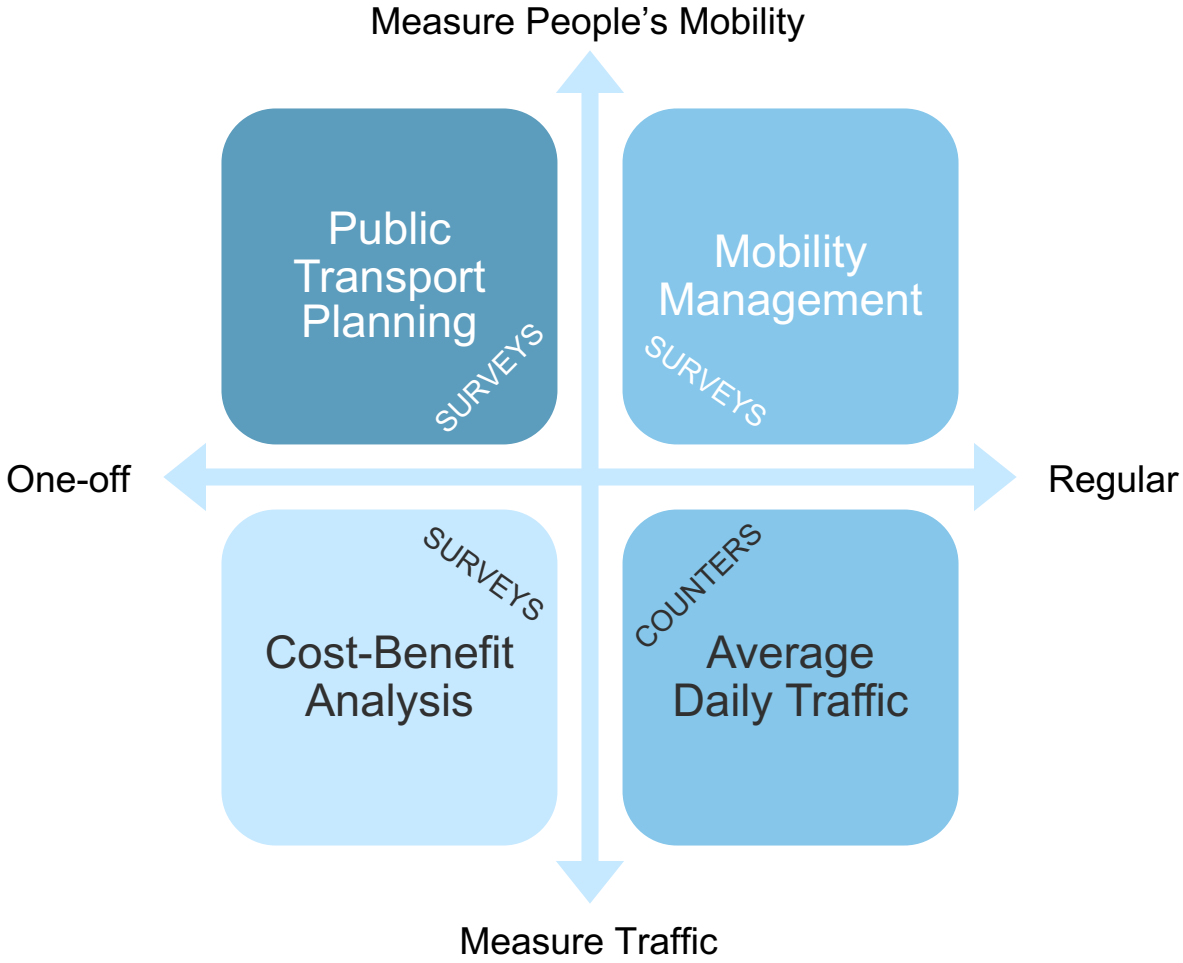
Siim Esko
International Project Development Lead
siim.esko@positium.com

Modelling for Transport:
Integration of Mobile Positioning Data with Other Datasets

Big Data integration with data for public transport and traffic counter data

UNECE Working Party on Transport Statistics (75th session)
25 April 2024

Why?


Measure People's Mobility
Public Transport Planning
SURVEYS
Mobility Management
SURVEYS
One-off
Regular
SURVEYS
Cost-Benefit Analysis
COUNTERS
Average Daily Traffic
Measure Traffic

positium

Fundamental questions in urban planning and transport

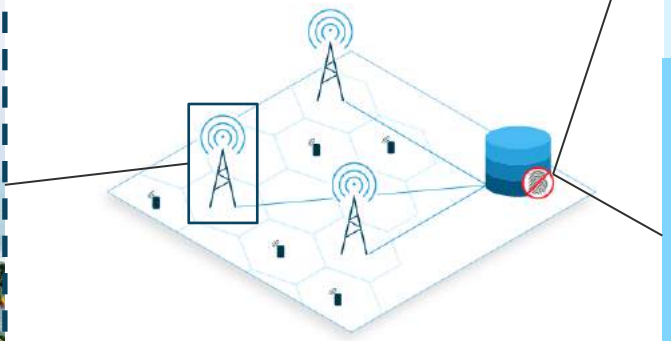
	Where are people?	Where do people move?
Long-term	Where people live and work	Home-work commuting
Short-term	Where people spend the day and night	Daily movement
Micro-term	Where people spend every 15 minutes	15-minute movement
Data sources	Registries & census (static) Surveys (static)	Traffic loops PT smart cards Mobile apps

Mobile positioning data (MPD)

MPD - the best dataset to connect dynamic population and origin-destination data on a strategic scale

Mobile Positioning Data (MPD)

Mobile Positioning Data is defined as any type of mobile network event data that are stored by the mobile network operator (MNO) that includes a subscriber identifier, time attribute and location.



Call Detail Records (CDR)

- Passively generated when a subscriber:
  Makes or receives a call
  Sends or receives an SMS
  Uses mobile data
- Routinely stored by MNOs for billing purposes

Passive Signalling Data

- Passively generated when a mobile device connects to the network
- More frequent than CDR data
- Very high volume storage

