15th ITS European Congress, Lisbon, Portugal, 22-24 May-2023

Paper ID #205

Visual traffic data analytics & digital traffic infrastructure

Juho Kostiainen[1*], Janne Rinne[2], Bart Adams[3], Michiel Van Hove[4], Leo Mäkelä[5], Janne Jokinen[6]

1. City of Helsinki, Finland, juho.kostiainen@hel.fi

2. Forum Virium Helsinki, Finland

3. xyzt.ai, Belgium

4. Geo Mobility, Belgium, michiel@geomobility.eu

5. Normiopaste, Finland

6. Marjetas Academy, Finland

Abstract

This paper presents results of data-related pilots in two topics: visual traffic data analytics and digital traffic infrastructure. The discussed pilots were done in the fall of 2022 between Mobility Lab Helsinki and several different companies. The first topic is addressed by a proof-of-concept pilot on the possibilities of data-driven mobility analytics, done with xyzt.ai and Geo Mobility. It explored how different floating vehicle data sets and modern visual analytics solutions can address the needs of traffic planners and researchers. Use cases included route analysis, delay times, identifying origins and destinations of trough traffic and speeding as well as looking into the coverage and applicability of the different data sets themselves. The other topic addresses the need to digitize and manage information about traffic control devices. The cases discussed look at comparing reality with plans as well as the use of machine vision for updating traffic sign information.

Keywords: traffic data, visual analytics

Introduction

Mobility Lab Helsinki

Mobility Lab Helsinki – the city's testbed for smart mobility – enables and helps companies and developers to experiment and co-create their novel smart mobility solutions in the real urban environment [1]. It is coordinated by the city and done in collaboration with Forum Virium Helsinki, the city's innovation company. A key focus area for the mobility lab is Digital Twin for Mobility – how traffic environment, traffic itself, and related conditions and context can be described with diverse data sources for different use cases [2].

City's needs related to traffic analytics

Understanding traffic flow and patterns is the basis for traffic planning and management as well as various smart mobility services. With more and more traffic data available, automation and analytics are needed to turn the data into relevant information. In 2019, City of Helsinki published the Helsinki Intelligent Transport System Development Programme 2030 [3]. It highlights this need for better data and tools for making use of it as a development priority, both in terms of collection of traffic data as well as the development of tools for situational and statistical awareness, monitoring, and reporting. During the past couple years, the city has developed data capabilities and harmonization of existing data sources as well as defined more specific user needs related to analysis tools.

Before getting to the analytics and the best ways obtain insights, you first need the data. There are various solutions and technologies available for this such as different sensors (e.g., cameras) and detectors (e.g., induction loops) as well as probe vehicles (i.e., Floating Car Data, FCD). With any data, it is important to know the quality and coverage. Different studies have proposed that the minimum probe share of total vehicles should be between 2-10% to provide reliable results. [4] When the penetration rate is sufficient, FCD can be a cost-effective option due to its scalability and insights beyond vehicle counts such as travel times and route choices [5].

The city's needs related to digital traffic infrastructure

The Finnish traffic code (729/2018) which came into effect in June 2020 requires information about new traffic control devices, such as traffic signs, to be delivered to the Finnish Transport Infrastructure Agency. The City of Helsinki is developing a City Infrastructure Platform ("Cityinfra") for managing all the information about the city's traffic control devices, which addresses this requirement.

The Cityinfra platform covers a wide range of data for different needs, including traffic signs, guideposts, barriers, traffic light plans and realizations, and is connected to planning tools (Figure 1). The platform consists of a database, an open REST-API and an admin interface. It has been developed especially for the needs of urban space processes; urban planning, construction works, maintenance, resource management and daily operations such as parking control. The platform's test environment includes example sets of plan documents and street view descriptions that can be used for testing purposes.

While the primary users of the Cityinfra platform are planners and civil servants first, the ultimate goal is to provide data to third parties, such as private companies and researchers. The users can range from those maintaining and repairing street infrastructure to ones developing new digital services, such as navigation, logistics and autonomous driving, that need to understand the context in which they operate. Besides managing the data, it of course first needs to be produced with proper attributes included. And the data also needs to be accurate and up-to-date, i.e., the reality should align with the documentation.

Focus of this paper

To support city's traffic information developments and to learn and experiment and build on the data needs of digital twins, Mobility Lab Helsinki did a proof-of-concept pilot in the fall of 2022 with xyzt.ai and Geo Mobility. The pilot aimed to experiment with and about state-of-the-art visual mobility analysis tools and building knowledge on the possibilities of traffic data sets available for the Helsinki region.

At the same time, to tackle the needs related to data on traffic control devices, Mobility Lab Helsinki's agile piloting process was used to try out potential solutions. Agile piloting is a flexible way to experiment innovative solutions in an urban environment [6]. In this process, the challenge for producing traffic control device data was defined based around the needs and future developments of the Cityinfra platform, after which an open call was published. Of the 20 proposals received, two were selected to be piloted between September and December 2022.

This paper is divided into two sections presenting the pilots and their results. The next section focuses on the visual traffic analytics while the latter one looks at the two pilots focused on digital traffic infrastructure. The sections cover the pilot setups and their more specific objectives as well as presents the findings and possibilities for different use cases. Finally, the conclusions section summarises the findings for both topics.

Visual traffic analytics – pilot description and use cases

Pilot setup

Mobility Lab Helsinki initiated and facilitated the pilot. The City of Helsinki traffic analysis experts and researchers participated as test users and in defining the use cases of their main interest. xyzt.ai provided the mobility analytics platform while Geo Mobility obtained data sets to use. The platform allows bringing in different data sources and provides a dashboard for visual, trend, distribution and origin-destination analytics (see Figure 1). The companies provided training and support for understanding and using the platform and data sets. The pilot started in the beginning of fall 2022. The data sets were acquired for the period of September 2022, and, in October-November, training sessions were organized and testing done. In December, the identified and explored use cases as well as the suitability of the data sets for different purposes were summarized.

