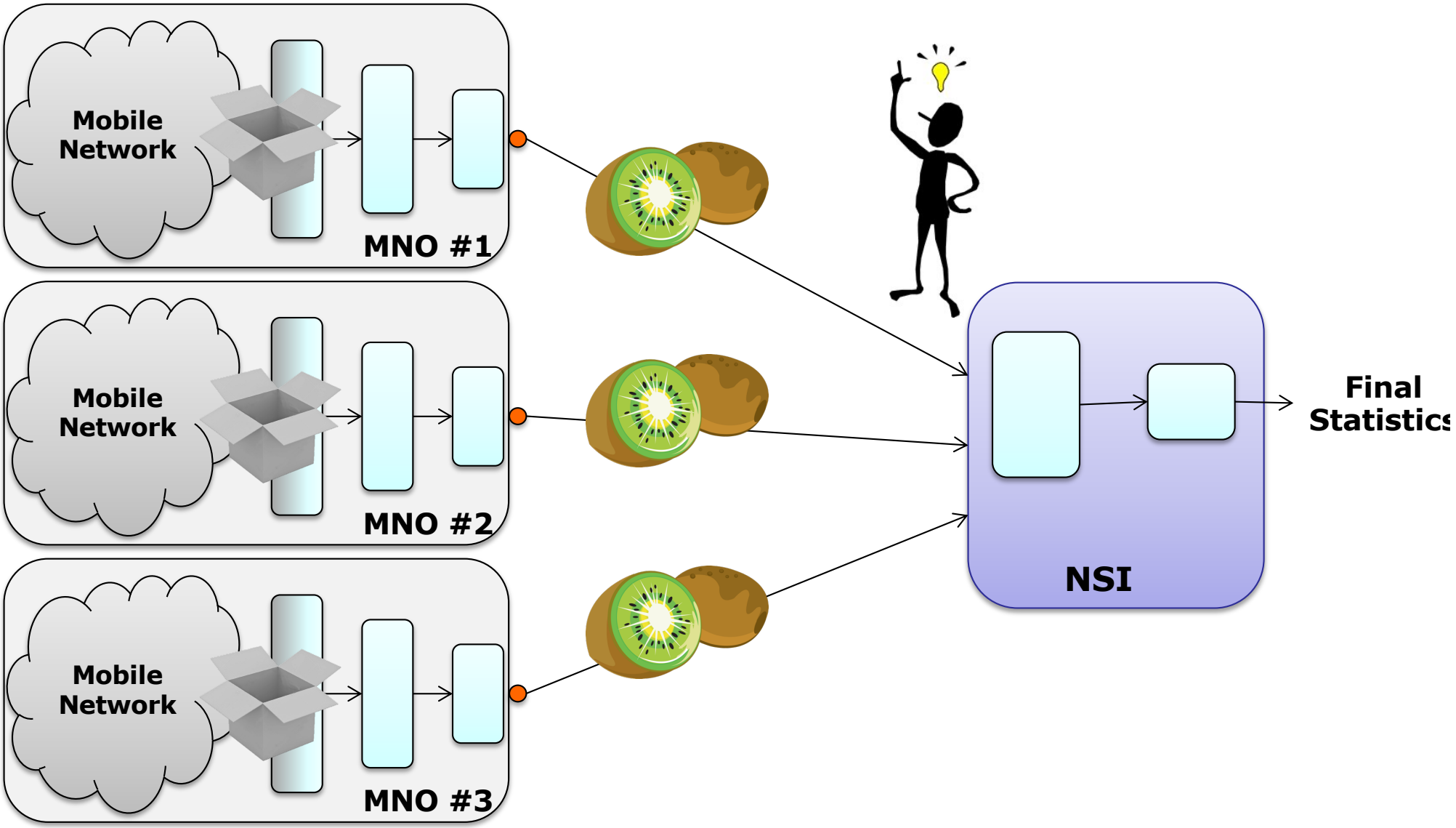
Go to page 17

European Commission

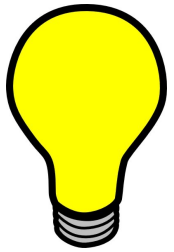
Black boxes not comparable, not composable