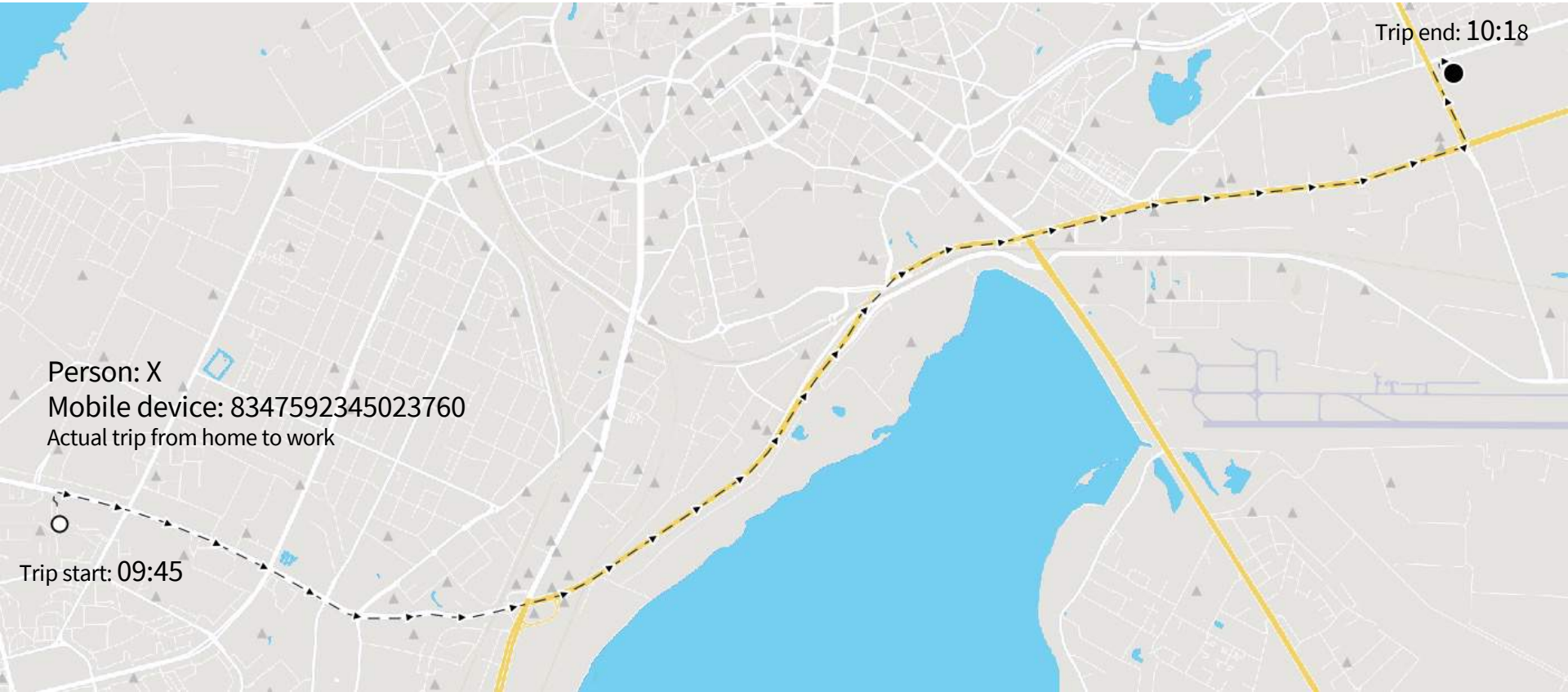
Model Reality in Statistics

Make a data model of the real mobility of people using data from mobile networks

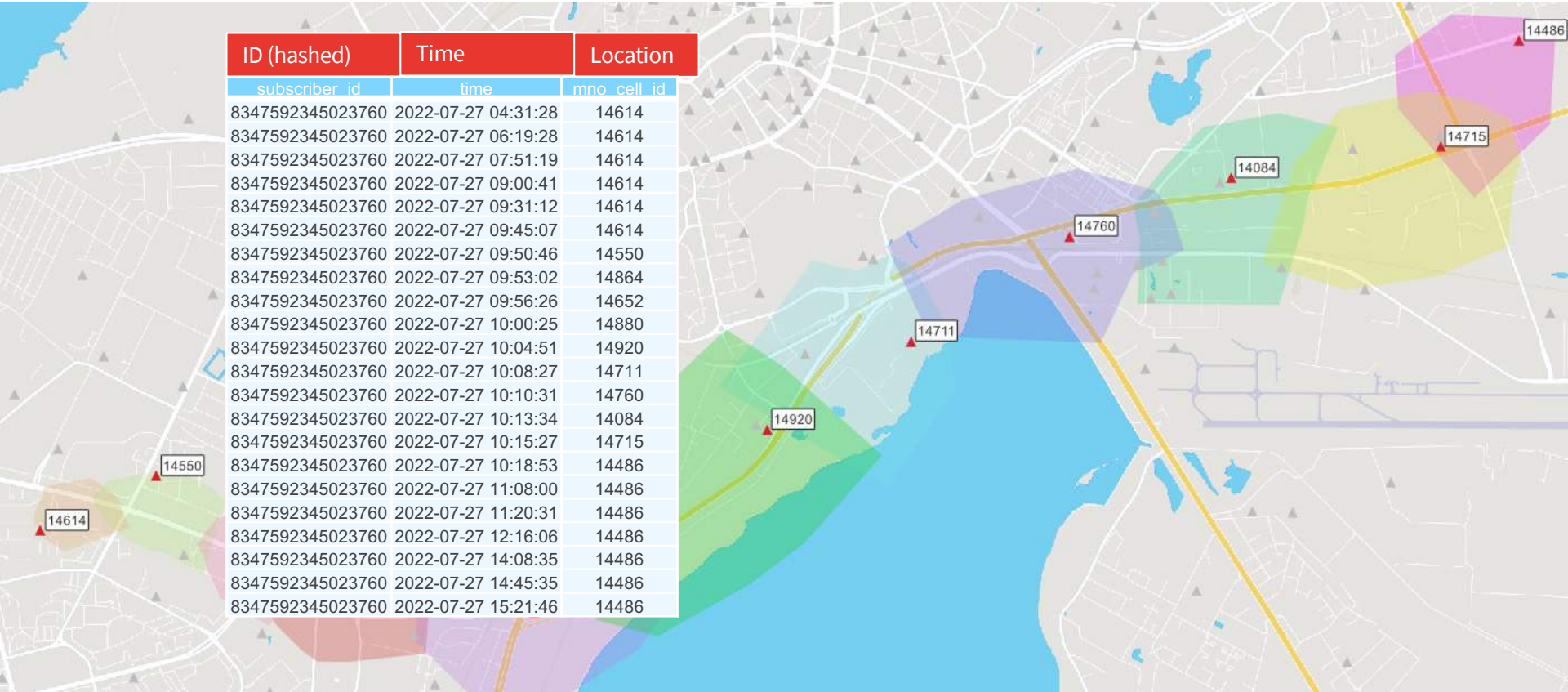
Video of Reality: Positium

Animation of MPD Model: Estonian National Museum

What is Mobile Positioning Data (MPD)?



What is Mobile Positioning Data (MPD)?

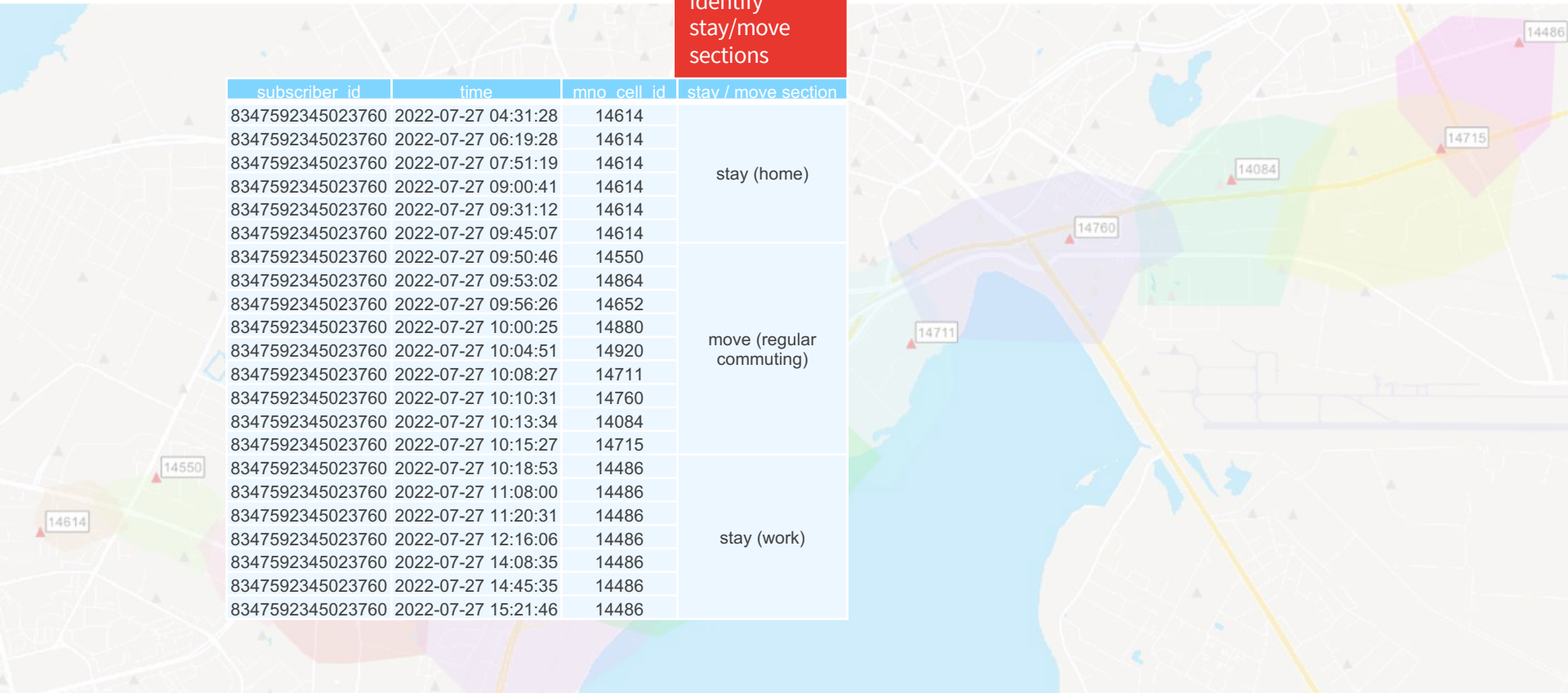


ID (hashed)	Time	Location
subscriber_id	time	mno_cell_id
8347592345023760	2022-07-27 04:31:28	14614
8347592345023760	2022-07-27 06:19:28	14614
8347592345023760	2022-07-27 07:51:19	14614
8347592345023760	2022-07-27 09:00:41	14614
8347592345023760	2022-07-27 09:31:12	14614
8347592345023760	2022-07-27 09:45:07	14614
8347592345023760	2022-07-27 09:50:46	14550
8347592345023760	2022-07-27 09:53:02	14864
8347592345023760	2022-07-27 09:56:26	14652
8347592345023760	2022-07-27 10:00:25	14880
8347592345023760	2022-07-27 10:04:51	14920
8347592345023760	2022-07-27 10:08:27	14711
8347592345023760	2022-07-27 10:10:31	14760
8347592345023760	2022-07-27 10:13:34	14084
8347592345023760	2022-07-27 10:15:27	14715
8347592345023760	2022-07-27 10:18:53	14486
8347592345023760	2022-07-27 11:08:00	14486
8347592345023760	2022-07-27 11:20:31	14486
8347592345023760	2022-07-27 12:16:06	14486
8347592345023760	2022-07-27 14:08:35	14486
8347592345023760	2022-07-27 14:45:35	14486
8347592345023760	2022-07-27 15:21:46	14486

What is Mobile Positioning Data (MPD)?

