MNO MINDS | Eurostat CROS Portal
Project e-mail: tss-meth-too@istat.it

Integrating MNO Data into Official Statistics: challenges and innovation

Book of Abstracts of the MNO-MINDS Project Conference

23-25 September 2025
OECD Conference Centre room CC12, 2 rue André Pascal, Paris, France

Co-funded by the European Union

Contents

Conference objective 4

Overview of MNO-MINDS project 4

Venue 5

Wi-Fi access 6

Food 6

Conference Committees 7

Conference Programme 8

Keynote Session 14

Dr. Fabio Ricciato 14

Reusing Mobile Network Operator data for Official Statistics: a European perspective 14

Professor Li-Chun Zhang 14

Methods for integrating MNO and non-MNO data: insights from the MNO-MINDS project 14

Abstracts 16

Session 1: MNO and non-MNO data 16

Integration of MNO data with further sources: Results from a landscaping analysis to identify the most promising non-MNO data sources 16

Measuring telework at high spatiotemporal resolution: combining official statistics with MNO data 17

A comparative analysis of MNO and GPS data for spatiotemporal population assessment using commercial, off-the-shelf data 18

Session 2: MNO and Survey data 19

A Traditional Survey Design for Bias Mitigation in Mobile Network Operator Data 19

Mobile Network Operator Data as a New Source for Transport Planning and Mobility Analyses 20

Estimating monthly internal migration in Ghana and Haiti from CDR and survey data 21

Lost Without a Signal? A Roadmap for Using MNO Data in Tourism Statistics 22

Session 3: MNO data for official statistics, public-private partnership and quality framework 22

Legitimacy Tensions of Mobile Network Operator Data Sharing for Official Statistics: Lessons learned from the Netherlands and Finland 23

From metrics to meaning: How OECD data on connectivity services and infrastructures can inform policy making 24

Building a comprehensive quality framework for the use of Mobile Network Operator (MNO) data in Official Statistics: results from Multi-MNO project 25

From experimental to official migration statistics: A conceptual and actionable framework for quality assurance in mobile-network-operator-derived estimates 26

Session 4: MNO-MINDS project: Methods at work 27

Integration of Mobile Network Operator data, survey data and administrative data for tourism statistics: a state space approach 27

Flash Estimates of Regional Tourism Using Mobile Network Operator Data 28
Disaggregated trip statistics combining MNOdata and Travel Survey 29
Experiments of quasi-randomisation estimation for official statistics based on aggregated mobile network operator data 30
Session 5: Software and tools 31
Statistical open source software for integrating MNO and non-MNO data sources 31
An R Shiny Assistant for Visualizing Multi-Source Data and Supporting New Methodology Development in the application of Spanish tourism statistics 32
Metadata analysis for combining data sets with an application to MNO data 33
Open MNO Mobility Data in Spain: A Reproducible Sharing Approach via R package 'spanishoddata' 34
Session 6: Mobility and Transportation 36
Transport Planning based on MNO Data in Spain 36
Combining Mobile Network Data with Floating Car Data and Traditional Data Sources to Estimate Mobility Demand in High-Resolution: a Case Study in the Bologna's Metropolitan Area 37
From Mobile Phone Data to Mobility Patterns 38
Analysis of HSR foreign passenger at Firenze S. Maria Novella Station in 2024 39
Session 7: Exploring the use of MNO data in different fields 40
Mobility data assimilation to forecast electricity demand in a sobriety context in France 40
An approach to identify people trips and stops for mobility analysis using mobile network data 42
Spatial Interactions of Consumption, Presence, and Mobility: Evidence from Mobile Phone and Transaction Data 43
Harnessing geospatial innovation: Insights from the OECD Geospatial Lab 44
Session 8: MNO data for Tourism 45
Improving Tourism Measurement in Spain Using Mobile Network Operator Data 45
Exploring the Use of Mobile Network Data for Official Tourism Supply-Side Statistics in Italy. Vodafone-Istat Project Foundation 47
From signals to statistics: measuring same-day visits with MNO data. An Italian use case .. 48
Enhancing Tourism Analytics through Mobile Network Data: A Case Study of Venice and the Veneto Coast 49
Session 9: Population & migration 50
An application of the quasi-randomization approach to migration statistics derived from MNO data 50
Using Mobile Positioning Data for Validation and Signs-of-Life in a Register-Based Population Census 52
Measuring internal migration using mobile phone data: evidence from The Gambia 53
Enhanced Dasymetric Mapping of Mobile Phone Data: An EM Algorithm Approach with Deep Geographic Features 54

Conference objective

The conference focuses on the integration of Mobile Network Operator (MNO) data with other data sources to produce official statistics.

Topics of interest include, but are not limited to: *augmentation of official statistics by MNO data, methods to combine MNO and non-MNO data, use of MNO data in different domains (e.g., transport, tourism, population, regional statistics, commuting), methods to assess the quality of statistics produced from MNO data, open-source implementation, practical experiences in combining MNO data and statistical data.*

Overview of MNO-MINDS project

Within the framework of the Single Market Programme (SMP) of the European Union, Eurostat has declared the project MNO-MINDS as the winner of the SMP-ESS-2022-TSS-METH-TOOLS-IBA. The acronym MNO-MINDS stands for Mobile Network Operator Methods for Integrating New Data Sources. The project is led by Istat, as coordinator, together with nine other European National Statistical Institutes as partners (Austria, Germany, Spain, France, the Netherlands, Norway, Romania, Sweden and Portugal).

The project fits into the framework of the Trusted Smart Statistics Strategy and Roadmap adopted in 2019, by implementing the Bucharest memorandum on official statistics in a datafied society and moves its steps drawing on all the previous experiences on big data in the ESS.

Its objective is to propose new methodologies for the integration of Mobile Network Operator (MNO) data with other non-MNO data sources for the production of regular official statistics and foresees the production of guidelines and the provision of training courses to prepare the European Statistical System for the use of the new methodologies for such integration.

Started on 1 November 2023, the project MNO-MINDS will end on 31 October 2025.

The project activities are structured into four work-packages (WPs):

- WP1 - coordination, management, dissemination;
- WP2 - landscaping analysis of possible non-MNO data sources to be integrated with MNO data;
- WP3 - development of methodologies and open-source tools for integrating MNO and non-MNO data sources;
- WP4 - proof-of-concept of an ad-hoc survey to improve MNO data.

The project is co-funded by the European Union.

Venue

The Conference takes place at OECD Conference Centre Room CC12, 2 Rue André Pascal, Paris, France.

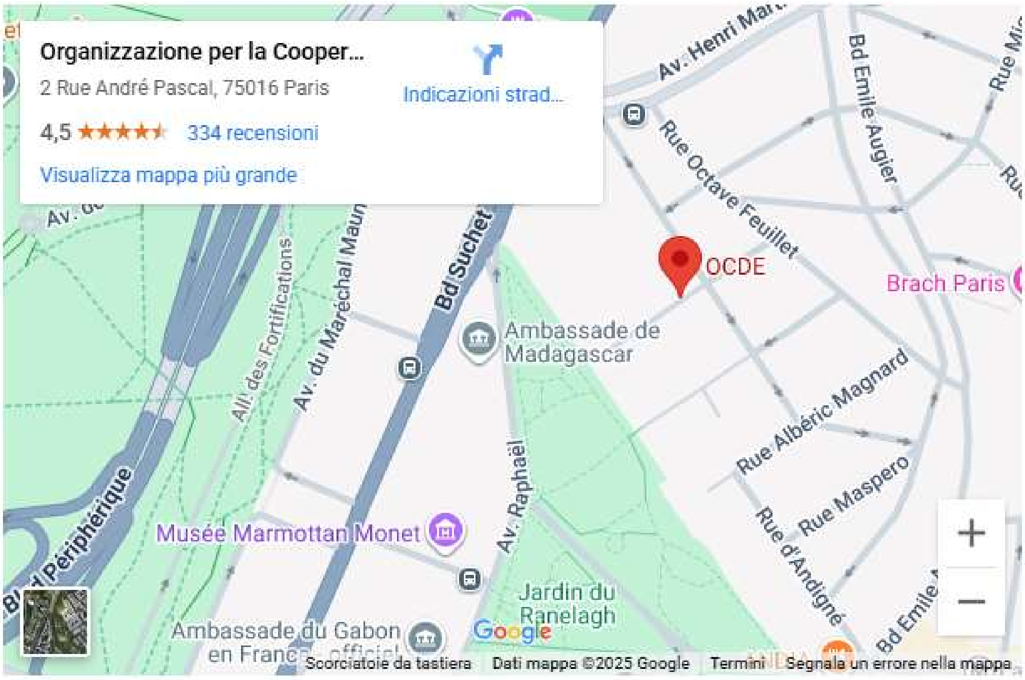
The OECD Room CC12 can be reached[1]:

Metro/RER
» Line 9, La Muette station (a 9 minute walk from the station)
» RER C, Avenue Henri Martin (a 7-minute walk) or Boulainvilliers station (11-minute walk)

Buses
» 22, 32, 52, 63, PC1

Bus stops
» La Muette-Boulainvilliers
» Octave feuillet
» Porte de La Muette

Bike rentals
An alternative way to get around Paris is by using the bicycle hiring option, more information can be found at Vélib-métropole.



On arrival at the OECD Paris Headquarters and Conference Centres, you will be required to register at the Welcome desk to obtain a visitor's badge, which will give you access to the OECD premises you will be visiting.

An email with a QR code will be sent to participants a few days before the event; they will need to present it together with their ID at the reception desk on arrival. They will then receive their badge; it will be valid for the entire duration of the event.

[1] https://www.oecd.org/en/about/plan-your-trip.html#:~:text=The%20OECD%20has%20two%20locations,from%20the%20centre%20of%20Paris

Accepted ID documents are as follows:

- **national identity card**
- **passport**
- **diplomatic card (issued by the French Ministry of Foreign Affairs)**
- **residence permit or card issued by the French Government**
- **French driving license (new format since end 2013)**

Please arrive at least 30 minutes before the start of your meeting or event, to allow time for registration.

For security reasons, visitors must wear their badges, clearly displayed at all times.

Wi-Fi access

The building is equipped with public Wi-Fi which is available for everyone.

Network: OECD-Public-WiFi
Password: NetAccess@6323

Food

Catering services in the OECD Conference Centre "La Muette"

- The Buffet du Parc restaurant is open from 12:00 to 14:00, offering seasonal lunch menus with locally sourced and organic produce, including vegetarian options;
- The Benugo cafeteria (also known as the Red Chairs café) is open from 08:00 to 17:00, offering a range of breakfast, lunch and snacks options as well as a variety of hot and cold beverages for purchase;
- The Conference Centre café features Benugo café and is open from 08:30 to 16:00. A selection of hot and cold beverages, breakfast, takeaway lunch options and snacks are available for purchase.

Local restaurants

Some interesting restaurants close to the conference venue are:

- Andia Paris
- Restaurant La Rotonde de la Muette à Paris
- Restaurant Bon - Best Restaurant Paris
- Le Bistrot des Vignes / Cuisine Française / PARIS
- Le Bois / Restaurant Bistronomique / PARIS
- Le Chalet XVI - Brasserie contemporaine dans le 16ème
- SETTEBELLO / Restaurant Gastronomique / Paris

More restaurants can be found at the following link Restaurants - Google Maps.

Conference Committees

The Scientific Program and Organising Committees are composed by the following members:

Scientific Program Committee

Name	Affiliation
Tiziana Tuoto (Chair)	ISTAT – Italy
Carlos Sáez Calvo	INE – Spain
Maria Grazia Calza	ISTAT – Italy
Pedro Cunha	INE – Portugal
Gloria Deetjen	DESTATIS – Germany
Johan Fosen	SSB – Norway
Gabrielle Gambuli	INSEE – France
Johannes Gussenbauer	STAT – Austria
Kira Gylling	SCB – Sweden
Remy Kamali	SCB – Sweden
Alexander Kowarik	STAT – Austria
Sónia Quaresma	INE – Portugal
Natalie Rosenski	DESTATIS – Germany
Nicolas Toulemonde	INSEE – France
Johan Van der Valk	CBS – The Netherlands

Organising Committee

Name	Affiliation	Role
Maria Grazia Calza	ISTAT – Italy	Co-chair Event Coordinator
Romain Lesur	INSEE – France	Co-chair Local Supervisor
Dongwook Choi	OECD – France	Local Assistant
Maria Francesca D’Ambrogio	ISTAT – Italy	Social Media and Web Manager
Virginie Elgrably	OECD – France	Local Assistant
Immacolata Fera	ISTAT – Italy	Relations and Communication
Leonora Lynch-Stein	OECD – France	Local Assistant
Valérie Mathey	INSEE – France	Local Assistant
Laura Tosco	ISTAT – Italy	Program and Editing
Tiziana Tuoto	ISTAT – Italy	Program Planner
Tamara Zangla	ISTAT – Italy	Event Coordination Assistant

Conference Programme

Conference

Co-funded by the European Union

OECD

ESSnet project

MNO-MINDS

Co-funded by the European Union

Integrating MNO Data into Official Statistics: challenges and innovation

Agenda

Day 1 - Tuesday 23/09/2025

OECD Conference Centre room CC12
2 rue André Pascal - Paris, France

13:00 – 14:00	**Registration and welcome coffee**
14:00 – 14:30	**Opening Session**
	Welcome address by Steve MacFeely – OECD Welcome address by Orietta Luzi – Istat Italy
14:30 – 15:30	**Keynote Session**
	Reusing Mobile Network Operator data for Official Statistics: a European perspective Fabio Ricciato – Eurostat Chair: **Orietta Luzi** – Istat Italy
15:30 – 16:30	**Session 1 – MNO and non-MNO data**
	1. Integration of MNO data with further sources: Results from a landscaping analysis to identify the most promising sources Gloria Deetjen – Destatis Germany Remy Kamali, Wictoria Widén – SCB Sweden | Mélina Hillion – Insee France Sónia Quaresma Gonçalves, Pedro Cunha – Statistics Portugal **2. Measuring telework at high spatiotemporal resolution: combining official statistics with MNO data** Gabrielle Gambuli – Insee France (SSPLab), Université Gustave Eiffel Etienne Côme – Université Gustave Eiffel | David Bounie, Chloé Breton – Télécom Paris John Galbraith – McGill University **3. A comparative analysis of MNO and GPS data for spatiotemporal population assessment using commercial, off-the-shelf data** Filipe Batista e Silva – European Commission Joint Research Centre, Ispra Michele Tucci, Sergio Freire, Lewis Dijkstra – European Commission Joint Research Centre, Ispra Chair: **Marcello D'Orazio** – Istat Italy

