Experiences from smartphone based travel data collection – System development and evaluation

Final report from the SPOT-project

Version 1.0

2016-03-09

Authors: Andreas Allström, Sweco*
Györö Gidofalvi, KTH
Ida Kristoffersson, Sweco
Adrian C. Prelipcean, KTH
Clas Rydergren, LiU
Yusak Susilo, KTH
Jenny Widell, Sweco

* Project manager (andreas.allstrom@sweco.se, 070-374 65 32)

SUMMARY

Using traditional methods for travel data collection is not feasible for gathering the large volumes of detailed, high quality data that are needed by the recent models used for transportation systems analysis. The numerous drawbacks of these traditional methods led to the study of technologies that can replace or complement the traditional methods. From numerous options, one stands out: capturing raw data from devices that can use any positioning technology (e.g., GPS, WiFi positioning, GSM, etc.), followed by transforming the raw data into meaningful travel data. Since most smartphones are equipped with sensors that can collect the smartphone's location, and since these devices are integrated in the daily life of most people, they provide an unprecedented opportunity for large-scale data collection.

The aim of this research project is to study the feasibility of replacing the traditional method for obtaining travel data, i.e., declared activity-travel diary, with a suite of tools that make use of smartphone collected travel data to generate activity-travel diaries. This is done by understanding the advantages and disadvantages of the traditional method and the proposed method when collecting travel diaries. For a fair comparison, both methods have been tested in parallel, i.e., in the same city, at the same time, and with the same respondents.

To achieve the objectives of the project, MEILI, a system that consists of a mobile application for capturing the movement of users and a web application for allowing the users to annotate their movement to generate travel diaries, has been deployed. MEILI has been tested and evaluated in both a small pilot and a larger field trial, although in different development stages.

The user experience for the web application of MEILI varies between the different participants in both the pilot and the large field trial. This can be due to the level of expertise of the users, which is higher for users participating in the pilot than for the users participating in the large field trial, or due to collection difficulties that are specific to some phone models. Besides rare problems with the installation and registration, the smartphone application functioned as expected for the majority of the participants. The solution proposed for overcoming battery limitations worked as expected and the majority of the users did not experience the battery consumption as a problem. Furthermore, the users rarely turn off the application due to integrity issues and the majority of the users found the traditional travel diary just as intrusive as the smartphone application. The main problem for several users, both in the pilot and in the large field trial, has been the annotation process and the website component of MEILI.

The recruitment of respondents is a critical phase for traditional travel diaries and, as expected, this was the case also for the smartphone based method. A lesson learned was that it is important to simplify the registration process as much as possible. It is also critical for the system to work as expected and, in case of failure, the communication with the users not to be delayed, otherwise the drop-out rate increases rapidly.

During the large field trial, 171 users annotated 2142 trips (1250 trips from 51 users for a duration longer than a week), hence it can be seen as one the most successful smartphone-based trials in the world. In total, about 1 million GPS readings were collected. Some tasks were successfully automated with a high accuracy, i.e., stop detection with a 97% accuracy, trip leg detection with a 70% accuracy, resulting in detecting trip chains with approximately 79% accuracy. Travel mode detection is done with an accuracy of 54%, which means that the current approaches need improvement before replacing the user's annotation, but when the top 3 most likely travel modes were presented to the user, the correct travel mode is in this list in 82% of the cases, which reduces the user's effort when choosing her travel

mode. The destination and purpose inference achieved accuracies under 50%, which implies that studying and improving the used methods are critical for a fully automated system.

This report shows that MEILI is a needed complement to traditional travel diary collection methods since it appeals to a different age group and collects more detailed travel data, and, in order for MEILI to completely replace traditional travel diaries, the efforts should be aimed at improving the web application interface and proposing better automated methods for travel mode, destination and purpose inference.

Despite its weaknesses, from a transport modelling and policy design point of view, MEILI opens new possibilities to provide better, more accurate, and realistic, human-centered analysis. The data that it produces enable us to address issues which cannot be answered with a traditional survey, such as how the individual's activity travel patterns vary across different days and time of day and more accurate data for (the value of) travel time variability. Furthermore, it enable us to understand better and simulate the interactions between individuals and his/her constraints in time and space dimensions. This knowledge and ability will help us to influence travellers to adopt a more sustainable travel behaviour in the way that is most convenient for them; thus increasing the possibility for success of the given transport policy implemented.

SVENSK SAMMANFATTNING

Traditionella metoder som används för insamling av resdata har en del nackdelar och samtidigt kräver de modeller som används för att analysera transportsystemet allt mer detaljerad data av hög kvalitet. Det finns således ett stort behov av nya metoder för insamling av resdata och mest framgångsrikt har det visat sig vara att samla rådata från enheter som kan använda olika positioneringsteknologier (t.ex. GPS, WiFi-positionering, GSM, etc.) och omvandla denna information till användbar resdata. Eftersom de flesta smartphones är utrustade med olika positioneringsteknologier och eftersom dessa enheter är integrerade i det dagliga livet för de flesta människor, ger de ett unikt tillfälle för insamling av resdata i stor skala.

Syftet med forskningsprojektet som presenteras i denna rapport är att studera om en applikation som installeras på en smartphone kan användas som ersättare eller komplement till en traditionell resvaneundersökning. Detta görs genom att samla in data och studera fördelar och nackdelar med respektive metod. Data har samlats i samma stad, vid samma tidpunkt och med samma svarande med båda metoderna för att göra resultaten så jämförbara som möjligt.

För att uppfylla projektets mål har ett system, MEILI, bestående av en smartphone-applikation som loggar telefonens rörelser samt ett webbaserat gränssnitt där användaren kan se, korrigera och komplettera insamlad data utvecklats. Systemet har, i olika utvecklingsstadier, testats och utvärderats i både ett litet pilotförsök och ett större fältförsök.

Som väntat varierade användarnas åsikter och upplevelse av systemet en hel del både i piloten och i det stora fältförsöket. Detta beror förmodligen främst på grund av att graden av IT-vana varierar, men också på grund av att systemet fungerar bättre för vissa telefonmodeller än andra. Förutom vissa sällsynta problem med installationen och registrering har funktionaliteten hos appen fungerat smidigt för majoriteten av deltagarna. Den algoritm som utvecklats för att begränsa batteriförbrukningen visade sig fungera bra och majoriteten av användarna upplever inte batteriförbrukningen som ett problem. Dessutom är det få användare som har slutat använda appen på grund av integritetsfrågor och majoriteten av användarna ansåg att den traditionella resvaneundersökning var lika integritetskränkande som appen. Det största problemet för användarna, både i piloten och i det stora fältförsöket, har varit den webbplats som använts för att komplettera och korrigera insamlad resdata.

Rekryteringen av användare är en kritisk fas för traditionella resvaneundersökningar och som väntat var så även fallet för den smartphonebaserade metoden. En lärdom var att det är viktigt att förenkla registreringsprocessen så mycket som möjligt. Det är också mycket viktigt att systemet och kommunikationen med användarna löper smidigt, framförallt vid eventuella problem, annars ökar bortfallet snabbt.

Under det stora fältförsöket samlade 171 användare in 2142 korrigerade och kompletterade resor (1250 resor samlades in från 51 användare som använde systemet under mer än en vecka). Försöket kan därmed betecknas som ett av de mest framgångsrika smartphone-baserade försöken som har genomförts i världen. Totalt samlades runt 1 miljon GPS-punkter in. En del av den automatiska klassificeringen gav hög noggrannhet, t.ex. detektering av stoppunkter (97%) och indelning i reselement (70%), vilket resulterade i att en korrekt reskedja detekterades med 79% noggrannhet. Färdmedel klassificerades med 54% noggrannhet vilket innebär att nuvarande metod behöver förbättras för att användarens arbetsbörda ska reduceras ytterligare. Det bör dock noteras att i rankingen av färdmedel hamnade korrekt färdmedel bland topp 3 i 82% av fallen. Klassificeringen av målpunkt och ärende gav en noggrannhet under 50% vilket innebär att dessa metoder behöver utvecklas vidare för att skapa ett helt automatiskt system.

Denna rapport visar att MEILI är ett välbehövligt komplement till traditionella resvaneundersökningar eftersom det tilltalar en annan åldersgrupp och samlar in mer detaljerad resdata. För att ett system som MEILI helt ska kunna ersätta traditionella resvaneundersökningar måste webbsidan för korrigering och komplettering av insamlad data förbättras och bättre metoder för att automatiskt bestämma färdmedel, målpunkt och ärende utvecklas.

Ur ett trafikmodellerings- och planeringsperspektiv skapar MEILI, trots sina brister, nya möjligheter för bättre, mer realistiska och med verkligheten mer överensstämmande analyser. Data som samlas in ger oss dels möjlighet att svara på frågor som vi inte kan besvara med data från en traditionell resvaneundersökning; exempelvis hur individers resmönster varierar mellan olika veckodagar. Men denna typ av data ger oss också möjligheten att bättre förstå och simulera sambanden mellan olika individers restriktioner i tid och rum, samt deras olika värderingar av bland annat restid. Denna kunskap kan ge oss ett bättre underlag för att hitta effektiva metoder för att påverka människors resande till att bli mer hållbart utan att resenärernas uppoffringar blir onödigt stora.

LIST OF ABBREVIATIONS AND ACRONYMS

Table 1: List of abbreviations/acronyms

Abbreviation/Acronym	Explanation
ASC	Application Server Component
CAPI	Computer-Assisted Personal Interview
CATI	Computer Assisted Telephone Interview
CRUD	Create, Read, Update and Delete operations
DSC	Data Storage Component
GPS	Global Positioning System
GSM	Global System for Mobile communications
MCC	Mobile Collection Client
MVC	Model-View-Controller
OS	Operating Systems
OSM	Open Street Map (https://www.openstreetmap.org/)
PAPI	Paper And Pencil Interview
POI	Point Of Interest
PP	Paper-and-Pencil method, by which we mean the same as traditional travel diary
WAC	Web Annotation Client

GLOSSARY

Table 2: Explanation of frequently used terms

Term	Explanation
Trip	A trip is in this context defined by a purpose. Hence, several modes can be used during the same trip and a new trip purpose initiates a new trip.
Trip leg	A trip can be divided in a number of trip legs. Several modes can be used during the same trip and each part of a trip using one mode is called a trip leg.
Trip destination	A trip destination is here defined by a Point of Interest where the user ends his trip. In MEILI the POI can either come from a predefined data base of POIs or be defined by the user.
Trip purpose	Categorizing trips into different purposes is very important for estimation of transport models, since there are major behavioural differences depending on trip purpose. For example, user valuation of travel time differs substantially between work trips and leisure trips.
Temporal quality	The temporal quality measures how well MEILI captures the timeline of a trajectory associated with a trip. The quality measure is based on a maximum allowed time difference between two consecutive locations.
Spatial quality	The spatial quality indicator measures how well MEILI captures the geometry of a trajectory associated with a trip. The quality measure is based on a maximum allowed distance between two consecutive locations.

TABLE OF CONTENTS

1 INTRODUCTION

1.1 BACKGROUND

Implementing sustainable transport solutions, reducing transport emissions and energy consumption, and achieving a modal shift away from the conventional private car are the main goals for many major metropolitan areas in the world. To achieve these goals, it is crucial to understand the individual travel decision-making behaviours and implement the correct urban and transport intervention policy measures. The success of a transportation policy itself depends on an accurate description and prediction of aggregate flows, as well as the disaggregate travel behaviour of individuals.

Currently, there is a great need of new methods for collecting travel data. Traditional methods used for collecting travel data have considerable drawbacks and, at the same time, the models used to analyse the transport system require more and more detailed and high quality data. One of the main barriers of collecting accurate and detailed disaggregate travel data is the limitation of the method used to follow the travellers' choices through space and time. In the report *Nya indatakällor för trafikprognoser* (Allström et al., 2013) GPS (receiver) equipped devices were identified as one of the most interesting technologies to investigate further. Given that GPS devices and smartphones with GPS functionality are now becoming more widespread, it is easier to capture the users' spatial movement and behaviour over time. This is also mentioned in the report *Utveckling av samhälls-ekonomiska metoder och verktyg, effektsamband och modeller inom transportområdet* (Trafikverket, 2012, Chapter 9.1.2). The use of a web-interactive survey, combined with visualizing the tracked movements, means that the complexity of individual travel decision making processes is now more observable and analysable. Because of this, it is believed that GPS-based data collection is one of the better alternatives in travel behaviour research, compared with traditional travel diary methods such as PAPI (Paper and Pencil Interview), CAPI (Computer-Assisted Personal Interview) or CATI (Computer Assisted Telephone Interview). Earlier studies (e.g. Wolf et al., 2004; Stopher et al., 2008; NCHRP, 2014) show that a GPS-based data collection has a great potential to solve the problems related to the estimation of distance/travel time, geographic coding of departure/destination locations and forgotten trips. It will also provide a more detailed data set, something that is essential for the next generation of transport models.

However, at the same time, this GPS technology is not without problems. Collecting GPS traces without any semantics attached to them requires further data processing and analysis. Furthermore, given the limitation of GPS related to the infrastructure and the built environment, there will always be situations and areas where GPS will fail to record locations.

Given that Trafikverket needs to find and test a better method for collecting travel data, this project has deployed and evaluated a prototype of a smartphone application that is jointly developed by the KTH Transport Science and Urban Planning Departments. The application is a promising alternative for travel data collection and is compared with data collected with a web-based version of a traditional travel diary method. Indeed, there are several commercial apps that provide similar information, however, it is hard to tailor a commercial app's output as an input into a detailed and advanced transport model, such as an activity based model. More importantly, it is impossible to control how these commercial apps use the respondents' socio-demographic and movement information, which can breach the Swedish research ethical, safety and security standards.

This project has been funded by Trafikverket (TRV 2014/10422) and the participating organizations have been Sweco, KTH and Linköping University.

1.2 AIM

The aim of this research project is to study if an application installed on a smartphone equipped with GPS can be used as a replacement or supplement to a traditional travel diary. Measurements of travel behaviour are carried out for both methods - smartphone and traditional travel diary - within the project. Their advantages, disadvantages and accuracy in measurement has been compared. The methods are implemented in the same city, at the same time and with the same respondents using both methods to make the results as comparable as possible.