Identify stay/move sections

subscriber_id	time	mno_cell_id	stay / move section
8347592345023760	2022-07-27 04:31:28	14614	stay (home)
8347592345023760	2022-07-27 06:19:28	14614	
8347592345023760	2022-07-27 07:51:19	14614	
8347592345023760	2022-07-27 09:00:41	14614	
8347592345023760	2022-07-27 09:31:12	14614	
8347592345023760	2022-07-27 09:45:07	14614	
8347592345023760	2022-07-27 09:50:46	14550	move (regular commuting)
8347592345023760	2022-07-27 09:53:02	14864	
8347592345023760	2022-07-27 09:56:26	14652	
8347592345023760	2022-07-27 10:00:25	14880	
8347592345023760	2022-07-27 10:04:51	14920	
8347592345023760	2022-07-27 10:08:27	14711	
8347592345023760	2022-07-27 10:10:31	14760	
8347592345023760	2022-07-27 10:13:34	14084	
8347592345023760	2022-07-27 10:15:27	14715	
8347592345023760	2022-07-27 10:18:53	14486	stay (work)
8347592345023760	2022-07-27 11:08:00	14486	
8347592345023760	2022-07-27 11:20:31	14486	
8347592345023760	2022-07-27 12:16:06	14486	
8347592345023760	2022-07-27 14:08:35	14486	
8347592345023760	2022-07-27 14:45:35	14486	
8347592345023760	2022-07-27 15:21:46	14486	



What is Mobile Positioning Data (MPD)?

Assign to transportation zone

subscriber_id	time	mno_cell_id	stay / move section	transportation zone
8347592345023760	2022-07-27 04:31:28	14614	stay (home)	A
8347592345023760	2022-07-27 06:19:28	14614		
8347592345023760	2022-07-27 07:51:19	14614		
8347592345023760	2022-07-27 09:00:41	14614		
8347592345023760	2022-07-27 09:31:12	14614		
8347592345023760	2022-07-27 09:45:07	14614		
8347592345023760	2022-07-27 09:50:46	14550	move (regular commuting)	B
8347592345023760	2022-07-27 09:53:02	14864		
8347592345023760	2022-07-27 09:56:26	14652		
8347592345023760	2022-07-27 10:00:25	14880		C
8347592345023760	2022-07-27 10:04:51	14920		
8347592345023760	2022-07-27 10:08:27	14711		
8347592345023760	2022-07-27 10:10:31	14760		
8347592345023760	2022-07-27 10:13:34	14084		D
8347592345023760	2022-07-27 10:15:27	14715		E
8347592345023760	2022-07-27 10:18:53	14486	stay (work)	
8347592345023760	2022-07-27 11:08:00	14486		
8347592345023760	2022-07-27 11:20:31	14486		
8347592345023760	2022-07-27 12:16:06	14486		
8347592345023760	2022-07-27 14:08:35	14486		
8347592345023760	2022-07-27 14:45:35	14486		
8347592345023760	2022-07-27 15:21:46	14486		

What is Mobile Positioning Data (MPD)?

Routing to road segments

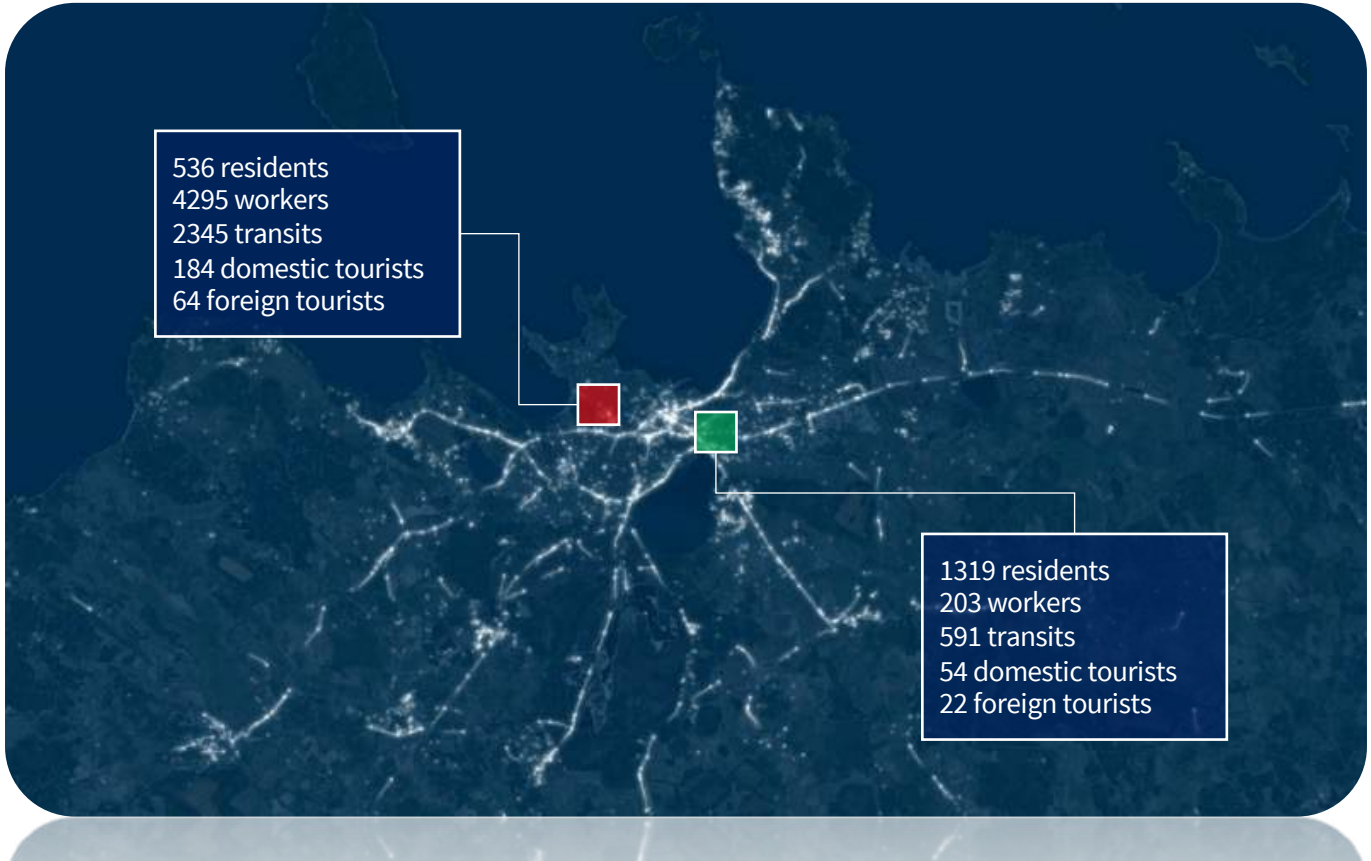
subscriber_id	time	mno_cell_id	stay / move section	transportation zone	road move segment
8347592345023760	2022-07-27 04:31:28	14614	stay (home)	A	
8347592345023760	2022-07-27 06:19:28	14614			
8347592345023760	2022-07-27 07:51:19	14614			
8347592345023760	2022-07-27 09:00:41	14614			
8347592345023760	2022-07-27 09:31:12	14614			
8347592345023760	2022-07-27 09:45:07	14614			
8347592345023760	2022-07-27 09:50:46	14550	move (regular commuting)		Local 513
8347592345023760	2022-07-27 09:53:02	14864		B	Local 513
8347592345023760	2022-07-27 09:56:26	14652			E23
8347592345023760	2022-07-27 10:00:25	14880		C	E23
8347592345023760	2022-07-27 10:04:51	14920			E23
8347592345023760	2022-07-27 10:08:27	14711			E23
8347592345023760	2022-07-27 10:10:31	14760			E23
8347592345023760	2022-07-27 10:13:34	14084		D	Local 823
8347592345023760	2022-07-27 10:15:27	14715		E	Local 8243
8347592345023760	2022-07-27 10:18:53	14486	stay (work)		
8347592345023760	2022-07-27 11:08:00	14486			
8347592345023760	2022-07-27 11:20:31	14486			
8347592345023760	2022-07-27 12:16:06	14486			
8347592345023760	2022-07-27 14:08:35	14486			
8347592345023760	2022-07-27 14:45:35	14486			
8347592345023760	2022-07-27 15:21:46	14486			

What is Mobile Positioning Data (MPD)?