Common open standard methodology → comparable, composable



The vision



- *In 202x MNO data are (re)used for regular production of **official statistics***
  - Not merely "experimental statistics"…
  - Data from multiple MNOs in each country and across countries - **Multi-MNO**
  - Processed according to **standard methodologies** and transparent **quality criteria** defined at EU level (by the ESS in collaboration with industry)
  - Evolvable methodological framework
  - Processed at MNOs premises (with financial compensation)
  - Built-in privacy protection measures defined at EU level
  - Combined with non-MNO data for calibration/stabilisation/validation

Methodologies – why standard?

- *Methodologies for transforming raw MNO data into official statistics to be as standardised as possible across EU*
  - One open ESS methodology applied (possibly with some degree of adaptation) to all MNOs vs. many proprietary/closed methodologies
  - Standard cannot be too rigid: some degree of flexibility is needed to adapt to different MNO settings (raw data are NOT standardised!)

- *Comparable and combinable results*
- *Transparency and Evolvability – MNO networks evolve, methodologies must do too*

Steps towards there …

- *MultiMNO project - **co-development partnership** NSI - industry*
  - *Open end-to-end methodological framework + quality criteria*
  - *Open-source reference software pipeline implementing the proposed methodological framework;*
  - *Practical demonstration of the processing pipeline across 5x MNOs in 4x EU countries*
  - *Present population included in the target use-cases*
- *Project started in January'23 for 2.5 years, until mid-2025*
  - *Project consortium:*

    GOPA (Germany, consortium leader)

    2x Industry partners: NOMMON (Spain), POSITIUM (Estonia)

    2x NSI: CBS (Netherlands), ISTAT (Italy)

    5X MNOs: Orange Spain, Vodafone Spain, Vodafone Italy, A1 Slovenia, POST Lux.

Steps towards there …

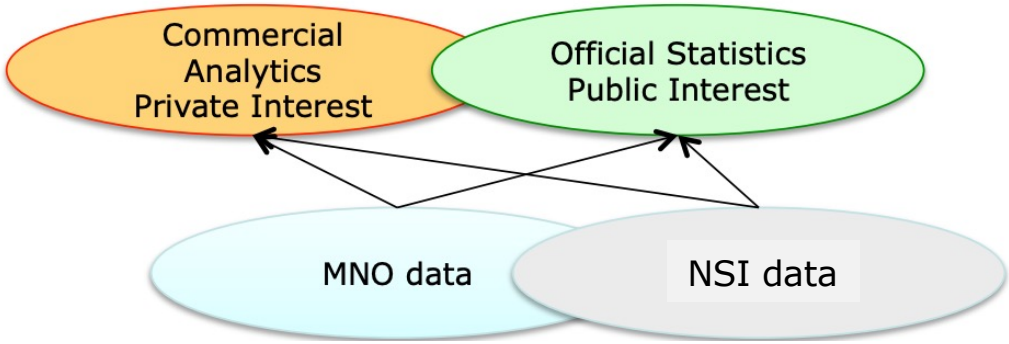
- *Rresearch Grant (ESSnet) on combination of MNO and non-MNO data*
  - ***Landscaping** of candidate non-MNO data sources with critical analysis of **costs-vs-benefits***
    - *Other Big Data sources, e.g., train tickets, event participants, flight passengers …*
    - *Census data, register data, existing surveys*
    - *New ad-hoc survey …*
  - *Developing formal methods for combination/fusion/calibration/validation*
  - *Designing and assessing the costs of new ad-hoc survey*

