EXPLORING INTELLIGENT MOBILITY

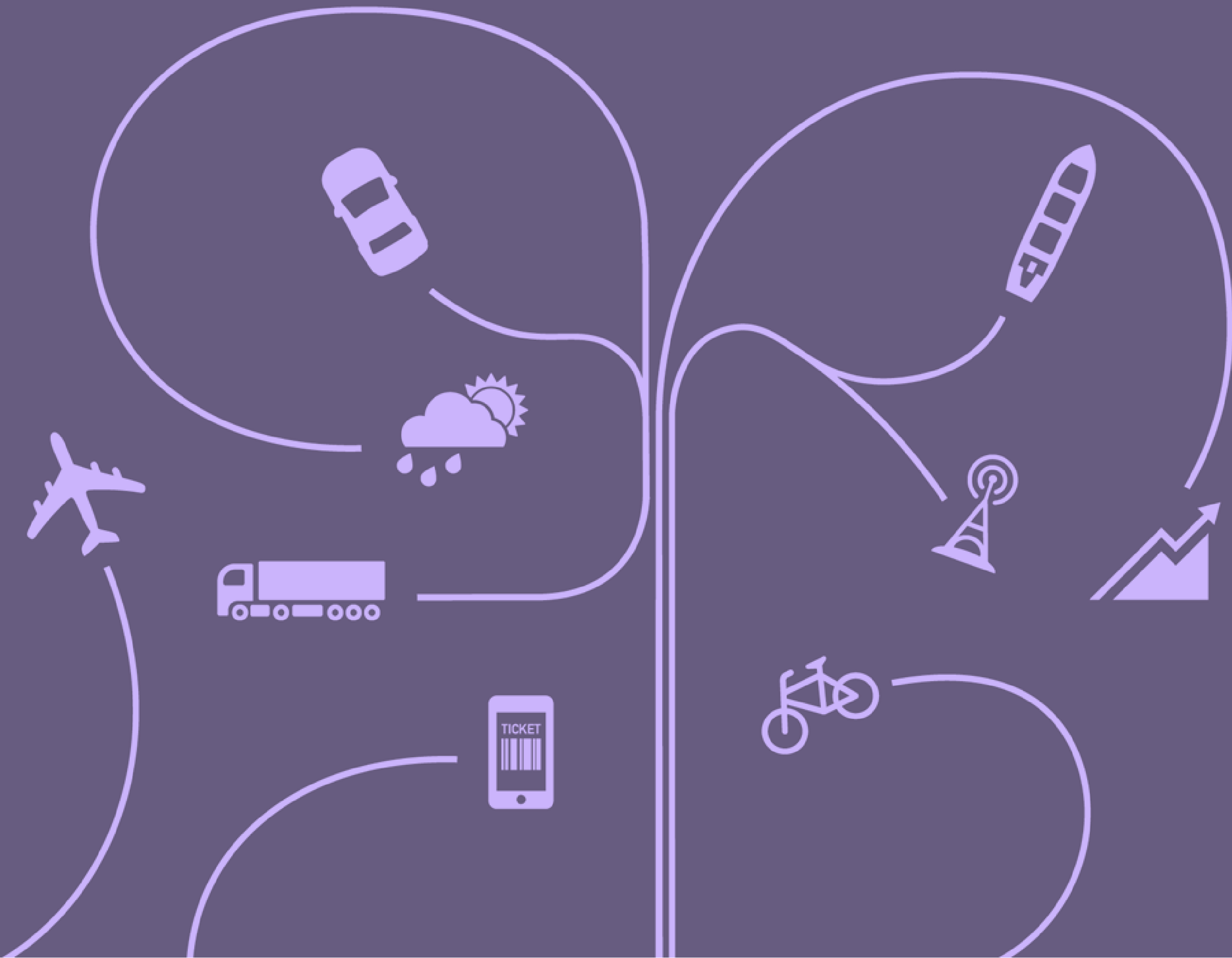
CATAPULT Transport Systems

Modernising the National Travel Survey

WP1 - Review of Current and Future Technologies and Data Sources Report

September 2016

Updated in February 2017



In this document "£" refers to United Kingdom pound sterling unless otherwise specified.