1

Co-funded by the European Union

ESSnet project MNO-MINDS

16:30 – 17:00	**Coffee break**
17:00 – 18:15	**Session 2 - MNO and Survey data**
	1. A Traditional Survey Design for Bias Mitigation in Mobile Network Operator Data Alexander Kowarik – Statistics Austria Loredana Di Consiglio, Tiziana Tuoto – Istat Italy | Gloria Deetjen, Maurice Brandt – Destatis Germany Remy Kamali – SCB Sweden **2. Mobile Network Operator Data as a New Source for Transport Planning and Mobility Analyses** Lorenz Ade – Destatis Germany Gloria Deetjen – Destatis Germany **3. Estimating monthly internal migration in Ghana and Haiti from CDR and survey data** Roland Hosner – Flowminder Foundation Zachary Strain-Fajth, Veronique Lefebvre – Flowminder Foundation **4. Lost Without a Signal? A Roadmap for Using MNO Data in Tourism Statistics** Sónia Quaresma Gonçalves – Statistics Portugal Pedro Cunha – Statistics Portugal Chair: **Gabriele Ascari** – Istat Italy

Agenda

Day 2 - Wednesday 24/09/2025

9:00 – 9:30	**Registration**
9:30 – 10:45	**Session 3 – MNO data for official statistics, public-private partnership and quality framework**
	1. Legitimacy Tensions of Mobile Network Operator Data Sharing for Official Statistics: Lessons learned from the Netherlands and Finland Sofie de Wilde de Ligny – Utrecht University Iryna Susha, Koen Frenken – Utrecht University **2. From metrics to meaning: How OECD data on connectivity services and infrastructures can inform policy making** Charles Laverdure – OECD, Connectivity Services and Infrastructures Unit (STI/DCES) Frédéric Bourassa, Lauren Crean, Alexia Gonzalez Fanfalone – OECD, Connectivity Services and Infrastructures Unit (STI/DCES) **3. Building a comprehensive quality framework for the use of Mobile Network Operator (MNO) data in Official Statistics: results from Multi-MNO project** Giorgia Simeoni – Istat Italy Gabriele Ascari, Erika Cerasti – Istat Italy **4. From experimental to official migration statistics: A conceptual and actionable framework for quality assurance in mobile-network-operator-derived estimates** Francisco Rowe – University of Liverpool Paul Baptiste Blanchard, Sveta Milusheva – The World Bank Chair: **Fabio Ricciato** – Eurostat

2

Co-funded by the European Union

10:45 – 11:15	**Coffee break**
11:15 – 12:30	**Keynote Session**
	Methods for integrating MNO and non-MNO data: insights from the MNO-MINDS project Li-Chun Zhang – SSB Norway & Southampton University Chair: **Tiziana Tuoto** – Istat Italy
12:30 – 14:00	**Lunch**
14:00 – 15:15	**Session 4 – MNO-MINDS project: Methods at work**
	1. Integration of Mobile Network Operator data, survey data and administrative data for tourism statistics: a state space approach Luis Sanguiao Sande – INE Spain **2. Flash Estimates of Regional Tourism Using Mobile Network Operator Data** Cristina Faricelli – Istat Italy Alessandro Piovani, Tiziana Tuoto – Istat Italy **3. Disaggregated trip statistics combining MNO data and Travel Survey** Li-Chun Zhang – SSB Norway & Southampton University **4. Experiments of quasi-randomisation estimation for official statistics based on aggregated mobile network operator data** Marcello D'Orazio – Istat Italy Li-Chun Zhang, Johan Fosen, Jens Haug – SSB Norway | Tiziana Tuoto – Istat Italy Chair: **Sónia Quaresma Gonçalves** – Statistics Portugal
15:15 – 16:30	**Session 5 – Software and tools**
	1. Statistical open source software for integrating MNO and non-MNO data sources Marcello D'Orazio – Istat Italy **2. An R Shiny Assistant for Visualizing Multi-Source Data and Supporting New Methodology Development in the application of Spanish tourism statistics** Jordi Verdú Naranjo – INE Spain Diego De la Puente Alonso – University of Valladolid **3. Metadata analysis for combining data sets with an application to MNO data** Yvonne Gootzen – CBS The Netherlands & Eindhoven University of Technology Johan Van der Valk – CBS The Netherlands **4. Open MNO Mobility Data in Spain: A Reproducible Sharing Approach via R package 'spanishoddata'** Egor Kotov – Max Planck Institute for Demographic Research, Universitat Pompeu Fabra Eugeni Vidal-Tortosa – Universitat Autònoma de Barcelona & University of Leeds Oliva Cantu-Ros, Javier Burrieza-Galán, Ricardo Herranz – Nommon Solutions and Technologies S.L. Tania Gullón Muñoz-Repiso – Ministry of Transport and Sustainable Mobility of Spain Robin Lovelace - University of Leeds Chair: **Mátyás Mészáros** – Eurostat
16:30 – 17:00	**Coffee break**

Co-funded by the European Union

ESSnet project MNO-MINDS

17:00 – 18:15	**Session 6 – Mobility and Transportation**
	1. Transport Planning based on MNO Data in Spain Tania Gullón Muñoz-Repiso – Ministry of Transport and Sustainable Mobility of Spain Ana Maldonado, César Rodriguez Cano, Pablo de la Presa – Ministry of Transport and Sustainable Mobility of Spain **2. Combining Mobile Network Data with Floating Car Data and Traditional Data Sources to Estimate Mobility Demand in High-Resolution: a Case Study in the Bologna Metropolitan Area** Daniele Di Antonio – GO-Mobility Giorgio Fiorillo, Eugenio Margelli – SRM - Reti e Mobilità Srl | Daniele Mancuso – GO-Mobility **3. From Mobile Phone Data to Mobility Patterns** Victor Tuekam – LMU Munich & ifo Institute Sebastian Wichert, Oliver Falck– ifo Institute | Göran Kauermann – LMU Munich **4. Analysis of HSR passengers at Firenze S. Maria Novella Station in 2024** Lorenzo Vannacci – FS Research Centre - Ferrovie dello Stato Italiane & Luigi Mancino – Fastweb + Vodafone Luigi Galieni, Chiara Bonvicini, Mario Tartaglia – FS Research Centre - Ferrovie dello Stato Italiane Bruno Zamengo, Alessandro De Cicco – Motion Analytica | Dario Di Sorte, Andrea Zaramella – Fastweb + Vodafone Chair: **Gloria Deetjen** – Destatis Germany

Agenda

Day 3 - Thursday 25/09/2025

8:30 – 9:00	**Registration**
9:00 – 10:15	**Session 7 – Exploring the use of MNO data in different fields**
	1. Mobility data assimilation to forecast electricity demand in a sobriety context in France Stefania Rubrichi – Orange Innovation Research, SENSE lab Nathan Doumeche, Yann Allioux, Yannig Goude – Électricité de France R&D **2. An approach to identify people trips and stops for mobility analysis using mobile network data** Chiara Bonvicini – FS Research Centre - Ferrovie dello Stato Italiane & Marco Agnolon – Motion Analytica Luigi Galieni, Lorenzo Vannacci, Mario Tartaglia – FS Research Centre - Ferrovie dello Stato Italiane Enrico Bontorin, Gianluca Busatta – Motion Analytica | Andrea Zaramella, Luigi Mancino, Dario Di Sorte - Fastweb + Vodafone **3. Spatial Interactions of Consumption, Presence, and Mobility: Evidence from Mobile Phone and Transaction Data** Chloé Breton – CREST & Insee France David Bounie – CREST, Telecom Paris | John Galbraith – McGill University Gabrielle Gambuli – Université Gustave Eiffel, Insee France **4. Harnessing geospatial innovation: Insights from the OECD Geospatial Lab** Claudia Baranzelli – OECD Alexandre Banquet, Josep Espasa Reig – OECD Chair: **Giorgia Simeoni** – Istat Italy
10:15 – 10:45	**Coffee break**

4

10:45 – 12:00	**Session 8 - MNO data for Tourism**
	1. Improving Tourism Measurement in Spain Using Mobile Network Operator Data Marta Sixto Neira – INE Spain Alfonso Fernandez Bes, María Velasco Gimeno, Ana Belén Marín Valverde – INE Spain **2. Exploring the Use of Mobile Network Data for Official Tourism Supply-Side Statistics in Italy Vodafone-Istat Project Foundation** Maria Teresa Santoro – Istat Italy Lorenzo Cavallo, Silvia Di Sante – Istat Italy **3. From signals to statistics: measuring same-day visits with MNO data. An Italian use case** Mascia Di Torrice – Istat Italy Barbara Dattilo, Mariangela Sabato – Istat Italy **4. Enhancing Tourism Analytics through Mobile Network Data: A Case Study of Venice and the Veneto Coast** Bruno Zamengo – Motion Analytica & Andrea Zaramella – Fastweb + Vodafone Luigi Mancino, Dario Di Sorte – Fastweb + Vodafone | Denis Cappellari – Motion Analytica Chair: **Alexander Kowarik** – Statistics Austria
12:00 – 13:15	**Session 9 - Population & migration**
	1. An application of the quasi-randomization approach to migration statistics derived from MNO data Paul Baptiste Blanchard – The World Bank Sveta Milusheva – The World Bank | Francisco Rowe – University of Liverpool **2. Using Mobile Positioning Data for Validation and Signs-of-Life in a Register-Based Population Census** Kaisa Vent – Positium OÜ Gerttu Pilsas, Erki Saluveer, Margus Tiru, Ivan Vasilyev, Oliver Bollverk – Positium OÜ Anto Aasa – University of Tartu Mobility Lab **3. Measuring internal migration using mobile phone data: evidence from The Gambia** Lamin L Dibba – Gambia Bureau of Statistics, GBoS Ayumi Arai – University of Tokyo, Reitaku University | Moritz Meyer – The World Bank Apichon Witayangkurn – Thammasat University | Tafsiru Loum – Public Utilities Regulatory Authority **4. Enhanced Dasymetric Mapping of Mobile Phone Data: An EM Algorithm Approach with Deep Geographic Features** David Crompton – Telcofy Alexey Sirotkin – Telcofy | Tor-Morten Grønli – Kristiania University of Applied Sciences Chair: **Johan Van der Valk** – CBS The Netherlands
13:15 – 13:30	**Closing Session**
	Closing Remarks **Maria Grazia Calza** and **Tiziana Tuoto** – Istat Italy **Corinne Prost** – Insee France **Fabio Ricciato** – Eurostat

5

Co-funded by the European Union

ESSnet project MNO-MINDS

Scientific Program Committee

Tiziana Tuoto (Chair)	ISTAT - Italy
Carlos Sáez Calvo	INE - Spain
Maria Grazia Calza	ISTAT- Italy
Pedro Cunha	INE - Portugal
Gloria Deetjen	DESTATIS - Germany
Johan Fosen	SSB - Norway
Gabrielle Gambuli	INSEE - France
Johannes Gussenbauer	STAT - Austria
Kira Gylling	SCB - Sweden
Remy Kamali	SCB - Sweden
Alexander Kowarik	STAT - Austria
Sónia Quaresma Gonçalves	INE - Portugal
Natalie Rosenski	DESTATIS - Germany
Nicolas Toulemonde	INSEE - France
Johan Van der Valk	CBS - The Netherlands

Organising Committee

Maria Grazia Calza (Co-Chair) Event Coordinator	ISTAT - Italy
Romain Lesur (Co-Chair) Local Supervisor	INSEE - France
Dongwook Choi Local Assistant	OECD - France
Maria Francesca D'Ambrogio Social Media and Web Management	ISTAT - Italy
Virginie Elgrably Local Assistant	OECD - France
Immacolata Fera Relations and Communication	ISTAT - Italy
Leonora Lynch-Stein Local Assistant	OECD - France
Valérie Mathey Local Assistant	INSEE - France
Laura Tosco Program and Editing	ISTAT - Italy
Tiziana Tuoto Program Planner	ISTAT - Italy
Tamara Zangla Event Coordination Assistant	ISTAT - Italy

Special thanks to

OECD

for hosting the event and making their facilities available for the conference

Final **Conference** of the **MNO-MINDS** project

cros.ec.europa.eu/mno-minds

tss-meth-too@istat.it

Istat · Statistik Austria · Destatis Statistisches Bundesamt · Instituto Nacional de Estadística · Insee · SCB · cbs · Statistisk sentralbyrå Statistics Norway · INS · Instituto Nacional de Estatística Statistics Portugal

Graphic design by M. Francesca D'Ambrogio

Keynote Session

Dr. Fabio Ricciato

European Commission – Eurostat, Big Data Task Force, Luxembourg

Reusing Mobile Network Operator data for Official Statistics: a European perspective

Keynote session Chair: Orietta Luzi - Istat

Since more than two decades, researchers have shown that the location data collected by Mobile Network Operators (MNO) primarily for network operation purposes may be (re)used to analyse mobility and presence patterns at the level of the entire population of mobile subscribers. Several MNOs, often in partnership with specialised startups, non-profit organisations or spin-off from research organisations, offer commercial statistics based on MNO data to public and private customers. More recently, statistical authorities worldwide have started to consider MNO data as a potential source for official statistics, and some of them have already released "experimental statistics" based on MNO data. The appetite for statistics based on MNO data boomed during the covid-19 crisis, due to increased demand by public authorities for timely information about presence and mobility at the level of the entire population, in support of designing and evaluating emergency response measures.

Moving beyond one-off research projects and "experimental statistics" towards regular production of official statistics requires statistical authorities to ensure target levels of quality and reliability. Towards this aim, sound and transparent methodologies are needed along with quality assurance mechanisms and business processes, underpinned by sustainable partnership models between statistical authorities and MNOs and with the highest data protection standards. These challenges are best addressed when statistical authorities join their forces, and towards this aim the European Statistical System (ESS) is developing a common European approach.

The talk will provide an overview of the challenges involved in the transition from "experimental statistics" to official statistics based on MNO data and the interdependencies that exist between methodological and non-methodological aspects (legal compliance, business models), with a focus on the European situation. The approach taken by the European Statistical System (ESS) to address these challenges will be presented along with highlights and key results from the recently completed Multi-MNO project, which complement the work done by the MNO-MINDS project, and a brief outline of future steps.

Professor Li-Chun Zhang

Southampton University & Statistics Norway

Methods for integrating MNO and non-MNO data: insights from the MNO-MINDS project

Keynote session chair: Tiziana Tuoto - Istat

Mobile network operator (MNO) data have great potentials for producing official statistics on population, tourism, mobility and environment. However, MNO data would not suffice on their own whenever the target statistical unit or the measurement unit is not mobile device per se.