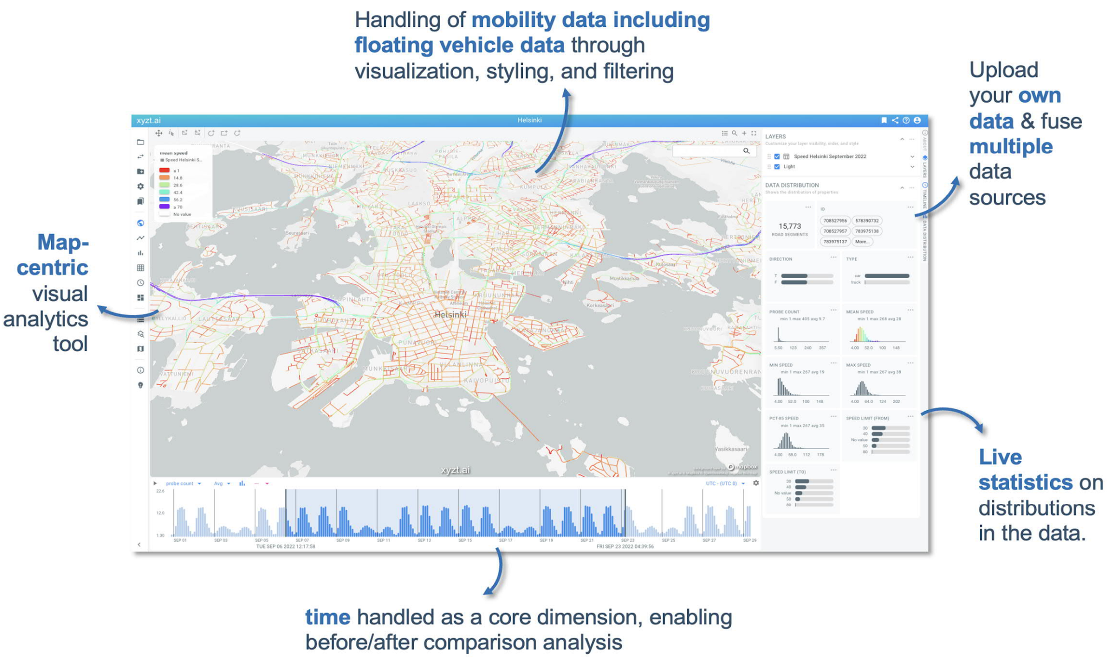

Figure 1 – The xyzt.ai visual analytics platform with temporal mobility data for the Helsinki region.

Main objectives

The underlying needs and interest are related to the city's ongoing work on developing its capabilities to better utilise data in traffic planning and management. For that work, user stories have been defined from the point of view of traffic planners and researchers through interviews, workshops and service design processes. The needs, on a general level, are about having a single flexible, visual user interface for bringing in, managing and understanding various data sets, and being able to analyse them, e.g., spatially, temporally and by transport modes.

While this pilot addresses much of these needs, it was neither done to directly find the answer, nor was it to do a deep dive into assessing specific situations of cases in the city. Rather, the goal was to gain knowledge and understanding of what can be done when the need arises, and what should be expected of modern solutions. The pilot was an exploratory proof of concept intended to learn about the possibilities of modern, data-driven mobility analysis tools.

The key interests and benefits of the pilot were the following:

- For the city of Helsinki: Get to understand the data quality and coverage and possible use cases and ROI for the city, and in addition being able to communicate the findings towards the wider mobility and smart city industry.
- For Geo Mobility and xyzt.ai: Build knowledge on regional data sources, coverage, quality, and challenges.

Use cases

While a wider range of more detailed analysis cases were discussed, the main ones explored that were consider most relevant for the users involved (traffic planners and researchers) and suitable to analyse were the following:

- Route choices: What routes are chosen and advised at different times. When and how often are different shortcuts or alternative routes taken.
- Congestion: When and where are delays occurring on the road network.
- Through travel: Where the traffic comes from and goes to on specific road sections.
- Speeding: Where and when does speeding occur.

Data sets

The focus was on floating vehicle data sets from HERE, TomTom and Google for the period of September 2022. Besides using the data for the previously described use cases, on point of interest was in understanding the coverage and applicability of data sources for the different needs. In addition, mobile network data from Telia was also incorporated to consider its possibilities. The city's traffic counter data from select locations was used as a reference and baseline to compare data coverage and quality. In addition, certain other data sets such as atmospheric, rainfall, and other types of local data sets were tested to look at the broader possibilities for analysing various data. Figure 2 below summarises the explored data sets.

Data Set	Summary	Origin
HERE road stats	For each road segment, contains information on number of vehicles, average speeds,… with hourly statistics.	Computed as aggregate floating vehicle probe data by HERE.
TomTom traffic stats	For each road segment, contains information on the number of vehicles, average speeds,… aggregated over a period of time (e.g., commute hours). There is no time series anymore in the resulting data set.	Downloaded from the TomTom platform after selecting a time period. Based on floating vehicle data.
TomTom origin destination	Trip counts and percentages between different regions. Statistics are aggregated data for periods selected by the user, e.g., morning/evening/entire day aggregated	Extracted from the TomTom platform after selecting an area and time period. Based on floating vehicle data.
ODIQ static link data	Fixed routes with information on how traffic flows (e.g., delays in seconds/km) along the routes during several days or weeks with an interval of between 1 and 15 minutes.	Obtained by querying the Google routing API over time for user selected origins and destinations.
Traffic counts	Contains (mostly) hourly statistics on traffic counts (number of vehicles) and average speeds at fixed locations.	Helsinki provided this information.
Telia origin destination	Contains movements (number of trips) between different areas in on grid cells or administrative boundaries.	Data is aggregate data obtained from cell phones on the Telia network.
Copernicus atmospheric monitoring data	Hourly measurements of different atmospheric particle concentrations such as NO2.	European Copernicus program, obtained through satellites.
Otonomo, Wejo, Bridgestone, INRIX,…	Raw floating vehicle data providers (GPS coordinates with associated velocities, accelerations, vehicle types, etc. of cars, trucks,…). This data was not used for the Finland region in the project, though there are sample projects with raw FVD data from INRIX that could be consulted.	From connected vehicles through OEMs (e.g., BMW, VW,…) or through fleet tracking (e.g., Webfleet technology).