Executive Summary

This paper describes the work and overall research that was undertaken for the Department for Transport's (DfT) research project 'Modernising the National Travel Survey'. The aim of this report is to provide an in-depth review of existing and new technologies and data sources that could be used to improve the National Travel Survey (NTS). The technologies examined were: smartphones, dedicated GPS devices, mobile network data, web-based travel diary surveys, social media data with an emphasis on the Twitter platform, and travel smartcards.

This report includes, for each technology:

- A rigorous review of each technology expanding on specific characteristics, the way it works and the prevalence in the market, in terms of ownership, usage etc.
- Examination on the potential of a technology to improve the quality of the NTS methods and the data that can be collected.
- Identification of the potential costs and limitations of introducing new technologies to the NTS. Research was undertaken on the costs that affect the survey directly and indirectly.
- Case studies where the different technologies had been used, focusing on travel or activity surveys.

After a thorough technology review and engagement with researchers and practitioners, the following points were concluded:

- Smartphones have a variety of sensors that can passively collect large volumes of data which in conjunction with a prompted recall approach can potentially replace the current NTS paper travel diary. The risk of data loss can be managed by the ability to upload data through mobile networks and Wi-Fi, and the potential utilization of participants' smartphones could reduce both the costs and the respondent burden of participating in the survey. A big limitation that can affect smartphone based surveys is the battery life of the devices. Although smartphones can add value in travel surveys, there is still a part of the population (mostly elderly people) who don't own a device. Therefore, it was concluded that it may take up to 5 years before the bias in the age profile of smartphone owners has been sufficiently removed for it to be a viable option.

- Dedicated GPS devices can collect travel activity data passively and in combination with a prompted recall survey could potentially replace the current NTS method. However, they do require participants to carry a device with them at all times, which could add to the respondent burden and possibly affect the data collection process if the participants forget to carry the device with them, forget to keep it charged etc. In ideal conditions data quality can be good, but there are situations where signal problems can skew the quality of the data.
- Mobile network data is becoming increasingly popular in the field of transport modelling, in the context of building origin / destination matrices and understanding mobility patterns. Although data needs to be acquired from network operators and their price range varies, they are potentially a good source of mobility data: using them reduces cost and effort; there is no need to recruit people and conduct expensive and time-consuming surveys. However, due to privacy reasons, this data cannot be associated with individual users so personal characteristics of those making the journeys will not be included in the data provided by mobile network operators.
- Web-based travel diary surveys have great potential for the collection of travel activity data. The developments in web-based technology suggest several features that can reduce the respondent burden and ease the survey management, something that is not very possible with the paper version of the survey. In addition, they utilise computers and smartphones which nowadays many UK households have, along with an internet connection. Although there are still some households that do not have internet access, in the future this number is expected to decrease.
- Twitter geolocated data is an alternative cost-effective data source that can indirectly provide travel activity information. This data can be used to create origin-destination matrices and if further processing is performed on the text message (tweet) additional information could be uncovered. Although this data can be acquired relatively easily, Twitter provides APIs and tools to do that, they can only be used to support the NTS and not replace it.
- Smartcard data provide direct travel activity data but may present difficulties in accessing them: most of the operational smartcard systems are implemented and maintained by private transportation companies and local councils. They can only be used as complementary to travel surveys because they are restricted to public transport journeys and geographic area since only larger cities such as London, Manchester and Birmingham have smartcard schemes.

In summary, the findings suggest: smartphones will be a sensible option for conducting travel surveys in around 5 years' time, dedicated GPS devices can collect travel activity data but the participant's interaction is still required, and other data sources such as mobile networks and Twitter feeds may provide interesting information but only to support not replace the current NTS. The recommendations that emerged from the technology review and the desk based research are as follows:

- For the use of smartphones or dedicated GPS devices, a prompted recall approach is advised to be followed: the participant's interaction is still needed for the validation of the data.
- A mixed mode data collection approach (web, computers, smartphone etc.) can cover a diversified and representative sample of the population such as the elderly, people without an internet connection or the necessary equipment.
- If a technology mentioned in the report is used, then a helpline or help desk should be considered.
- The mentioned data sources can help validate assumptions and findings from the survey:
    - Mobile network data and Twitter data can validate findings such as travel activity in specific regions, frequency of travelling and other travel activity characteristics.
    - Smartcard data can add value to the NTS by providing elements needed for the survey but only restricted to the public transport part of the survey. Barriers that should be considered are: the architecture of the system, the ease of access of the data and the restriction to certain cities.