Moreover, the Official Statistics Agency (OSA) by and large do not have access to MNO data at the device level, due to confidentiality, commercial interest and technology reasons. What is being made available to the OSA is anonymised macro data (see the TSS Multi-MNO project for reference processing pipelines of such data), aggregated over multiple devices, which nevertheless may be subject to measurement errors, population domain misclassifications, device duplication noises, user ambiguity and target population coverage errors.

In this talk, we present the developments and insights from the ESSnet project MNO-MINDS, regarding statistical methods for combining MNO macro data and other non-MNO sources in order to produce official statistics. Depending on how the associated uncertainty is defined, we distinguish three broad approaches.

In the randomisation approach, a specialised device usage survey is used to convert the MNO macro counts into the target statistical outputs. The uncertainty of the results is considered to be dominated by the survey sampling error.

In the quasi-randomisation modelling approach, MNO users in different subpopulations are treated as if they were randomly sampled from the respective subpopulations, which allows one to expand the MNO aggregates accordingly. The uncertainty is considered to be dominated by the observation process of the MNO data.

In the super-population modelling approach, MNO macro data are used as given features for target totals that are either known or can be estimated in non-MNO sources, in order to predict the other unknown target totals or to improve the existing target total estimates. The uncertainty of prediction here is due to the statistical relationship between the target totals and the given MNO features, rather than the observation process of the MNO macro data.

The key assumptions and practical challenges of the different approaches will be explained, as well as how the MNO macro data are used. Application scenarios will be used to illustrate various estimation methods under each of the three approaches.

Abstracts

Session 1: MNO and non-MNO data

Session chair: Marcello D'Orazio – Istat

Integration of MNO data with further sources: Results from a landscaping analysis to identify the most promising non-MNO data sources

Gloria Deetjen[1], Remy Kamali[2], Wictoria Widén[2], Mélina Hillion[3], Sónia Quaresma Gonçalves[4], Pedro Cunha[4]

Gloria.Deetjen@destatis.de, remy.kamali@scb.se, wictoria.widén@scb.se, melina.hillion@insee.fr, sonia.quaresma@ine.pt, pedro.cunha@ine.pt

[1]Destatis, [2]SCB Sweden, [3]Insee, [4]Statistics Portugal

Keywords: non-MNO data, data integration, landscaping analysis

Application scenarios with Mobile Network Operator (MNO) data for official statistics usually comprise integration with further (non-MNO) data sources because by combining these sources, more precise results can be achieved. For example, population statistics can be improved by higher temporal and spatial precision. Moreover, the scope of issues covered by official statistics can be broadened.

But non-MNO data that can potentially be linked with MNO data are heterogenous and can differ in e.g. granularity, availability and in the necessary resources to process non-MNO data. Therefore, linkage with MNO data come with various challenges and limitations but also opportunities.

With the help of a landscaping analysis, potential non-MNO data sources can be identified and each non-MNO data source can be examined for its data access, availability and quality as well as for its relevance to concrete application scenarios. A distinction between data types as well as whether sources are official statistics data or new digital data sources is useful to summarize the main aspects.

The analysis presented in this contribution consists of an assessment matrix that provides the base for further analysis and a primary scoring of sources. In addition, selected quality requirements and results from a short survey support collecting information on data availability at National Statistical Institutes in the European Statistical System.

Finally, a list of data sources for integration with MNO data is suggested and categorized in three levels: From the most promising sources, to promising sources that require substantial work or that are not widely available yet and to – at the present time – less relevant sources.

Measuring telework at high spatiotemporal resolution: combining official statistics with MNO data

Gabrielle Gambuli[1,2], Etienne Côme[2], Chloé Breton[3], David Bounie[3], John Galbraith[4]

gabrielle.gambuli@insee.fr, etienne.come@univ-eiffel.fr, chloe.breton@telecom-paris.fr, david.bounie@telecom-paris.fr, john.galbraith@mcgill.ca

[1]Insee (SSPLab) - France, [2]Université Gustave Eiffel, [3]Télécom Paris, [4]McGill University

Keywords: Telework; Mobile phone data integration into official statistics; High-Frequency Measurement; Municipality-Level Analysis

The COVID-19 pandemic has significantly reshaped work practices, notably through the expansion of teleworking. In France, the share of workers reporting at least one day of telework per week increased from 3% in 2017 to 20% in 2024 (Enquête Emploi). While national surveys provide estimates of telework prevalence over longer time horizons, there is a notable absence of data capturing daily telework intensity. No existing dataset or survey currently allows researchers to monitor day-to-day telework practice, at a fine spatial scale. This paper addresses this gap by proposing a novel methodology to estimate daily telework exposure at the municipality level, using a combination of labor force survey data, population census data, and anonymized mobile phone location data.

Mobile phone data reveal the number of residents present at home during working hours on weekdays, but they do not allow to distinguish between teleworkers and non-working individuals (e.g., inactives, unemployed, part-time workers on day off).

First, we estimate the number of teleworkers at the municipality level, by combining labor force survey data with census population data. Specifically, we compute telework rates by occupation and by location type (urban core, inner suburbs, outer suburbs) using the labor force survey. These rates are then applied to the employed population by occupation and municipality from the population census, yielding an estimate of the potential teleworking population in each area. In parallel, we estimate the number of part-time workers likely to be on their day off.

Then, we estimate daily working-from-home shares using an OLS approach, aligning predicted home presence with mobile phone-based home counts. Our estimates of daily telework intensity align well with national benchmarks: the weekly sum of daily shares equals approximately 2.4, matching the average number of teleworking days per week observed in Q4 2022 (Enquête Emploi). We find strong day-of-week patterns, with, for example, 24% of teleworkers working from home on Thursdays and 78% on Fridays (in Lyon FUA, September 2022). These patterns mirror those of part-time workers, though they are even more pronounced.

By combining these daily working-from-home patterns, telework propensities by occupation, and census-based population structures, we produce estimates of telework exposure for each municipality and day. We compute two indicators, distinguishing between the residence municipality (where additional presence occurs) and the workplace municipality (where absence occurs), reflecting the spatial redistribution of work-related presence due to teleworking. While we acknowledge that our estimates of telework behavior are not sufficiently robust to serve as official statistics, our approach nonetheless constitutes a meaningful advancement in approximating telework exposure by combining the best available data sources.

This methodology supports our broader research objective: assessing the impact of telework on local consumption, using bank card transaction data to measure in-store consumption at the municipality-date level. Our case study focuses on the Lyon Functional Urban Area (FUA), the second-largest FUA in France, in September 2022.

A comparative analysis of MNO and GPS data for spatiotemporal population assessment using commercial, off-the-shelf data

Filipe Batista e Silva[1], Michele Tucci[1], Sergio Freire[1], Lewis Dijkstra[1]

filipe.batista@ec.europa.eu, michele.tucci@ec.europa.eu, Sergio.FREIRE@ext.ec.europa.eu, Lewis.DIJKSTRA@ec.europa.eu

[1]European Commission Joint Research Centre, Ispra - Italy

Keywords: population, spatiotemporal, MNO, GPS, off-the-shelf, commercial, data, Europe

The spatial distribution of population is not static. Variations in population distribution over time and space occur as a result of constant mobility of population such as commuting to work or study, shopping, recreation, tourism, healthcare, visiting friends and relatives. Despite the substantial progress in the field of geospatial data and methods, knowledge of population distribution is very limited. It relies on conventional population maps based on place of residence. These maps are typically updated every 10 years, a lag that starkly contrasts with the fast-paced, increasingly digitalized world. Areas such as urban and infrastructure planning, disaster risk management, public security, would all benefit from timely and accurate spatiotemporal population assessments.

Mobile Network Operator (MNO) data has emerged as a promising but underutilized solution. While numerous academic studies have tested MNO data for one-off analyses, and statistical offices have conducted limited pilot projects, MNOs and resellers are now marketing such data, offering proprietary, closed-source solutions to the private sector. However, these still lack some requirements of official statistics, namely concerning the transparency and consistency of processes and outputs.

This short paper presents preliminary findings from a study comparing night-time population estimates from two distinct mobile phone data sources — MNO and GPS-based mobile phone data — and in relation to the 1 km population census grid 2021, using Spain as a test case. In addition, we examine the potential and limitations of GPS data vis-à-vis MNO data. Considering the higher spatial granularity of GPS data, as well as the access to a multi-year time series, we are exploring and testing the potential and robustness of this data source to assess: 1) diurnal-to-nocturnal population shifts, 2) annual persistence of night-time population proportions, 3) the temporal profiles of land uses and Points of Interest and relative affluence levels. The paper wraps-up with practical challenges in working with commercial data providers, including data quality, customization requirements, and lessons learned in adapting off-the-shelf datasets for policy-relevant applications.

These analyses are ongoing and are part of a broader European Commission Joint Research Centre and Eurostat project aiming to explore the potential of non-traditional data to improve population and housing statistics.

Session 2: MNO and Survey data

Session chair: Gabriele Ascari - Istat

A Traditional Survey Design for Bias Mitigation in Mobile Network Operator Data

Alexander Kowarik[1], Loredana Di Consiglio[2], Gloria Deetjen[3], Tiziana Tuoto[2], Maurice Brandt[3], Remy Kamali[4]

alexander.kowarik@statistik.gv.at, diconsig@istat.it , Gloria.Deetjen@destatis.de, tuoto@istat.it, Maurice.Brandt@destatis.de, remy.kamali@scb.se

[1]Statistics Austria, [2]Italian National Institute of Statistics – Istat, [3]DESTATIS – Germany, [4] SCB - Sweden

Keywords: Survey, MNO, questionnaire

We investigate whether we can create a purpose-built survey that can systematically reduce these biases found in mobile phone data. We develop and deploy an ad-hoc survey instrument to gather explicit information on the number of devices and SIM cards per respondent, discrepancies between contractual account holders and actual users, and demographic characteristics of usage behavior. We conducted a preliminary user-testing phase to assess the survey's clarity and effectiveness in capturing the desired information, and we report initial insights into device use patterns and user–contract holder distinctions in the not representative respondents of the friendly user test.

We further discuss methodological extensions, including area-based surveys, aimed at enhancing data representativeness. The results provide a foundation for further development of traditional surveys to be included into MNO-based statistical estimations, with the goal of improving accuracy and reducing systemic biases.

Mobile Network Operator Data as a New Source for Transport Planning and Mobility Analyses

Lorenz Ade[1], Gloria Deetjen[1]

Lorenz.Ade@destatis.de, gloria.deetjen@destatis.de

[1]Federal Statistical Office of Germany (Destatis)

Keywords: MNO data, transport planning, mobility analyses, survey, statistical simulation

Past research has shown that current methods of extrapolating Mobile Network Operator (MNO) data to the total population suffer from a lack of additional information to compensate existing structural differences between MNOs. This refers especially to the need for better information on the socio-demographic background of mobile phone users. To prepare an implementation of this data source in official statistics, these qualitative challenges need to be addressed. Building on the work of previous years, the current research project "NeDaMo" which is co-funded by the German Government aims to improve efficiency of extrapolation methods for MNO data, e.g. in order to map traffic networks more precisely. The Federal Statistical Office of Germany (Destatis) is leading the national research consortium of mobile network operators, data providers and universities for the project's duration of three years.

One of the main achievements will be to collect the abovementioned missing data by a large-scale survey conducted within the research project and the obtained information is linked with MNO data from two network operators: By analysing their different customer structures, it will be possible to gain important insights into the selectivity of data by region and by socio-demographic characteristics. Through combining survey data with traditional (official) data, MNO data and small area estimation methods, the research consortium aims to compensate the known quality limitations of mobility data as best as possible.

A second strand of research involves the use of statistical simulation in the context of MNO data to investigate how the consistency of complex traffic forecasts can be improved. Microsimulations are applied to investigate how sample information can contribute to a significant improvement in the estimation quality of the resident population and mobility behaviour at regional level using MNO data. To this end, findings from the aforementioned dedicated survey will help to optimise simulations, which in turn can benefit the design of future sampling concepts. Furthermore, methodological and practical questions regarding the optimisation of existing large-scale mobility surveys in Germany are integrated as well into the research project.

Lastly, the findings from the large-scale survey, the resulting analyses of distortions and the possibilities for compensation become altogether the basis for the development of a quality concept for traffic data that are based on MNO data. On the one hand, this concept analyses the pre-processing of (raw) MNO data by data providers. On the other hand, the limits within which MNO data-supported traffic data can be used for different transport planning issues are described in detail. As a result, the consortium will provide regionally detailed origin-destination matrices that are adjusted for distortions and selectivities as far as possible, having leveraged the results of the large-scale survey and the comparative calculations for both mobile network operators. These quality-optimised and quality-assured origin-destination matrices – segregated by mode of transport – are finally made available as part of a pilot application for mobility simulations and forecasts in selected regions.

Estimating monthly internal migration in Ghana and Haiti from CDR and survey data

Roland Hosner[1], Zachary Strain-Fajth[1], Veronique Lefebvre[1],

roland.hosner@flowminder.org, zachary.strain@flowminder.org, veronique.lefebvre@flowminder.org,

[1]Flowminder Foundation

Keywords: Call Detail Records, internal migration, data integration

While population censuses, register data and large-scale household surveys continue to be the indispensable cornerstones of population statistics, the advent and use of so-called “Big Data” from sources such as social media, satellite imagery or mobile phone data has opened up discussions both around the analytical and statistical potential of such data, as well as debates around the methodological implications and limitations of using these data.

Some of the key challenges associated with the use of Big Data revolve around the non-random nature of the data generation processes. Individuals leaving data footprints online or through the use of their mobile devices do usually not represent the general population, and the resulting selection bias does fundamentally affect the validity and the usefulness of insights derived from such data (Meng 2018). What is more, assessing and quantifying these biases is usually only possible when corresponding ground-truth data or standard estimates exist, yet the potential of these “Big Data” is often cited in contexts where existing data and estimates are lacking or even absent.

Our early work focused on analyses of Call Detail Records - the event data automatically generated by mobile network operators for billing purposes whenever mobile phone users make calls, send texts or use data - for humanitarian and development use cases (for example Bengtsson et al. 2011, Bengtsson et al 2015, Lai et al. 2019). These efforts have laid the groundwork for the work in recent years which has started to integrate multiple types and sources of data (notably CDR, survey and census data) for the production of estimates in the fields of travel, mobility, migration and population statistics.