1.3 METHOD

To achieve the objectives, a prototype of a travel tracking application has been deployed to understand the benefits, challenges and feasibility of using a smartphone application in collecting one week travellers' travel data. The results of the smartphone data collection have been compared with the results collected via a traditional travel diary. The criteria that has been used to measure the performance are the number and the accuracy of reported/recorded travel engagements on the given day and other practical matters that serves the interests of Trafikverket. This includes forgotten trips, estimation of distance/travel time, geographic coding of departure and destination locations and burdens for the respondents and administrators for the different survey methods. A large part of the project consists of the development of the application. The application has been tested in a small pilot during 2014 and a larger field trial during 2015.

1.4 PROJECT DISSEMINATION

During the project, four meetings with the steering group of the project have been organized where the results and progress of the project have been presented and discussed. The results have also been presented at a number of conferences and been published in journal papers.

Conferences

- Transportforum 2015, 2016
- Nationella konferensen i transportforskning 2015
- MT-ITS 2015
- WCTR 2016 (submitted)
- European Transport Conference 2016 (submitted)

Journal papers

- A. C. Prelipcean, A. C., G. Gidófalvi, and Y. O. Susilo. (2014). "Mobility Collector." Journal of Location Based Services 8 (4).
- International Journal of GIS, submitted 2015
- Transport Reviews, submitted 2015

1.5 OUTLINE

This report is outlined as follows. Chapter 2 covers a literature review of previous work in the area of GPS-supported travel data collection. Chapter 3 describes the development of the software applications and in Chapter 4 an overview is given of the machine learning algorithms used for mode and purpose detection. The evaluation method is described in Chapter 5 and results of the pilot trial and large field trial are described in Chapters 6 and 7, respectively. Chapter 8 concludes the report and discusses future work.

2 PREVIOUS WORK

The success of a transportation policy depends on an accurate description and prediction of aggregate flows as well as the disaggregate travel behaviour of individuals. In doing this, good quality of data is paramount. The traditional methods used for collecting travel data have substantial drawbacks and the public participation towards travel diaries is in continuous decline, especially in Sweden (VTI, 2002; Trafa, 2015). At the same time, communication and computing technologies together with location-, orientation-, and motion sensors (e.g., GPS, Wi-Fi positioning, accelerometer) in smartphones enable the large-scale collection of movements of individuals. Consequently, there has been a surge in the number of trials and studies that investigate the potential use of these new technologies to complement and replace the traditional travel diary acquisition. Whereas widely implemented travel diaries typically depend on self-reported trips and activities, the subject temporal and spatial movement now can be tracked effortlessly in detail (Wolf et al., 2004; Stopher et al., 2008; NCHRP, 2014; Cottrill et al., 2013).

The current mobile positioning technology, however, is not without problems (Anderson et al., 2009; Cottrill et al., 2013; NCHRP, 2014). Given the limitations of positioning technologies (e.g., no indoor GPS, low resolution/reduced availability of Wi-Fi and GSM in rural areas), and the built environment constraints (subways, tall buildings, urban canyons), a user's trace is not guaranteed to be recorded at a stable resolution (space and/or time) everywhere. Applications using this technology are often collecting GPS traces, together with accelerometer readings, which have to be further processed to derive the needed entities, i.e., trips and trip legs, and their attributes, i.e., transport means, and trip destination and purpose. Although positioning technology can be used to directly record accurate time and geographic information of travel (e.g., Chung and Shalaby, 2005; Gong and Chen, 2012; Feng and Timmermans, 2014; Rasouli, 2014), the participants are still needed to be heavily involved by providing/verifying the entities and their attributes. To collect information that cannot be derived from GPS data alone, various prompted recall methods may be used, including paper-based (e.g. Bachu et al., 2010), mobile-phone based (e.g. Ohmori et al., 2005), and web-based (e.g. Ali and Lui, 2011; Bourbonnais and Morency, 2013).

Recently, there have been attempts to automate some of the travel diary generation tasks (Abdulazim et al., 2013; Cottrill et al., 2013; Greaves et al., 2014; Ellison et al., 2014). The automation of the activity-travel diary generation using a smartphone mobile application, as opposed to dedicated GPS tracking devices, is expected to reduce the survey cost and the minimize the cases in which the users forget to carry the tracking device. A complete review of the literature is beyond the scope of this document, but the related work sections of Prelipcean et al., (2014, 2015b and 2015a) provide an in-depth review of automatic transport mode detection and travel diary generation and comparison. The following paragraphs highlight the most important issues.

Although automating activity-travel diary generation has been the main objective of previous research, be it academic or industrial, the main drawbacks that prevented the full automation are: 1) the lack of thorough methodology, 2) the lack of collection and annotation tools, and 3) the separation of tasks without keeping in mind the objective.

First, the lack of a thorough methodology is found in studies that compare travel diaries obtained by different techniques (Bricka and Bhat, 2006; Forrest and Pearson, 2005; Stopher and Li, 2011; Prelipcean et al., 2015a). Whereas there is an agreement regarding the definition of associating the trips recorded by both systems, i.e., by using a spatio-temporal purpose-based join, the results mostly report on the percentage of the trips that were joined (Bricka and Bhat, 2006; Forrest and Pearson, 2005). Some authors (Stopher and Li, 2011; Prelipcean et al., 2015a) also provide the rationale behind the lack of recorded trips by either system, but this type of analysis does not offer any meaningful insight regarding the question: “Is it possible to replace surveys as a way of obtaining travel diaries?” This report shows that the question itself might be flawed since neither system can perfectly collect activity-travel diaries, but rather the data provided by an automated travel diary collection system can complement the data provided by traditional travel diaries.

Second, the focus on data gathering tools has shifted from tools that collect GPS traces and process the data in the absence of user annotations (Axhausen et al., 2003; Wolf et al., 2001, Wolf et al., 2003), to systems that allow for the user to annotate their collected trips (Prelipcean et al., 2014; Montini et al., 2013; Bohte and Maat, 2009). The change in focus also lead to new considerations for the development of such products, which are either private contractor developed products (Bohte and Maat, 2009), in-house proprietary products (Montini et al., 2013), or open-source products (Prelipcean et al., 2014). Unfortunately, there does not seem to be an agreement regarding which technique is preferred, which in turn leads to results that are not comparable due to different collection strategies and methodology.

Third, it is common for researchers to split the activity-travel diaries generation into smaller tasks such as travel mode inference, destination inference, and purpose inference. Each of these tasks have specific approaches and used data types.

Travel mode inference is approached as either a point-based classification, where each location is classified into its travel mode (Stenneth et al., 2011; Prelipcean et al., 2014), or a period-based classification, where a sequence of locations is grouped into trip legs based on heuristic rules and the trip legs are further classified into transportation modes (Chung and Shalaby, 2005; Stopher et al., 2008). Furthermore, there are different types of data used for travel mode inference, such as GPS only datasets (Stenneth et al., 2011), GPS fused with accelerometer datasets (Prelipcean et al., 2014; Reddy et al., 2010), and accelerometer only datasets (Hemminki et al., 2013; Yu et al., 2014). Furthermore, these datasets can be complemented with GIS auxiliary data (Stenneth et al., 2011) to distinguish between modes that have similar movement characteristics, e.g., cars and buses. While the authors report comparable accuracy values such as: 90.8% for seven classes (Prelipcean et al., 2014), 93.6% for five classes (Reddy et al., 2010), 93.5% for five classes (Stenneth et al., 2011), or 90.6% for five classes (Yu et al., 2014), a deeper investigation into mode detection performance evaluation showed that these precisions are not comparable to one another and they over-estimate the achievable accuracy (Prelipcean et al., 2015b).

Destination and purpose inference are closely intertwined since at least one of the features used in purpose inference is derived from a given destination. Most destination inferences are based on proximity to a point-of-interest (POI), which implies the need of a well-defined external POI dataset (Axhausen et al., 2003; Wolf et al., 2001). Bohte and Maat (2009) identify the closest point of interest to a trip's end and, based on its type, derive the purpose of the trip, with an accuracy of 43% for 13 purposes. Oliveira et al., (2014) evaluate two methods, which rely on GIS land use and POI datasets, for purpose inference that are based on choice modelling and decision tree analysis achieving an overall accuracy above 70% for 12 categories. Wolf et al., (2001) show in their pilot study that it is feasible to derive trip purpose by combining GPS point data with a spatially accurate GIS land use database, reporting an accuracy of over 90% for a small data set of 151 trips. Montini et al., (2014) use random forests to infer trip purposes for a one week travel survey in Switzerland 2012, which involved 156 participants, with an accuracy varying between 80% and 85%.

While these tasks have been studied independently, they do not offer an answer regarding how well the travel diary collection can be automated or complemented, which is mostly due to non-uniform error measures used by different approaches, and due to the surface study of precision, e.g., there are no methods that allows us to understand what happens when combining an 70% travel mode inference method with a 59% destination inference method and a 90% purpose inference method.

This report presents a comparison between a traditional travel diary collection system and a semi-automated one, discusses the system used to collect data with the semi-automated travel diary collection system, presents different methods for inferring travel diary specific tasks, and illustrates these concepts in a large field case study on 300 users. Compared to the existing approaches, this research focuses on automating all inference tasks, providing an open source travel diary collection and annotation system, and understanding the differences between the traditional approach and the new suggested approach.

3 THE MEILI SYSTEM

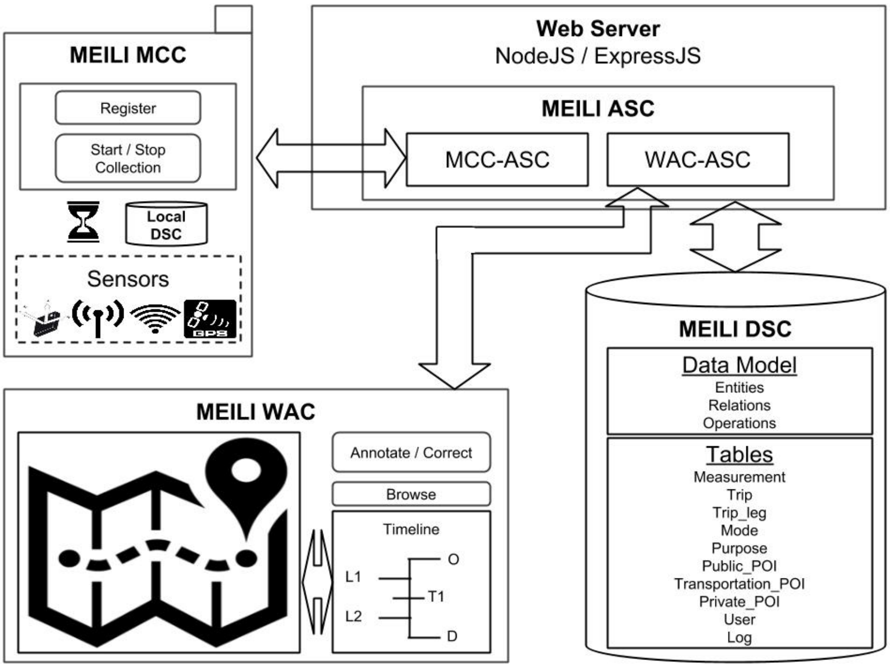
3.1 MEILI DESIGN DBJECTIVES

The ambitious design objective of MEILI is to be an open-source, smartphone based software system that in a fully automatic fashion can effectively collect the accurate travel diaries of its users using state of the art mobile computing-, communications-, and positioning and sensor technologies, auxiliary spatial information (e.g., public transport and POI datasets), and sophisticated machine learning algorithms to infer trips, trip legs, travel modes, and trip destinations and purposes. To be able to collect ground truth information for verification and machine learning algorithm training purposes, this design objective is further extended to develop a web based interface where users can annotate/correct/verify their travel diaries. The following subsections explain the issues that were encountered and the solutions that were taken during the design of MEILI. Those interested in the actual code and the open source license can contact the authors of this report.

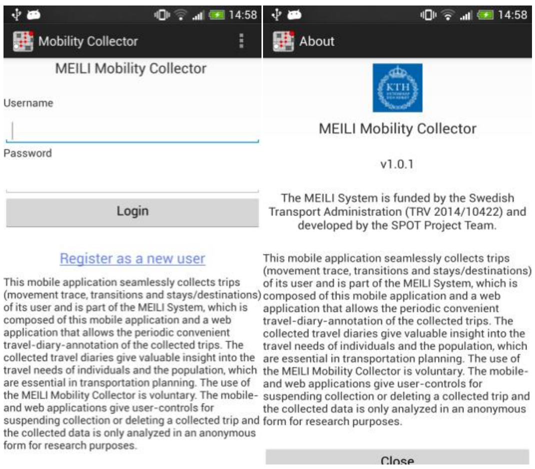
3.2 MEILI ARCHITECTURE

MEILI, an activity-travel diary collection, annotation and automation system was developed to achieve the objective of the project. MEILI's architecture is a typical, three-tier, Model-View-Controller (MVC) that has two types of clients: a Mobile Collection Client (MCC) and a Web Annotation Client (WAC). The primary task of the MCC is to collect movement information from a user's smartphone in a seamless and battery efficient fashion. The primary task of the WAC is to allow users to annotate their movement information (collected by MCC) with travel semantics (i.e., trips, trip legs, travel modes, trip destinations and purposes). To reduce the user's burden, MEILI performs inferences about the travel semantics, which the user can verify and, if necessary, correct.

Both client components connect via a web server to an Application Server Component (ASC) that allows for the bi-directional data transfer between the clients and the Data Storage Component (DSC), which is a PostgreSQL/PostGIS database. The current system architecture of MEILI and the interfaces of the trialled mobile and web clients are shown in Figure 1. The architecture of this system has been continuously improved, leading to a well-tested data model, and an interactive and intuitive user interface.



a. MEILI system architecture



b. Interface of MEILI on respondents' mobile phone unit

Figure 1: MEILI system architecture and the interface of the mobile component

3.3 MEILI MCC - SMARTPHONE APPLICATION

MEILI MCC (Mobility Collector, Prelipcean et al., 2014) was designed as a highly configurable, open source, battery-conscious, mobile tracking and travel annotating framework that is specifically designed for transport research purposes. The task of MEILI MCC is to collect movement information from the smartphone of the user in a seamless and battery efficient fashion. To develop this seemingly straight-forward functionality has proven to be a non-trivial task that required iterative design-development-testing cycles. The major issues encountered and the innovative solutions provided are briefly described below, for details the reader is referred to Prelipcean et al., (2014).