Figure 2 – Overview of the data sets explored. The pilot investigated aggregate data sources in combination with other data sets. Raw floating vehicle data (last line) was outside the scope of the pilot.

Visual traffic analytics – results

Here we describe the key insights from the pilot with examples of the use case analysis possibilities. First, we look at the use cases. Then, the applicability of the different data sets is discussed. The results are indicative since they have not been thoroughly analysed with baseline research. The validity of the results is based on the city's experts' assessments and "sanity checks" as well as comparison of results between different data sets, including some of the local loop detector measurements.

Use cases

Route choice analysis helps understand how, for example, rush hours affect driver behaviour and delays on different route alternatives. Figure 3 below illustrates how travel times vary according to Google data throughout the day using alternatives routes. This can help traffic planners understand how congestion on main roads affect traffic on nearby streets, for example. The data also allows comparing the navigation recommendations with the reality of route choices.

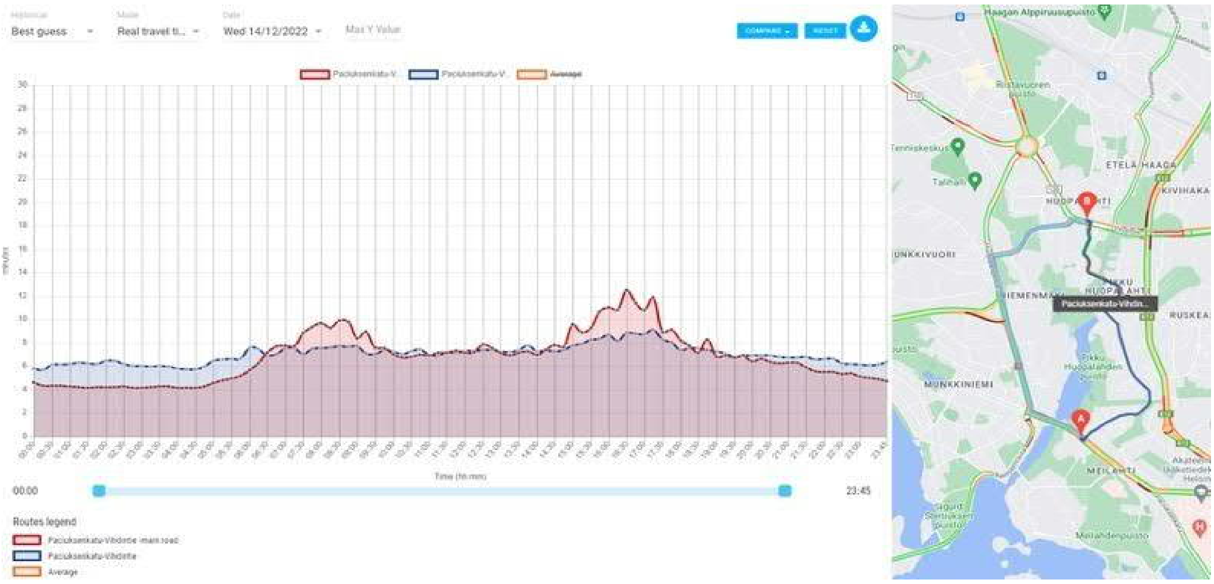

Figure 3 – Drive-through and route choice, analysing whether shortcuts through smaller streets are faster. Data used: ODIQ (Google Maps based proprietary platform).

Measuring congestion and delay times is another function of route analysis. Figure 4 below illustrates how the platform can be used to visualise delays on each location depending on time and day, based on Google Maps data. For this, we used colour coding on the map itself visualized it on a timeline graph.