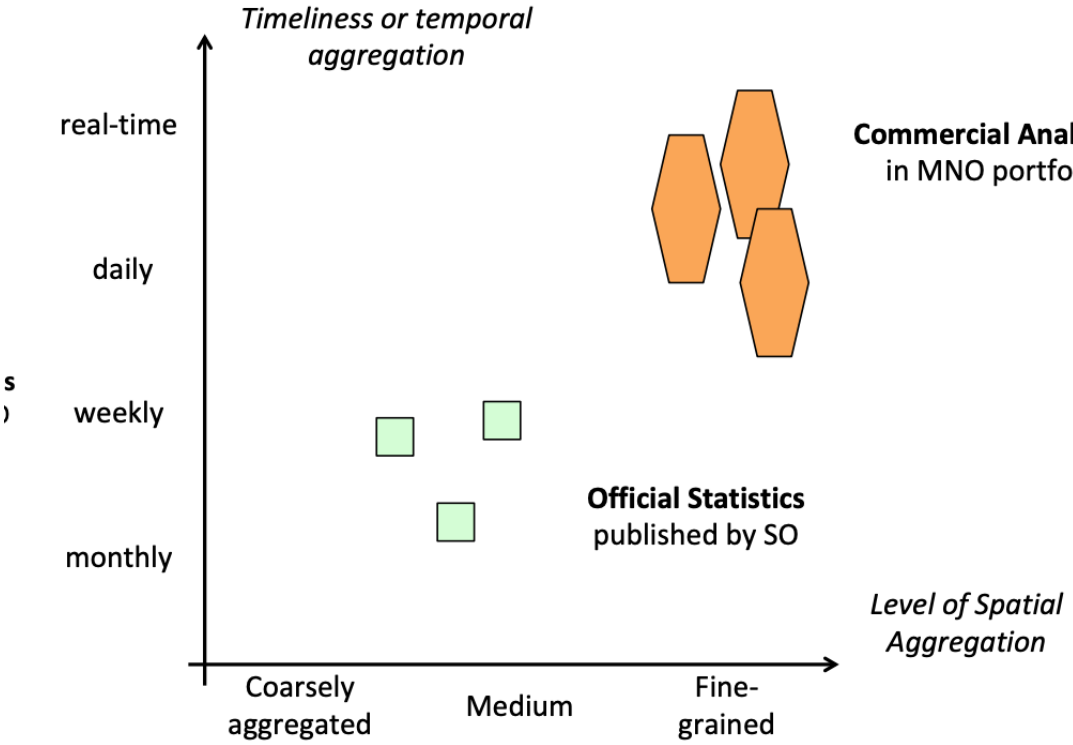
- *Project starting in Nov'23, duration 2 years, until end-2025*
- *10x NSI, coordinated by ISTAT, Italy*

European Commission

Official Statistics and Commercial Analytics

- *Partnership?*
- *Official statistics stimulating the market for commercial analytics?*





See: **Processing of Mobile Network Operator data for Official Statistics: the case for public-private partnerships**, DGINS 2018 conference https://cros-legacy.ec.europa.eu/system/files/dgins2018_mno-so_ricciato_0.pd