Based on the emerging discussion around integrating data from Big Data sources and other traditional sources such as surveys (Rao 2021, Zhang et al. 2024), we have identified the Data Integration Estimator suggested by Kim and Tam (2021) as a viable approach to provide blended estimates of internal mobility within countries, derived from CDR and survey data.

We draw from unique data sources in Ghana and Haiti. For Ghana, we commissioned a dedicated telephone survey in 2022 with a view to derive parameters that can be used for mobility and population statistics from CDR data. The CATI telephone survey in Ghana was implemented in November/December 2022 and included a range of questions on mobility and phone use, from detailed questions on the current place of residence and changes of place of residence from month to month to mobile phone usage questions, for example on the use of multiple SIM cards or duration and frequency of usage. The dedicated survey data we collected in Haiti in 2023 comprise a range of mobility and phone use indicators, including detailed questions on the current place of residence and changes of place of residence from month to month.

We will analyse and show estimates based on the Data Integration estimator, and contrast them with estimates from CDR data only and survey data only. In addition, we will attempt to run and show the same estimates on the admin2 level - districts in Ghana, communes in Haiti - for admin2 areas with sufficient sub-sample size.

Lost Without a Signal? A Roadmap for Using MNO Data in Tourism Statistics

Sónia Quaresma Gonçalves[1], Pedro Cunha[1]

sonia.quaresma@ine.pt, pedro.cunha@ine.pt

[1]Statistics Portugal

Keywords: MNO Data, Tourism Statistics, Total Error Framework

Making effective use of Mobile Network Operator (MNO) data for tourism statistics can feel like navigating an unfamiliar terrain, full of potential but also methodological uncertainty. To support this journey, we propose a roadmap that guides National Statistical Institutes (NSIs) through the essential phases of applying a Total Error Framework to MNO-based statistics, structured around six foundational elements:

1.Target Statistics – defining what we aim to measure, such as the number of visitors or nights spent;

2. Statistical Unit – specifying who or what is being measured (person, trip, SIM, etc.);

3. Data Configuration – understanding the structure and granularity of available data, including macro vs. micro formats and temporal resolution;

4. Nature of MNO Data – clarifying whether the MNO data serves as the primary source or plays an auxiliary role in estimation;

5. Methods – identifying and applying appropriate methodologies and corrective techniques;

6. Total Error Analysis – systematically assessing, quantifying, and addressing the most relevant sources of error in light of the specific target statistics.

The process begins with a clear destination: defining the target statistics (Step 1). What exactly are we trying to measure: visitor flows, overnight stays, domestic vs. international movement? From there, we turn to defining the statistical unit and population of interest (Step 2), ensuring that we know who or what is being counted: people, SIMs, trips? Equally important is the role of MNO data itself (Step 3) is it the main source driving the estimates or an auxiliary source supporting other data pipelines?

These early decisions shape the next phase, where the focus shifts to the data context and structure. We begin by mapping the observation units and available data sources (Step 4), MNO and non-MNO alike (e.g., accommodation records, border surveys), and then characterizing the configuration of these data (Step 5) macro vs. micro, spatial and temporal resolution, frequency of updates.

The third phase addresses the heart of the Total Error Framework: error identification and mitigation. We start by identifying and classifying potential sources of error (Step 6), then quantifying the most critical ones, where feasible (Step 7). With a better understanding of the error landscape, we can select or develop appropriate correction methods (Step 8), from reweighting and imputation to model-based adjustments. Finally, we validate the results and review assumptions (Step 9), ensuring transparency, replicability, and relevance to the original statistical objectives.

Together, these steps offer a structured way to think through the complexity of MNO data use, helping statistical institutions determine not only what to do, but when and why. By situating this process within a clearly defined conceptual frame, the roadmap helps turn a technically demanding process into a coherent and navigable workflow.

Session 3: MNO data for official statistics, public-private partnership and quality framework

Session chair: Fabio Ricciato – Eurostat

Legitimacy Tensions of Mobile Network Operator Data Sharing for Official Statistics: Lessons learned from the Netherlands and Finland

Sofie de Wilde de Ligny[1], Iryna Susha[1], Koen Frenken[1]

s.e.dewildedeligny@uu.nl

[1]Utrecht University

Keywords: business-to-government data sharing, B2G, B2G4S, big data for statistics, mobile network operator (MNO) data, legitimacy, data partnerships

National Statistical Offices (NSOs) show a growing interest in the use of privately held data (e.g., scanner data, satellite data, financial transaction data) to enhance the timeliness and granularity of official statistics, but also to explore new statistical opportunities (e.g., Kitchin, 2015). As a result, partnerships with Mobile Network Operators (MNOs) have emerged as a prominent model within Business-to-Government Data Sharing for Statistics (B2G4S) (Vespe et al., 2021, Eurostat, 2022). While these partnerships promise new and timely statistical opportunities (e.g., tourism in and outbound, population measurements), they also raise concerns related to privacy, legal compliance, and public trust (Rahmandian et al., 2023). In this context, legitimacy - understood as the perceived appropriateness and justification of certain practices - becomes a key analytical lens to understand the tensions of such partnerships.

As such, this research investigates the question: How do different assessments of legitimacy of MNO data sharing partnerships unfold in B2G4S over time, and how do they shape the dynamics of the partnerships? To explore this, we analyze case studies of two NSOs in the EU engaging with MNOs in data sharing partnerships, one in the Netherlands (CBS) and one in Finland (Tilastokeskus). Subsequently, we draw on semi-structured interviews, document analysis, and participant observation in project activities, aiming to capture the legitimacy assessments across various audiences, such as NSOs, businesses, governmental institutions, and media.

Our analysis draws on institutional theory to conceptualize legitimacy in data partnerships. Building upon the work of Deephouse et al. (2017, 32), we understand legitimacy in B2G4S data sharing as the perceived appropriateness of a data sharing practice to a social system in terms of rules, values, norms, and definition. Hence, we understand legitimacy tensions as contrasting assessments of legitimacy by various assessors within a B2G4S data sharing practice. These assessments are not static; they evolve through aspects such as temporal conditions (e.g., COVID-19), institutional decisions (e.g., governance structures), changes in regulatory frameworks (e.g., GDPR), and public controversies (e.g., location data ethics). We show that the different legitimacy assessments are not just something that needs to be taken into account while establishing data partnerships; it is something that develops and changes over time, and it actively influences how these partnerships are set up, managed, and whether they become stable or eventually fall apart.

By unpacking how legitimacy assessments unfold over time and how they shape the dynamics of partnerships, this study contributes to a deeper understanding of the social and institutional tensions inherent in B2G4S initiatives. The findings highlight the importance of building legitimacy not only through compliance or performance but through continuous negotiation among various audiences in a social system with diverging values and mandates. In doing so, this study contributes to both an academic and practical level: it offers a framework for analyzing the various assessments of legitimacy in data partnerships in B2G4S, and it provides actionable insights for NSOs seeking to navigate the evolving landscape of data sharing practices.

From metrics to meaning: How OECD data on connectivity services and infrastructures can inform policy making

Charles Laverdure[1], Frédéric Bourassa[1], Lauren Crean[1], Alexia Gonzalez Fanfalone[1]

Charles.LAVERDURE@oecd.org, frederic.bourassa@oecd.org, lauren.crean@oecd.org, alexia.gonzalezfanfalone@oecd.org

[1]OECD, Connectivity Services and Infrastructures Unit, STI/DCES

Keywords: broadband connectivity, MNOs, communication service prices, digital transformation

Broadband connectivity is critical infrastructure underpinning digital transformation. It is essential that everyone is connected—and connected well. This means having access to communication services (availability), at affordable prices (affordability) and of a high quality (quality). Measuring these components provides the data policy makers need to design targeted connectivity policies.

For decades, the OECD has been a global reference for broadband metrics, including mobile connectivity, through its Broadband Statistics. The organisation has also developed key partnerships to provide additional insights into broadband performance to enforce its motto of evidence-based policy making.

The talk will open by presenting the OECD's innovative approach to measuring broadband connectivity, showcasing the OECD Broadband Statistics and the OECD interactive Data Kitchen. It will also highlight OECD data on communication service prices, the data collection of which was internalised in 2025 to ensure data continuity of this critical dataset. The OECD's leading methodologies enable robust international price comparisons to measure affordability of communication services.

As part of the discussion, the presentation will consider some “lessons learned” from data collection and validation of the OECD's various broadband indicators to ensure international comparability. It will also discuss the OECD's data partnerships and established relationships with relevant industry organisations (Ookla, Opensignal, GSMA Intelligence) that complement OECD data.

Then, we will highlight recent work published over the past few months. A notable example is the "Closing broadband connectivity divides for all" report published on 10 July 2025, which illustrates how connectivity data can be combined with innovative data sources to provide meaningful insights at the subnational level. The report uses data from Ookla, Opensignal, and GSMA Intelligence and applies harmonised spatial definitions to assess broadband performance across the urban-rural continuum at a subnational level. As another example of recent work, the Digital Economy Outlook of 2024 displays the OECD's data on communication services prices and provides key insights into the affordability of communication services across the OECD. Finally, the OECD report on "Financing broadband networks"(2024) of the future showcases a variety of financial metrics of the communication sector using data, including Mobile Network Operators (MNOs) from Bloomberg and GSMA Intelligence, among others, to assess trends in broadband network financing and future implications for network investment.

Building a comprehensive quality framework for the use of Mobile Network Operator (MNO) data in Official Statistics: results from Multi-MNO project

Giorgia Simeoni[1], Gabriele Ascari[1], Erika Cerasti[1]

simeoni@istat.it, gabascari@istat.it, erika.cerasti@istat.it

[1]Italian National Institute of Statistics – Istat

Keywords: Quality framework, privately held data, quality metrics

The Multi-MNO project, funded by Eurostat, aimed at the “Development, implementation and demonstration of a reference processing pipeline for the future production of official statistics based on Multiple Mobile Network Operator data”.

Official statistics is subject to strict quality requirements, that, in the European Statistical System (ESS), are represented by the Principles of the European Statistics Code of Practice (ES CoP). In order to develop a reference methodological pipeline (RMP) based on MNO data that could be used in the official statistics production, it was needed to equip it with a comprehensive quality assurance framework. Task 3 of the Multi-MNO project was focused on the definition of such quality framework.

The developed proposal takes stock from existing literature and previous international projects and considers the different quality-related aspects (institutional, organisational, methodological and technical) that arise when using MNO data for statistical purposes.

First, all the principles and indicators of the ES CoP, and the methods of the ESS Quality assurance system, were analysed. The analysis was aimed at identifying, on the one side, what quality requirements should still be fulfilled for MNO data and, on the other side, how the ESS CQF could be extended and improved to support the use of MNO, or more generally privately held, data. This resulted in a set of recommendations, like the importance to invest in partnerships with MNOs, or the need for a focus on input data quality. It also emerged that the definition itself of the standard RMP could represent a tool to satisfy many quality requirements. In other cases, specific tools were developed ad hoc, including a business process model to illustrate how organisational aspects can be managed.

Afterwards, the quality assurance system for the RMP was defined, organised in input, throughput and output quality. Main sources of errors were identified starting from quality issues arising in input data; quality metrics to monitor their occurrence in the input and throughput data were then defined. Particular attention was given to define quality metrics and warnings for the data processing to be carried out by MNOs, out of the direct control of the NSI. A template for a quality report for the pre-processed intermediate data transmitted from MNO to National Statistical Institutes (NSI) was developed. Finally, output quality was developed, with proposals for validation methods. Here quality reporting is considered with a proposal for updating and adapting the ESS standards (the Single integrated metadata structure SIMS and the ESS Handbook for quality and metadata reporting). It should be mentioned, however, that a relevant part in assuring the accuracy of statistics based on MNO data can be managed only through the integration of MNO data with not MNO-data, which was out of the scope of the Multi-MNo project and the focus of another parallel project, the MNO-MINDS ESSnet.

To conclude, also the quality of the technical side of the RMP was taken into account, with the analysis of software quality and how it can be maintained.

From experimental to official migration statistics: A conceptual and actionable framework for quality assurance in mobile-network-operator-derived estimates

Francisco Rowe[1], Paul Baptiste Blanchard[2], Sveta Milusheva[2]

fcorowe@liverpool.ac.uk, pblanchard@worldbank.org, smilusheva@worldbank.org

[1]University of Liverpool, [2] World Bank

Keywords: MNO data; migration; error framework; quality assurance

The integration of Mobile Network Operator (MNO) data into official statistical frameworks holds significant promise for augmenting population and migration statistics, particularly in data-scarce contexts. MNO data provide unique benefits for real-time data collection, enhancing the ability of national statistical offices (NSOs) to provide timely and disaggregated statistics at high spatial and temporal granularity. However, significant methodological challenges must be addressed to move from experimental to official statistics. Specifically, systematic methods are critically needed to better inform data users on the nature and scale of potential discrepancies between statistics derived from MNO data and true values. While it is commonly assumed that this can only be achieved through a comparison of phone-based estimates with survey-based ground truth, direct validation is often unfeasible due to the limited availability of such data, misalignment in timing or definitions, or their complete absence.

As an alternative, this paper proposes a conceptual framework for unpacking the various sources of error in migration statistics derived from MNO data. The proposed framework classifies the sources of error into two broad domains: (i) errors at the device-level and, (ii) errors related to sample composition. The first domain includes three key sources of error: (a) the spatial accuracy associated with MNO data; (b) methodological choices for the detection and definition of migration and (c) device usage behaviour. The second domain covers three other key sources of error: (a) sample selection (due to both the non-probability nature of MNO data samples and the selection of active users); (b) out-of-scope devices and (c) device duplication.

The paper provides a thorough description of each identified source of error and generally discusses three aspects. First, standard methods with relevant secondary data sources allowing to evaluate the significance of a source of error are provided. Second, the paper reviews available methods for mitigating these sources of error. Third, critical knowledge gaps are identified, paving the way towards future research that would allow to better quantify errors in phone-based migration statistics, improve methods to mitigate them, and evaluate their performance in quantitative terms. To illustrate many of these aspects, the paper draws on a multi-year dataset of Call Detail Records with national coverage.

The proposed framework seeks to contribute to ongoing efforts to establish methodological standards for integrating MNO data into official statistical production. It offers a conceptual structure and practical guidelines for NSOs and researchers. By articulating the ingredients needed to develop a rigorous approach for error diagnosis and correction, the framework contributes to the broader agenda of rendering MNO data fit-for-purpose in official migration statistics. The proposed framework could ultimately be integrated within a broader quality assurance framework, improving the quality and reliability of migration estimates derived from MNO data.