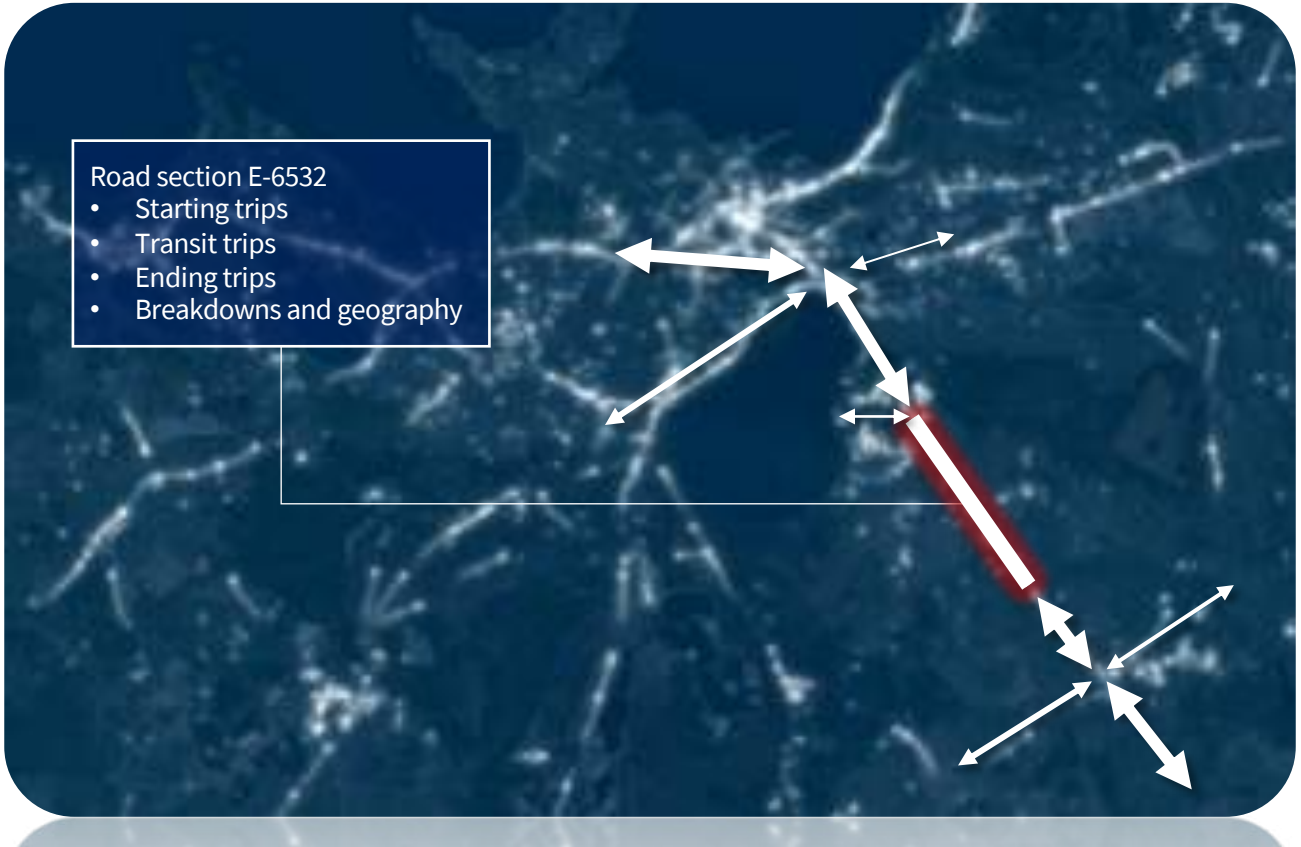
Assign origin, destination, transit

subscriber_id	time	mno_cell_id	stay / move section	transportation zone	road move segment	otd
8347592345023760	2022-07-27 04:31:28	14614	stay (home)	A		origin
8347592345023760	2022-07-27 06:19:28	14614				
8347592345023760	2022-07-27 07:51:19	14614				
8347592345023760	2022-07-27 09:00:41	14614				
8347592345023760	2022-07-27 09:31:12	14614				
8347592345023760	2022-07-27 09:45:07	14614				
8347592345023760	2022-07-27 09:50:46	14550	move (regular commuting)		Local 513	transit
8347592345023760	2022-07-27 09:53:02	14864		B	Local 513	
8347592345023760	2022-07-27 09:56:26	14652			E23	
8347592345023760	2022-07-27 10:00:25	14880		C	E23	
8347592345023760	2022-07-27 10:04:51	14920			E23	
8347592345023760	2022-07-27 10:08:27	14711			E23	
8347592345023760	2022-07-27 10:10:31	14760			E23	
8347592345023760	2022-07-27 10:13:34	14084		D	Local 823	
8347592345023760	2022-07-27 10:15:27	14715		E	Local 8243	destination
8347592345023760	2022-07-27 10:18:53	14486	stay (work)			
8347592345023760	2022-07-27 11:08:00	14486				
8347592345023760	2022-07-27 11:20:31	14486				
8347592345023760	2022-07-27 12:16:06	14486				
8347592345023760	2022-07-27 14:08:35	14486				
8347592345023760	2022-07-27 14:45:35	14486				
8347592345023760	2022-07-27 15:21:46	14486				

Insights based on MPD

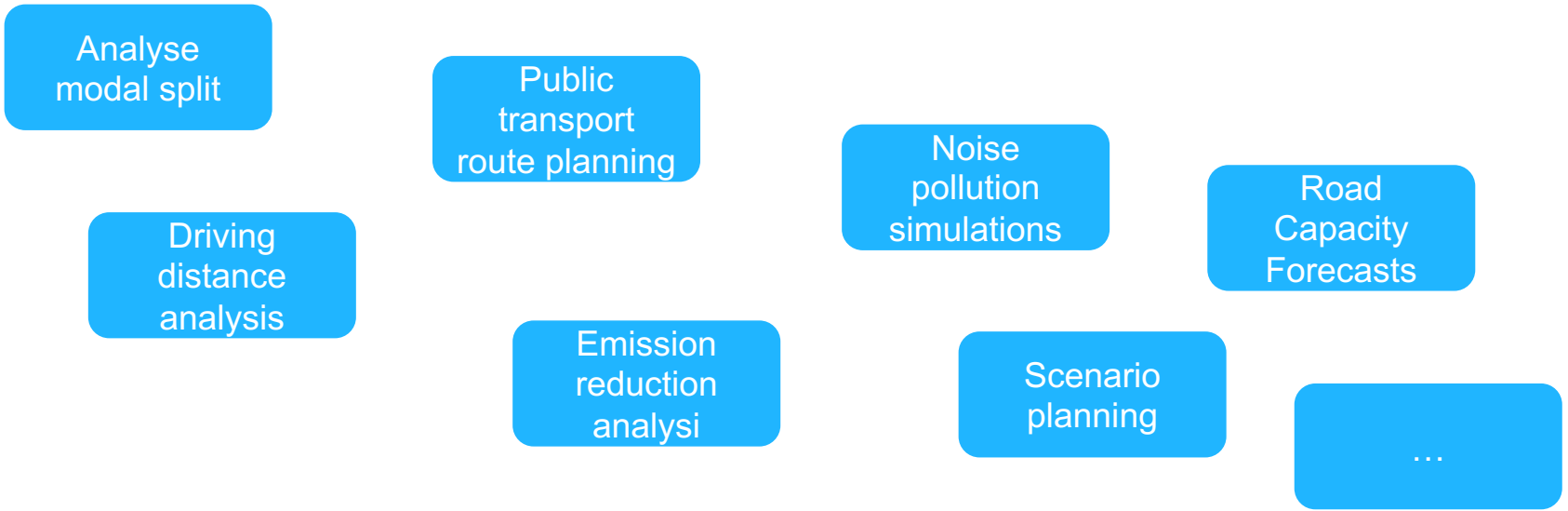


Road Network Density / Demand



Integration Possible with Transportation Tools and Models

CUBE, VISSUM, OPENTRACK, EMME, Remix, custom models



Mobile Positioning Data Projects Globally – Positium

Key stats:
20+ mobile operators
15+ countries
300+ projects
20+ use cases
1 Billion+ population

Inbound, outbound and domestic tourism methodology and technology

Official statistics production

Transport use cases

Where the next level of the railway and city public transport efficiency was achieved thanks to the analysis and decisions based on MPD

Republic of Estonia
Ministry of Economic Affairs and Communications

Inter-city transport analysis: Riga-Tallinn-St.Petersburg Railway Line

Client: Estonian Ministry of Economic Affairs and Communications:

Objective: Analyse the demand for potential Tallinn-Tartu-Riga and Tallinn-St. Petersburg railway lines and forecast changes from 2015 to 2030.