Contents

1 Introduction

1.1 Background

The National Travel Survey (NTS) is a survey designed to provide information about personal travel and to monitor long-term trends in travel behaviour. The NTS is the primary source of data on personal travel patterns by residents of England within Great Britain. The survey collects information on how, why, when and where people travel as well as the factors affecting travel (e.g. car availability and driving licence holding).

The first NTS was conducted in 1965/1966 by the Ministry of Transport, and since then it was repeated in an ad-hoc fashion until 1988 when it became a continuous survey. Since 2002 the Department for Transport (DfT) has commissioned NatCen Social Research (NatCen), an independent social research institute, as the contractor with sole responsibility for the implementation of the NTS. NatCen is responsible for questionnaire development, sample selection, data collection and editing, data file production and building the database. DfT is responsible for data analysis, publication and archiving. DfT uses the data to identify changes in travel patterns and inform the development of policies.

The 2015 NTS used a randomly selected sample of 7,005 private households in England and used two data collection methods. Face to face interviews using computer assisted personal interviewing (CAPI) were conducted with all the members of chosen households. Information was gathered about a variety of topics, such as demographics, driving licence holders, vehicle ownership and long distance journeys that have recently been made. The second data collection method used was a self-completion travel diary, in the form of a notebook, over a seven-day period, which was provided to all household members at the end of the first interview.

1.2 Scope & Objectives

The current paper based method of the travel diary requires each survey participant to describe or recreate their travel behaviour for a specified week. Because the diary can be completed at any point in the week, this can lead to trips being under-reported. Furthermore, respondents are naturally inclined to roundup trip times and distances, leading to potential data quality issues.

The following report is part of the 'Modernising the National Travel Survey' research project led by Transport Systems Catapult (TSC) for DfT. The purpose of this research is to identify and investigate new and emerging technologies suitable to accommodate and potentially replace the current methods used in the NTS. This research aims to identify the potential some technologies have, their ability to be used for surveying purposes and any technical issues or limitations related to their usage.

The technologies that are examined in this report are those with a focus on location-based services or the ability to capture a location. This report considers the use of GPS based devices as technologies with location based capabilities. Mobile network data and smartcard data are widely known data sources that can provide location data, among other things. In the context of future proofing the NTS, social media, with an emphasis on Twitter data, were investigated as a potential data source that can add value to the current methods of the survey.

The goal of the report is to provide a set of recommendations and insights on the ability of a technology to completely replace the paper based travel diary in the short-term to medium-term and its ability to complement the current methods of the survey. This will give a good indication to DfT if certain technologies are worth pursuing further.

The choice of technologies will be assessed based on the quality and type of data they can collect, the burden they introduce to the participants of the survey, in terms of inconvenience when using the technology, completing the survey etc., and any extra cost on the survey organisers and respondents.

1.3 Research Methodology

The activities that took place during this project were as follows:

- **Inventory development.** Generating an inventory of advanced technologies and data sources suitable for the collection of travel data. Research was conducted mainly using the Internet as a primary research tool. Technologies with similar characteristics have been grouped together (Technology Groups 1-4).
- **Market research.** Research was undertaken on the market penetration of a technology with a focus on ownership statistics, usage and customer behaviour analysis.
- **Literature review.** An extensive literature review was conducted of experiments and studies from universities and organizations on the collection of travel data and the use of travel surveys.
- **Researchers engagement.** Discussions were held with researchers and practitioners who have tested the use of a new technology in travel surveys and travel activity collection frameworks.

- **Stakeholder engagement.** Household Travel Survey reports from around the world were reviewed and discussions with NatCen[1] staff took place, to formulate a good understanding of the current methodologies and the problems they face in data collection and processing.

1.4 Report Structure

The overall structure of the report is as follows:

- Assessment on the ability of the technology, based on a literature review, to improve the data collection methods of the current NTS. Also, an analysis was conducted on the data that can be collected from this technology.
- Analysis on the costs of using this technology and the limitations that come with it
- Overview of the literature findings and practical issues on the use of this technology in experiments and travel surveys around the world.
- Conclusions and recommendations on technology and the ways it could be implemented in the NTS.