Session 4: MNO-MINDS project: Methods at work

Session chair: Sónia Quaresma Gonçalves – INE-PT

Integration of Mobile Network Operator data, survey data and administrative data for tourism statistics: a state space approach

Luis Sanguiao Sande[1]

luis.sanguiao.sande@ine.es

[1]INE (Spain)

Keywords: MNO data, data integration, state space models, Kalman filter

Mobile Network Operator (MNO) data is a promising source for tourism statistics, however when compared to survey data, it shows a systematic bias, which varies seasonally and regionally. This seems to discourage the use of this source alone. We propose an integration approach, using state space models, of three sources: MNO data, survey data and administrative data.

Each source has its own strengths: MNO data has small variance and survey data is unbiased. Administrative data does not distinguish between tourism trips and other trips but provides a sound total for trips. The state space model incorporates uncertainty and temporal dependencies, providing a flexible framework to correct bias, and estimate variances. This way, we can not only provide a combined estimation but also its variance.

Several simulation studies were carried out to evaluate the potential of the proposed method. In particular, some simulations were designed with reduced sample sizes to evaluate whether the integration of mobile data can compensate for the resulting loss in precision.

Flash Estimates of Regional Tourism Using Mobile Network Operator Data

Cristina Faricelli[1], Alessandro Piovani[1], Tiziana Tuoto[1]

cristina.faricelli@istat.it, alessandro.piovani@istat.it, tuoto@istat.it

[1]Italian National Institute of Statistics – Istat

Keywords: Aggregated MNO data; non-probabilistic samples; selection error, device duplication, user ambiguity

The ability to provide timely information is increasingly important for supporting decision-making in the tourism sector, especially in regions characterized by strong seasonal dynamics. Official statistics on the number of nights spent by tourists are typically available with a delay of two months from the end of the reference period, due to the time required for collecting and validating data from local accommodation providers. However, some data sources become available earlier and can serve as valuable proxies for anticipating key indicators. Among these, mobile network operator (MNO) data offer promising opportunities due to their high frequency and near real-time availability.

This study explores the feasibility of producing early regional estimates of the total number of nights spent by tourists by leveraging data available before the official figures are released. The modeling approach is based on information at the municipality level, where data from different sources are integrated and is applied to Emilia Romagna region. The models rely on MNO data, historical values of the target variable (i.e., overnight stays), and a rich set of auxiliary variables. Among the most relevant auxiliary data are those describing the tourism characteristics of the area—such as the typology of tourism (coastal, urban, rural), seasonality, and the accommodation capacity of each municipality.

Several models and estimation strategies have been tested and compared, with the goal of identifying the most accurate and robust solutions for short-term forecasting. The modeling strategies vary in terms of complexity and the way they exploit historical and proxy data. They were evaluated across different months to test their consistency and adaptability. Special attention has been paid to municipalities with high tourism relevance, where early information can have the greatest impact.

The results show that integrating MNO data with historical and contextual information can significantly improve the timeliness and precision of estimates. In particular, the availability of MNO-derived indicators—representing the inferred number of overnight stays—helps anticipate trends and fluctuations in the official data, even though the MNO values themselves can differ from administrative records due to the nature of the signal captured. The performance of the models was assessed using multiple evaluation metrics, both at the individual municipality level and in terms of aggregated regional totals. The most promising approaches achieved good accuracy, with a notable reduction in the average error when compared to baseline models that do not use MNO data.

Overall, this work confirms the strategic value of mobile data for producing flash estimates of tourism activity and provides a flexible framework that can be further adapted and extended to other domains where timely indicators are needed.

Disaggregated trip statistics combining MNOdata and Travel Survey

Li-Chun Zhang[1,2]

li.chun.zhang@ssb.no

[1]Statistics Norway, [2]University of Southampton

Household or individual National Travel Survey (NTS) is conducted in a number of European countries. Typically, data are collected on trips of various distance/duration, modes of travel and purposes, for few days or only a single day according to the survey design (so as to limit the response burden). While the NTS is designed to provide official statistics for the total resident population or major population domains, they simply cannot support various temporal or spatial disaggregation that is of great interest to the users, due to the limited number of trips observed in the survey.

Meanwhile, given the high penetration rate of mobile phones now-a-days and their utility in daily life, MNOs can produce very detailed trip counts by time or geography, which are derived from signalling data directly. But there are also various obvious issues that prevent such MNO trip counts to be accepted as official statistics directly, such as coverage error -- due to the non-users and varying market shares, measurement or processing error – related to trip start, stop and delineation.

Through the collaboration between the Swedish NTS and MNO-MINDS, we have developed some methods for disaggregating the given yearly trip total to monthly statistics, as well as the given total of local trips to all the 21 counties in Sweden. The approach here is that of super-population modelling, where we treat the survey estimates as unbiased but subject to large sampling variances, and the corresponding MNO trip totals as biased but with negligible variances in comparison.

As the Official Statistics Agency responsible for the Swedish NTS, Trafikanalys has already published some results as experimental statistics. In this talk we shall present the underlying estimation methods, illustrate and discuss the relevant considerations when using MNO data for temporally or spatially disaggregated statistics.

Experiments of quasi-randomisation estimation for official statistics based on aggregated mobile network operator data

Marcello D'Orazio[1], Li-Chun Zhang[2,3], Johan Fosen[2], Tiziana Tuoto[1], Jens Haug[2]

madorazi@istat.it, li.chun.zhang@ssb.no, Johan.Fosen@ssb.no, tuoto@istat.it

[1]Italian National Institute of Statistics – Istat, [2] Statistics Norway, [3]University of Southampton

Keywords: Aggregated MNO data; non-probabilistic samples; selection error, device duplication, user ambiguity

Data from mobile network operators (MNOs) have great potential for producing official statistics on population, tourism, mobility, and the environment. However, statistical agencies have no access to such MNO data due to confidentiality concerns, commercial interests, and technological limitations. What can be accessed is anonymized macro MNO data, which refers to summary measures over multiple devices within a specified time period, such as the number of devices that moved from city A to B on a given date. However, since the target statistical or measurement unit is not the mobile device itself, it is necessary to convert from the device to the target unit (persons, trips, nights spent abroad, etc.) in order to reuse MNO data for statistical purposes. For this task, integration with other non-MNO data is necessary, depending on the work domain (population, tourism, mobility, etc.). According to Deliverable 3.2 from the ESSnet MNO-MINDS project, three inferential approaches can be distinguished in this context depending on how the associated uncertainty is conceptualized and measured: (i) randomization, (ii) quasi-randomization (QR), and (iii) super-population modeling. In randomization, the inference is based on a probabilistic sample survey that collects specific mobile device usage data, which is subsequently integrated with external MNO data; this is the commonly used approach in NSI, also known as the design-based approach in official statistics, and quasi-randomization (QR) is a known way of handling inference from observational studies or nonprobability samples, which are not generated according to some known probabilities. QR assumes a model for the underlying observation mechanism, treating the data as if it had been obtained through designed randomization. The advantage of this approach is that the same mechanism applies to all statistical variables associated with the detected devices, unlike superpopulation modeling, where the model is tailored to a specific target variable (response or dependent variable).

This work presents the outcomes of two similar proof-of-concept experiments developed and conducted in Norway and Italy. These experiments used available data from statistical registers and sample surveys (lacking MNO data). These experiments explore viable methods and highlight potential complications that can arise, how they affect the data, and how they can be solved, if possible. The estimation methodology and experimental approach to non-survey big data can be useful for others with similar statistical interests, regardless of whether the data arise from mobile phones, transactions, etc. These experiments are important because they identify key aspects fundamental to negotiating access to MNO data, with the ultimate goal of using such data to produce official statistics.

Session 5: Software and tools

Session chair: Matyas Meztaros – Eurostat

Statistical open source software for integrating MNO and non-MNO data sources

Marcello D'Orazio[1]

madorazi@istat.it

[1]Italian National Institute of Statistics - Istat

Keywords: synthetic MNO data; R environment; open source statistical tools

The MNO-MINDS project reviewed various statistical methods for using MNO data alone or with other data sources to produce official statistics (see Deliverable D3.2). To illustrate some of these methods, the deliverable included application scenarios[1]. These case studies provided an opportunity to understand how the methods can be implemented. The good news is that many of the presented methods are already implemented in open-source statistical tools, particularly in the R environment. The R environment benefits from a vast number of additional packages that are freely available on CRAN and implement most of the considered methodologies. Only a couple of case studies required writing new R functions to implement specific methods/estimators. R was also the environment used for most data preprocessing/treatment, which is not part of the project deliverable.

Unfortunately, one of the main problems was the unavailability of real MNO data with which to test some of the presented methods. For this reason, an open-source software that simulates real-world mobile network events was developed. This tool uses an agent-based simulation model that positions agents on a digital map. The agents can then move based on predefined mobility patterns. The tool then records the resulting network events triggered by interactions between mobile devices and the network infrastructure. The tool is not available as an R package, but its open source code includes a Docker image (a snapshot of the application and its environment, including all necessary code, runtimes, system tools, system libraries, and settings), as well as a Windows installation kit.

An R Shiny Assistant for Visualizing Multi-Source Data and Supporting New Methodology Development in the application of Spanish tourism statistics

Jordi Verdú Naranjo[1], Diego De la Puente Alonso[2]

jordi.verdu.naranjo@ine.es

[1]National Institute of Statistics (INE, StatSpain), [2]University of Valladolid

Keywords: Tourism, MNO data visualization, R Shiny

Multi-source data integration is an essential element for improving the accuracy of the statistical aggregates estimation, especially when such integration involves combining more common statistical sources such as surveys or administrative data with new sources such as MNO data. The new sources of data can provide opportunities to improve the accuracy of previous estimates. However, before selecting an integration method, it is crucial to perform an in-depth exploratory data analysis to determine the characteristics and quality of our sources. Visualization tools can help in this task, even more to subject matter experts who have the domain knowledge. While there are useful available packages such as ggplot and matplotlib, for R and Python respectively, these tools become insufficient on their own when dealing with multiple large, high-granularity datasets. This is why packages such as Shiny and Arrow are no longer optional, but rather essential to streamline visualizations.

In this work, a dashboard developed with Shiny was created to visualize administrative, survey, and MNO data as well as the results of the integration process related to tourism statistics in Spain. The datasets are standardized in parquet format. It includes data from various border entry points, such as airports and roads, along with survey data and mobile network operator SIM counts from multi-MNO.

Metadata analysis for combining data sets with an application to MNO data

Yvonne Gootzen[1], Johan Van der Valk[2]

yapm.gootzen@cbs.nl, j.vandervalk@cbs.nl

[1]Statistics Netherlands (CBS) and Eindhven University of Technology, [2]Statistics Netherlands (CBS)

Keywords: Metadata analysis, open-source implementation, combining data sources

In the context of statistical production, combining data from multiple sources, such as data from Mobile Network Operators (MNOs), presents both opportunities and challenges. This contribution presents the application of a framework that facilitates metadata analysis to support the combination of diverse data sets with MNO data, assisting in the creation of new statistical outputs. The presented framework is designed to assist National Statistical Institutes (NSIs) in evaluating the feasibility of generating desired target outputs based on available data sets and models.

For given user input, the framework is able to answer questions such as “Can an intended output be created from a given set of input data?” and “If so, what sequence of processing steps is then needed?”. The required input consists of several parts. First, a list of available data sets, including administrative records, survey data, and other data sets such as sensor data, along with the variables included in each data set. Second, the relations between granularities of the variables across different data sets. Third, available models defined by input and output data sets only, without the need to specify algorithmic details. Lastly, the target output is required in the form of a single data set, specified by a set of variables that the user wishes to generate. A python program automatically searches if the available models and data sets can lead to the desired target output. If possible, a path will be provided consisting of all analysis steps in chronological order required to create the target output. The resulting software is currently being prepared for open-source release as a python package.

Additionally, a user-focused Jupyter notebook has been developed, specifically tailored an application to MNO data. This notebook both illustrates the capabilities of metadata analysis, as well as enables those interested in combining MNO data with other data sets to analyze various scenarios. Metadata from sources such as census data, MNO data, national travel survey, and population register are preloaded. Metadata of models such as calibration from sim to person, modality choice and shortest path are preloaded. The user can make a selection from these preloaded data sets and models, creating a specific scenario to be explored. Users may extend the framework by adding new data sets and models, thereby broadening the range of scenarios that can be analyzed.

The framework enables users to make a systematical overview of available data sets and models. Metadata descriptions can be shared and discussed between NSIs without privacy concerns, thus promoting collaboration and knowledge exchange across statistical organizations. This research supports the broader goal of improving data integration practices and enabling NSIs to more effectively combine data sets through metadata analysis.

Open MNO Mobility Data in Spain: A Reproducible Sharing Approach via R package 'spanishoddata'

Egor Kotov[1,6], Tania Gullón Muñoz-Repiso [2], Ricardo Herranz[3], Robin Lovelace[4,5], Eugeni Vidal-Tortosa[4,7], Oliva Cantu-Ros[3], Javier Burrieza-Galán[3]

kotov@demogr.mpg.de, tgullon@transportes.gob.es, ricardo.herranz@nommon.es, R.Lovelace@leeds.ac.uk

[1]Max Planck Institute for Demographic Research, [2]Ministerio de Transportes y Movilidad Sostenible, [3]Nommon Solutions and Technologies S.L., [4]University of Leeds, [5]Institute for Transport Studies, [6]Universitat Pompeu Fabra, [7]Universitat Autònoma de Barcelona

Keyword: Spain, public MNO data, origin-destination matrices, R, data-package

The integration of Mobile Network Operator (MNO) data with traditional statistical sources offers unprecedented opportunities to enrich official statistics in transport, tourism, population dynamics, and regional commuting. The Spanish Open Mobility Big Data — a publicly available MNO-derived dataset published by the Spanish Ministry of Transport and Sustainable Mobility — consists of hourly origin-destination matrices aggregated to daily tables, enriched with sex, age, income, and trip-purpose attributes, covering over 3,500 mobility zones across Spain and beyond. To understand and overcome adoption barriers, we systematically reviewed 36 peer-reviewed studies (2020–2024) that employ this dataset and, in response, provide a fully open-source R toolkit ('spanishoddata', https://ropenspain.github.io/spanishoddata/) that automates robust, reproducible workflows for data discovery, acquisition, storage, and citation.

Our literature analysis uncovered three pervasive issues:

1. Ad hoc data acquisition: Nearly all studies relied on manual downloads or one-off scripts, preventing external researchers from reliably executing published analyses, because they cannot acquire the data in the same way.