Data sources: The analysis considered passenger counts, border statistics, and other sources to estimate existing travel demand.

Result: Full Cost-Benefit Analysis and specific recommendations for maximizing economic benefits

Urban transport analysis: Tartu City Public Transport

Client: Tartu City Government

Objective: Big data-driven public transport remodelling

Data sources: 20 layers of data, including mobile positioning data, registries, public transportation check-ins, land use data, survey.

Result: From data collection to start of operation of new bus lines in 2 years, resulting in +15% increased trips and 86% satisfaction rate

Key to Getting Data on Journeys – MPD for Mobility Demand

Challenge: Tackle data scarcity with big data as the core mobility dataset

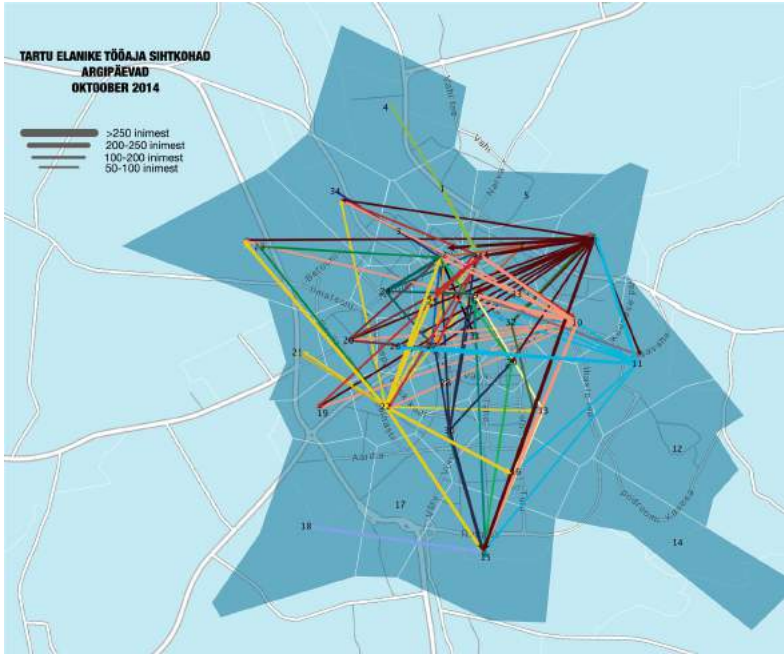
Data sources:

1. **Mobile positioning data** was the main data source for mobility demand, and for home and workplace figures

Tartu City

Population: 100,000

Movement combinations: 6,000,000



Origin-Destination Matrices between city districts from MPD

Key to Solving Data Scarcity – Data Integration

Challenge: Tackle data scarcity by using 20 layers of data in an integrated way

Data sources: Mobile positioning data is used to compare to:

2. **Registries** to validate population data (home, work, school),
3. **Public transportation check-ins** to measure unmet demand,
4. **Land use data layers** to characterize transport zones,
5. **Surveys** to add qualitative aspects,
6. **Journey planner application** to gather user feedback about old vs new network

Data layers integrated through location attributes

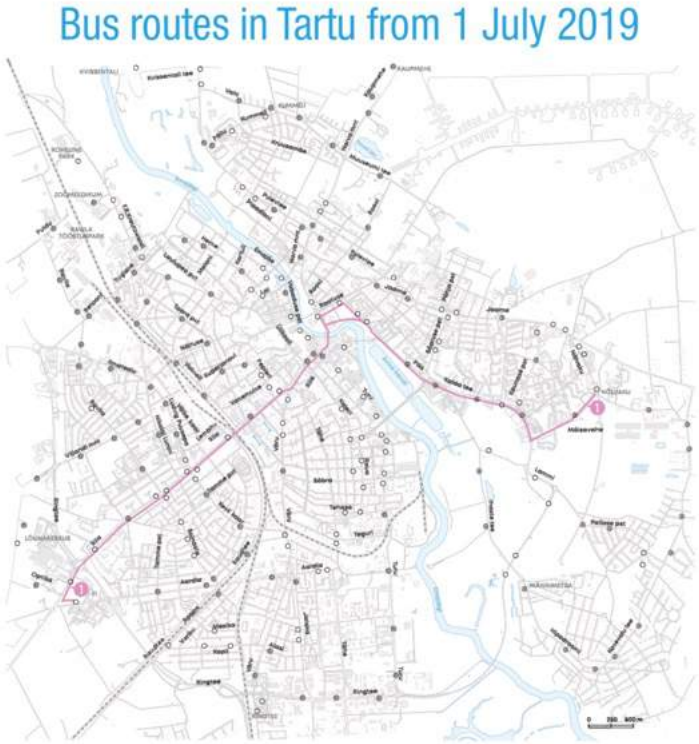
Key to a Time-Efficient Process – Data-First Approach

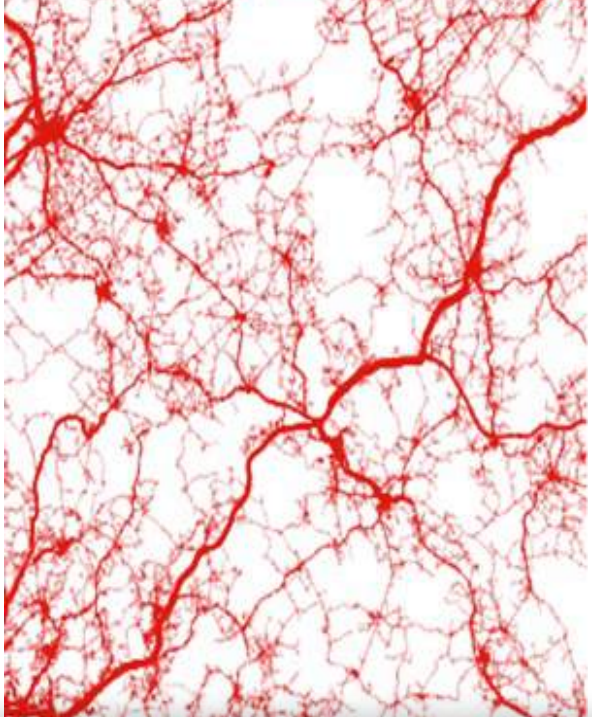
Challenge: 2 years from data collection to new routes in operation

With a data-driven process, followed up with qualitative feedback, it was possible to achieve results:

- **2 years from data collection to starting** the new bus line network
- **30% more people** use public transport regularly
- **86% are satisfied** with the new network, across all age groups

And after 3 more years, the lines were expanded to nearby municipalities with a repeat of the process.





AADT use case

Estonian Transport Administration road maintenance plan with AADT based on MPD and counter data

Challenge: Independent tool to measure road traffic on all roads, not only those with permanent counters

The Estonian Transport Administration, in collaboration with us and Telia, a leading Mobile Network Operator (MNO) in Estonia, has developed an innovative solution to estimate the Annual Average Daily Traffic (AADT) for the entire region's roads.

Machine learning model was applied as an independent tool for AADT calculation, combining the strengths of both datasets.

Transportation planners rely on AADT numbers to make informed decisions about infrastructure investments and improvements for more than 4000 road segments.

“From the beginning of our partnership, the team at Positium demonstrated a strong understanding of our objectives and challenges, and their expertise in combining existing counter data with mobile positioning data resulted in a cost-effective, efficient, and accurate solution.”