European Commission

Internal research work at Eurostat specifically on present population

On the Estimation of Spatial Density From Mobile Network Operator Data

Fabio Ricciato and Angelo Coluccia, *Senior Member, IEEE*

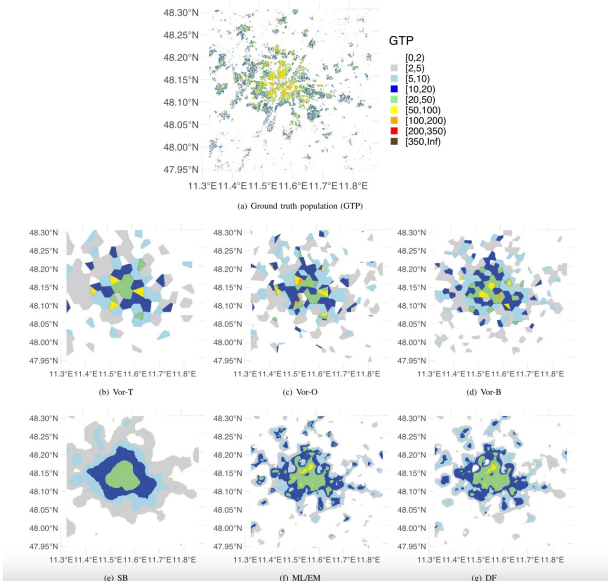
Abstract—We tackle the problem of estimating the spatial distribution of mobile phones from Mobile Network Operator (MNO) data namely Call Detail Record (CDR) or signalling data. The process of transforming MNO data to a density map requires geolocating r cells to determine their spatial footprint. Traditional geolocation solutions rely on Voronoi tessellations and approximate cell footprin by mutually disjoint regions. Recently, some pioneering work started to consider more elaborate geolocation methods with partially overlapping (non-disjoint) cell footprints coupled with a probabilistic model for phone-to-cell association. Estimating the spatial dens such a probabilistic setup is currently an open research problem and is the focus of the present work. We start by reviewing three dif estimation methods proposed in literature and provide novel analytical insights that unveil some key aspects of their mutual relation and properties. Furthermore, we develop a novel estimation approach for which a closed-form solution can be given. Numerical res based on semi-synthetic data are presented to assess the relative accuracy of each method. Our results indicate that the estimator based on overlapping cells have the potential to improve spatial accuracy over traditional approaches based on Voronoi tessellation

Index Terms—Mobile network data, call detail records, spatial density estimation, present population

1 Introduction

Most people nowadays carry a mobile phone. Mobile phones interact several times a day with the mobile network infrastructure, and every interaction reveals the approximate location of the phone to the network, at least at radio cell level. Such interactions are recorded by the Mobile

statistical products derived from MNO data in var cation domains, from humanitarian support to tourism flows, and statistical organizations are lo increasing interest at MNO data as a potential sour piling new official statistics [7], [8], [9]. However, great volume of research literature on the top methodological aspects remain open along th



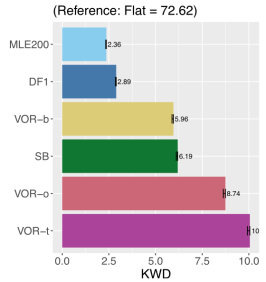


Fig. 8. Kantorovich-Wasserstein Distances (KWD) between the truth density and different estimates. The KWD values are expres tile units. As the length of a single tile is 100 me, the KWD value represents an average spatial error of 1 km.

https://ieeexplore.ieee.org/document/9647984

https://doi.org/10.1016/j.pmcj.2020.101263

Pervasive and Mobile Computing 68 (2020) 101263

Contents lists available at ScienceDirect

Pervasive and Mobile Computing

journal homepage: www.elsevier.com/locate/pmc

Towards a methodological framework for estimating present population density from mobile network operator data

Fabio Ricciato [a,*], Giampaolo Lanzieri [b], Albrecht Wirthmann [a], Gerdy Seynaeve [c]

[a] *Eurostat Unit B1 - Methodology; Innovation in Official Statistics, Luxembourg*
[b] *Eurostat Unit F2 - Population and migration, Luxembourg*
[c] *Proximus EBU Innovation, Belgium*

ARTICLE INFO

Article history:
Received 10 January 2020
Received in revised form 22 June 2020
Accepted 15 September 2020
Available online 19 September 2020

Keywords:
Mobile network operator data
Signalling data
Present population
Spatial density estimation
Experimental statistics

ABSTRACT

The concept of *present population* is gaining increasing attention in official statistics. One possible approach to measure present population exploits data collected by Mobile Network Operators (MNO), from simple Call Detail Records (CDR) to more informative and complex signalling records. Such data, collected primarily for network operation processes, can be repurposed to infer patterns of human mobility. Two decades of research literature have produced several case studies, mostly focused on to CDR data, and a variety of ad-hoc methodologies tailored to specific datasets. Moving beyond the stage of explorative research, the regular production of official statistics across different MNO requires a more systematic approach to methodological development. Towards this aim, Eurostat and other members of the European Statistical System are working towards the definition of a general Reference Methodological Framework for processing MNO