3.3.1 Experiences with developing MEILI MCC on multiple mobile Operating Systems

To be able to provide MEILI to a large fraction of the mobile user population, MEILI MCC has been developed for the two most popular mobile Operating Systems (OS), Android and iOS. Due to its simplified publication and deployment procedures, the initial versions of MEILI MCC have been developed for the Android platform versions 1.6 and up. The development for multiple Android versions has been greatly simplified by the backward compatibility between different Android versions. In later stages of the development, MEILI MCC was ported to the iOS. Although the number of different types of iOS devices on the market is only a fraction of the number of different types of Android devices, the publication and deployment has been more cumbersome than initially envisioned. In particular, the backward compatibility between different iOS versions is weaker than in the case of Android, which means that essential functionality has to be rewritten for every major release of iOS, if the release includes changes to the aforementioned functionality. Additionally, due to the more stringent security/permission model of iOS, essential functionality like running collection processes in the background and starting up such processes in the case of system restarts had to be implemented using custom solutions.

3.3.2 Adaptive, Equidistance, Power-conscious Sampling

Although both the Android and the iOS platforms provide functionality to track the movements of the phone user by sampling the location of the mobile unit, the default sampling strategies do not provide equidistance samples—which is desirable from a mapping and machine learning perspective, as well as from an information theoretical perspective—at an optimal power consumption (Prelipcean et al., 2014).

Consequently, MEILI MCC adopts an adaptive, equidistance, power-conscious sampling strategy that is motivated by the lack of adequate default sampling strategies, and two additional observations. First, the battery consumption of an unsuccessful location request inside a building (where the user/phone is stationary, i.e., a large fraction of the time, see Klepeis et al., 1996) is significantly higher than the cost of a location request outdoors (where the user moves, i.e., the relevant information that MEILI tries to capture). Second, the cost of obtaining accelerometer readings is only a fraction of the cost of obtaining a location reading. Using these observations, depending on which readings are available (location- or accelerometer readings) given the characteristics of the physical environment, MEILI MCC switches between two timer-triggered sampling loops: location sampling loop (when location readings are available) and accelerometer sampling loop (to initiate the location sampling loop when the accelerometer readings suggest movement). To provide equidistance samples given the dynamic movement characteristics of users using different transportation modes, MEILI MCC adaptively adjusts it location sampling frequency based on the distances between recently observed location samples. As a result, MEILI MCC can collect equidistance location samples at a battery consumption that is significantly lower than that provided by the default sampling strategies. In particular, on current mobile phones, MEILI MCC can collect the movements of an average mobile user for a period of 30 hours. In addition to the location samples, MEILI MCC also utilizes the accelerometer readings that it collects simultaneously and derives various features (i.e., descriptive statistics of the readings and number of peaks/steps) from these readings between consecutive location samples. These features, fused together with period/sequence based features derived from location samples, are used in subsequent mobility inference tasks like trajectory segmentation and transportation mode detection.

The detailed algorithm of this battery sampling strategy, together with the effect of each battery saving strategy, can be seen in Prelipcean et al., (2014). Field tests showed that 70% of the battery is discharged after MEILI MCC runs for 13.5 h, during which the GPS was enabled for approximately 2.5

h (with a discharge rate of 13% of battery capacity every hour for bus, and 5% of battery capacity every hour for walking) and included all the user's movements. These measurements were performed on an HTC One X+ smartphone model with a battery life of approximately 18 hours without running MEILI MCC, and vary with each phone model.

3.3.3 Continuous seamless sampling

In order to guarantee that the sampling process, which can be manually started and stopped by the user, once started, keeps on collecting the desired location and accelerometer measurements with minimal interference with the normal use of the phone, a number of design measures were necessary. First, the sampling process had to be implemented as an independent background process without a UI component that gets automatically restarted by the system after a system restart. Additional, application logic had been implemented on both systems to detect and ask the user to confirm when the user manually tries to disable the positioning capabilities of the mobile device when the sampling process is running. Finally, to ensure that the collected data is never lost and is efficiently saved, periodically, the data collected by each MEILI MCC is replicated to the server-side Data Storage Component (DSC), MEILI has been designed to periodically upload its not-yet-uploaded data in a chunk as an automatic asynchronous task.

3.4 MEILI WAC - WEB APPLICATION

The design of MEILI's Web Annotation Component (WAC) has undergone significant changes throughout the lifetime of MEILI. The latest version of the WAC is shown in Figure 2. While some of these changes were technical changes that were fuelled by general web technology developments, others were fuelled by the need to provide user-friendly interfaces and functions interact with the collected data, i.e., annotate the collected location measurements with travel diary information.

To guide the user through the annotation process an introduction step-by-step explanation of the system is available when the users login for the first time. There is also a help function available, which can guide the users when annotating any trip.

3.4.1 Data model change from points to periods

A major change in the design of the MEILI WAC was due to a recent redesign of the data model used in MEILI. While in the early versions of MEILI WAC it was individual or sequences of location measurements that the user annotated, which then subsequently were explicitly marked as being either transition, stop, or regular points with a given transportation mode, in the current version most of the annotations are designed to take place on time periods.

3.4.2 CRUD (create, read, update, delete) operations for travel diary entities

Apart from these major changes that are a consequence of the redesign on the data model, a number of annotation tools have been implemented, some of which were mainly necessary to deal with artefacts of the not-always-perfect data collection and travel inference methods. First, it has been found useful to implement tools for deleting, altering, and adding points to deal with noisy/incorrect, inaccurate, and missing measurements, respectively. In the earlier versions of MEILI these point-based update operations have also been extended to sequences of points that define polylines that represent trip legs or whole trips. In particular, these extended tools were used to manually set the transportation mode of a trip leg or to delete an incorrectly identified trip that is composed of noisy location measurements or a trip that the user deems sensitive for privacy reasons.

3.4.3 Tools for POI definition

Although the database of MEILI contains public places and transport related POIs from different sources, as these datasets are neither complete nor always accurate, tools have been implemented to define spatial and non-spatial aspects of new POIs, thereby allowing the crowd-sourced collection of POIs. The same tool has also been used to allow the definition of personal POIs such as the home, work, and addresses of friends of the user. Notably, while geocoding, i.e., the automatic mapping of a partial address string to a geographical location, could have been used to define such POIs, due to potential errors as well the limited ability of users to remember precise addresses (even if one can be associated with a place) the designed tools allow to define POIs by dropping place marks on the map.

3.4.4 Trip browsing and revision functionality

To allow the user to view his or her past trips in the context of one another, a browsing functionality has been implemented to allow the user to go back in time and review previously annotated trips. The same trip navigation has also been used to implement the trip revision functionality that allows the user to change travel annotations of a previously annotated trip. As such changes can in principle affect several preceding and succeeding trips, the trip revision functionality has been designed to minimize these propagation effects by only allowing 1) the backward temporal extension of a trip that to fall inside the stop period that precedes the trip that is being altered and 2) the forward temporal extensions a trip that falls inside period that starts with the start time of the stop period that succeeds the trip that is being altered and ends with the end of the stop period that succeeds the trip that succeeds the trip that is being altered.

3.4.5 Sequential annotation process

To ensure that the user annotates/verifies all of the collected data, MEILI is designed to direct the user to the first not fully annotated trip and does not allow the unconstrained browsing of not-yet-annotated trips, i.e., MEILI enforces a sequential annotation process. It is needless to state that the development and implementation of these functionalities, especially the correct maintenance of temporal relationships, have not been trivial or simple.

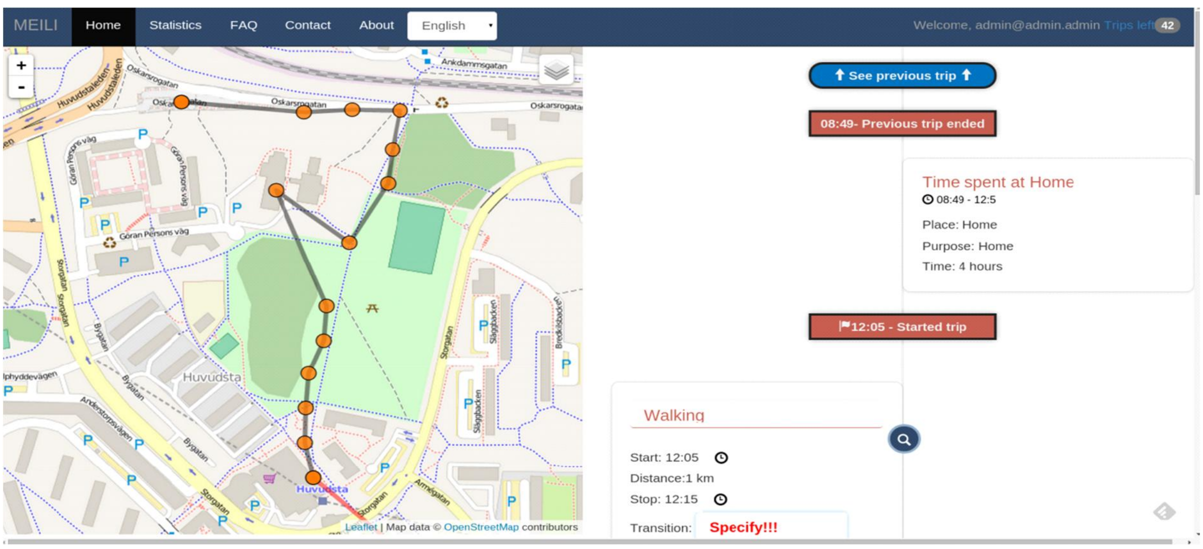


Figure 2: The new interface of MEILI WAC

3.5 MEILI DSC - DATA STORAGE COMPONENT

The design of the MEILI DSC accommodates the unique characteristics of each travel diary entity. This section briefly describes the logic behind the MEILI DSC.

3.5.1 MEILI MCC data replication

The data collected by each MEILI MCC is periodically uploaded to the MEILI DSC for fall back and accessibility considerations. In the current MEILI design, the inference is performed by the server, which requires the MEILI MCC collected data to be available on the server. Even in a completely automated setting, where MEILI MCC would infer the entire travel diary entity set, the fall-back mechanism ensures that any MEILI MCC does not lose all its collected and generated data.

3.5.2 MEILI inferred travel diary entities

To greatly reduce the burden of users annotating their data, MEILI learns how users perform the annotations for trips and trip legs, together with their attributes, and then classifies every new batch of MEILI MCC data into travel diary entities (see Section 4.3 for more detailed explanations). One of the key considerations for the inference methods was the type of output, which, regardless of the inferred entities, is not a single most probable item, but a list of possible items ordered by their likelihood. The ordering has two main functions: 1) it compares the probability of the most likely inference with a threshold value and if it is greater than the value, then the option is "preselected" for the user, if not then the user has to actively select the correct option, and 2) it aids the user in the selection process since the correct option is most of the times in the top 3 items in the ordered list (this is discussed at length in the case study). These inferences are also cached for later performance measurements.

3.5.3 User confirmed travel diary entities

Users interact with the MEILI WAC to perform CRUD operations on the proposed trips and trip legs, together with their entities, to correct any mistakes on the MEILI inferred entities. After the user's interaction via either CRUD or confirm operations, the entities are regarded as ground truth for later performance measurements.

3.5.4 Static POI data

Populating the public places and transportation related POI tables has been far from simple. Namely, different subsets of data from different data sources needed to be selected and integrated. In particular, a relevant subset of POIs from the crowd-sourced data of OpenStreetMap (OSM, https://www.openstreetmap.org/) had to be identified based on functionality-, type-, and name labels that are very liberally defined by the OSM users/maintainers. The selections contained duplicates that needed to be automatically identified and eliminated by custom clustering and merging of POIs based on spatial proximity and string similarity between labels. Using a similar methodology, the OSM-based transportation POIs had to be additionally merged with official data sources from the local public transport authority SL (http://sl.se/), which was also incomplete, a process that proved to be cumbersome.

3.5.5 Crowd-sourced POI data

Since the POI dataset available from either OSM or SL is incomplete, users can enter personal POIs, which are visible to, and editable only by, the user that entered them, or public POIs, which are visible to, and editable by, all users. The public POIs are, for example, transportation stations, and users can edit the type of public transportation modes available, together with the lines. The personal POIs are any non-transportation POI, such as home, work or restaurant locations.

4 ALGORITHM DEVELOPMENT

4.1 TRIP QUALITY INDICATORS

To analyse the performance of a system that collects travel diaries, we use the analysis framework proposed by Prelipcean et al., (2015a), which allows for the calculation of quality indicators for each recorded trip and trip leg. The framework proposes two indicators, a spatial quality indicator, which measures how well a system captured the geometry of a trajectory associated with a trip, and a temporal quality indicator, which measures how well a system captures the timeline of a trajectory associated with a trip. In order to define what "how well" means, one has to specify the maximum allowed distance between two consecutive locations (for the spatial quality indicator), and the maximum allowed time difference between two consecutive locations (for the temporal quality indicator).

As mentioned in Prelipcean et al., (2015a), computing quality indicators allows for discriminating between different types of entities proposed as trips such as: well captured trips (high spatial and temporal indicator values), sparsely captured trips (low spatial and temporal indicator values), noise (low spatial and high temporal indicator values), and possibly merged trips (high spatial and low temporal indicator values). Furthermore, the well captured trips can be considered candidates for ground truth, which is useful when comparing trips as captured by different travel diary collection systems from the same users during the same time periods.

4.2 TRIP MATCHING

When inspecting whether the travel data collection can be automated, one has to collect travel data with the traditional travel diary, as well as with the new tools, on the same user set, for the same study period. The difficulty of collecting the same data by using two different collection systems is finding the correspondence between trips as collected by both systems, as well as identifying trips that have been collected by one system only. The aforementioned framework (Prelipcean et al., 2015a) proposes a method to match trips from two systems by using two constraints: 1) temporal co-occurrence, which implies that the difference between the start/stop time of a trip as recorded by one system has to be within a threshold value of the start/stop time of a trip as recorded by the other system, and 2) identical purposes, which restricts two trips to be wrongfully matched solely on temporal co-occurrence (the case of short trips).

4.3 TRAVEL DIARY ENTITY INFERENCE

Providing a system for the fully automatic collection of travel diaries implies the need for automatic inference methods that associate travel information with the MEILI MCC collected data. The following subsections describe the principles behind the four travel inference methods that have been developed for MEILI.