– Reimo Tarkiainen, Head of Data and Analysis Dept, Transport Administration

1.5M
OD pairs from mobile network

100+ permanent counters
and 1100 temporary counter locations

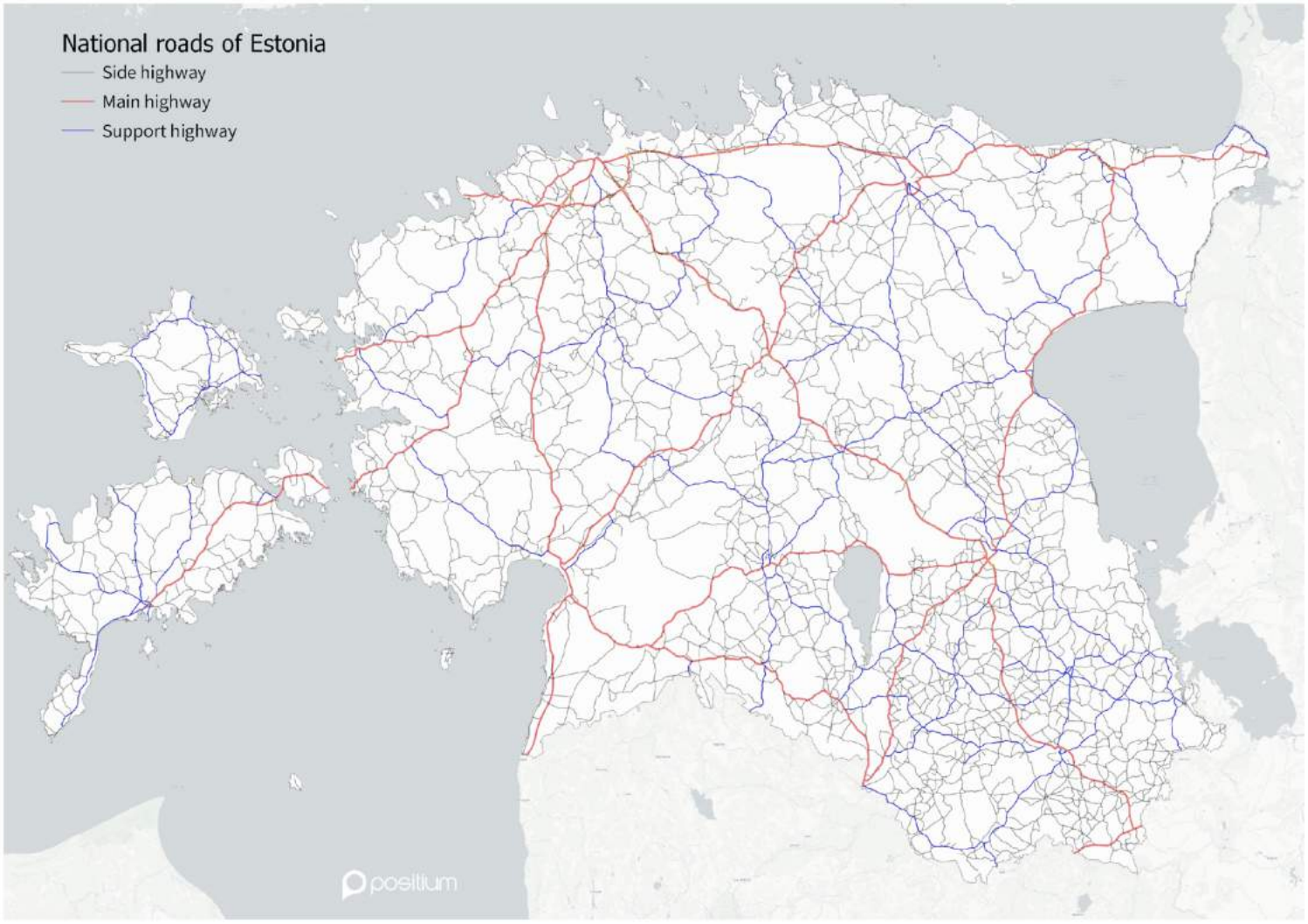
4000
road segments

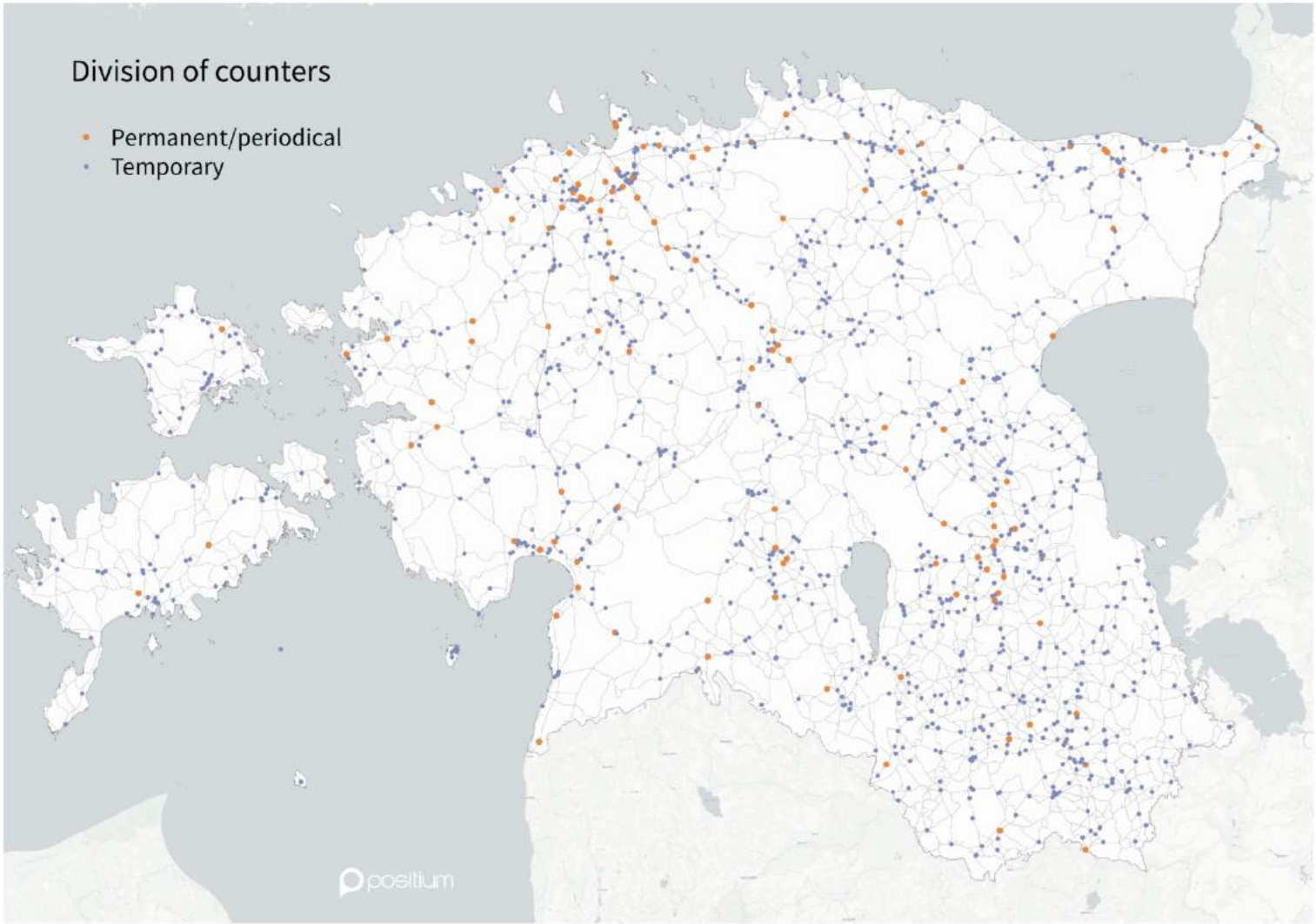
DATA FROM MNO AND SENSORS

AADT FOR THE COUNTRY

METHODOLOGY DOCUMENT

QUALITY ASSESSMENT


National roads of Estonia
Side highway
Main highway
Support highway
positium


Division of counters
Permanent/periodical
Temporary
positium

ADT delivery approach – 4 to 5 months

1. Data Collection

~ 6 weeks

2. ML Model Development

~ 5 weeks

3. ADT Estimation and Validation

~ 4 weeks

4. Final Results

~ 1-2 weeks

ACTIVITIES

1. Collect and store the data
2. Data qualification into types
3. Input data cleansing:
   - OD matrices
   - Counter locations
   - Counter results
   - Roads topology
4. Data prioritization methodology
5. Combination and grouping of different data types