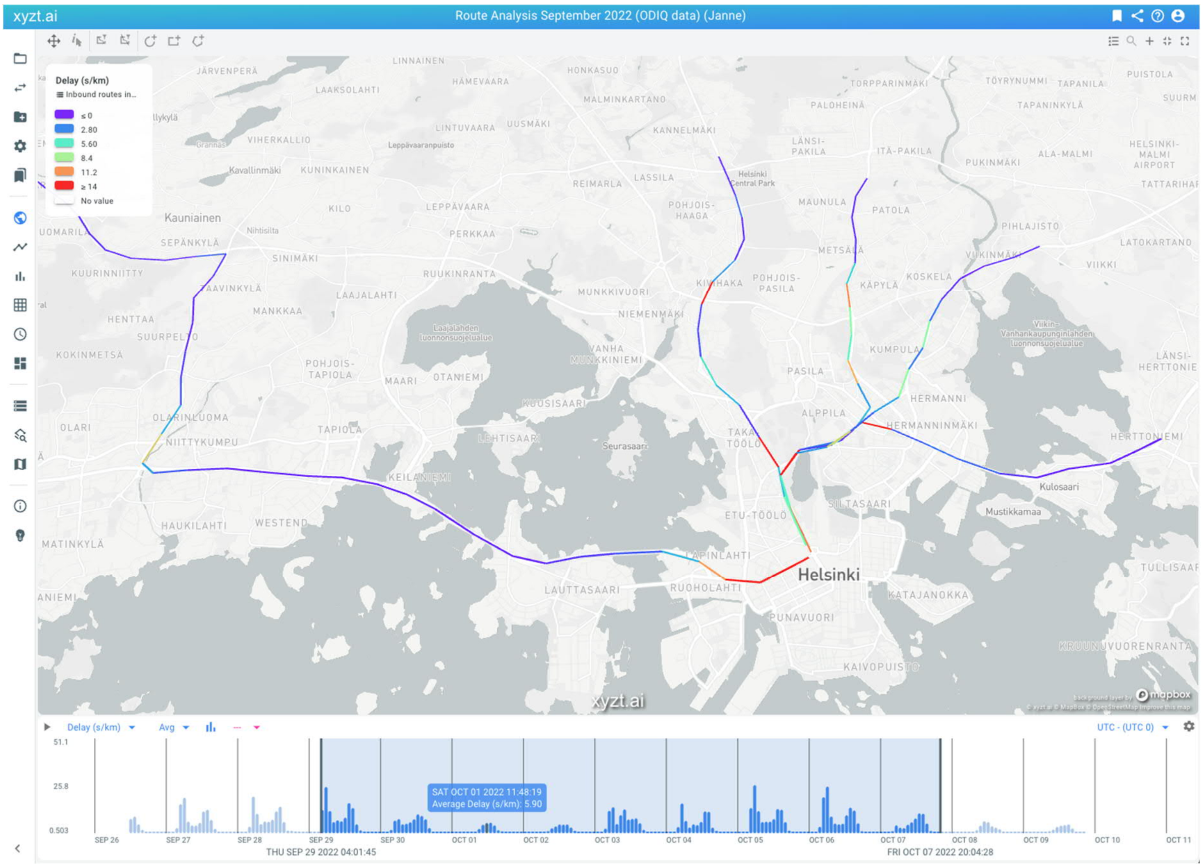

Figure 4 – Route analysis and delays at different times of day and along different routes. Data used: ODIQ

Through travel analysis is an interesting case for understanding where traffic is coming from and where it is going. It helps visualise who are using certain road sections and why, allowing planners to figure out how changes or alternatives to a certain section might affect the surrounding traffic as well. Besides planning, it is suitable for before-and-after comparisons to see which routes were affected if changes were done. The figures below indicate one interesting section (Figure 5) and its cut-through traffic (Figure 6, using TomTom data) in the very centre of Helsinki where a trial is planned for reducing the number of vehicle lanes.

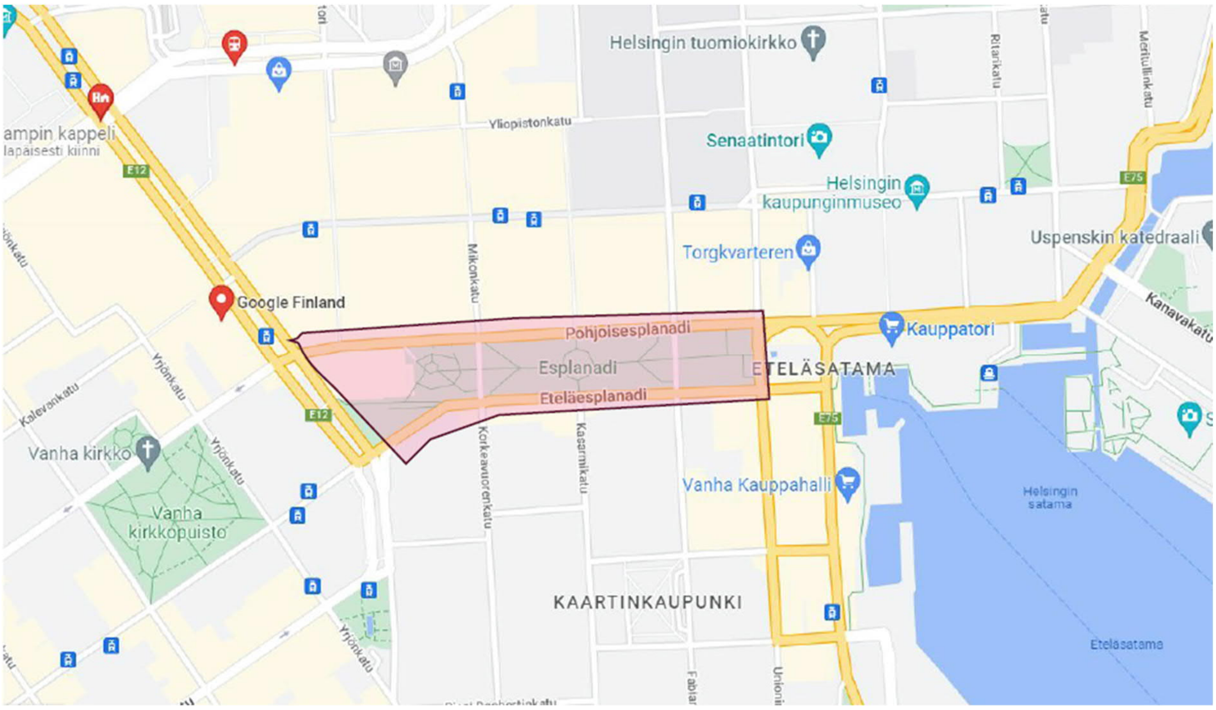

Figure 5 – Drive-through origin and destination analysis, defining the section of interest. Data used: TomTom Move.