4.3.1 Trip and trip leg segmentation

Segmenting a sequence of locations into trips and trip legs is not an easy task, and it can be approached in multiple ways - three of the most widely used ways are presented in Prelipcean et al., (2015b). Based on the prototype implementations and evaluation of the three approaches MEILI currently adopts the explicit-consensus-based transport mode segmentation approach. With respect to trip inference, stop periods are detected by finding the longest periods during which the movement characteristics based on relatively accurate positions, as well as the accelerometer readings, suggest that the user is stationary for a period of at least 5 minutes. With respect to trip leg inference, start and end times are detected as time instances that separate two segments of location measurements that have similar movement characteristics internally but differ from one another.

4.3.2 Travel mode

During the development of MEILI, a number of different machine learning techniques as well as different general approaches for travel mode inference have been developed and tested (see Prelipcean et al., 2014; Prelipcean et al., 2015a). To infer trip legs, the current method of travel mode inference uses the explicit-consensus-based transport mode segmentation approach with a nearest-neighbour classifier. To accommodate for a nearest-neighbour classifier, MEILI DSC periodically generates and stores statistics on all annotated travel modes for both trip legs and GPS readings. These statistics are computed both for each user, and for the population. The classifier then compares each inferred trip leg and the readings within its period with the period-generated statistics and provides an ordering based on the similarity of the entities. If the user has annotated trips before, the classifier compares the inferred trip leg and its readings with the user history-generated statistics, and provides an ordering based on the similarity of the entities. When proposing the mode, the classifier combines the two orderings and puts a higher weight on the user history ordering. The output of the classifier is a list of all modes, ordered by the mode probability, which is then presented to the user and, if the highest probability is lower than a threshold value, the user is prompted to select the travel mode in the MEILI WAC.

4.3.3 Trip destination

Different types of methods have been trialled for the task of trip destination inference, such as: 1) personal history based methods based on the previously observed destination (POIs) of the user, 2) proximity based methods based on the proximity of possible destinations to the end point of the trip, 3) conditional probability methods based on conditional probabilities of POI types that are primarily based on the time-of-day and day-of-week of the trip, and 4) spatial significance methods based on the importance of a destination given the relative spatial distribution of different types of POIs in the vicinity end point of the trip.

Based on empirical evaluations the chosen final method for MEILI's trip destination inference is as follows. MEILI detects a set of candidate POIs near the user's destination by using a distance buffer around the end of a detected trip. If the user has previously specified destinations via MEILI WAC that are within the candidate set, then MEILI returns the list of candidate POIs ordered by the number of times the user visited each location previously. In the absence of prior annotations on the candidate set, MEILI returns the candidate POI set ordered by the distance to the last recorded point of the trip. Based on the number of previous visits, MEILI computes the probability for each POI and, if the highest probability is lower than a threshold value, the user has to select the destination in the MEILI WAC.

4.3.4 Trip purpose

Different types of methods have been trialled for the task of trip purpose inference, such as: 1) personal history based method that infers the most likely purpose for a given destination that the user revisits and 2) conditional probability based trip purpose inference methods make inferences based on conditional probabilities of trip purposes that are primarily based on the time-of-day and day-of-week of the trip. The used classifier is a combination of both methods that leverages the information available regarding the user's history to choose between the two.

Based on empirical evaluations the chosen final method for MEILI's trip purpose inference is as follows. MEILI detects the trip's purpose as a function of destination, a direct consequence of this is the fact that the purpose is re-inferred if a new destination is selected by the user. Given a destination, the trip purpose is the most often occurring purpose at the given destination, if it was visited before. In the absence of a previous visit to the destination, the purpose inference uses a Naive Bayes classifier based on the user population most probable purpose within the time frame (the inference takes into account the day of week and the hour of day). MEILI computes the probability for each purpose, returns the purpose list ordered by the probability, and, if the highest probability is lower than a threshold value, the user has to select the purpose in the MEILI WAC.

5 EVALUATION METHOD

Two methods for collecting travel data have been compared in this project via two field studies – one smaller pilot in 2014 and a larger field trial in 2015. In both field studies the comparison is made by letting *the same users* collect travel data with a traditional web-based travel diary and the smartphone application MEILI. Data comparison is made for *the same one day* with both methods. During the day chosen for comparison, users have the app running, but they do not see any results of logged trips. The day after comparison day, users first fill in the traditional travel diary. Only after submitting the traditional travel diary, users get the link to the webpage for review and annotation of the travel data collected with the MEILI system. This way they are not influenced by the GPS data when filling out the traditional travel diary. After comparison day, users are encouraged to have the app running for a whole week, but do not fill out any more traditional travel diaries. The data set for comparison is thus smaller than the data set collected by MEILI for the whole week.

There are many different dimensions along which the two data collection methods – web-based traditional travel diary and the smartphone based MEILI system – could be compared. This chapter describes the dimensions that have been chosen in this project for evaluation and comparison of the two travel data collection methods.

5.1 RESPONDENTS

5.1.1 Bias in selection of respondents

Bias exists in all surveys and need to be corrected to create a representative selection and result. This project will not be able to fully describe the bias in the two methods, since that would require two completely separate studies in order to see the bias in each method. In this project, the question of bias will be discussed based on experiences from the recruitment.

5.1.2 Workload for the respondents

Workload is often seen as one of the largest drawbacks of traditional travel diaries. This report will discuss and compare the workload of the two methods, but one should note that the workload of MEILI is continuously reducing as the automatic identification of mode, purpose, destination etc. is improved.

5.1.3 Integrity

The project will investigate whether there is a difference in experience of personal integrity depending on collection method.

5.1.4 Other information about the respondents

Estimation of transport demand models require information about the respondent's age, sex, driver's license, access to car in household etc. However, this study will only collect basic information about the respondents in the traditional travel diary since the same respondents participate in both surveys.

5.2 QUALITY IN COLLECTED TRAVEL DATA

5.2.1 Measured variables using both methods

The travel data listed below will be collected using both methods and their results will be compared in coming sections of the report.

- Start and end time of each trip
- Travel distance per trip and mode
- Travel time per trip and mode
- Waiting time per travel mode (pilot only)
- Travel mode aggregates
- Origin and destination
- Purpose
- Number of trip legs per trip

5.2.2 Definitions of measured variables

For the comparison of different collection methods it is very important that common definitions of the measured variables are used and that these definitions correspond to the definitions most common in national travel diaries and used in transport models. It will not be possible to collect data for all variables with both methods, which will be one of the results of the project. The lists in the tables below should be seen as wish-lists over which variables we want to collect data for.

Trip and trip leg

In this study, a trip is defined by its purpose. Several modes can be used during the same trip. Each part using one mode is then called a trip leg. The respondent will in the traditional travel diary be asked about all modes used to perform a trip for a certain purpose. The GPS-based collection will use the stop time at a destination to define the end of a trip. It might therefore be difficult for the application to differentiate between transfer point and end point of a trip, but the respondent will be able to correct eventual errors via the webpage.

Purpose

Categorizing trips into different purposes is very important for estimation of transport models, since there are major behavioural differences depending on trip purpose. For example, user valuation of travel time differs substantially between work trips and leisure trips.

The following thirteen trip purposes are common in traditional diaries and will be used in this project, both in the web-based traditional diary and in the MEILI system:

- Travel to work
- Travel to school
- Business travel
- Leisure travel (e.g. go to cinema, theatre)
- Food/grocery shopping
- Non-food shopping
- Personal business (e.g. medical visit, bank, cutting hair)
- Pick-up or drop-off children/other persons
- Visit relatives and friends
- Sport/hobby related travel
- Restaurant/Café
- Return home
- Other (incl. walk/travel without specific purpose)

Mode

For comparison with the traditional travel diary a detailed list of modes is used. The fifteen modes listed below are commonly used in traditional travel diaries in Sweden. It is however very difficult for the MEILI system to differ between e.g. car as driver and car as passenger. The respondent will have to review and correct eventual misclassifications via the web interface.

- Walk
- Bicycle
- Bus
- Car as driver
- Car as passenger
- Commuter train
- Ferry boat
- Flight
- Moped/Motorcycle
- Paratransit
- Subway
- Taxi
- Train
- Tram
- Other

5.3 LIMITATIONS OF THE STUDY

A description of the cost of conducting a survey of each type (traditional travel diary contra smartphone-application) is not included; this is a research project and the aim is to develop and test methods.

The focus of the project is on travel data collection, which means that questions about preferred departure time, reasons for mode choice and travel costs are not included. Attitude questions will also not be asked.

6 PILOT

The users in the pilot study were recruited from SLL, Trafikverket, TK Stockholm stad, Trafiktekniska föreningen, Sweco and KTH. This of course creates a bias towards people working with transport related questions, but given that the main purpose of the pilot is to collect learning data for the machine learning algorithm and evaluate the actual system this was not seen as a problem. By the time of the pilot, the smartphone app was only available for Android.

6.1 RESULTS

Out of the original 51 persons that showed interest for the pilot, 42 answered the travel diary and 39 downloaded the app and managed to collect data. Among the drop-outs four people realized they had an iPhone, two had problem downloading the app, one person had sent his phone for service and two were on vacation.

Of those 39 who collected data 30 managed to annotate their data and 24 persons annotated all of their data. The reason that they didn't annotate all of their data was different problems with the website. The main problem was that the respondents had trouble understanding how to annotate their data.

In total, around 1050 trips were collected and 720 of those where annotated.

6.1.1 Method evaluation

After the pilot, a follow-up questionnaire was distributed to the users. There were 34 users that answered the questionnaire and the results are presented below.

Installation and use of the app

In general, the participants managed to install the app without any problems, 85% (29 respondents) claim that they had no problem installing the app. Those who had problem had never installed an app before or hadn't registered a Google account. 95% (32) managed to install the app without any help from other people.

85% (29) had the app installed during the whole week, see Figure 3.

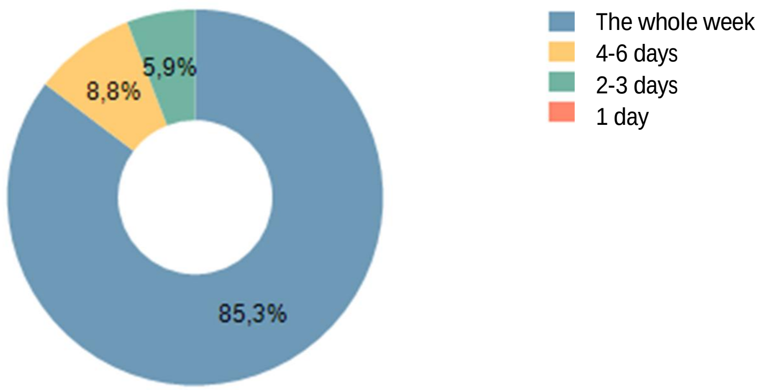


Figure 3: For how long did you have the app installed?

Of those who only used the app for a couple of days, 40% (2) claimed that it drained the battery too much and 60% (3) that it didn't work. However, if we look at the comments from those who said it didn't work, two respondents said that they didn't travel or were abroad while the last one had an app installed that cleared the memory of the phone and automatically turned off the data collection.

Looking at the battery life of the phones, 27% (9) didn't notice any big difference when the app was running while 24% (8) had to charge their phone a lot more often, see Figure 4. The reason for these differences is hard to analyse since there are so many things affecting the battery life. Which phone model, how long it has been used and how much the respondent travel are just some of the parameters affecting the battery life.

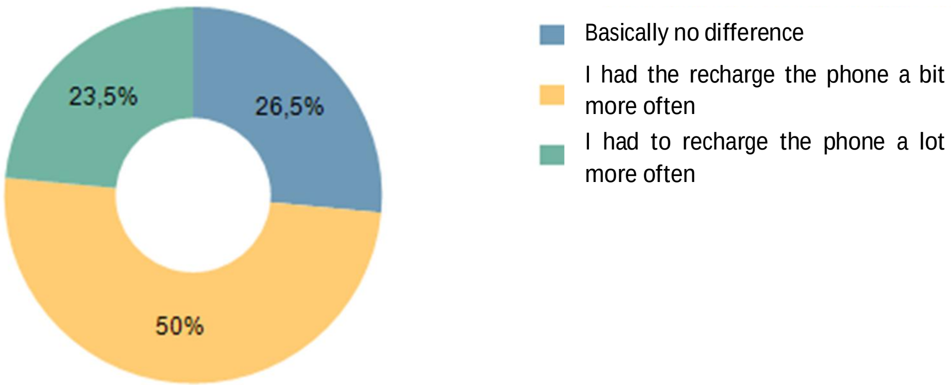


Figure 4: How did the app affect the battery life of your phone?

Website used for annotation

The website used for annotation caused a lot of problems for some of the users. A majority answered that they thought the website was difficult to understand, see Figure 5. However, there were a number of users that found it intuitive and had no problem annotating their trips.

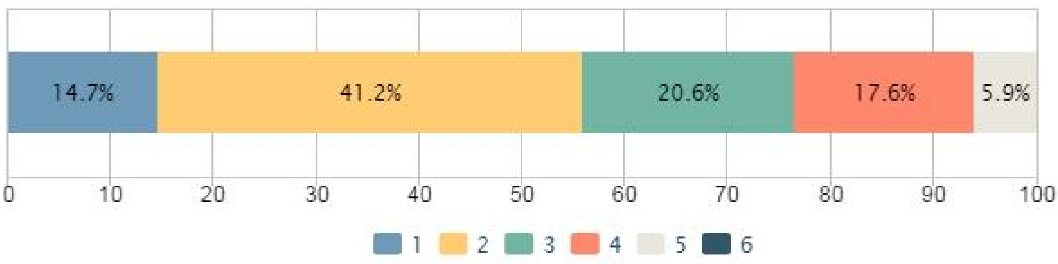


Figure 5: How did you experience the website used for annotation of the collected data?
1 - Very difficult to understand, 6 - Very intuitive

Among the comments there are a few issues regarding the annotation that are recurring:

- Difficult to get started and know where to begin.
- The system automatically detected to many transition points which takes time to correct
- Website a bit slow at times
- Difficult to define POI
- Difficult to know how to deal with trips where only part of the trip has been detected
- The pop-up sometimes disappears and you have to zoom out to see it
- Difficult to choose between two points when they are close to each other and when the trip is short

There were also some suggestions for improvement:

- Zoom to next trip leg
- Possibility to add points manually and/or move points that are in the wrong place
- Undo-button (today it is only available at certain stages of the process)
- Support for Internet Explorer and mobile/tablet
- Confirmation when a step and a complete trip is annotated
- A cleaner website where only relevant information for each step is visible
- Better guidance throughout the process
- Four step process. Points – Mode – POI – Purpose

Some respondent has reported that the system merge trips from several days, even though they have been at home during the whole night. Furthermore, the time stamp detected by the system was wrong and had to be corrected for some users.