2. Limited reproducibility: Only one in five papers shared any preprocessing code, and fewer than one-third provided full analysis scripts or preprocessed datasets, severely restricting result verification and extension.

3. Fragmented citation: Authors referenced the same dataset in over a dozen different ways — linking to multiple URLs, page titles, or secondary portals — hindering consistent attribution and obstructing data-provider impact assessment.

To address these challenges, our open-source R toolkit — fully independent of the data custodian — embeds best practices for MNO-based statistical workflows. It enables users to:

- Discover metadata programmatically, exploring spatial resolutions (district, municipal, provincial) and temporal coverage;

- Download compressed raw files in parallel with built-in resume and integrity checks;

- Transform large origin-destination tables into highly efficient, queryable columnar formats (Parquet and DuckDB) leveraging modern columnar engines;

- Cite the source in compliance with the official open-data licence, ensuring uniform acknowledgement.

Importantly, the entire data-acquisition and conversion pipeline can be reproduced in just a few lines of code, eliminating barriers to sharing complete, transparent analysis workflows. The toolkit is publicly hosted, encourages community contributions, and can be adapted to other national MNO datasets.

We demonstrate its practical utility with two case studies (materials available at https://doi.org/10.5281/zenodo.15289979 as part of a submitted journal article):

- Cycling potential in Valencia: Filtering intra-district trips of 0.5–10 km, overlaying the existing cycle-path network, and identifying spatial mismatches between demand and infrastructure.

- Commuting dynamics in Madrid: Comparing morning versus evening flows by distinguishing work-related and non-work trips, revealing temporal asymmetries with implications for public-transport scheduling.

By abstracting low-level ETL burdens, enforcing standardised citation, and harnessing efficient storage formats, our toolkit empowers researchers and statistical agencies to integrate public MNO data at scale—promoting reproducible, efficient mobility analytics. We invite conference participants to explore this example as one of the possible approaches of sharing MNO data with the end users via opinionated software interfaces that guide the users on how to cite the data, how to best acquire it and aggregated for analysis. At the same time, we are lowering the barrier for novice analysts, educating users on modern big-data tools, and enabling robust impact tracking through consistent data citation.

Session 6: Mobility and Transportation

Session chair: Gloria Deetjen – DESTATIS

Transport Planning based on MNO Data in Spain

Tania Gullón Muñoz-Repiso[1], Ana Maldonado[1], César Rodriguez Cano[1], Pablo de la Presa[1]

tgullon@transportes.gob.es, amaldonado@transportes.gob.es, cirodriguez@transportes.gob.es, ppresa@transportes.gob.es

[1]Ministry of Transports and Sustainable Mobility of Spain

Keywords: Mobility bigdata phone CDR transport Ministry

The Spanish Ministry of Transport has been monitoring mobility in Spain since 2020. To study the mobility flows of people, it analyzes anonymized mobile phone records of more than 30% of the population using big data technologies. This allows for the availability of origin-destination matrices at the level of census districts for the entire country, by hour, by purpose of travel, and by sociodemographic profile. Until a few years ago, having this vast amount of valuable information was unthinkable. This data source is essential for supporting decision-making in the Ministry of Transport, but it is also being used by hundreds of companies, administrations, universities, and researchers for numerous purposes, among which transportation planning and support for decision-making in both the public and private sectors stand out.

Combining Mobile Network Data with Floating Car Data and Traditional Data Sources to Estimate Mobility Demand in High-Resolution: a Case Study in the Bologna's Metropolitan Area

Giorgio Fiorillo[1], **Daniele Di Antonio**[2], Daniele Mancuso[2], Eugenio Margelli[1]

giorgio.fiorillo@srmbologna.it, diantonio@go-mobility.it, mancuso@go-mobility.it, eugenio.margelli@srmbologna.it

[1]SRM- società Reti e Mobilità - Pubblic transport authority in Bologna, [2]Go-Mobility

Keywords: MND, FCD, SUMP

High-resolution Origin-Destination matrices are key to support efficient and inclusive transport planning, particularly in low-demand areas when designing user-centric services such as Demand Responsive Transit. This study presents an integrated methodology for estimating travel demand in the Metropolitan City of Bologna, leveraging anonymized mobile network data (MND) in combination with additional sources such as mobility surveys, socio-economic datasets, and Floating Car Data (FCD). Unlike traditional approaches that rely solely on travel behaviour surveys, which are limited by sample size, respondent bias, and lack of data on non-residents, this methodology integrates traditional datasets with MND to overcome these constraints. Among others, MND adds the capacity to estimate movement flows from non-residents and tourists and allows for a deep classification by user type, enhancing the granularity of demand estimation.

The proposed analytical framework adopts a fine-grained zoning system, dividing the metropolitan area into over 1,000 traffic analysis zones. O-D matrices are generated for three distinct months in May, July, and November, with differentiation by weekday vs. weekend and enriched by targeted analyses on large-scale events such as concerts, sport matches and trade fairs. Mobility surveys contribute additional insights on trip purposes and modal split, which, coupled with FCD, enable the identification of car-dependent patterns, supporting the design of public transport policies in line with Bologna's Sustainable Urban Mobility Plan (SUMP) targets for reducing private car use. The methodology is implemented via a modular data-processing framework, designed for full replicability and periodic updates. This enables the system to serve not only as a decision support tool for service planning, but also as a long-term monitoring instrument for evaluating the impact of SUMP measures.

From Mobile Phone Data to Mobility Patterns

Victor Tuekam[1,2], Sebastian Wichert[2], Oliver Falck[2], Göran Kauermann[1]

tuekam@lmu.de, wichert@ifo.de, falck@ifo.de, goeran.kauermann@stat.uni-muenchen.de

[1]Department of Statistics LMU Munich, [2]ifo Institute

Keywords: human mobility, mobile phone data, generalized additive models, hurdle models, functional principal components analysis, longitudinal network analysis

This study demonstrates how large high-frequency mobile phone data can be transformed into flow networks to analyze human mobility patterns. Using anonymized data provided by a large mobile network provider in Germany, we utilize digital traces generated through user interactions with mobile networks, such as calls, messages, and automatic connectivity updates, to infer movements between locations. The spatial resolution is thereby recorded at the level of cell tower locations, while mobility patterns are of interest at the level of administrative districts (e.g., postcode areas). We demonstrate a large-scale data analysis that is still privacy-preserving, examining population mobility without relying on GPS or user-consented tracking.

Our approach focuses on two main tasks: preprocessing the raw mobile phone records and conceptualizing them as network representations of spatial flows. Since the data reflect connections to cell towers rather than precise locations, the spatial resolution is defined by the geographic positions of the towers and resulting individual trajectories, which themselves are imprecise. We address challenges related to trajectory reconstruction and data sparsity by applying a scalable, spatio-temporal linear interpolation technique to estimate user movement between towers. These estimated flows are then aggregated to the postcode level to create interpretable mobility networks.

This work contributes a practical method for transforming infrastructure-based mobile phone data into meaningful representations of urban and regional mobility. By emphasizing data preprocessing and network construction—while avoiding user-level tracking—the study supports the development of scalable, privacy-conscious methods for analyzing large-scale human movement patterns. For example, we demonstrate movement patterns in Munich and apply a dynamic gravity model to explain them. We also tackle the question of obtaining information on the mode of transportation.

Analysis of HSR foreign passenger at Firenze S. Maria Novella Station in 2024

Lorenzo Vannacci[1], **Luigi Mancino**[3], Luigi Galieni[1], Chiara Bonvicini[1], Bruno Zamengo[2], Alessandro De Cicco[2], Mario Tartaglia[1], Dario Di Sorte[3], Andrea Zaramella[3]

l.vannacci@fsitaliane.it , l.galieni@fsitaliane.it, c.bonvicini@fsitaliane.it, bruno.zamengo@motionanalytica.com, alessandro.decicco@motionanalytica.com, m.tartaglia@fsitaliane, dario.disorte@vodafone.com, andrea.zaramella@vodafone.com, luigi.mancino@vodafone.com

[1]FS Research Centre - Ferrovie dello Stato Italiane, [2] Motion Analytica, [3]Fastweb + Vodafone

Keywords: tourism, Big Data, rail, mnd

Tourism is a key industry in Italy, and 2024 marked a record year for the country, largely due to international visitors. This study offers a detailed examination of the mobility patterns of foreign travellers who arrived or departed by train from Firenze S. Maria Novella station, utilizing SIM card data to track their movements within Italy and Tuscany. Travelers are categorized by nationality based on their SIM card's country code, with all data anonymized in accordance with EU privacy regulations.

Methodology: The study observed all unique foreign travellers who completed their vacation in Italy by March 31, 2024, and travelled by train to or from Firenze S.M.N. station in January 2024. The analysis includes various aspects such as entry and exit points, itineraries, and key destinations. The nationality of foreign travellers is identified through their SIM card.

Results: The research investigates the primary chains of overnight destinations during their stay in Italy, linking the overnight stay to the municipality where the traveler spent most of the day. Special attention is given to excursions within Tuscany, pinpointing the municipalities where travelers spent at least an hour, regardless of overnight stays.

Conclusions:
• Most travellers enter and exit Italy through major airports, with over one-third using Fiumicino airport. However, half of the travellers leave Italy from a different border point than their entry.
• The most connected Italian and Tuscan railway links with Firenze station are Roma, Venezia, Pisa, and Siena.
• The average mobility range and number of overnight stays vary by nationality, with non-European travellers exhibiting a higher mobility range and spending more time in Italy.
• Only 39% of travellers who stayed overnight in Firenze had the city as their main vacation destination, compared to 55% for Roma.

This summary presents the key points and findings of the document, providing an overview of the mobility behaviour of foreign travellers at Firenze S. Maria Novella station and their travel patterns within Italy.

Session 7: Exploring the use of MNO data in different fields

Session chair: Giorgia Simeoni – Istat

Mobility data assimilation to forecast electricity demand in a sobriety context in France

Stefania Rubrichi[1], Nathan Doumèche[2], Yann Allioux[2], Yannig Goude[2]

stefania.rubrichi@orange.com, nathan.doumeche@edf.fr, yann.allioux@edf.fr, yannig.goude@edf.fr

[1]Orange Innovation Research – SENSE lab, [2]Électricité de France R&D

Keywords: mobile phone data, energy consumption, human behaviour, socio-economic indicators

Energy is at the very core of modern economies and politics. In addition, its impact on climate change is forcing our society to change its consumption patterns. As a result, there is a growing interest in energy savings and in the transition to sustainable energy sources.

In France, electricity is the main source in the energy mix. The French ecological transition plan is based on a massive electrification of the uses powered by nuclear energy. The recent context of high energy prices due to the post-covid industrial renewal and to the war in Ukraine has had an important impact on European economies and has been accompanied by energy savings in Europe.

In this work, we document electricity savings in France during the winter of 2022-2023, which various media have described as a sobriety period. However, there is no well-established modeling of the impact of price increases and government incentives on electricity consumption patterns. In addition, traditional statistical models based on meteorological data struggle to adapt to such brutal ruptures.

Recently, machine learning techniques have been applied to electricity load forecasting to ensure the balance of the electricity grid and to reduce electric waste. Indeed, because electricity storage capacity is limited and expensive, the supply of electricity must match demand at all times. As a consequence, electricity load forecasting at different forecasting horizons has attracted a growing interest in recent years. This work focuses on 24-hour ahead load forecasting, which is particularly relevant for operational applications in industry and the electricity market. Recent research in electricity load forecasting has demonstrated the advantages of adaptive methods that adapt to changes in regimes. In particular, state-space models and online aggregation of experts have proven successful in capturing changes in patterns caused by the COVID-19 crisis.

Most state-of-the-art models rely on historical data of past electricity loads, calendar data such as holidays or the position of the day in the week, and meteorological data such as temperature and humidity. However, such data cannot accurately describe the complex human behaviors that affect the variability of energy demand. As a result, traditional models have struggled to account for brutal societal events such as COVID-19 lockdowns. A deeper knowledge of human behaviours is thus necessary to better model the electricity demand. Over the past decades, complementary datasets generated from cell phone networks, location-based services (LBS), and remote sensors in general have emerged to provide a more fine-grained description of human behaviour.

This work relies on the use of the adjusted high-quality data on human presence provided by the Orange business service Flux Vision to model electricity demand during the sobriety period in France in 2022-2023. This dataset is based on measurements of mobile antenna traffic, while most datasets are based on users' consent to share their geolocation with specific applications. The resulting signal is very stable compared to other datasets where the user base may vary as users unsubscribe from the application. We start by characterizing electricity savings during the sobriety period in France. We then show that mobility data from Orange are correlated with socio-economic indicators (eg tourism, workplace attendance).

Finally, we show that models using mobility data outperform the state-of-the-art in electricity demand forecasting. Our results show how sequential learning methods are able to capture the changes in mobility behaviors and their impact on the electricity demand.

An approach to identify people trips and stops for mobility analysis using mobile network data

Chiara Bonvicini[1], **Marco Agnolon**[2,] Luigi Galieni[1], Enrico Bontorin[2], Lorenzo Vannacci[1], Andrea Zaramella[3], Luigi Mancino[3], Dario Di Sorte[3], Mario Tartaglia[1], Gianluca Busatta[2]

c.bonvicini@fsitaliane.it, l.galieni@fsitaliane.it, marco.agnolon@motionanalytica.com, enrico.bontorin@motionanalytica.com, l.vannacci@fsitaliane.it, Andrea.Zaramella@vodafone.com, Luigi.MANCINO@vodafone.com, Dario.DISORTE@vodafone.com, m.tartaglia@fsitaliane.it

[1]FS Research Centre - Ferrovie dello Stato Italiane, [2]Motion Analytica, [3]Fastweb + Vodafone

Keywords: big data, mobile network data, mobility analysis, new data sources, FS Research Centre

The widespread deployment of connected devices, such as smartphones and in-vehicle systems, has made available an increasing amount of data generated automatically by users.

Among these, a particularly important role is played by mobile network data, i.e. information derived from interactions between mobile phones and cellular network infrastructure. Although not specifically designed for movement tracking, these data are now a strategic resource for the study of mobility, allowing for detailed, up-to-date and large-scale spatio-temporal reconstructions. Their use has opened up new perspectives in the analysis of individual and collective mobility behavior, overcoming many of the limitations of traditional methods.

The methodology used to process these data makes it possible to convert chronological sequences of network events, characterized by high temporal resolution but spatial uncertainty related to cell coverage, into meaningful trip profiles, this overcomes the limitations of traditional techniques in terms of cost, coverage, timeliness and sample size.