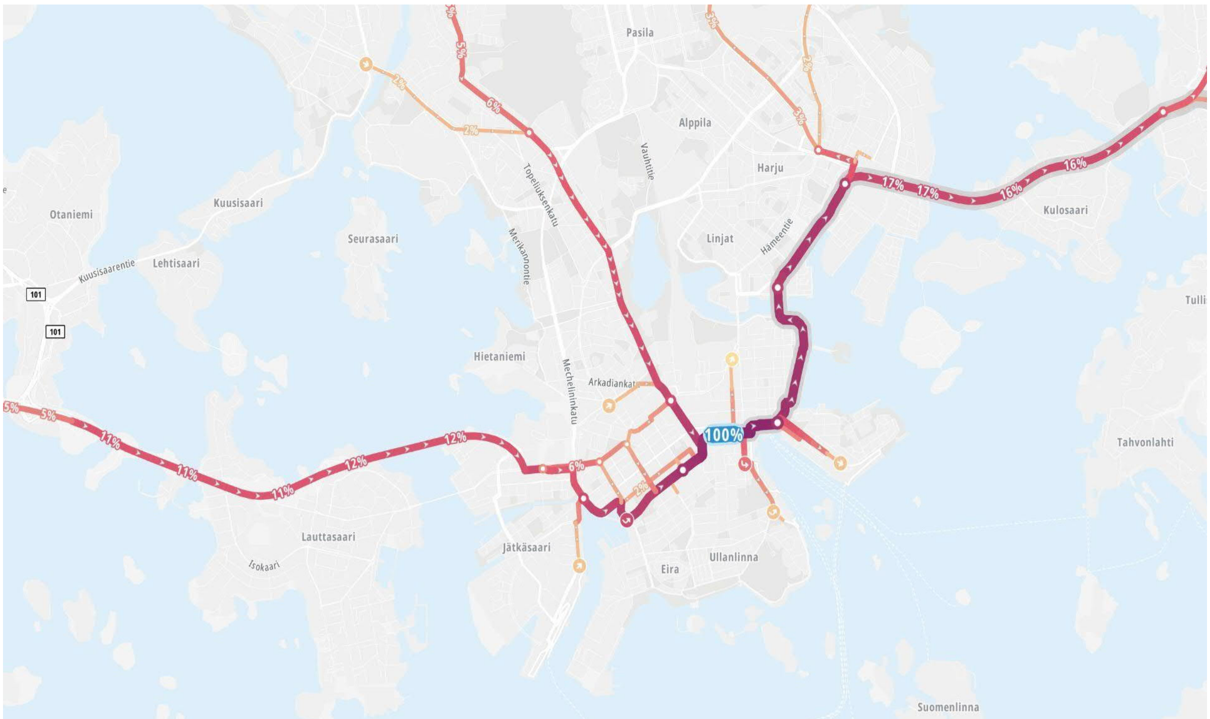

Figure 6 – Drive-through origin and destination analysis, visualising where traffic comes from and goes to. Data used: TomTom Move.

With speeding analysis, main aspects of interest are identifying where and when speeding happens, and zooming in on specific sections that are expected or known for such behaviour (e.g. through citizen feedback, incident reports or other insights). One of the interests we tested the platform for was to validate or compare the data with dubious looking results of an earlier camera-based speed detection study. Figure 7 below is an example of how the distribution of mean speeds and speeding occurrence at different times and over different sections of a road can be visually analysed on the platform with HERE data.

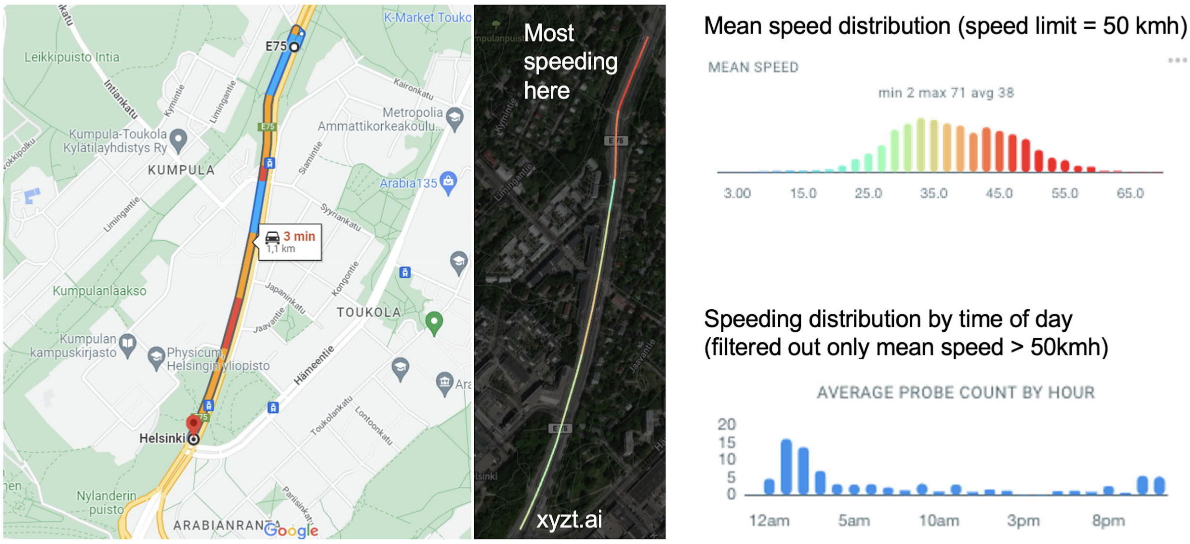

Figure 7 – Speeding analysis, allowing visualisation of (mean) speeds temporally and spatially. Data from HERE.

Data sets

Different data sets are better suited for different needs. For example, some are time series based with certain time intervals while others aggregated over a select period of time. Some are better for understanding origin-destination patterns while others can be more suited for analysing speeding, for example. Figure 8 below indicates the suitability of the explored data sets for a variety of mobility analysis use cases. It should be noted that the table is only indicative, and the right fit should be evaluated for specific needs.