[1] http://natcen.ac.uk/about-us/

2 Technology Group 1: GPS based devices and data sources

2.1 Smartphones

A smartphone is a mobile phone with an advanced operating system and hardware. It provides features that a normal mobile phone provides such as texting and making phone calls, with the addition of gaming, Web browsing, email etc. Smartphones have become the most ubiquitous and prevalent technology of the past few years. They came in to existence around 1994, the very first smartphone was the IBM Simon, but the release of the first iPhone in 2007 propelled smartphones into the mainstream. Since then, the market has seen a rapid increase in smartphone ownership (Figure 1, Figure 3) and consumer behaviour has changed. Smartphone ownership is increasing by age (Figure 3). More specifically, the rise is quicker for the age groups 16-24, although there is a steady yearly rise in older age groups. Evidently, in 2016, 42% of the 55+ group own a smartphone, providing evidence towards the fact that in the future more people aged 65+ would own a smartphone. Additionally, people are spending increasing amounts of time interacting with their phones, either online or offline (Figure 4).

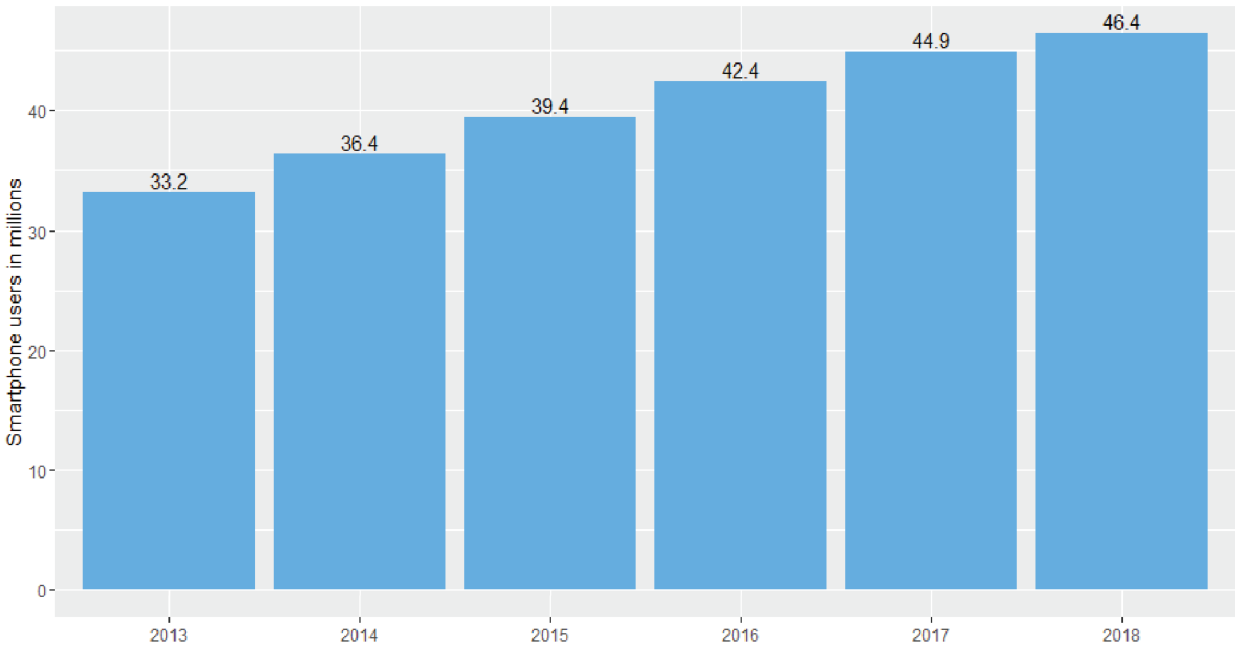


Figure 1: Number of smartphone users in the UK from 2013 with a forecast to 2018 [2]

[2] http://www.emarketer.com/Article/2-Billion-Consumers-Worldwide-Smartphones-by-2016/1011694

Currently, there are many smartphone manufacturers competing in the market. The latest research from Kantar WorldPanel in April 2016 showed Android, Google's operating system, had 58.5% of the market share whereas Apple's iOS had 35.1%[3]. There are other operating systems as well but they make up a very small percentage of the market: Microsoft's Windows Phone has around 5.8% and BlackBerry has 0.5%. One of the reasons for this big range in market share is the use of Android by a wider range of manufacturers in their devices. The most popular manufacturers in the UK are Samsung, who has the lead in the market, followed by HTC, LG and Sony with relatively equal percentages between them: In 2015, Samsung had 21.4% of the market, HTC and LG 9.4% and Sony 7.2%[4].