However, the raw data do not allow to discern, without appropriate processing, between stationarity and movement, and they do not indicate mode or motivation of travel. Hence the need to apply complex algorithms: cleaning operations, trajectory segmentation, inference of mode of transport, sample expansion techniques and anonymization according to privacy regulations.

In the present work, starting from the definition of trip as a sequence of ""points"" separated by ""stops"" we explore the crucial issue of using a temporal threshold, a minimum value of duration below which the interruption is considered negligible, that allows us to identify the significant stops made by the user during the journey. However, a fixed threshold applied to all trips, if too short or too long, can lead to false classifications: inclusions of fictitious breaks or exclusion of many real trips, causing high information loss and doubtful representativeness of the remaining sample.

To overcome these limitations, we propose an alternative data-driven paradigm, in which the significance of a stop is defined in relation to individual behavior and the actual route taken, compared to the expected route between origin and destination.

Trip by trip the deviation traveled is analyzed, comparing it with those of similar users on equivalent trajectories: a significant deviation is interpreted as an indication that the destination touched has a specific value (e.g. activity, specific purpose). This makes it possible to go beyond the rigid time threshold, adapting the distinction between stopping and moving to real human dynamics.

This approach allows for more authentic reconstructions of trips, respectful of behavioral diversity and differences between users: the result is a segmentation of raw data into meaningful movements that maintains high information fidelity, minimizing omissions and false positives.

Spatial Interactions of Consumption, Presence, and Mobility: Evidence from Mobile Phone and Transaction Data

Chloé Breton[1,2], David Bounie[1], John Galbraith[3], Gabrielle Gambuli[2,4],

chloe.breton@insee.fr, david.bounie@telecom-paris.fr, john.galbraith@mcgill.ca, gabrielle.gambuli@insee.fr

[1]CREST, [2]Insee – France, [3]McGill University, [4]Université Gustave Eiffel,

Keywords: mobile phone data, transaction data, data integration, spatial analysis

Network Operator (MNO) data and card transaction data to examine the relationship between presence and consumption in the Lyon Functional Urban Area (France) during the month of September 2022.

Our methodological approach integrates two high-frequency, spatially granular data sources: MNO data from Orange, which detect real-time population presence based on SIM activity, and card transaction data from Cartes Bancaires, which capture consumption through in-person purchases made by consumers physically present in stores (as opposed to online transactions). These datasets are not linked at the individual level. Instead, we match them at an aggregated geographical scale (Iris zones—the smallest statistical spatial unit in France) and harmonize them temporally (30-minute intervals).

We first assess the quality and representativeness of both data sources by comparing them to official census data from Insee. Although the underlying units differ—billing address for cardholders and residential location as inferred by the mobile operator—both datasets show strong correlations with population counts at the Iris level. This validates their use as proxies for local activity.
Our second objective is to examine whether these two sources reveal similar temporal and spatial patterns, and to identify when and where they diverge. While MNO and transaction data measure different aspects—where people are versus where they spend money—they often show similar patterns. For example, on Saturdays, both sources follow very similar hourly trends. We also observe close alignment around lunchtime and in the evenings on weekdays. However, discrepancies emerge at other times, such as weekday mornings and Sundays. These differences likely reflect structural factors such as limited time for spending during work or commuting, and shop opening hours. They highlight the complementary nature of the two data sources.

To quantify the link between presence and consumption, we estimate a Poisson regression model with high-dimensional fixed effects (zone, date, time). We find that a 1% increase in mobile presence is associated with a nearly proportional increase in transactions, suggesting a strong average elasticity close to one. However, this elasticity varies across contexts. It is highest in dense urban cores, where agglomeration effects likely lead to more concentrated consumption. In contrast, peripheral and rural areas show significantly lower elasticities, indicating that increased presence does not necessarily translate into spending. We also observe temporal heterogeneity, with elasticities rising toward the weekend and peaking on Saturdays.

Beyond correlation, we integrate public data on local retail equipment and services to explore spatial mismatches—areas where presence is high, but consumption remains low. These discrepancies reveal imbalances between demand and supply, offering diagnostic signals for public action. Such mismatches are not visible through conventional statistical sources.

Our study demonstrates that combining independent, anonymized, and aggregated data sources—mobile and bank—provides valuable insights into urban dynamics without compromising privacy. This framework opens new possibilities for statistical innovation and territorial decision-making, from commercial development to transport planning. It also highlights how high-frequency, geo-localized data can complement official statistics to monitor real-time shifts in mobility and consumption patterns.

Harnessing geospatial innovation: Insights from the OECD Geospatial Lab

Claudia Baranzelli[1], Alexandre Banquet[1], Josep Espasa Reig[1]

claudia.baranzelli@oecd.org, alexandre.banquet@oecd.org, Josep.ESPASAREIG@oecd.org

[1]OECD

Keywords: geospatial data, data integration, international collaboration, regional statistics, local statistics

The rapid growth of new data sources is both an opportunity and a challenge. On one hand, these data offer the potential to generate more timely, granular, and actionable evidence; on the other, their complexity and fast pace of development pose significant challenges for National Statistical Offices (NSOs) and international organisations, which must ensure trust, quality, and transparency.

Many of these emerging data sources are geospatial by nature, or can be used to enable geospatial analysis, such as data collected through personal mobile devices. The OECD Geospatial Lab was established to promote the widespread use of geospatial data and methods to inform effective policies. It does so also by developing methodologies and indicators that integrate geospatial information into official statistics.

Through practical examples, this presentation highlights how the Lab leverages geospatial data, including data generated by personal mobile devices, to generate fresh insights into social, economic, and environmental trends, while safeguarding privacy and supporting sustainable data pipelines. Key areas of focus include developing methods to combine high-resolution geospatial datasets with conventional statistical sources, fostering partnerships to enhance data access and interoperability, and strengthening collaboration across OECD bodies, international partners, researchers, and public and private stakeholders. These efforts ensure that geospatial innovation remains trustworthy and transparent. By showcasing recent applications and lessons learned, this contribution aims to stimulate discussion on how geospatial intelligence can help address pressing policy challenges, from regional inequalities and urban development to climate resilience.

Session 8: MNO data for Tourism

Session chair: Alexander Kowarik – STAT Austria

Improving Tourism Measurement in Spain Using Mobile Network Operator Data

Marta Sixto Neira[1], Alfonso Fernández Bes[1], María Velasco Gimeno[1], Ana Belén Marín Valverde[1]

marta.sixto.neira@ine.es, alfonso.fernandez.bes@ine.es, maria.velasco.gimeno@ine.es, anabelen.marin.valverde@ine.es

[1]National Statistic Institute of Spain

Keywords: Tourism, MNO, data integration

Tourism is a key sector in Spain, representing a significant share of the national economy. Accurate, timely, and granular data is essential for policymakers and stakeholders to make decisions. However, traditional survey-based methods lack sufficient geographical granularity. They are also costly and impose a high burden on respondents. We found in MNO data a possible solution. We have been working with this data since 2020 to integrate it.

Over the past five years, a collaborative effort between our NSI and MNO teams has focused on designing methodologies to transform raw MNO data into indicators aligned with international tourism definitions. This cooperation enabled continuous improvement of algorithms, ensuring compatibility with international definitions and enhancing the usability of mobile data for tourism analysis.

Our framework covers all types of tourism: inbound, outbound, and internal. In this paper, we focus on inbound tourism. Mobile phones act as proxies for individuals and the SIM card's country of origin as proxies for country of residence.

The general definitions are as follows. A trip is considered when a mobile device is detected in Spain between 10:00 p.m. and 6:00 a.m. and finishes the last day it is seen in the country. Overnight stays are counted based on the number of nights the device is present. Excursionists are devices detected during the day (no night) for more than three hours.

Foreign SIM cards likely belonging to residents in Spain are filtered out. Devices observed more than three months in the last six months are excluded. We estimate stages and the number of overnight stays associated. These variables support data validation and improve accuracy.

Several corrections were made to the algorithm over time based on experience. For example, Spain has a high level of nightlife-related mobility. To avoid misclassifying nightlife movements as tourists, we introduced a requirement: the device must remain in the same municipality between 6:00 a.m. and 10:00 a.m. following a night-time detection.

Knowledge gained during the project showed that not all limitations can be solved by algorithms alone. It is necessary to complement mobile data with other sources. In our case, administrative data—such as traffic loops and cameras, and passenger data from airports, ports, and trains—proved essential. Additionally, data on the number of beds in each accommodation type was needed. These sources help refine estimates of tourist flows, correct misclassifications near borders, and improve destination identification.

New questions were added to our inbound tourism survey to gather more information on mobile phone usage and SIM card types. This aims to identify non-personal or local SIMs acquired upon arrival (e.g., in

European airports) and improve country-of-residence assignments. Surveys remain essential for collecting variables not accessible via mobile data, such as travel purpose, accommodation type, and trip organization.

This five-year experience shows that mobile data is valuable for enriching tourism statistics, but its full potential is realized only through a mixed-source strategy. Integrating MNO data with traditional surveys and administrative records enables more detailed and accurate tourism statistics.

Exploring the Use of Mobile Network Data for Official Tourism Supply-Side Statistics in Italy. Vodafone-Istat Project Foundation

Maria Teresa Santoro[1], Lorenzo Cavallo[1], Silvia Di Sante[1]

masantor@istat.it, cavallo@istat.it, disante@istat.it

[1]Italian National Institute of Statistics – Istat

Keywords: MNO data; Official statistics; Tourism flows measurement; Alternative data; Methodological challenges; Public–private collaboration

The use of alternative data sources for the production of official statistics has gained significant momentum in recent years, particularly in response to the increasing demand for timely, granular, and cost-effective data. This paper presents an overview of the Project Foundation (PF), a collaborative initiative between Vodafone, a major Mobile Network Operator (MNO), and Istat, the Italian National Institute of Statistics. The project aimed to explore the feasibility and methodological implications of using mobile network data to produce official tourism statistics, with a thematic and conceptual focus on measuring arrivals and nights spent, both for domestic and inbound tourism.

The PF project built upon the results of an earlier, short-term exploratory phase known as Sprint, conducted between September and November 2020. The Sprint phase was designed as a proof of concept to assess the potential of mobile data for capturing mobility patterns relevant to tourism flows in Italy. While this initial effort revealed a number of limitations, particularly regarding alignment with official statistical definitions, it also provided encouraging indications that mobile data could complement traditional sources, especially for trend detection and early warnings.

A second phase of the project, Project Foundation (September-November 2023), focused on addressing these limitations by deepening the methodological work. Key areas of investigation included the harmonization of conceptual definitions and classifications used in official statistics with those that can be derived from mobile network data. The project also aimed to improve the overall robustness and reliability of mobile-based estimates through iterative testing, model refinement, and validation procedures.

This paper reports on a targeted case study involving two Italian regions over a 16-month observation period. During this timeframe, estimates produced by Vodafone were systematically compared to official statistics from Istat's "Occupancy in collective accommodation establishments" survey. The comparative analysis highlighted both the strengths and limitations of using MNO data for tourism statistics. In particular, the results showed a reasonably consistent trend alignment between mobile-derived estimates and official data, suggesting potential for timely and cost-efficient monitoring of tourism dynamics at the regional level.

Nevertheless, the study also underscored several persistent challenges. These included issues of data representativeness, transparency of proprietary algorithms, privacy safeguards, and the conceptual interpretation of mobility signals in terms of statistical phenomena. While the findings justify a cautiously optimistic outlook on the integration of mobile data into the statistical production process, the paper emphasizes that further methodological development and institutional cooperation are required before such data can be scaled to national-level production or used in regulatory contexts.

In conclusion, Project Foundation contributed to the international debate on the role of new data sources in official statistics. It offered a practical, empirically grounded case of public–private cooperation and shed light on both the opportunities and limitations of mobile data, not as a replacement, but as a valuable complement to traditional statistical instruments.

From signals to statistics: measuring same-day visits with MNO data. An Italian use case

Mascia Di Torrice[1], Barbara Dattilo[1], Mariangela Sabato[1]

maditorr@istat.it, dattilo@istat.it, sabato@istat.it

[1]Italian National Institute of Statistics – Istat

Keywords: Tourism, usual environment, same-day visits, mobile phone data, MNOs

The integration of data from Mobile Network Operators (MNOs) into official statistics presents a significant opportunity to enhance statistical production. MNO data offer a wide population coverage, high spatial and temporal granularity, and near real-time availability—features particularly valuable in domains such as tourism, where traditional surveys often struggle to capture dynamic, short-duration mobility patterns, such as same-day visits (SDVs).

However, transforming MNO data into meaningful statistical indicators poses several conceptual and operational challenges, as data collected originally for non-statistical purposes lack direct information on motivations or visit characteristics. Producing tourism indicators thus requires inferring behaviours from large volumes of anonymized cell phone location data.

A critical issue in tourism statistics is the interpretation of MNO data to ensure alignment with international definitions, particularly those set by the International Recommendations on Tourism Statistics (IRTS 2008). A key concept in this field is the Usual Environment (UE), which distinguishes tourism-related movements from routine ones. Defining and operationalizing the UE using MNO data is particularly challenging in the absence of explicit information on the purpose, frequency, duration, and destination of each movement.

To address this challenge, Istat collaborated in 2023-2024 with one of the major Italian MNOs (Vodafone) by conducting an experimental study aimed at estimating domestic same-day visits using mobile phone data. The approach relied on the definition of specific algorithms and related metrics to infer the UE from mobility patterns in MNO data and to align the definition of same-day visits with international statistical standards. The algorithm implemented by Vodafone was designed to reproduce key parameters for detecting SDVs and to reduce the entropy of raw MNO signals.

This approach was applied to a specific use case: the estimation of domestic SDVs made by residents visiting Tuscany and Latium during the third quarters of 2022 and 2023. Then, the resulting estimates were compared with official data from Istat's Trips and Holidays Survey (THS), which is the Italian household sample survey that provides estimates of tourism events by residents, including SDVs.

This work outlines the long and complex process of adapting and operationalizing tourism definitions and statistical concepts within the context of MNO data. It describes the metrics used to determine the Usual Environment and identify same-day visits, as well as the implementation steps and the iterative refinement of different scenarios through various simulations. Furthermore, it discusses the results of comparing MNO-based estimates with official figures from the THS. Lastly, it provides a critical assessment of the strengths and limitations of this experimental approach, reflecting on the potential future use of MNO data in tourism statistics, particularly as a complementary resource alongside traditional surveys.