	Speeding/Congestion analysis	Origin-destination analysis	Travel time analysis	Delay analysis	Traffic density analysis	Multi-modality analysis	Dangerous events (acceleration, breaking, cornering,...)	Traffic light performance analysis	Emission analysis
	Analyze where and when traffic is speeding or where and when traffic is congested.	Analyze where traffic comes from and goes to, e.g., analyzing how many people move between different areas during different times of the day.	Analyze how long trips take between different areas and possibly analyse how travel times are impacted due to for example road works, detours, etc.	Analyze where most delay is seen along different trips.	Analyze the amount of traffic and change in amount of traffic at certain locations and times.	Analyze the modal split, such as what percentage of people drive by car, bicycle,...	Where and when do most dangerous events happen, such as people cornering on a highway ramp.	Where do traffic lights perform poorly, i.e., where do people have to wait a long time in front of a red light.	What is the impact of a LEZ, what is the impact of a change in traffic plan on the emissions?
HERE road stats	✅				✅ (only relative)				
TomTom MOVE	✅	✅	✅	✅	✅ (only relative)			✅ (Junction analytics)	
ODIQ	✅		✅	✅					
Traffic counts	✅	✅			✅	✅			
Telia origin destination		✅				Telia is working on this			
Copernicus atmospheric monitoring data									✅ (very coarse though)
Otonomo, Wejo, Bridgestone, INRIX,...	✅	✅	✅	✅ (less straightforward)	✅ (only relative)	Some data providers are working on this	✅ (e.g., Bridgestone abc data)	✅	✅ (model based on vehicle type, speed,...)

Figure 8 – Suggested suitability of different data sets for use cases.

By nature, floating vehicle data allows for greater spatial coverage and availability – and essentially immediate scalability – compared to physical sensors. The accuracy of the data depends on the market share which is generally between 5 and 25%. We compared the traffic counts and speeds of TomTom and HERE data with a select set of loop detector measurements the city has. Based on that, the share of data points in the traffic flow was well above what is generally considered necessary for these types of analyses.

In one of the compared locations, the share was significantly higher than in others. This is explained by the fact that there's a higher coverage on main roads compared to smaller streets. In Helsinki, the coverage was between 16 and 40%. Since we know the number of probes per road segment, it is easy to compare to on-street counts. Geo Mobility regularly does these types of analysis to constantly assess the coverage in the different countries they are active in such as Belgium, the Netherlands, Spain, United Kingdom and Ireland.

Compared to the floating vehicle data which can provide very detailed information in terms road sections and time, mobile network data is much coarser. It can, however, provide macro-level insights into where masses of people are moving. For example, on a district or city level, where people are coming from.

The scope of the pilot was on analysing collected data. Therefore, more active use cases and the use of real-time data was not pursued here. The data sets in question did not provide a way to differentiate and analyse travel mode specific information. Alternative means for gaining insights on walking and cycling are a remaining future need.

Digital traffic infrastructure – pilot cases

This section presents the two pilot cases and their results. Both pilots were done in the Jätkäsaari district of the West Harbour area of Helsinki, covering roughly the same routes (Figure 2). The area is new, which means that its planning documentation should be as modern and up-to-date as possible. At the same time, the area is still under construction, which may result in the use of temporary traffic signs and other oddities.

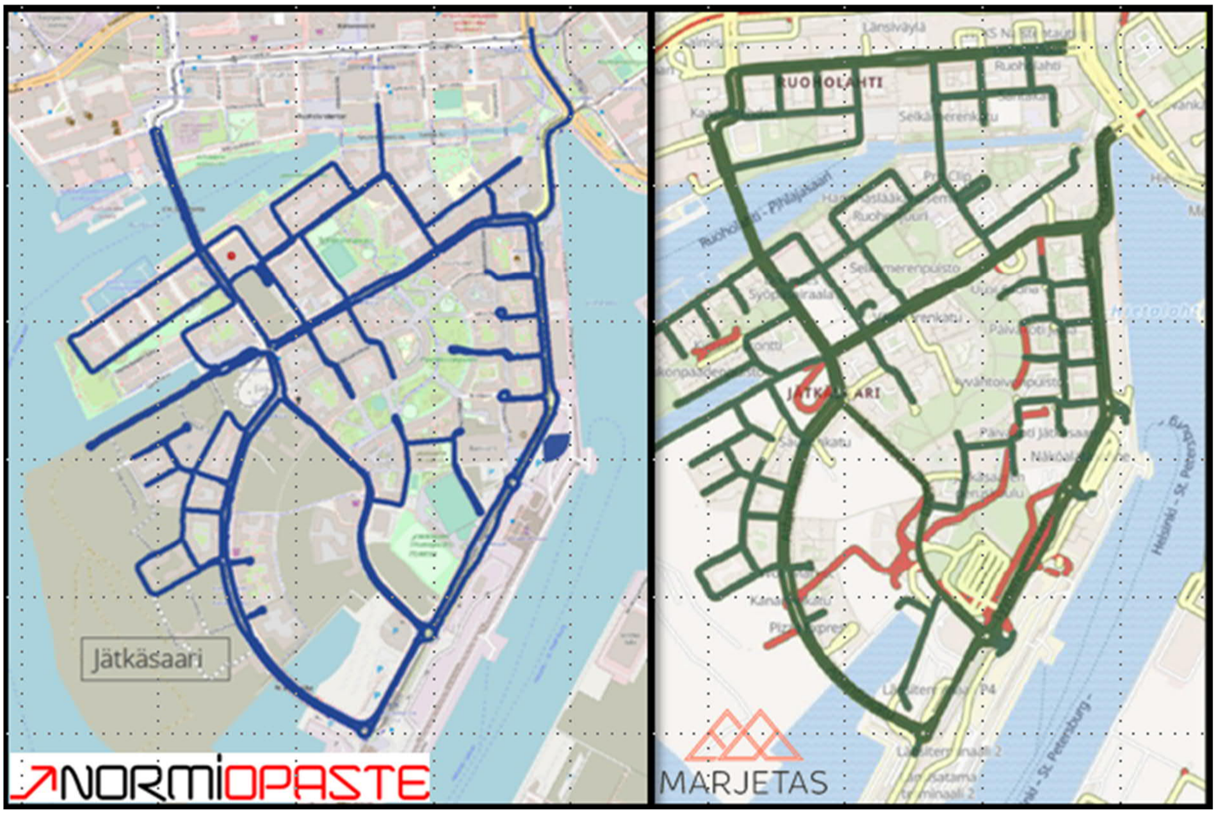

Figure 2 – Routes mapped by Normiopaste (left) and Marjetas Academy (right) in the pilots.