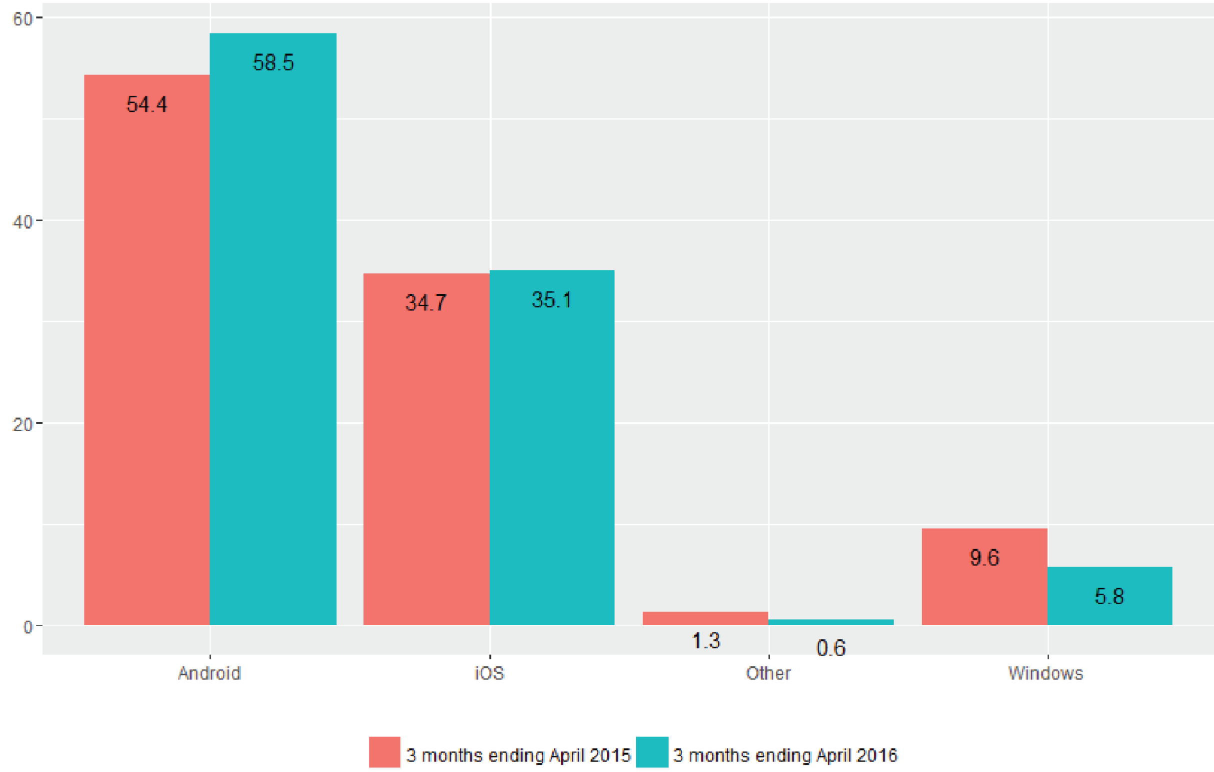


Figure 2: Great Britain Smartphone OS Sale Share (%)[5]

The increased proliferation of smartphones within the population and in combination with the increased capabilities offered, made them essential to everyday life. Consumers are using their smartphones for work, shopping, during travel, for entertainment and many other aspects of everyday activities. According to research from the Traveller Needs Survey, 54% of the respondents consider smartphones essential for their travelling

[3] http://www.kantarworldpanel.com/global/smartphone-os-market-share/
[4] http://www.statista.com/statistics/387227/market-share-of-smartphone manufacturers-in-the-uk/
[5] http://uk.kantar.com/tech/mobile/2016/june-great-britain-smartphone-data/

experience and 57% would consider sharing their data if it has the potential to optimize their commute[6]. Nowadays, people understand that transport as an industry can be improved if they actively contribute instead of just purchasing or consuming mobility services and products.