Regarding the list with POIs 47% (14) found it hard to find an appropriate POI and 20% (6) would have liked a broader range of alternatives to choose from. The rest, 33% (10) had no problem finding the right POI. Among the comments there were some people that had problem understanding the concept of POI and how it relates to the trip purpose and the location of home and work.

Detection of trips

For more than half of the participants (52% (15)), the system did not detect one or more trips that they had performed. However, it is difficult to find a pattern among the undetected trips. In some cases, they are shorter walk or bicycle rides, in other subway journeys underground. There are also examples of when the data collection has stopped unexpectedly without a reason, but there are also cases when the user has forgotten to turn on the app.

On the question whether the app had detected a trip that the respondent hadn't done 60% (18) answered yes. Basically all of these trips are short trips around their home or work place.

Travel diary

Regarding the travel diary that the participants had to fill in for the first day of the pilot, the opinions are very varying. Most people found it ok, some people preferred the app but others preferred the travel diary. Some people claimed that it was hard to remember all the trips and the details of the trips.

Integrity

On the question of which method was the most intrusive 67% (23) answered the smartphone app and 6% (2) answered the travel diary, see Figure 6.

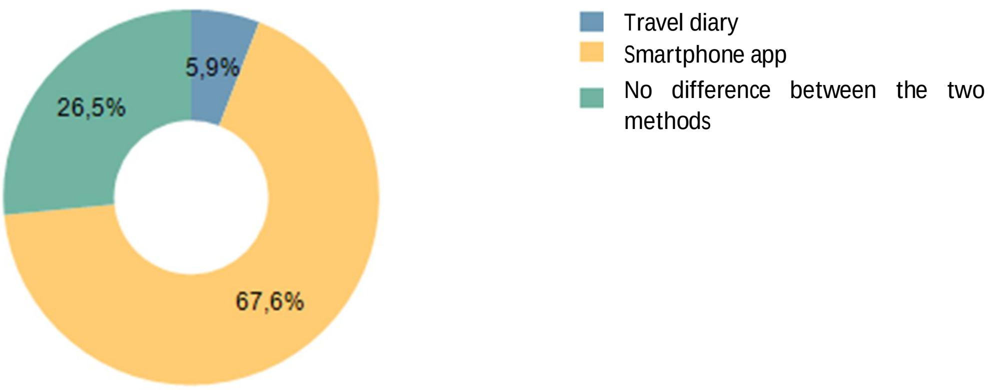


Figure 6: Which data collection method was most intrusive?

In the comments the following opinions were recurring a few times

- Since the user can choose whether to participate or not the integrity is not a big issue
- Important to state what will happen to the data after the trial and how it will be used
- The app collects all movements and the user has no control over which data that is transmitted (not entirely true…)

Possible improvements to the system

The final question in the feedback questionnaire was an open question where the respondents were asked to list possible improvements to the system besides the ones that they have written in earlier questions. Below is a summary of the more useful and interesting ones:

- Improve the automatic detection of transition points, especially in the subway network and when biking
- Improve the automatic split between different trips
- Improve the annotation process
- Longer and more describing help texts in the WAC
- Limit the zooming by divide the process into four stages Points – Modes- POI – Purpose
- More POIs and trip purposes to choose from
- Let the users understand how much work it is to annotate the trips
- Improve the interface of the webpage
- Make it possible to annotate my trips in the app
- The app should provide some interesting statistics or some other functionality to further motivate the user to use it

Even though many participants have struggled and had to put a lot of time and effort into annotating their data 68% (23) are positive to participate in focus group discussions on how the system can be improved.

Feedback received through e-mail during and after the pilot

From the feedback and support related matters received through e-mail during the pilot the following are the most frequent and urgent issues to solve:

- Internet Explorer (IE) – Several users haven't noticed the information about supported browsers and have tried to annotate their data in IE. Some people prefer to do the annotation at work and have only access to IE. This could be solved by implementing support for IE (preferable) or implement a pop-up that occurs when a user is using IE.
- One-point issue – The one point issue that make the web app crash/get stuck in loading has occurred for a number of users and needs to be solved.
- Loading of webpage – Even though there is no data to annotate the webpage seem to be loading data the first time the users logs in. Also, the loading takes some time for long trips. We should include a text under the loading bar that this process might take while.
- Forgotten password and user-name – Several people have forgotten their username and/or password. We need a function for this where the user can get the username/password sent to their e-mail.
- Server breakdown at numerous occasions – The server has crashed at numerous occasions. This needs to be solved. Stress test of database? An alarm-function needs to be implemented.
- The app stops collecting data – For some users the app has stopped collecting data even though the app says it is collecting data. We need to take a closer look at this. A restart has solved the problem but the user won't notice the problem until it is time to annotate the data.
- Trouble understanding the annotation process – The annotation process has to be improved. Maybe it could be done in four steps: Points – Mode – POI – Purpose. The web app needs to communicate with the user and tell him/her what to do next and when a trip is finished.
- Time stamp – The time stamp of previous trips that is showed in the list is from the original trip, not from annotated data. Also, the time stamps seem to be wrong for many users and they have to correct it.
- iPhone – Several people has shown interest for the project but can't participate since they have an iPhone. The development of an iPhone app would probably more than double the number of users and the data that is being collected.

Besides these issues there are a number of smaller issues/possible improvements to the system that has been reported and identified:

- Better to collect data Tuesday – Monday than Monday to Sunday. If we start on Tuesday, we have Monday available to help the users install the app and prepare for the collection period.
- Easier access to support e-mail, include a link in the web app.
- Possibility to zoom using Ctrl++
- It should be possible to restart the annotation of the current trip, as it is now it is only possible to go back at certain steps.
- It should be possible to also see next trip. This would make it easier to set the right end point and whether or not trips should be merged.
- What happens with the trip chain when a time stamp is changed?
- If the user undoes a trip that was part of a trip chain before it was annotated, then the whole original trip chain is undone.

Summary

As expected the user experiences vary a lot between the different participants in the pilot. This is probably mainly because the level of IT-knowledge varies, but also because the system works better for some phone models than others. A trip that spans over several days, wrongly detected transition points and trips where only parts of the trip has been detected aggravates the annotation process which affects the user experience. Besides some rare problems with the installation, registration and the functionality of the app the collection has worked relatively smoothly for the majority of the participants. The main problem for several users has been the annotation process and the website used for this. The main conclusion from the pilot is that this process has to be improved and simplified.

Finally, to summarize the feedback there are a number of issues that has been raised both from the feedback questionnaire and through e-mail that has to be addressed before the main field trial takes place.

- The annotation process has to be simplified and the website has to be cleaner where only relevant information for each step is visible. A four step process has been proposed: Points – Mode – POI – Purpose.
- Improve the automatic detection of start, end and transition points
- A function where the user can get the username/password sent to their e-mail
- Support for Internet Explorer or a pop-up that tell the user to change browser
- Undo-button that is available throughout the process and a restart-button
- Possibility to add points manually and/or move points that are in the wrong place
- Solve the one-point issue that makes the webpage crash
- The time stamp of previous trips that is showed in the list is from the original trip, not from annotated data
- Find out why the server has crashed at numerous occasions and fix this

6.1.2 Data comparison

In the pilot study, the number of trips captured by both MEILI and the traditional travel diary (PP[1]) is 43 trips, the method used for matching trips is previously described in section 4.2. The number of trips that were captured only by MEILI (44) is roughly equal to the number of trips captured only by PP (51). The algorithms on how the trip, trip purposes, activity locations and travel modes have been inferred from the app readings can be seen at Section 4.3. More detailed algorithms on this matter, please see Prelipcean et al., (2014), and Prelipcean et al., (2015a). As can be seen in Table 3, the characteristics of the trips captured by both systems do not differ considerably as derived from the MEILI MCC data, and as declared by users in PP. In the pilot, the temporal and spatial quality indicators were computed only for the MEILI dataset, but they can also be computed for the PP dataset, as shown in the large field study.

[1] PP is short for Paper-and-pencil method, by which we mean the same as traditional travel diary

Table 3: MEILI and PP trip statistics

	PP	*MEILI*
Duration(mins)	23±15	24±19
Length (m)	6300±7600	6300±6000
Trip legs	1.8±1.1	1.8±1
Waiting time	-	6±10
Temporal quality	-	85%±30%
Spatial quality	-	61%±36%

Analysing the difference between trips captured by both systems and trips captured by either system (see Table 4), it is noticeable that users forget to declare mostly short trips (median duration of 20 minutes), travelled with only one mode and with very low waiting times, which indicates that the user could be travelling via a non-public/non-shared travel mode such as walking, bike, or personal vehicle. The trips recorded only by PP are of similar duration to the trips recorded by both systems, the only main difference being the length of the trips, which are roughly 1.5 km shorter than those travelled by both systems. These might be trips during which the users did not carry their phone or during which the data collection process was not working adequately.

Table 4: Trip statistics for PP only, MEILI only and for trips captured by both.

	PP and MEILI	*PP Only*	*MEILI Only*
Duration (mins)	24±19	23±20	64±85 - median 20
Length (m)	6300±6000	4500±5000	3800±5100
Trip legs	1.8±1	1.7±1.1	1.2±0.3
Waiting time	6±10	-	1±3
Temporal quality	85%±30%	-	83%±29%
Spatial quality	61%±36%	-	60%±32%

The data collected by MEILI can be aggregated to generate more insight regarding trip legs than can be derived from PP. As seen in Table 5, the median duration of a mode can be computed, together with the percentage of time travelled with that mode in a trip. This can offer more insight regarding how people choose travel modes and which travel modes are considered primary. Similarly, one can compute the waiting time that is associated with each mode (very low waiting time for personal modes such as walking, bicycle, car, and taxi), and the percentage of the time spent while waiting for each mode during a trip (high percentages for public transportation modes such as bus, subway and train). This information can be further analyzed to provide insight regarding how people chain their travel modes during trips, or how people travel under different purpose assumptions, but this type of analysis is outside of the scope of this project.

Table 5: Trip leg data collected with MEILI.

	Duration (min)	*Duration (% of trip duration)*	*Length (m)*	*Length (% of trip length)*	*Waiting time (min)*	*Waiting time (% of trip duration)*	*Number of trip legs*
Walk	6	41	488	31	1	3	414
Bicycle	17	86	2510	91	2	6	113
Moped	14	71	6624	99	-	-	6
Car (driver)	15	88	5981	97	5	5	140
Car (passenger)	13	95	4616	97	3	5	32
Taxi	7	54	3445	59	0	0	2
Bus	6	18	1389	30	5	13	63
Subway	8	29	2955	53	8	19	104
Tram	45	85	8207	96	-	-	3
Commuter train	13	25	11156	51	17	27	26
Train	73	61	174082	89	20	23	6

7 FIELD TRIAL

Recruitment for the field trial was made via the large travel diary send out to 130000 people in Stockholm County in September/October 2015. There was a text at the end of the *missiv-letter* and also a possibility to enter one's e-mail address at the end of the web-based travel diary. The text in the *missiv-letter* was given only in Swedish, whereas there was a Swedish and an English version in the web-based travel diary. The English translation of the text included is:

Do you want to participate in a research project where you would use your smart phone to collect travel data? The project is financed by the Swedish Transport Administration and aims to improve the methods by which we gather travel data. If you are interested, please enter you e-mail below and we will send you more information.

At the time of recruitment, no incentive was given. After installing the app and stating to collect data, respondents got information about a lottery with three prices for those who annotated at least three days of travel data.

Descriptive data about the respondents that participated in the field trial shows that there is bias in recruitment. For example, more men than women signed up. This was however not seen as a problem since the purpose of the field trial is to evaluate a method to collect travel data.

7.1 RESULTS

1559 people showed interest to participate in the field trial, most recruited from the web-based travel diary (only 30 people from the *missiv-letter*). 495 signed up for the field trial. It is not evident why so many respondents dropped off already before installing the app, but three major reasons are likely:

- Some of the users possibly just wanted to have more information about the research project, they did not see it as signing up for a field trial.
- iPhone-users first got the information that they could not participate. One day later – when the app was suddenly accepted and published on Appstore – they got new information that they could participate. Probably some users read only the first information about participation not possible for iPhone users.
- A handful of drop-outs were not Android or iPhone users and could therefore not participate (mainly WindowsPhone users).

Out of the users who signed up, 431 responded to the travel diary and 293 managed to collect data using the smartphone application. The reason for the drop-out in collecting data was mainly due to a new setting in iOS 9.0 that made the iPhone version of the app to misbehave. This new setting was released just a few days before the trial and there was no chance for the developers of the app to adapt to the change. It will however not be a problem to adapt to the new settings in the long term.

In total 2142 trips were collected and annotated by 171 users in the field trial. 51 of the users annotated trips covering more than a week. Those annotated trips were based on the results of trip, trip purposes, activity locations and travel modes inference algorithms (as described earlier at Section 4), thus, the users were only required to approve and/or correct/edit the given suggestions whenever necessary.

7.1.1 Method evaluation

A follow-up survey was sent out to the users after the field trial. 303 people answered the follow-up survey and the results are given in this section.

Installation and use of the app

As Figure 7 shows, installing the app went smoothly for most respondents. 87% had no problems when installing the app. The respondents downloaded the app from GooglePlay (Android users) or Appstore (iPhone users), which is the common way to install an app and experienced by the vast majority of smartphone users.

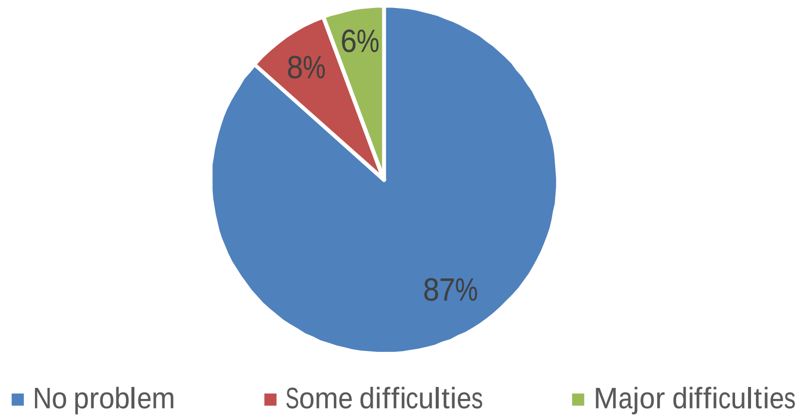


Figure 7: How did you experience installing the app?