European Commission

References

- *Overview of recent Eurostat work in the field (since 2018) MNOdata4OS page(*)*

https://cros-legacy.ec.europa.eu/content/mobile-network-operator-data-official-statistics-mnodata4os_en

- *TF MNO Position Paper*

https://ec.europa.eu/eurostat/en/web/products-statistical-reports/w/ks-ft-23-001

- *MultiMNO project(*)*

https://cros-legacy.ec.europa.eu/content/multi-mno-project_en

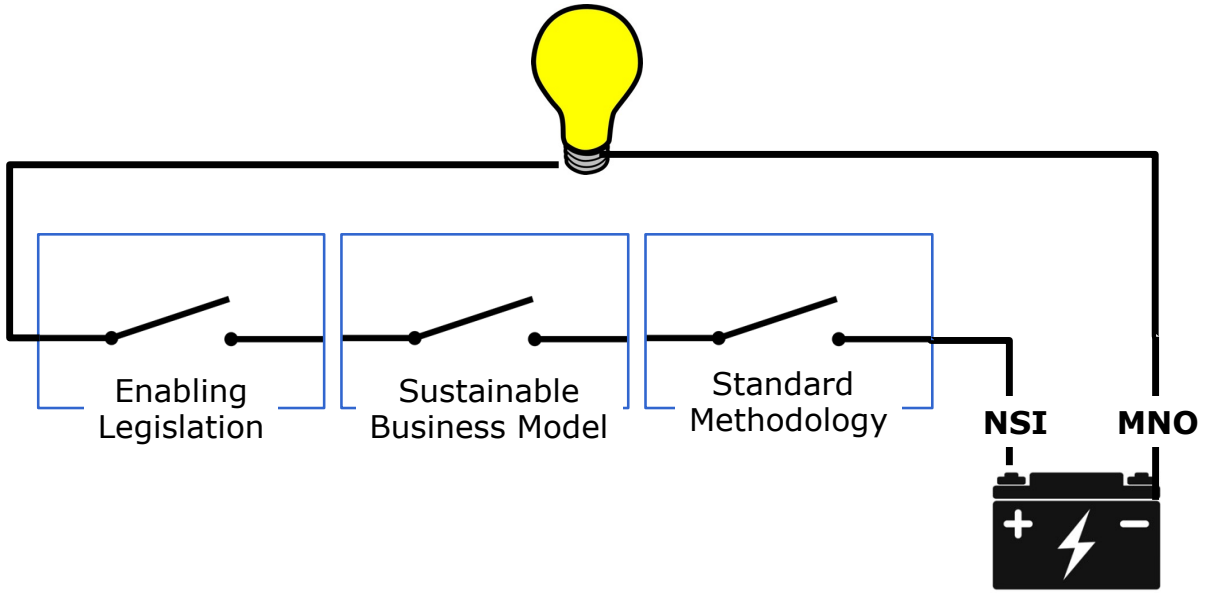
- *Scientific papers on present population estimation*

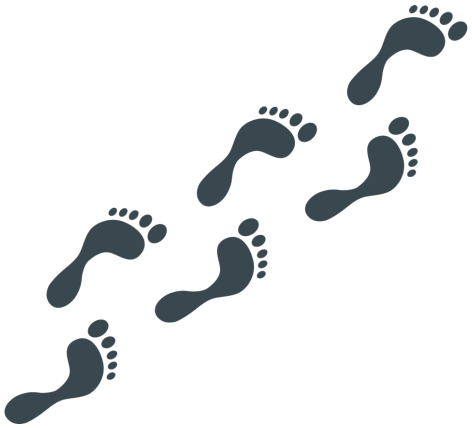
https://ieeexplore.ieee.org/document/9647984

https://doi.org/10.1016/j.pmcj.2020.101263

(*) Temporary page due to ongoing migration of Web platform for Eurostat research and collaboration activities

European Commission





Let's keep walking …

Thanks for your attention

For follow-up

Fabio.Ricciato@ec.europa.eu

European Commission