Comparison of reality vs. plans: Pilot with Normiopaste

The pilot with Normiopaste focused on creating an inventory of traffic signs using geo-located, immersive panoramic 360 images, and then comparing the observed reality (i.e., the true location of traffic signs) with plans (i.e., the planned locations specific in documentations).

The scanning was done using a Trimble MX7 mobile imagine system, and the Applanix POSPac MMS Software was used for GNSS/INS trajectory processing. The geo-referenced images were then analysed and compared with the plan documentation available in the city's Cityinfra platform to assess the accuracy of the information.

The preliminary results of the analysis reveals the true current status of the traffic signs in relation to the respective plans. There were more traffic signs installed than then plans were indicating. The locations of the instalments followed the plans very closely, although in some individual cases the exact installation of the sign differed from the plans. In a limited number of cases, some traffic signs were missing. Some older design versions of temporary traffic signs were being used instead of up-to-date ones. In addition, some temporary traffic arrangements related to the large-scale construction in the area had been left uninstalled, potentially creating unclear situations for road users.

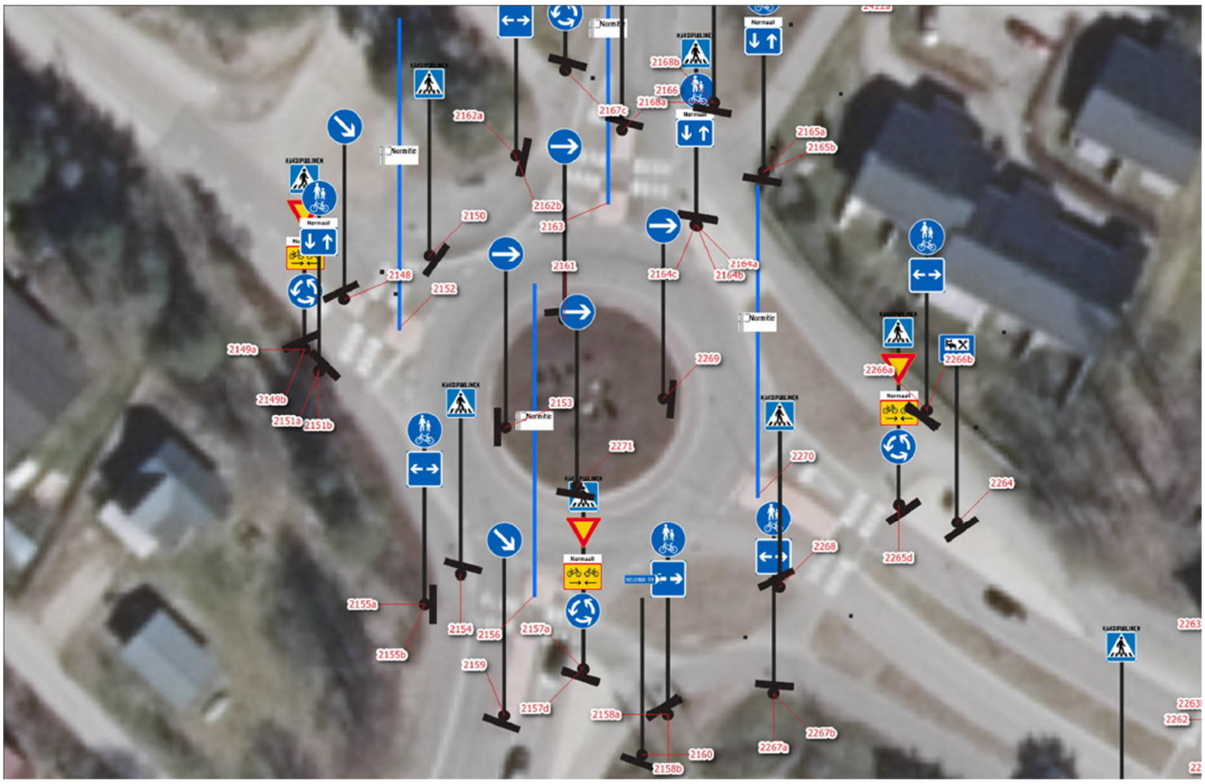

Figure 3 – QGIS visualisation of reality (Normiopaste)

Updating traffic sign information: Pilot with Marjetas Academy

The pilot with Marjetas Academy focused updating traffic control device data using machine vision. First, high accuracy machine vision (Cyclomedia DCR10) was used to produce a digital inventory of traffic signs along the selected route. Then, the same route was scanned using another system (Cyclomedia Street Hive) to analyse and update any differences with the previously documented data. The first mapping (DCR10) was done in September 2022, and the second on (Street Hive) in October 2022. The identification of traffic signs with machine vision took place in December 2022, and the comparison between the two traffic sign inventories in January 2023. The DCR10 system uses a 100 Mpx camera with 10 cm positioning accuracy while the Street Hive system has a 24 Mpx camera with 30 cm positioning accuracy. The idea is that the high-accuracy scanning can produce a reliable baseline of data whereas the Street Hive scanning is sufficient for checking for and updating changes in signs. The Street Hive system can be used by quickly trained users rather than professional, and with integrated processing and analysis, it can produce the results relatively quickly (essentially updated in a day rather than in a month). The system delivers the data to the city's Cityinfra platform so that it is interoperable and complies with the requirements set by the national authorities.