Regarding battery life of the phones as many as 62% did not notice any major difference in battery consumption (Figure 8). This is major improvement compared to 27% not noticing any difference in battery consumption during the pilot in 2014.

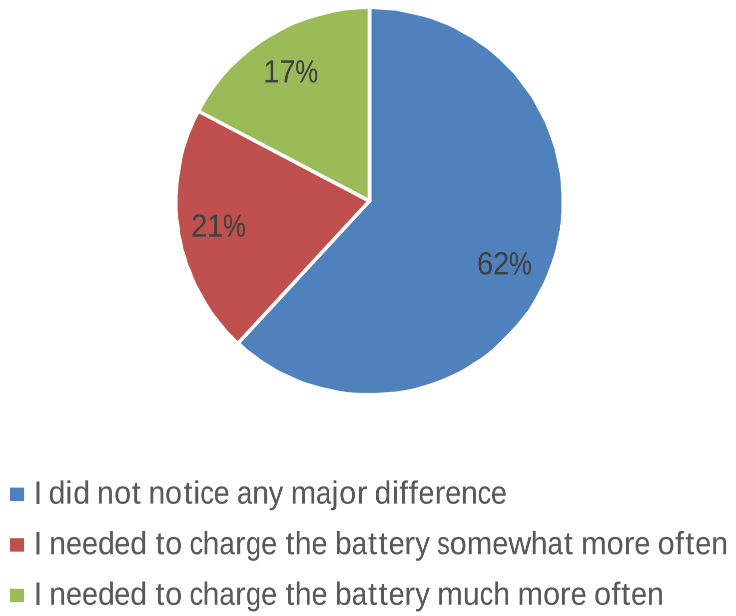


Figure 8: Did you experience any change in battery life time when you had the app running?

60% of the respondents state that they did have the data collection running for the whole week. Figure 9 shows the reasons given for not having the app running the whole week. Surprisingly, only 7 percent state that they turned off the app due to integrity reasons. Most users that did not have it running for the whole week state that this was due to that it did not work as expected. Most of the users stating that the app did not work as expected were iPhone users (50 out of 144 iPhone users compared to 11 out of 148 Android users).

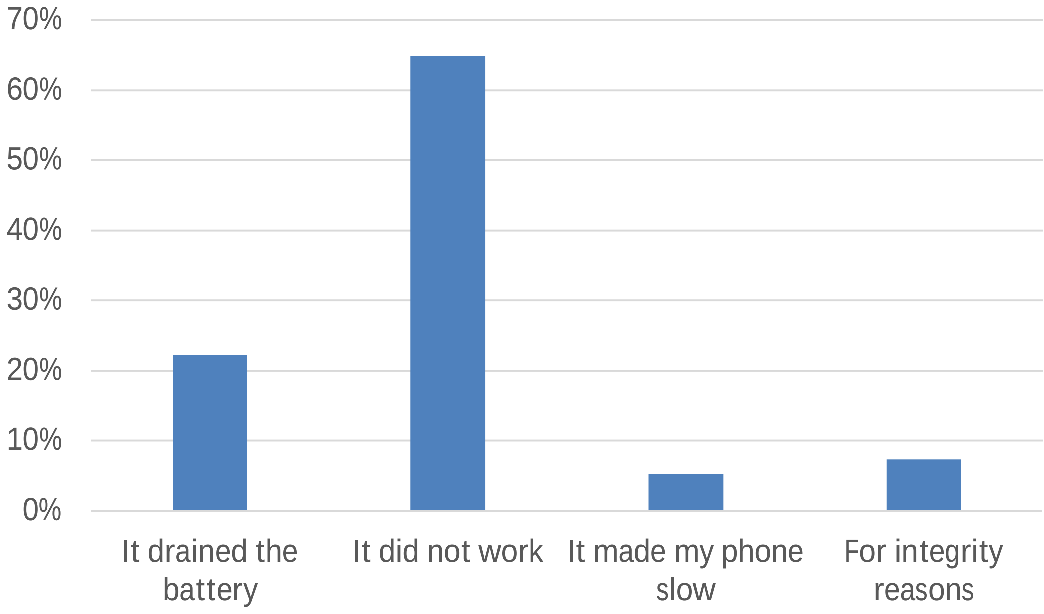


Figure 9: Why didn´t you have the app running the whole week?

Website used for annotation

Figure 10 shows that, even though major development work was put into the website used for annotation, it still worked poorly for many users. From the user feedback one can see that complaints during the field trial are more about website performance (slow website, cannot move on to next trip etc.), and less about the site being un-intuitive and difficult to understand.

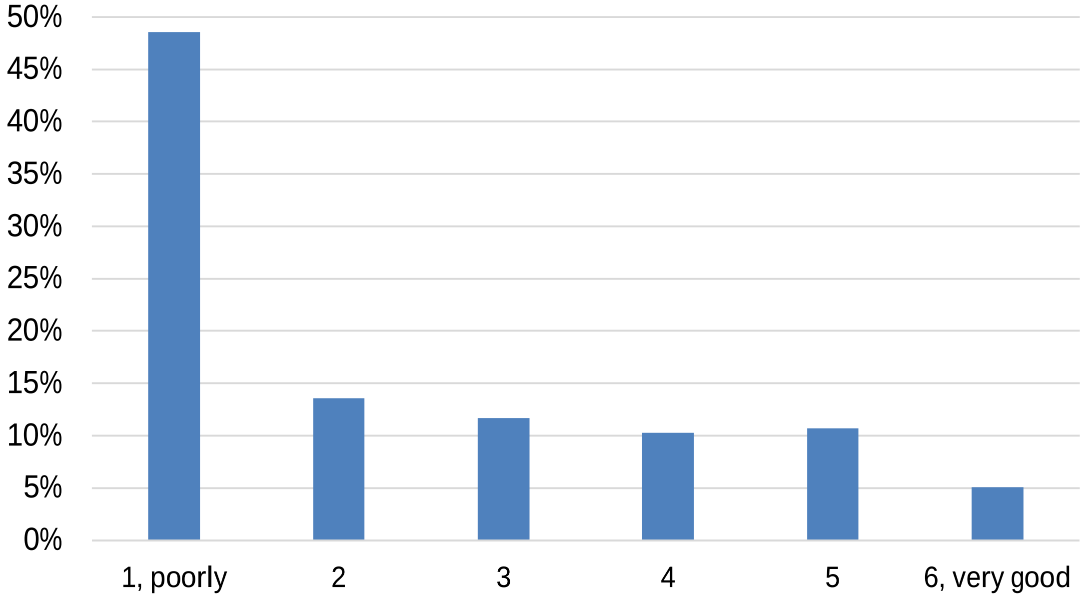


Figure 10: How did the webpage where you were asked to review and annotate your travel data work?

Detection of trips

Of the users, 27% state that they made at least one trip that the app did not register and 19% say that the app registered a trip that they did not undertake. These percentages in faulty detected trips are much lower than in the pilot. Especially the percentage of users saying that the app registered a trip they did not make has gone down from 60% to 19%. However, in the follow-up survey for the field trial, as many as 57% answered “Don´t know” on both questions.

Travel diary

Figure 11 shows that about 40% had no problem with the traditional travel diary. About 30% found it acceptable, but time consuming.

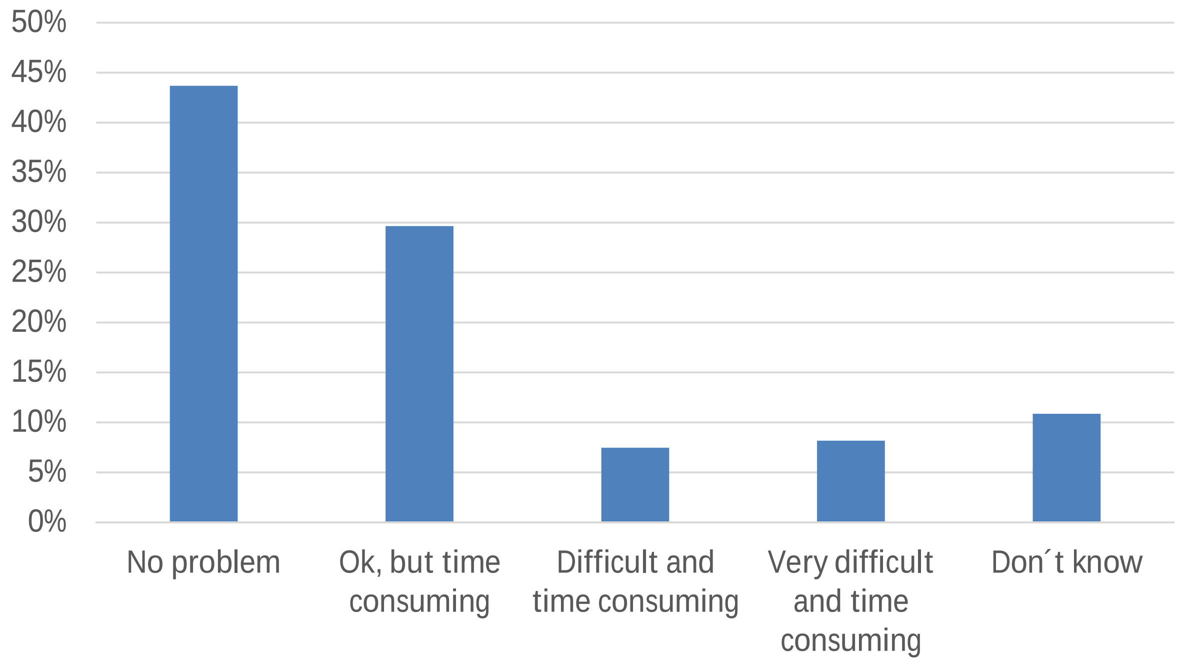


Figure 11: How did you experience the traditional travel diary?

Integrity

Regarding which of the two methods that the users found most intrusive, 43% answered the smartphone app (Figure 12). This number has gone down substantially compared to the pilot were 67% answered the smartphone app.

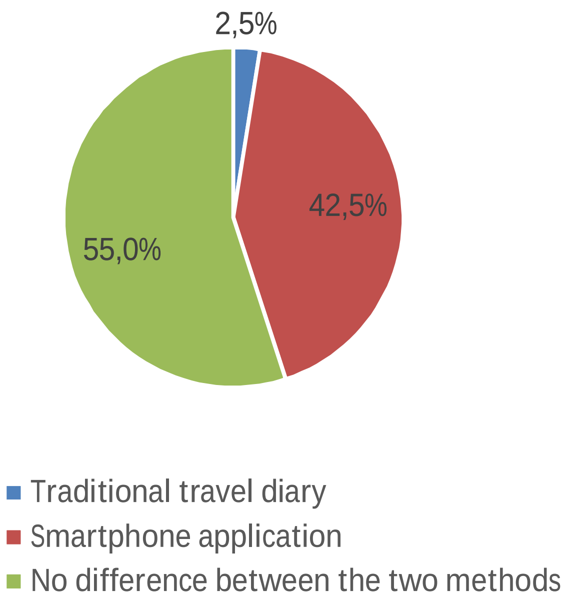


Figure 12: Which travel data collection method do you find most intrusive?

Common support issues during the field trial

- Need new username and/or password
- No possibility to use Chrome
- Problem to log in/nothing is shown when logging in on webpage
- Is stuck on a trip and cannot move on to annotate more trips
- ”Invalid state” because of too many clicks on webpage

Common feedback during the field trial

- Questions about bias and the project in general
- Constructive comments about how the webpage can be improved
- Complaints about the webpage

Common feedback from the follow-up survey - comments

- The webpage is too slow
- User is thrown out from the webpage and have to log in again
- Users loosing annotations when login in again
- Webpage crashing
- It should be possible to change the password without contacting support
- It would be good to have an overview of all trips during the day
- It would be good to have an overview of both annotated and not annotated trips
- Webpage experience improved after 2-3 days of the trial

Summary

As in the pilot, the installation and functionality of the app worked smoothly for most participants during the field trial except for the iOS 9.0 issue (see above). Furthermore, the battery consumption is improved compared to the pilot and only 17% now say they need to charge their phone much more often compared to when not using the app.

The main problem is still the annotation process and the webpage used for annotating trips. Since a lot of time was spent on developing an iPhone version of the app, there was not enough time to improve the webpage for annotations. However, the conceptual change of the webpage that was made seems to be favoured by the users. Complaints that used to be about the webpage being un-intuitive now mainly concern bugs and other performance issues.

7.1.2 Data comparison and analysis

In the large field trial, the number of trips captured by both MEILI and PP, according to the matching method presented in Section 4.2, is 189. Different from the pilot study, where the number of trips captured only by PP is roughly equal to the number of trips captured only by MEILI, in the large field trial, the number of trips captured only by MEILI (166) is larger than the number of trips captured only by PP (112). As seen in Table 6, the characteristics of the trips captured by both systems do not differ considerably as collected by MEILI or by PP. One of the attributes that varies most is the length, which can be obtained in different ways. In the case of MEILI, the length of a trip is defined as the distance covered by the consecutive points that form a trip. In the case of PP, the length of a trip is either declared by the user in a web-form (estimated), or it is computed as the distance covered by the consecutive points that fall within the period declared by the user as a trip in PP (as an interval). The disagreement between the lengths collected via the two methods is caused by approximation errors (users tend to round up the declared values) and by the lack of attention to measurement units, which results in errors that are orders of magnitude higher than approximation errors. The length of trips captured by MEILI is similar to the interval length captured by PP. The spatial quality indicator values for both PP and MEILI are similar, but the temporal quality indicator values differ, which could be caused by the approximations that people make when declaring the intervals during which they were travelling.

Table 6: PP and MEILI statistics from large field trial.

	PP	*MEILI*
Duration(mins)	29±24	26±24
Length (km)	16.9±57.7 (estimated) 10.9±18.9 (interval)	11.8±18
Trip legs	1.9±1.3	1.7±1.3
Temporal quality	55%±29%	62%±31%
Spatial quality	71%±34%	71%±33%

Analysing the difference between trips captured by both systems and system specific trips (see Table 7) it is noticeable that users forget to declare mostly short trips in PP (median duration of 12 minutes), travelled with one mode. The trips recorded only by PP are of similar duration to the trips recorded by both systems but the PP only trips' length is roughly double, which suggests that the trips were performed via a high speed travel mode. These might also be trips that were difficult to annotate by the user, which might have caused the user to delete trip in the MEILI WAC. Interestingly, trips captured in MEILI only have a significantly lower temporal indicator values than the trips captured by both systems. This may suggest that the trajectories describing these trips were difficult to record due to the used travel mode (e.g., subway trips).