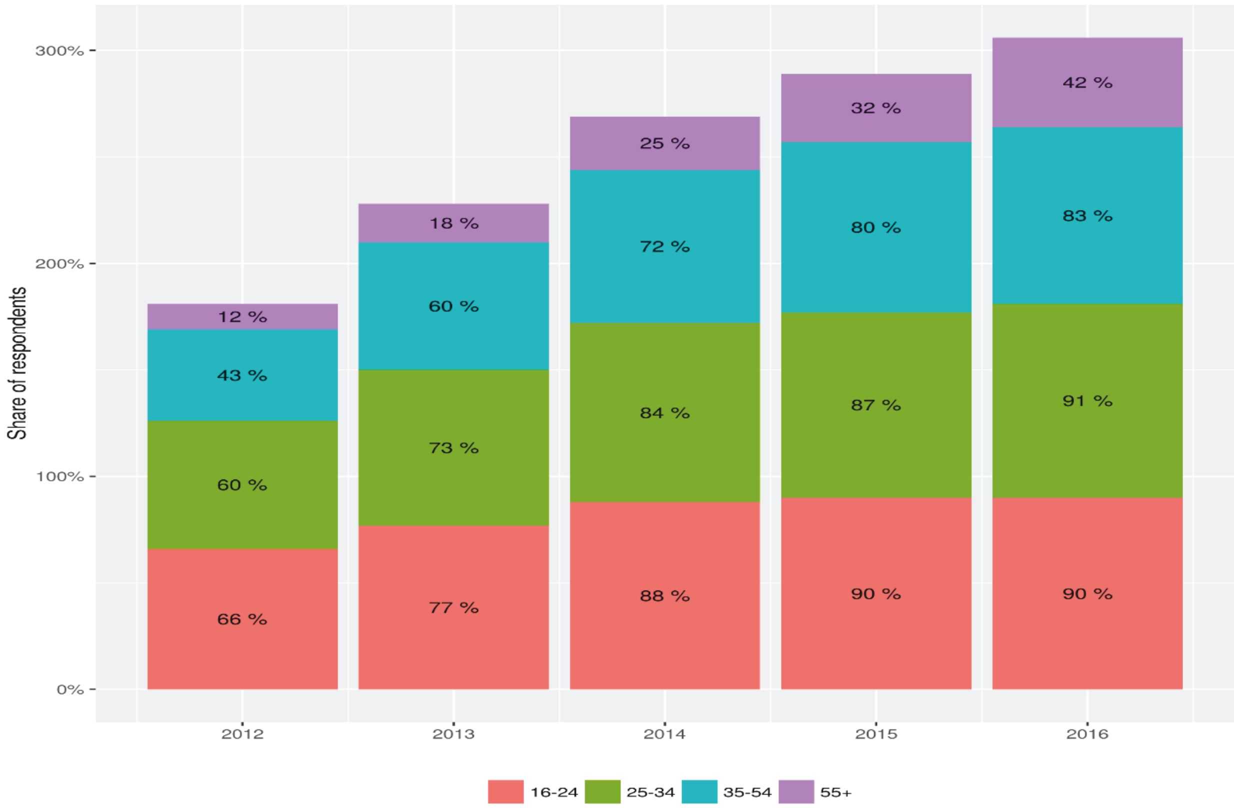


Figure 3: Smartphone ownership penetration in the United Kingdom (UK) in 2012-2016, by age [7]

This makes smartphones potentially suitable devices to conduct travel activity surveys to understand the behaviour of transportation systems users. There is a growing interest in the use of smartphones because of the variety of options they provide for participant interaction and data collection by tracking participants. More specifically, smartphones can be used to collect travel activity data via the variety of sensors they provide such as: GPS, Wi-Fi, and mobile network location. In addition, these devices have more processing power to process the data coming from the sensors and derive patterns and information from the collected data.