Enhancing Tourism Analytics through Mobile Network Data: A Case Study of Venice and the Veneto Coast

Bruno Zamengo[1], **Andrea Zaramella**[2], Luigi Mancino[2], Denis Cappellari[1], Dario Di Sorte[2]

bruno.zamengo@motionanalytica.com, andrea.zaramella@vodafone.com, luigi.mancino@vodafone.com, denis.cappellari@motionanalytica.com, dario.disorte@vodafone.com

[1]Motion Analytica, [2] Fastweb + Vodafone

Keywords: Mobile Network Operator (MNO) data; Tourism big data; Usual environment; Overtourism; Official tourism statistics; Tourism analytics methodology

Understanding tourist dynamics in high-density destinations is essential for public authorities, local businesses, and urban planners. This study presents an innovative approach based on Mobile Network Operator (MNO) data to monitor and analyze tourism flows across the Veneto region—focusing on the city of Venice, including its distinctive sub-zones (e.g., Cannaregio, San Marco, Lido, Murano), and the main seaside resorts of Cavallino Treporti, Jesolo, Eraclea, Caorle, and Bibione.

The project leverages anonymized and aggregated MNO data over a 24-month period (2023–2024) to generate detailed indicators on arrivals, overnight stays, and day trips. The Mobile Analytics service is fully GDPR-compliant, as it processes irreversibly anonymized and aggregated big data, referenced in time and space. Based on definitions previously developed in collaboration with ISTAT, the Italian National Institute of Statistics, the project applies both a consolidated methodology (“classic mode”) and an innovative classification system (“usual environment mode”). The latter introduces the concept of an individual’s usual environment, a dynamic, multi-location definition of habitual presence derived from long-term mobility patterns.

The collaboration with ISTAT, and later with Ca’ Foscari University, focused on aligning methodological frameworks and exploring how official statistical definitions could be mapped onto mobile user behavior. This joint work led to the refinement of the usual environment concept and its application to improve the consistency of MNO-based tourism estimates with official statistics.

In the sixth phase of the iNEST project, in partnership with Ca’ Foscari University, we produced tourism indicators for selected destinations in Veneto at both municipal and sub-municipal levels, using two scenarios: one excluding and one including the usual environment definition. These datasets were jointly examined and further analyzed by the University, comparing MNO-derived estimates against official tourism statistics and, for the city of Venice, against public data related to the “Contributo d’accesso”, a fee introduced for day-trippers on selected dates.

By distinguishing between habitual residents, commuters/students, overnight tourists, and day visitors, and classifying users by nationality and duration of stay, the analysis provides a nuanced understanding of visitor presence. While both methodological modes reveal broadly consistent trends, the usual environment mode aligns more closely with official sources, suggesting enhanced reliability.

This project illustrates how telco data, when processed through rigorous privacy protocols and grounded in robust definitional frameworks, can offer valuable insights into complex tourism phenomena. The case of Venice and the Veneto coast highlights the potential of mobile data to support evidence-based tourism management, particularly in areas under visitor pressure.

Session 9: Population & migration

Session chair: Johan Van der Valk – CBS

An application of the quasi-randomization approach to migration statistics derived from MNO data

Paul Baptiste Blanchard[1], Sveta Milusheva[1], Francisco Rowe[2]

pblanchard@worldbank.org, smilusheva@worldbank.org, fcorowe@liverpool.ac.uk

[1]World Bank, [2] University of Liverpool

Keywords: MNO data; migration; selection bias; quasi-randomization

Accurate migration statistics are essential for informing key policy decisions in developing countries. Timely insights into internal population movements can support a wide range of applications, from urban planning and resource allocation to crisis response. Yet, internal migration estimates typically rely on national censuses conducted only once every decade, leaving significant gaps in understanding the dynamics that occur between these infrequent data collections.

In this context, Mobile Network Operators (MNOs) data, particularly Call Detail Records (CDRs), offer significant advantages for tracking population movements. They not only enable the detection of long-term internal migration typically captured by censuses but also provide insights into temporary mobility patterns, with high spatial and temporal granularity. More importantly, the near-real-time availability of MNO data creates opportunities to produce more frequent migration estimates, enhancing the effectiveness of policy decisions.

However, key challenges have hampered the systematic integration of MNO data into national statistical systems. One major concern is that phone users do not constitute a representative sample of the general population, leading to biased migration estimates. While some ad hoc bias adjustment methods exist, the absence of a standardized framework remains a significant gap for the adoption of MNO data for official statistics. Building on recent methodological contributions, this paper aims to provide practical guidelines for applying a quasi-randomization approach that integrates MNO data with external sources to derive migration statistics from a non-probability sample. The approach is demonstrated using a multi-year CDR dataset from a leading MNO in a developing country, combined with representative survey data and census data.

The paper discusses a critical yet often overlooked issue: the selection of an appropriate target population. Migration estimates derived from MNO data are frequently adjusted to represent the general population. For some migration statistics, like seasonal migration, this approach can be misleading when the non-probability sample systematically excludes entire demographic groups – such as children – who both constitute a significant share of the population and exhibit distinct mobility patterns. The paper explores how observable characteristics within the MNO data, combined with insights from representative surveys, can guide the identification of a more plausible and methodologically sound target population.

Next, the paper introduces a structured statistical framework to address bias in non-probability samples. Within this framework, although selection probabilities are unknown, unbiased estimates for a carefully defined target population are obtained by inferring a latent selection model. This model estimates selection probabilities based on individual characteristics that must be observable in both the non-probability sample and the target population. In the case of CDR data, home location is typically the only available characteristic, prompting the use of spatial stratification to approximate the selection mechanism. The paper examines how to determine the appropriate geographic level for this stratification, weighing the trade-offs between persistent bias at coarser spatial resolutions and increased variance at finer ones. The paper generates estimates of migration applying sampling weights based on different levels of spatial stratification, using census data as ground truth to assess the performance and highlight the effect of these methodological choices. Finally, the paper explores complementing the geography-based selection model with a bounding exercise that tests higher-order stratification layers within geographic weighting units.

Using Mobile Positioning Data for Validation and Signs-of-Life in a Register-Based Population Census

Margus Tiru[1], Gerttu Pilsas[1], Erki Saluveer[1], **Kaisa Vent[1]**, Ivan Vasilyev[1], Oliver Bollverk[1], Anto Aasa[2]

margus.tiru@positium.com, gerttu.pilas@positium.com, erki.saluveer@positium.com, kaisa.vent@positium.com, ivan.vasilyev@positium.com, anto.aasa@ut.ee

[1] Positium OÜ, [2]University of Tartu Mobility Lab

Keywords: population census, signs-of-life, mobile positioning data anchor point, register-based census

The modern register-based census offers significant cost and efficiency savings compared to traditional field enumeration, but is wholly dependent on the quality of administrative data, which can be plagued by inaccuracies. This paper builds on the theoretical concept and practical experience based on the work Ahas et al, and Sõstra and Lehto and presents a conceptual approach on how data from mobile network operators (MNOs)—specifically mobile positioning data (MPD)—can serve as a powerful, cross-cutting tool to enhance the quality and cost-effectiveness of a register-based census. The framework focuses on two options of using MPD: validation of the usual place of residence and other geographical locations related to the respondents, and the use of MPD in a signs-of-life process along other registers.

This paper does not concentrate on the data privacy and protection aspects of using MPD. The technological and administrative methods such as privacy preserving enhancement technologies (PET) or national legislation related to such practice must ensure the privacy of the subscribers, and can be discussed in separate discussions.

Validation of the home address–a core challenge for any register–based census is a mismatch between the official registered address (de jure residence) and an individual's actual place of living (de facto residence). MPD provides a dynamic solution for the validation of the place of residence based on the population register or other locations based on signs-of-life. By processing longitudinal MPD location data, a probable home location (home anchor point) can be derived for each device based on diurnal and geographical patterns and activity over a defined reference period. This algorithmically determined location can be compared against the register-based address. Discrepancies between the de facto and de jure locations serve as a critical flag, highlighting records that may be outdated due to internal migration or the quality of the registers. Moreover, home anchor points can be used in cases where the place of residence is unknown.

Alternatively, MPD can be used within a signs-of-life process along population register and other registers to identify de facto residence and other relevant locations like work-place, secondary homes, etc., and to potentially be used in identifying household information based on similarities of geospatial behaviour of the household members.

By integrating MPD as a validation layer or a part of the signs-of-life process, national statistical offices can mitigate the principal weaknesses of a register-based system related to the home residency of the respondents. This hybrid approach moves beyond static administrative records to create a dynamic, continuously verifiable population frame. While careful methodological consideration must be given to privacy, data security, and representation biases, the use of MPD for residence validation and signs-of-life checks provides the necessary assurance to build a trustworthy, accurate, and modern register-based census.

Measuring internal migration using mobile phone data: evidence from The Gambia

Lamin L Dibba[1], Ayumi Arai[2], Moritz Meyer[3], Apichon Witayangkurn[4], Tafsiru Loum[5]

lamin.dibba77@gmail.com

[1]Gambia Bureau of Statistics – GBoS, [2]University of Tokyo, Reitaku University, [3]World Bank, [4]Thammasat University, [5]Public Utilities Regulatory Authority - PURA

Keywords: internal migration, Call Detail Records (CDR), mobile phone data, migration statistics, official statistics

Internal migration plays a major role in shaping demographic and economic dynamics in developing countries. Over the past decades, numerous studies have identified it as a key driver of economic development, largely due to persistent economic and structural disparities between rural and urban areas. However, accurately measuring internal migration remains challenging, as tracking individuals over time is costly and logistically difficult in settings with limited administrative infrastructure.

Innovations in mobile technologies have rapidly advanced data infrastructure by generating digital footprints of people at scales and resolutions that traditional surveys cannot achieve. Call Detail Records (CDRs), which are generated when individuals use mobile phones, capture the time and cell tower location of each phone interaction, offering a novel data source for population movement analysis.

This study uses 20 months of pseudonymized CDR data, spanning from January 2023 to August 2024, to assess internal migration in The Gambia. To address non-random selection and mobile ownership biases, we incorporate data from the 2024 Population and Housing Census to adjust and validate migration estimates. The combined dataset enhances the robustness of our analysis, particularly for understanding rural-to-urban migration patterns.

This study uses 20 months of anonymized CDR data from January 2023 to August 2024 to assess internal migration in The Gambia. To address non-random selection and mobile ownership biases, we incorporate data from the 2024 Population and Housing Census to adjust and validate migration estimates. The combined dataset enhances the robustness of our analysis, particularly for understanding rural-to-urban migration patterns.

This study demonstrates the potential of CDR data as a valuable source for producing frequent and granular migration statistics. We examine rural-to-urban migration patterns using 20 months of CDR data from January 2023 to August 2024. In addition, we update the measurement of internal migration in The Gambia using CDR data from 2020 to explore how migration patterns have evolved over time, particularly in response to post-pandemic recovery and changing economic conditions.

The results show intra-LGA migration is more common than inter-LGA migration in The Gambia, indicating the investment in decentralized services in the headquarters of rural LGAs could improve access to essential services and help reduce rural poverty and inequality. Strengthening rural institutions to enhance service delivery and expand economic opportunities—particularly in rural LGA headquarters—may improve household wellbeing, reduce unequal access to services, and increase incentives for rural investment. By comparing findings from the previous study with the updated analysis using the new dataset, we examine changes in the intensity, direction, and drivers of internal migration, with a focus on evolving roles of urban centres and emerging rural hubs.

Enhanced Dasymetric Mapping of Mobile Phone Data: An EM Algorithm Approach with Deep Geographic Features

David Crompton[1], Alexey Sirotkin[1], Tor-Morten Grønli[2]

david.crompton@telcofy.ai, alexey.sirotkin@telcofy.ai, Tor-Morten.Gronli@kristiania.no

[1]Telcofy, [2]Kristiania University of Applied Sciences

Keywords: Daysmetric interprolation Geographic features

The spatial distribution of human populations varies dramatically across space and time, yet traditional dasymetric mapping approaches for mobile phone data rely on static weight assignments and external surveys for temporal patterns. While recent advances like Deep Gravity have demonstrated the power of rich geographic features and non-linear modeling for predicting mobility flows, the inverse problem—inferring fine-grained population distributions from mobile phone data—remains inadequately addressed by existing methods.

Current dasymetric interpolation techniques, exemplified by Järv et al. (2017) and standardized in the EU-funded Eurostat MNO project, use predetermined weights based on ancillary data and depend on external time-use surveys for temporal modeling. These approaches suffer from three key limitations: (1) static weight assignments that cannot adapt to local variations in mobile phone usage patterns, (2) inadequate handling of spatial uncertainty in cell tower coverage areas, and (3) oversimplified treatment of residential areas without distinguishing between permanent and transient populations.

Building on insights from Deep Gravity regarding the importance of comprehensive geographic features and non-linear relationships in mobility modeling, we propose a novel Expectation-Maximization (EM) framework for dasymetric mapping that addresses these limitations through data-driven learning rather than predetermined assumptions. Our approach leverages the same rich OpenStreetMap-derived features used in Deep Gravity—including detailed land use, transportation networks, commercial facilities, and residential classifications—but applies them to the complementary problem of population inference from telecommunications data.

Our method introduces four key innovations: First, we implement an iterative EM algorithm that learns optimal population distribution weights directly from observed mobile phone patterns, eliminating dependence on static assumptions. Second, we test different approaches to modeling cell tower coverage areas and evaluate their impact on device placement accuracy, providing insights into optimal spatial uncertainty handling. Third, we distinguish between permanent residential areas (apartments, houses, dormitories) and transient residential areas (hotels, short-term rentals) to better capture tourism and temporary population dynamics. Fourth, we employ machine learning techniques to discover temporal patterns directly from the data rather than relying on external time-use surveys.

The algorithm alternates between an E-step that distributes observed mobile device counts to geographic grid cells based on current population estimates and coverage area modeling, and an M-step that updates population distribution parameters by learning from geographic features and temporal patterns. This iterative approach allows the model to discover non-linear relationships between geographic characteristics and actual mobile phone usage patterns, adapting to local contexts in ways that static methods cannot.

Our residential-focused grid design enables more accurate modeling of nighttime populations—critical for applications in urban planning, emergency management, and epidemiological modeling. Our approach represents the first integration of Deep Gravity-inspired feature engineering with dasymetric mapping principles, bridging mobility flow modeling and population distribution estimation.

By learning rather than assuming population-geography relationships, our method offers improved accuracy and generalizability compared to existing dasymetric approaches, particularly in areas with complex land use patterns or significant tourism activity. This work opens new possibilities for real-time, adaptive population mapping using telecommunications data.