Table 7: The statistical aggregates (average ± standard deviation) of the trips that were captured by both systems, only by PP and only by MEILI. The trips captured only by MEILI are either very short (the median duration is 12 minutes) or very long, which causes a large standard deviation, and mostly traveled with one travel mode. The trips captured by PP only are usually traveled by two travel modes at a higher speed than the trips captured by both systems.

	PP and MEILI	*PP Only*	*MEILI Only*
Duration(mins)	*25±23*	*26±26*	*173±429[2] - median 12*
Length (km)	11.8±18	27±127 (est) 20±7 (int)	13±46
Trip legs	*1.9±1.3*	*1.6±1*	*1.4±0.8*
Temporal quality	*62%±31%*	*47%±30%*	*48%±37%*
Spatial quality	*71%±33%*	*69%±34%*	*71%±35%*

The algorithms used by MEILI detect stops with a precision of 96.7% and a recall of 73.8%, and trip legs with a precision of 70.2% and a recall of 53.8%. Combining these two measures, the algorithm detects correct trips with the correct number of trip legs with a precision of 78.5% and a recall of 58.1%. These are classical error measures that are not very insightful regarding how the missed trips and trip legs affect the collection as a whole, but exploring such measures is part of future research.

When it comes to detecting the travel mode, the nearest-neighbour classifier mentioned in Section 4.3.2 detects the correct mode with 53.5% accuracy (Figure 13). This result can be improved although it is important to mention that the baseline is 35.8% and that the number of considered travel modes is 15, which makes this classification task difficult. However, a closer investigation of the top-3 classifications reveals that for 82.1% of the cases, the correct mode is in the top 3 predictions made by the classifier, which suggests that a hierarchical classifier that takes into account more complex features than the proposed one can be built for the top-3 ordering.

[2] The very long trip duration might be a consequence of the lack of noise filtering prior to data analysis. It might be an indicator of a user having difficulty adjusting the start / end time of a trip.

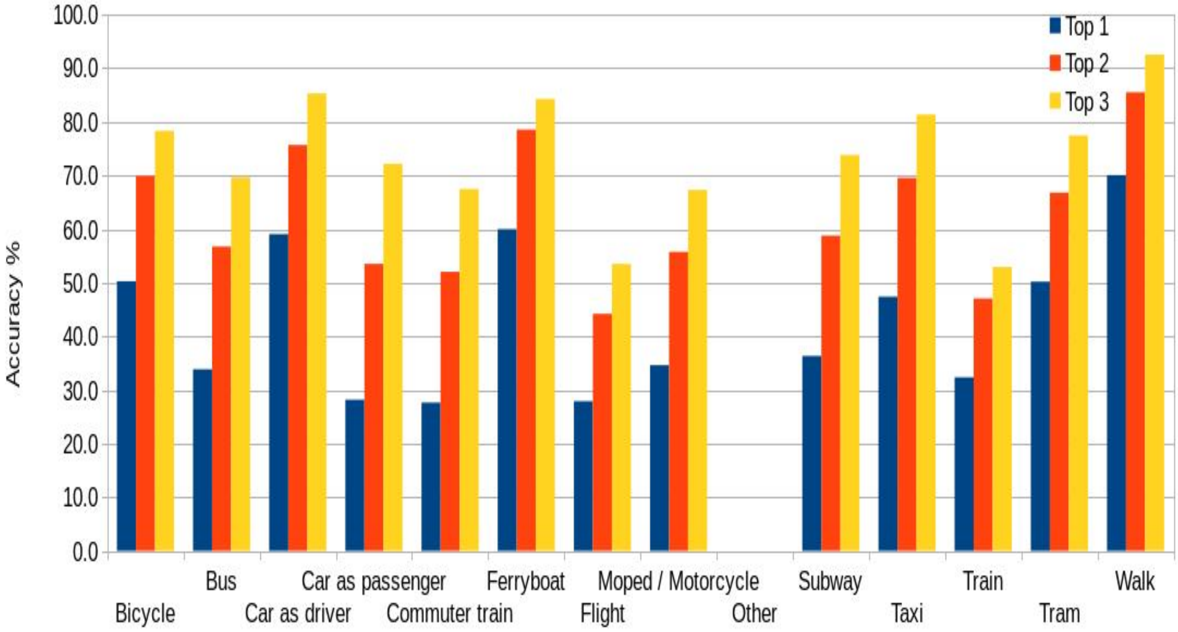


Figure 13: Travel mode accuracy

The destination inference is problematic given the large dataset and the number of users that only annotated few trips that do not share destinations. As opposed to travel mode, the baseline for destination inference is negligible due to the high number of POIs in the dataset, which is non-exhaustive because the home and work POIs of users have to be, in most cases, declared by the users. A top-3 classifier reaches an accuracy of 47 % (see Figure 14), which is insufficient for the considered task and has to be improved.

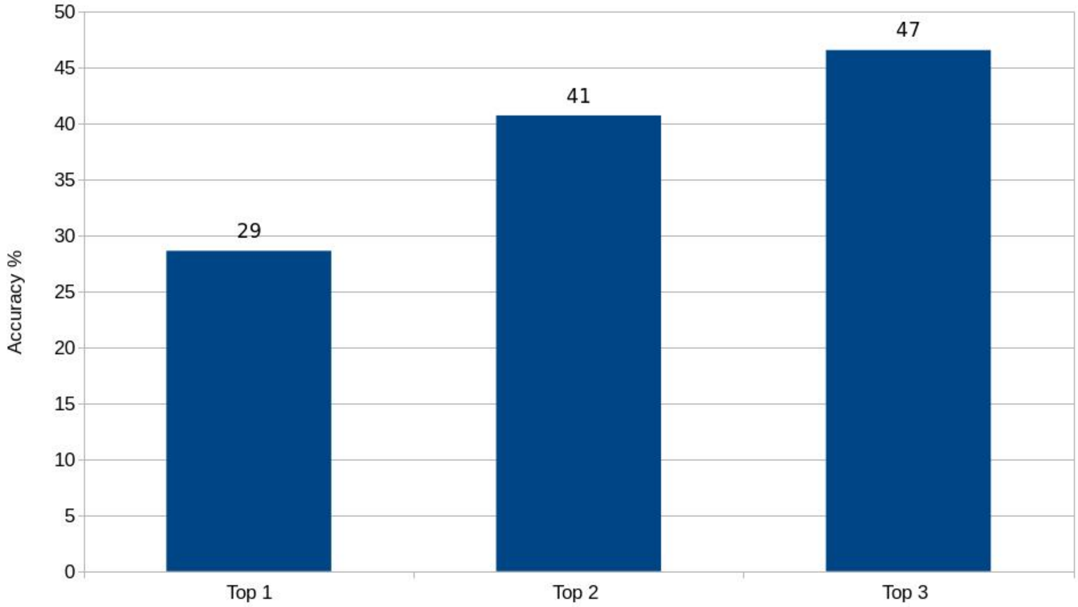


Figure 14: Destination inference accuracy for top-1, top-2 and top-3

The purpose inference is dependent on the destination inference, which affects the performance of the inference. The baseline for purpose detection is 27.6% and the obtained overall accuracy is 38%

for top-1 and 49% for top 2 (see Figure 15). When performing the studies while considering a correct destination, the purpose accuracy was over 80%, which suggests that the most critical part is the improving the destination inference.

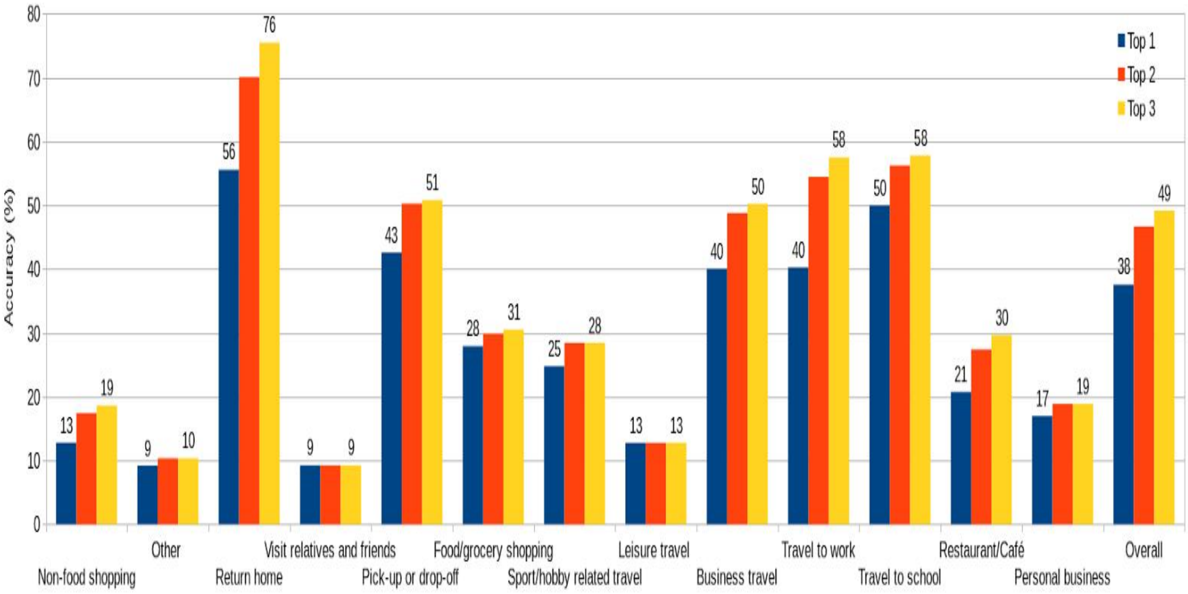


Figure 15: Trip purpose inference accuracy

Similar to the pilot travel mode dissemination, the large field trial also generated data relevant for trip leg and travel mode analysis. Table 8 shows that there is a large variability in the length and duration attribute of each mode, i.e., high standard deviation, which can be explained by people using modes different from one another (e.g., some people might walk longer distances than others) and by the difficulty some users might have encountered during the annotation process. However, the computed median values per mode are closer to what one might expect, which suggests that an outlier detection and removal analysis would be beneficial for this dataset. The spatial and temporal indicator values are as expected. There are low spatial indicator values for modes where one cannot collect data such as subways and flights and high values for the others, albeit the standard deviation of each mode related spatial indicator values is still large. The low temporal indicator values are also accompanied by a high standard deviation, which can be explained by similar reasons as for why the length and duration attributes have a high standard deviation. While the waiting time can be computed with the large field trial data, the lack of time and changes in the MEILI DSC did not allow for this to be computed on the large dataset.

Table 8: Trip leg data collected with MEILI. The large standard deviation values indicate that the travelers make use of the same travel mode but for different distances, and / or that the users have not paid attention to the mode specification or did not interact with the graphical user interface in the intended way.

	Duration (min)		*Length (km)*		*Spatial quality (%)*		*Temporal quality (%)*		*Number of trip legs*
	Avg ± Sd	*Median*	*Avg±Sd*	*Median*	*Avg±Sd*	*Median*	*Avg±Sd*	*Median*	
Bicycle	194±641	20	3.1±8	0.5	82±27	99	51±39	83	346
Bus	64±229	13	6.9±18	2.4	73±34	92	59±36	81	512
Car as driver	77±246	14	10.1±17.2	4.1	77±31	93	62±35	81	986
Car passenger	103±401	12	7.9±25	0.5	82±30	100	51±40	83	284
Commuter train	76±191	20.5	9.9±14.1	5.0	56±41	79	32±31	51	246
Ferryboat	42±80	17.5	2.5±17.1	0.1	84±32	100	8±15	17	70
Flight	94±36	80	601.4±385	444	1±2	0	30±46	21	5
Moped/ MC	64±198	18.5	6±8.7	0.4	84±27	100	49±38	89	52
Subway	53±180	19.0	9.8±69.2	4.1	37±41	55	23±28	32	431
Taxi	147±249	13.0	1.1±4.7	0.2	82±31	100	20±29	39	481
Train	526±940	427	16.7±83.8	0.2	78±36	100	18±28	55	137
Tram	58±145	12.0	1.7±12.4	0.1	80±34	100	24±32	96	253
Walk	54±216	5.0	1.5±22.7	0.3	81±29	100	50±35	67	2150
Other	1277±2842	235	70.3±178.9	0.4	75±35	91	9±8	7	8
Total					75±13	86	35±19	57	5961

8 CONCLUSIONS AND FUTURE WORK

The aim of this project was to develop and evaluate a prototype system for replacing or complementing the traditional methods for travel diary collection. The project shows that such a system is plausible to implement given the existing technologies, and that some of the most noticeable concerns regarding such a system have been addressed: battery consumption and user integrity concerns. The algorithm developed to limit battery consumption had a significant impact, which resulted in the majority of the users not experiencing battery consumption as a problem. Furthermore, the users rarely turn off the smartphone application due to integrity issues and the majority of the users found the traditional travel diary just as intrusive as the app.

The recruitment of users/respondents is a critical phase for traditional travel diaries and, as expected, this was the case also for the smartphone based method. A lesson learned was that it is important to simplify the registration process as much as possible. The difference between entering an e-mail address at the end of a web-based travel diary, and sending an e-mail to an address printed in a letter might seem small but it had a determinant effect on the participants. It is also critical that the system works as expected and, in case of failure, the communication with the users is not delayed, otherwise the drop-out rate increase rapidly.

Something that has only been briefly mentioned in this report is bias among the users. Depending on what the data will be used for this can be more or less troublesome. To avoid a bias towards people with technical knowledge, in general younger people, it is important that the system used is user friendly and without problems. In the conducted field trial, the webpage for annotation caused most problems for the users. While improvements to the annotation web application are planned, it is worth mentioning that if the user annotates a sufficient portion of the collected data, the need for annotations is minimal.

In the large field trial 2142 annotated trips from 171 users (1250 trips from 51 users for a duration longer than a week) were collected, hence it can be seen as one the most successful smartphone-based trials in the world. In total, about 1 million GPS readings were collected.