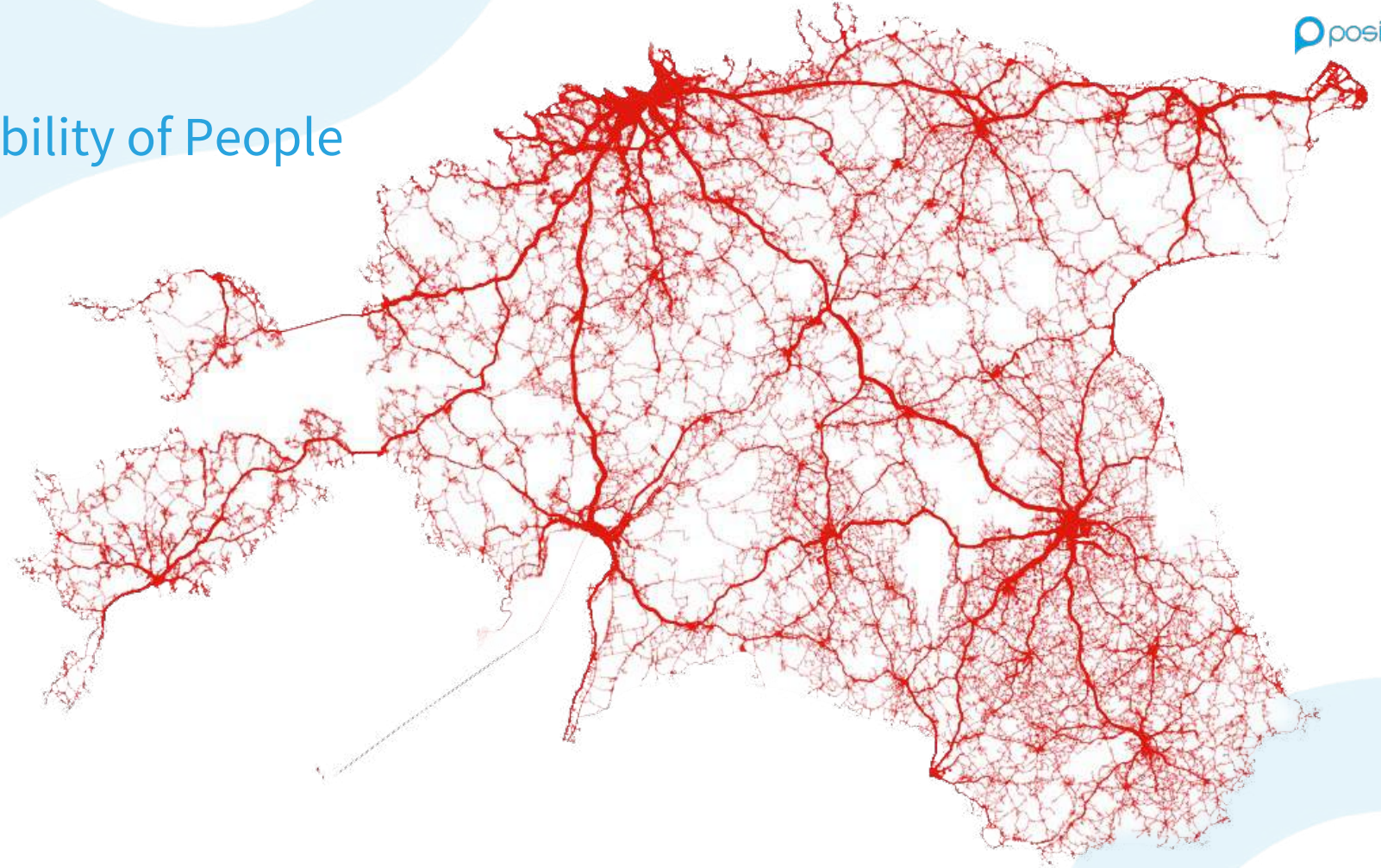
1. Develop the approach for utilization of sensors data, OD matrices and survey data in one model
2. Find ADT values for roads with all available data
3. Test various approaches to achieve the best quality and granularity of the results

> XGBoost

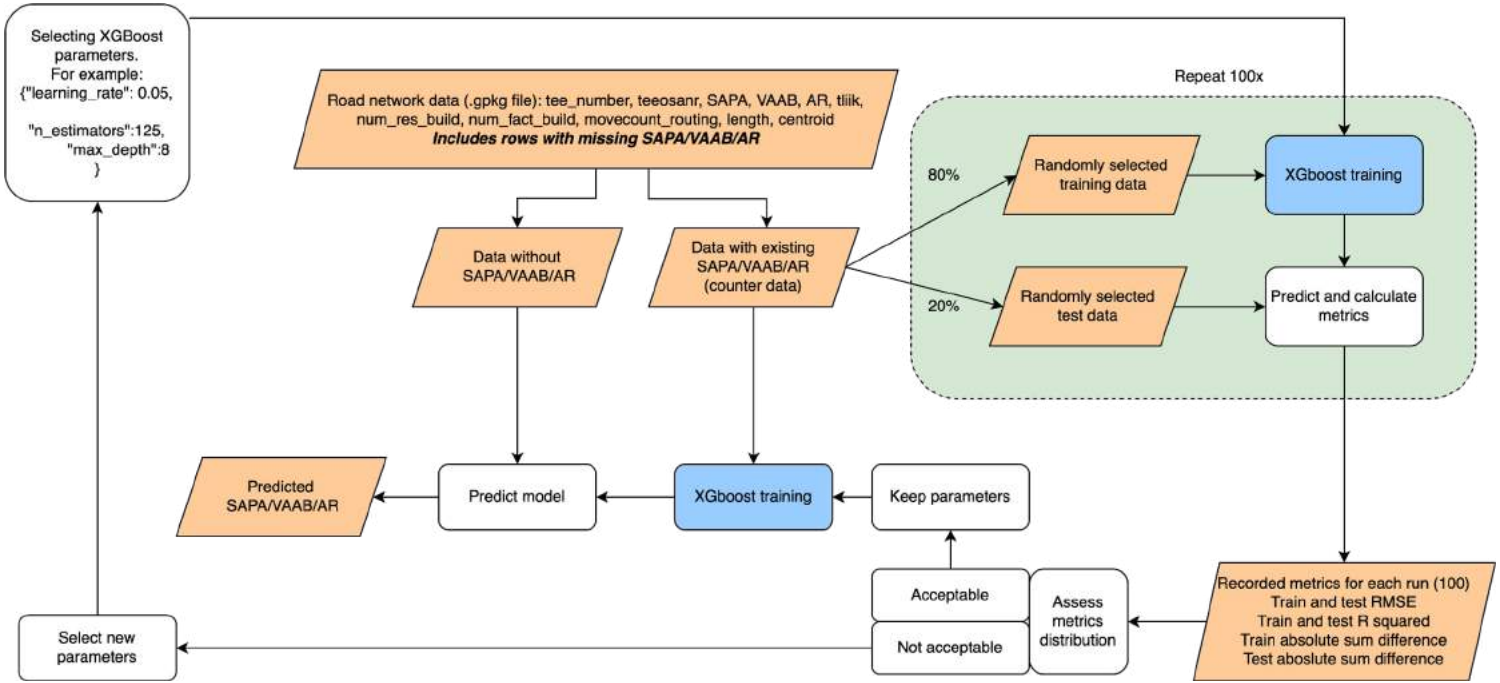
1. Iterative modelling and analysis generation
2. Validation of the results

1. Result delivery
2. Tool delivery
3. Assist with results analysis and using results for business purposes
4. Comprehensive training

Mobility of People



Modelling: Open Python-based Machine Learning Pipeline

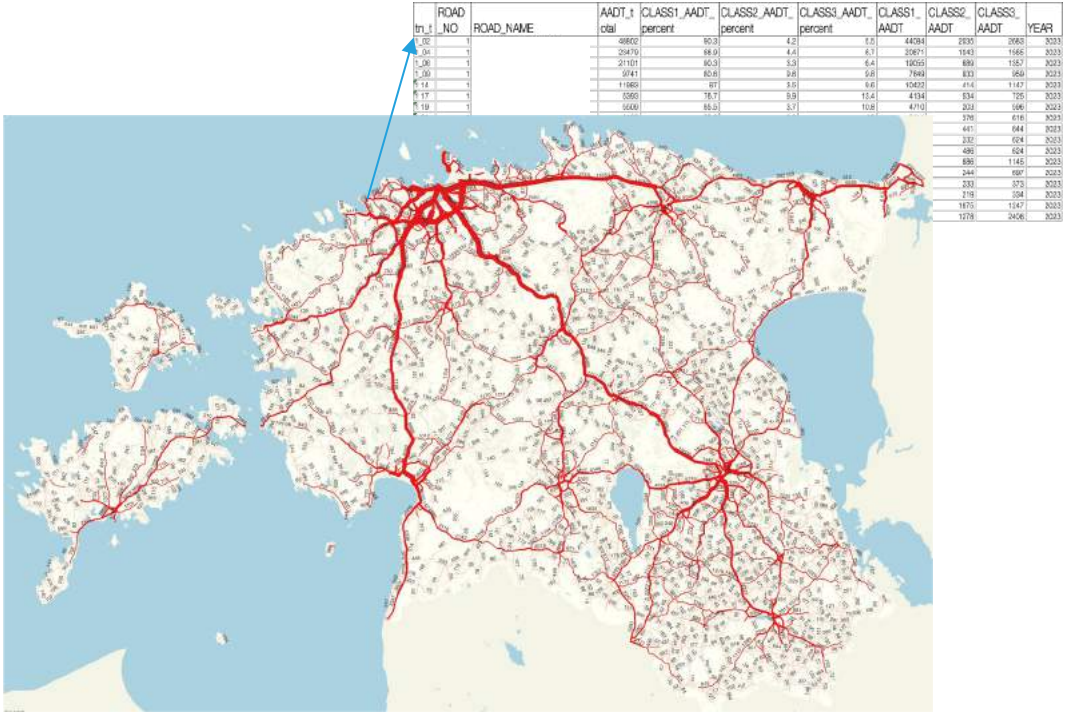


Result - Total AADT

Deliverables:

1. Results on AADT divided into vehicle classes for all roads
2. Self-service tool:

- Project methodology on 45 pages
- 4 Jupyter notebooks for Quality Assurance purposes
- 8 Python scripts



Evaluation Criteria

- ✓ Allowed difference on main roads +/- 8 %.
- ✓ Allowed difference on supporting roads +/- 15 %.
- ✓ Allowed difference on side roads +/- 30%

“The estimation of Annual Average Daily Traffic (AADT) using machine learning and mobile positioning data has been extremely successful and has significantly improved our ability to make informed decisions regarding infrastructure investments and improvements.”
– Reimo Tarkiainen, Head of Data and Analytics, Transport Administration

Key Takeaways

Recap: Tartu City's Transformation

Key Points:

- Use of MPD through transport modelling optimized the bus network.
- Increased ridership, improved efficiency, and user satisfaction.
- Demonstration of fast and effective urban transport remodelling.