The preliminary analysis indicates that the piloted solution is capable of identifying the traffic signs in the street space efficiently. By coincidence, the parking regulations in the study area changed between the two inventories, and parking-related traffic signs were replaced with the new ones. The change in over 40 parking signs was detected in the analysis, as well as individual cases of added new traffic signs, removed traffic signs, and relocated signs.

Conclusions

This paper covered pilots in two topics related to data: visual traffic analytics and digital traffic infrastructure. All the pilot cases described were done in the fall of 2022 as part of Mobility Lab Helsinki's work aiming to support companies in testing and co-developing their new smart mobility solutions. This section summarizes the conclusions, first for the visual traffic analytics topic, then for the digital traffic infrastructure cases.

Visual traffic analytics

This visual traffic analytics section of the paper described the pilot Mobility Lab Helsinki did with the companies xyzt.ai and Geo Mobility in the fall of 2022. It was an explorative proof-of-concept aiming to learn about modern data-driven analytics possibilities for the needs of traffic planners and researchers. Data from HERE, TomTom and Google as well as Telia was used for the period of September 2022 to test how different use cases can be analysed: route choices, congestion, through travel and speeding as well as the data quality itself.

Based on the results and, the different data sets suit a variety of use cases. Some suit certain needs better than others. The data coverage and quality of the different floating vehicle data sources in Helsinki seem sufficient for the explored use cases. While floating vehicle data allows detailed street-level analysis, mobile network data can be suitable for macro level analysis. By comparison of the data sets and by city's experts' estimations, the coverage and accuracy seem sufficient. While an indicative listing of best applicability of the data sources for different use cases was defined, a more detailed definition of a use case should be defined to choose the right data source.

The visual analytics platform itself enabled incorporating and analysing different data sets with a variety of features, allowing the user to pin-point the spatial, temporal and parameter details he wanted. Initially, the vast number of tools, functions and filters in the platform made it somewhat complex to find all the right functions and features needed. With training and getting familiar, however, it of course becomes easier, and more options adds to the versatility and possibilities of what can be done. The platform enabled, for example, useful ways of side-by-side comparisons of a situation at different times which seems very useful for, e.g., before-and-after analysis of impacts when there are road works or changes in the network, speed limits etc.

The focus of the tests was on the data sets collected in September. So, use cases for real-time monitoring of for the needs of active traffic management was not in the scope of the pilot. The data sets explored were focused on car traffic and did not really enable modal analysis. Better awareness about walking and cycling is one future interest for suitable data is currently lacking.

Digital traffic infrastructure

The digital traffic infrastructure pilots described in this paper aimed to provide insights about the possibilities of modern technologies and solutions for the city's need for digitising its infrastructure information, such as traffic signs. The two pilots conducted approached the same topic from different angles. They demonstrated the potential of using 360-imagery, machine vision and other technologies to produce accurate and up-to-date data on traffic control devices. The pilots showed that it is possible to compare the observed reality with plans and update differences using machine vision. The use of high-accuracy and lower-accuracy systems in combination was found to be a promising approach for delivering reliable data, addressing different needs.

The key observation from the two agile pilots was that 360-imagery can be utilised in identifying traffic signs in the street space. Machine vision, combined with other sensors, can cost-efficiently detect traffic signs and their exact location. Automated software not only helps to avoid manual work, but also enables safeguarding residents' privacy, for example automatically blurring peoples' faces and register plates of cars from the 360 images. The pilots revealed that the traffic environment can change fast, for example due to parking policy changes, roadworks or construction, and therefore finding cost-efficient means to update the data are necessary.

The pilots were part of the Cityinfra platform development by the city. The hands-on pilot provided insights to the development, for example regarding the data transfer (APIs, data dumps), traffic sign documentation, metadata structure, and other practical development needs for the Cityinfra platform.

The final results and conclusions of the two agile pilots will be available in April 2023.
The use of agile piloting allowed for the flexible testing of innovative solutions in an urban environment. Future approaches for keeping the digital information about traffic control devices and other city resources accurate and up-to-date could include the continued use of machine vision and other technologies such as digital tags for keeping track of different infrastructure items and objects.

References

1. Mobility Lab Helsinki – the city's testbed for smart mobility. http://mobilitylab.hel.fi [accessed 4.1.2023]

2. Rinne, J., Virtanen, J.-P., Sahala, S., Kostiainen, J. & Koskela, A. (2022). *Digital Twin for Mobility: Concept and baseline study.* Working paper 9 September 2022. Mobility Lab Helsinki. Available at: https://mobilitylab.hel.fi/digital-twin/ [accessed 4.1.2023]

3. City of Helsinki (2019). *Helsinki Intelligent Transport System Development Programme 2030: Developing traffic information, new mobility services and automation.* Urban Environment publications 2019:16. Available at https://www.hel.fi/static/liitteet/kaupunkiymparisto/julkaisut/julkaisut/julkaisu-16-19-en.pdf [accessed 4.1.2023]

4. Budimir, D., Jelušić, N. & Perić, M. (2019). *Floating Car Data Technology*. In Scientific Journal of Maritime Research 33, pp. 22-32.

5. Ceccato, R., Gecchele, G., Rossi, R. & Gastaldi, M. (2022). *Cost-effectiveness analysis of Origin-Destination matrices estimation using Floating Car Data. Experimental results from two real cases.* In Transportation Research Procedia 62, pp. 541-548.

6. Spilling, K. & Rinne, J. (2020). *Pocket Book for Agile Piloting: Facilitating co-creative experimentation.* Forum Virium Helsinki. Available at: https://fvh.io/pocketbook [accessed 12.12.2022]