Based on strict temporal co-occurrence and trip purposes matching rules, there were 189 trips of 355 reported trips (from 85 users who participated in both methods) matched. This highlights the capability of MEILI in matching the results reported by the traditional methods. Most common reasons of mismatch were: (1) missed movements/forgot to bring the smartphone, (2) reported with a different purpose (which is expected due to the multi-tasking nature of modern travellers' activities), and (3) miss-reported/undetected trip chaining. For example, the traditional travel diary method tends to miss "return home" trips, whilst MEILI tends to miss the "pick-up and/or drop-off" trips. Nevertheless, MEILI captured more detailed information about trip legs and the location attributes derived from the route (destinations) agrees with the user annotated data more than with the address specified in the traditional travel diary.

Using spatial-temporal quality measurements (for which details can be seen in Prelipcean et al., 2015a), it can be seen that, on average, MEILI has a lower error in time, but a higher error in length when making mistakes than the traditional method. Presumably because in traditional methods one consciously registered his/her home and work address, while in MEILI this information was inferred via methods that were described in Section 4. The accuracy of the current inference methods was considered as relatively high, as we can infer the stop locations with 96.7% accuracy, the trip-leg construct with 70.2% accuracy, and the complete trip chain about 78.5% accuracy. Inferring the trip purpose, however, only has about 47% accuracy since it relies on the user's history, the proximity to POIs, and on the hypothesis that a user visits a place for a singular purpose, which is problematic since for example one may visit a shopping centre not only to shop, but also to pick-up/drop-off somebody else, or without a specific purpose.

Nevertheless, the data collected from MEILI demonstrated that Stockholm travellers have unique travel patterns which vary from day-to-day. For example, it can be seen from the datasets that shopping trips were mostly done on Saturday, whilst family and social trips were carried out on Sunday. The sport and hobby related trips were mostly done on Wednesday and Saturday. This variability of activity participation indicates the importance of not only base analysis of travel demand on one-day cross-sectional data, but on multi-day data.

As mentioned earlier, the previous studies show that a GPS-based data collection has a great potential to solve the problems related to the estimation of distance/travel time, geographic coding of departure/destination locations, and forgotten trips. It will also provide a more detailed dataset, something that is essential for the next generation of transport models, such as activity-based analysis models. All of these also has been found and confirmed in this SPOT project. It is very clear that MEILI offers a real potential to collect data which can enable the development of more realistic decision support tools.

At the same time, however, it was also learned that deploying a smartphone application as a survey tool required more support than expected, which is similar to launching a product, which should involve a multi-disciplinary team, not only to design the questionnaire and support system, but also in making the user interface more user friendly. Furthermore, when deploying a product, dedicated teams that track the developments of different platforms (e.g. Android, iOS) are needed, since the dynamics of the mobile operating systems might require the source code to be adapted to new libraries.

It is confirmed that collecting individual travel diaries with MEILI is plausible and, with the progress of the technology, its reliability and user acceptance would increase. As for now, solutions such as MEILI, do offer a real alternative way to collect travel diaries that complement of the more traditional methods, since they capture additional trips with great level of detail and since they appeal to different group of respondents than the traditional methods.

There are numerous directions one can follow given the developed system and the collected dataset. With the more complete and comprehensive dataset that has been collected in the SPOT project, it is still unknown how much more understanding and detailed information of the traveller's behaviour and choices one-week smartphone travel diary data can really reveal than what we already know from standard one-day travel diaries. To investigate this, one of the possible future directions of the research would be using the data that have been collected by the MEILI travel diary collection system in the main field trial of the SPOT project to investigate the three sub-topics below:

1. The stability and variability of individuals' day-to-day choices. This includes analysing the tendency and degree of variability and stability of their chosen travel modes, activity locations, individual's time allocations and trip chaining behaviours, across different groups of travellers.

2. The comparability of standard transport model indicators and assumptions with GPS based observations. This includes comparing the individual's revealed multimodal and route choices with hypothetical shortest path and multimodal assumptions which are commonly used in urban transport modelling.

3. Exploring space-time accessibility factors that shape and influence individual travel decisions. The individual activity-travel pattern that was collected from the survey is a result from trade-off between individual's travel needs and constraints. These constraints include the physical space and time accessibility (e.g. travel time to reach shopping location at street A instead of at street B, no shopping

trips before opening time X, the requirement to be at work location C before time Z, and pick up kids at school D before time Y), which some are changes on daily basis, whilst others are not. Understanding how these factors shape and influence individual travel decision would give us an insight about how to influence/nudge travellers to adopt a (more sustainable) behaviour in the way that is most convenient for them.

These investigations would complement and support Trafikverket's wider agenda to provide a more realistic, accurate and inclusive transport model. The project relates to Trafikverket's development plan for socio-economic methods and tools, especially to the subtopics 3.3.1.2 Nya RES och VFU med nya metoder, 3.3.1.3 Valideringsdata person- och godstrafik and 3.7.2.5 Aktivitetsbaserade efterfrågemodeller. The results of this project will not only reveal the potential use of a cheaper, more efficient and richer GPS-based data collection, but will also provide us with a deeper understanding on the stability and the variability of Stockholm travellers' travel behaviour and also the accuracy of assumptions in the current SAMPERS model.

Thinking about the future of modelling needs, in many respects, activity based modelling has been considered as a promising alternative of the advancement of the current modelling practice. It considers travel as a derived demand; if we can influence the activity locations/engagements of the given agent, we will be able to influence their travel patterns. The given agent would have freedom to decide every single of their activity-travel engagements and decisions at very detailed time unit/interval. This includes (the variability of) departure times, combination of mode choices, trip chaining, re-scheduling activities, etc. The disadvantage of this model, however, is the detailed level of dataset needed which RVU cannot fully provide. In this sense, data that is produced by MEILI would give much more advantages.

In transport model and policy design in general, MEILI would provide us with various detailed performance information which we did not have, such as (1) real time performance of particular road segments and public transport fleets/routes on the given time, (2) waiting time, traveller behaviour and their time spent in interchanges points, (3) missing last mile information (detailed walking time/length to from public transport and from/to other short distance locations), and (4) multiday data without significant extra costs. These data would enable us to understand various different issues that cannot be answered with a traditional travel survey, such as how individual activity travel patterns that we have now vary across different days and time of day, and how accurate/inaccurate our analysis on (the value of) travel time variability and uncertainty, etc.

In activity based models in particular, on top of the advantages mentioned above, MEILI would provide us with: (1) more complete information on trip and mode chaining patterns, including their (2) detailed route choice data, and (3) detailed activity locations. This would enable us to develop a better scheduling algorithm over available activity and travel opportunities, better destination choice model, and better trip chaining and route choice models in the given city/area. Furthermore, we will also be able to separate between trip purpose and activity location (e.g. whether one's travel to the mall for shopping or for leisure trips (window shopping) or for social family event etc., which in a traditional travel survey not is recognised separately. Further, by using MEILI data, we can simulate individual time-space prisms which will help us to define the travel and activity opportunities, in the given time-and-space configuration, in much a more realistic and accurate (and justified) way than the current approach.

9 REFERENCES

T. Abdulazim, H. Abdelgawad, K.N. Habib, B. Abdulhai. (2013). Using smartphones and sensor technologies to automate the collection of travel data. Transportation Research Board Annual Meeting, Washington D.C., U.S.

A. Ali, S. Lui. (2011). Household Trip Patterns and Travel Characteristics in Lethbridge, Alberta, in Annual Conference of the Transportation Association of Canada. Edmonton, Alberta, Canada.

A. Allström, M. Fransson, D. Gundlegård, E. Jenelius, D. Gundlegård. (2013). Nya indatakällor för trafikprognoser, Trafikverket

T. Anderson, V. Abeywardana, J. Wolf, M. Lee. (2009). National Travel Survey - GPS Feasibility Study: Final Report. National Centre for Social Research, United Kingdom.

K. W. Axhausen, S. Schönfelder, J. Wolf, M. Oliveira and U. Samaga. (2002). 80 weeks of GPS-traces: Approaches to enriching the trip information, Arbeitsbericht Verkehrs- und Raumplanung, number 178.

P. Bachu, R. Dudala, S. Kothuri. (2001). Prompted Recall in Global Positioning Survey: Proof of Concept Study, Transportation Research Record, 1768, 106-113.

W. Bohte and K. Maat. (2009). Deriving and validating trip purposes and travel modes for multi-day GPS-based travel surveys: A large-scale application in the Netherlands, Transportation Research Part C: Emerging Technologies, number 3, volume 17, pp. 285–297.

P.-L. Bourbonnais, C. Morency. (2013). Web-based travel surveys, in Zmud, J., Lee-Gosselin, M., Munizaga, M., Carrasco, J.-A. (Eds.), Transport survey methods: best practice for decision making, 207–223. Emerald, Bingley, UK.

S. Bricka and C. R. Bhat. (2006). "Comparative Analysis of Global Positioning System Based and Travel Survey Based Data", Journal of the Transportation Research Board, number 1972, pp. 9–20.

E. H. Chung, A. Shalaby. (2005). A trip reconstruction tool for GPS-based personal travel surveys. Transportation Planning and Technology, 28(5), 381-401.

C. D. Cottrill, F.C. Pereira, F. Zhao, I.F. Dias, H.B. Lim, M. E. Ben-Akiva. (2013). The Future Mobility Survey: Experiences in developing a smartphone based travel survey in Singapore, in Annual Meeting of the Transportation Research Board, Washington, D.C., USA.

A. B. Ellison, R. Ellison, D. Rance, Greaves, S.P., Standen, C. (2014). Harnessing smartphone sensors for tracking location to support travel data collection, 10th International Conference on Transport Survey Methods, Leura, Australia.

T. Feng, H.J.P. Timmermans. (2013). Transportation mode recognition using GPS and accelerometer data. Transportation Research Part C, 37, 118-130.

T. L. Forrest and D. F. Pearson. (2005). Comparison of Trip Determination Methods in Household Travel Surveys Enhanced by a Global Positioning System, Journal of the Transportation Research Board, number 1917, pp. 63–71.

H. Gong, Chen, C. (2012). Automating Web Collection and Validation of GPS Data for Longitudinal Urban Travel Studies: Final Report. US DOT - Research and Innovative Technology Administration.

S. P. Greaves, A. B. Ellison, R. B. Ellison, D. Rance, C. Standen, C. Rissel, M. and Crane. (2014). A Web-Based Diary and Companion Smartphone app for Travel/Activity Surveys, International Conference on Transport Survey Methods, Leura, Australia.

S. Hemminki, P. Nurmi, S. Tarkoma. (2013). Accelerometer-based transportation mode detection on smartphones. In Proceedings of the 11th ACM Conference on Embedded Networked Sensor Systems, ACMm p. 13.

N. E. Klepeis, A.M. Tsang, J.V. Behar. (1996). Analysis of the National Human Activity Pattern Survey (NHAPS) respondents from a standpoint of exposure assessment. US EPA, Las Vegas, NV.

L. Montini, N. Rieser-Schüssler and K. W. Axhausen. (2013). Field Report: One-Week GPS-based Travel Survey in the Greater Zurich Area, 13th Swiss Transport Research Conference (STRC).

L. Montini, N. Rieser-Schüssler, A. Horni, K. W. Axhausen. (2014). Trip Purpose Identification from GPS Tracks. Transportation Research Record: Journal of the Transportation Research Board, (2405), 16-23.

NCHRP. (2014). Applying GPS data to understand travel behaviour: Background, Methods and Tests, Report 775, Transportation Research Board.

N. Ohmori, M. Nakazato, N. Harata. (2005). GPS Mobile Phone-Based Activity Diary Survey. Proceedings of the Eastern Asia Society for Transportation Studies, 5, 1104-1115.

M. S. Oliveira, P. Vovsha, J. Wolf, M. Mitchell. (2014). Evaluation of Two Methods for Identifying Trip Purpose in GPS-Based Household Travel Surveys, Journal of the Transportation Research Board, number 2405, pp. 33–41.

A. C. Prelipcean, G. Gidófalvi, Y.O. Susilo. (2014). Mobility Collector. Journal of Location Based Services, 8(4).

A. C. Prelipcean, G. Gidofalvi, Y.O Susilo. (2015a). Comparative framework for activity-travel diary collection systems. In 2015 International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS), IEEE, pp. 251-258.

A. C. Prelipcean, G. Gidófalvi, Y.O. Susilo. (2015b). Evaluation of novel point- and segment-based transport mode segmentation methods for trajectories. Forthcoming at International Journal of Geographical Information Science.

S. Rasouli, H.J.P. Timmermans. (2014). Mobile technologies for activity-travel data collection and analysis. New York: IGI.

S. Reddy, M. Mun, J. Burke, D. Estrin, M. Hansen, M. Srivastava. (2010). Using Mobile Phones to Determine Transportation Modes", ACM Transactions on Sensor Networks, number 2, volume 6, pp. 13:1–13:27.

L. Stenneth, O. Wolfson, P. S. Yu and B. Xu. (2011). Transportation mode detection using mobile phones and GIS information, Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, pp. 54–63.

P. Stopher, C. FitzGerald, J. Zhang. (2008). Search for a global positioning system device to measure person travel. Transportation Research Part C: Emerging Technologies, 16(3), 350-369.

P. Stopher and S. Li. (2011). In-depth Comparison of Global Positioning System and Diary Records, Journal of the Transportation Research Board, number 2046, pp. 32–37.

Trafikverket. (2012). Utveckling av samhällsekonomiska metoder och verktyg, effektsamband och modeller inom transportområdet (available online at http://www.trafikverket.se/contentassets/12f70193edcf4c5abe11abd073ae4a36/trafikverket_plan_utveckling_metoder_effektsamband_modeller.pdf).

VTI. (2002). Enkäten och intervjun som metod för att samla in information om trafikanters exponering i trafikmiljön. Notat 14-2002, Statens väg- och transportforskningsinstitut. Trafa. 2015. RVU Sverige 2011–2014: Den nationella resvaneundersökningen Statistik, 2015:10.

J. Wolf, R. Guensler and W. Bachman. (2001). Elimination of the Travel Diary: An Experiment to Derive Trip Purpose From GPS Travel Data, Journal of the Transportation Research Board, number 1768, pp. 125–134.

J. Wolf, M. Oliveira and M. Thompson. (2003). The impact of trip underreporting on VMT and travel time estimates: preliminary findings from the California statewide household travel survey GPS study, Journal of the Transportation Research Board, number 1854, pp. 189–198.

M. Yu, C. Lin, T. Yu, E. Y. Chang and S. Wang. (2014). Big data small foot-print: The design of a low-power classifier for detecting transportation modes, The Proceedings of the VLDB Endowment, volume 7, pp. 1429–1440.