2 years
to
86% satisfaction
rate

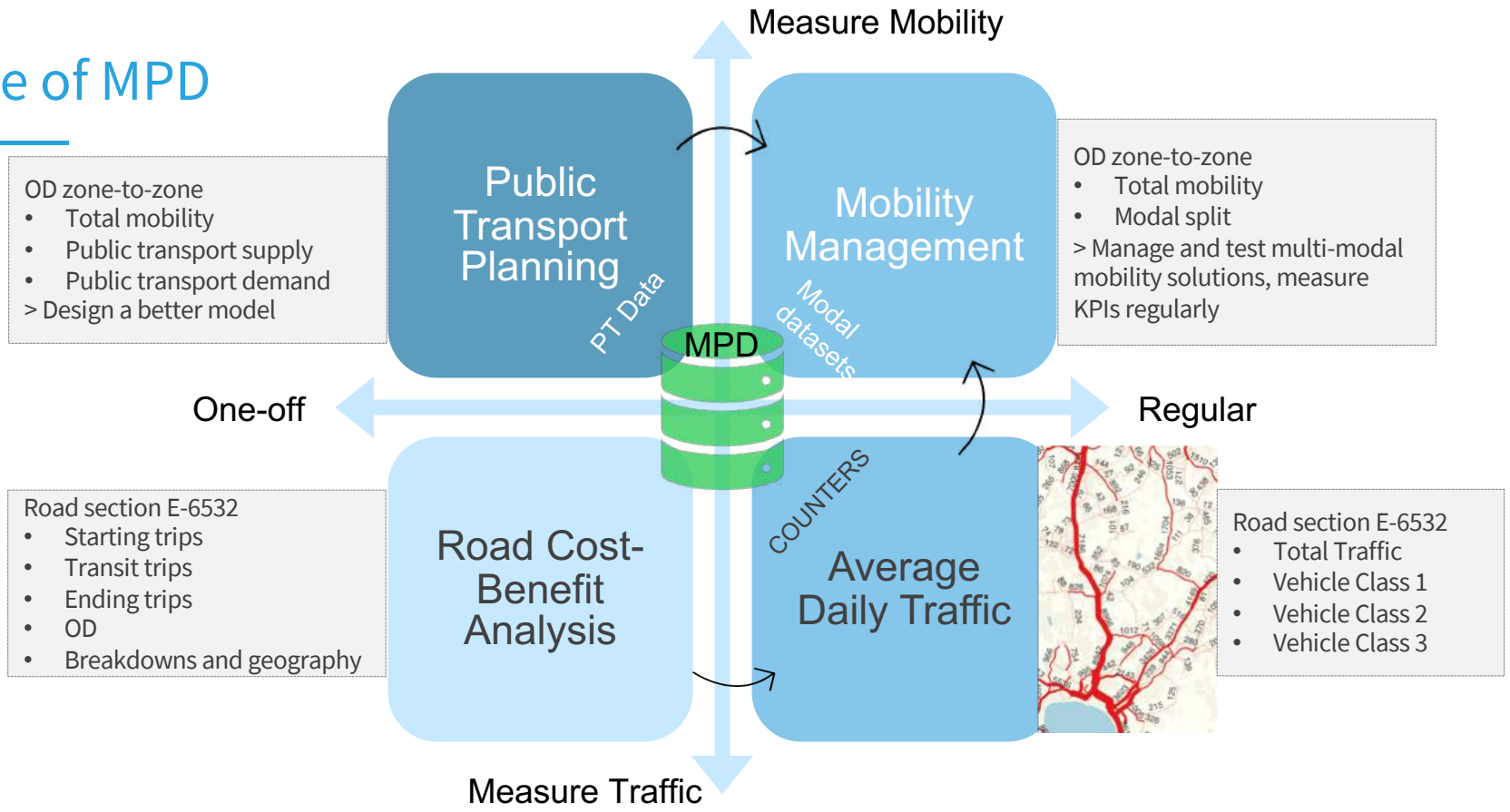
Recap: Innovations in AADT Estimation

Key Points:

- ML and MPD integration for comprehensive traffic analysis.
- Achieved high accuracy in AADT across diverse road types.
- Cost-effective and efficient approach for large-scale application.

Extend 100
counters to 4000
road segments
with MPD

Use of MPD



OD zone-to-zone
- Total mobility
- Public transport supply
- Public transport demand

> Design a better model

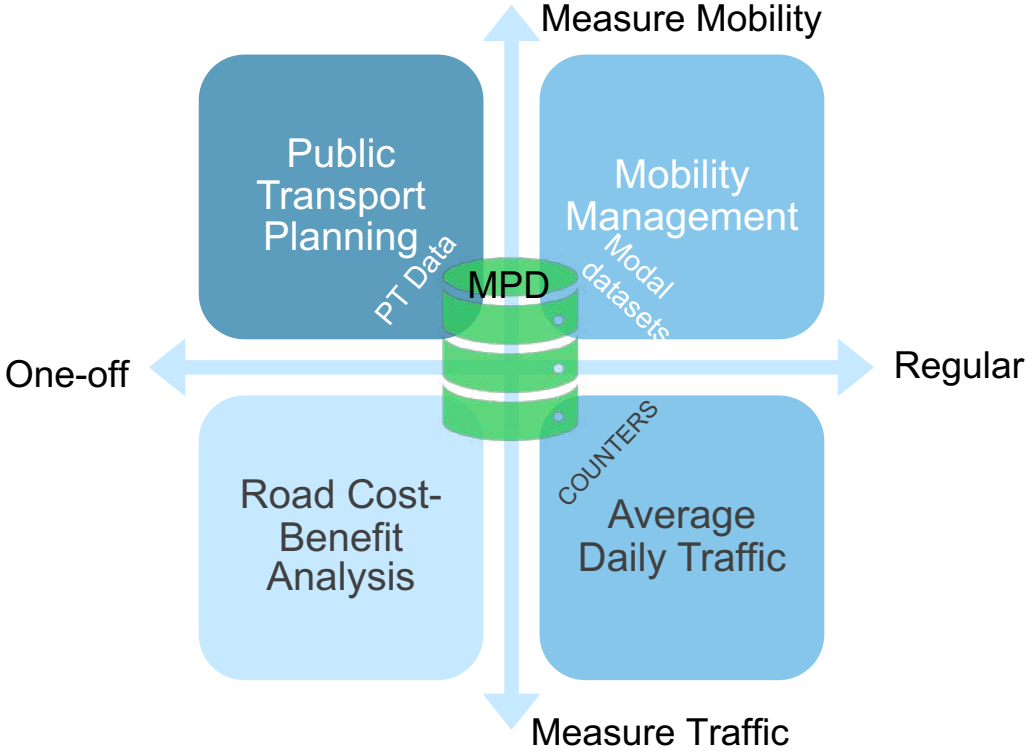
OD zone-to-zone
- Total mobility
- Modal split

> Manage and test multi-modal mobility solutions, measure KPIs regularly

Road section E-6532
- Starting trips
- Transit trips
- Ending trips
- OD
- Breakdowns and geography

Road section E-6532
- Total Traffic
- Vehicle Class 1
- Vehicle Class 2
- Vehicle Class 3

Digital Data Platform for Transportation



A Digital Data Platform with MPD as the basis allows **constant longitudinal mobility and traffic monitoring** across transport planning functions

Thank you!

We strive for the future where every country in the world benefits from mobile positioning data for the good of society

For further information, please contact:

Siim Esko: siim.esko@positium.com

Animation: Positium