[6] https://ts.catapult.org.uk/news-events-gallery/news/uks-largest-traveller-experience-study-identifies-key-themes-for-improving-transport-network/
[7] https://www.ofcom.org.uk/research-and-data/cmr/cmr15

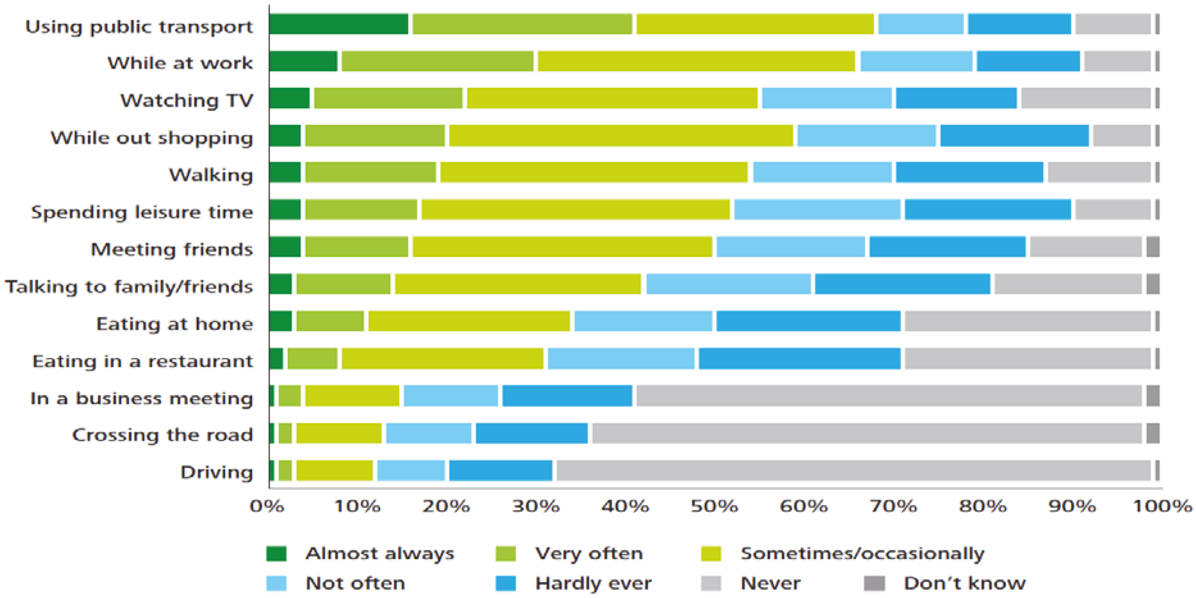


Figure 4: Usage of smartphones while doing other activities[8]

2.1.1 Mobile Sensors

Smartphones, now routinely, come packed with a variety of sensors that can provide different kinds of data and information regarding movement, orientation, and environmental conditions. These types of sensors are:

- **Accelerometer**: This is usually a 3-dimensional accelerometer which measures the acceleration force that is applied to the device. It is used to detect motion such as shake or tilt. For the purposes of this survey, accelerometer data can be used to identify movement and make assumptions about the mode of transport
- **Gyroscope**: measures angular rotation across three axes (x, y, z). This means that the gyroscope can track when a phone is spinning or turning.
- **Magnetometer**: Measures the geomagnetic field strength, basically acting as a compass.
- **Proximity sensors**: measures the proximity of an object to the screen of the phone. It is mostly used to determine the position of the phone to the face during a call.
- **GPS**: mostly used outdoors, it gives a record of the location of the phone at any point in time.
- **Other sensors**: light sensor, Bluetooth, Wi-Fi, cellular radios (GSM) and microphone

Some higher end models have more advanced sensors like thermometers, barometers (which measure the atmospheric pressure) and humidity sensors. Data collected from sensors can be used alone to understand and process information. However, research on

[8] Deloitte. Mobile Consumer 2015: The UK cut Game of phones

data fusion techniques have shown that data from the accelerometer, gyroscope and magnetometer can be combined and used to infer more reliable activity information[9]. Since Samsung is the most popular manufacturer in the UK market, the table in chapter 9.1 in the Appendix B, lists some Samsung devices with the sensors and GPS types.

2.1.2 Smartphone applications

A smartphone application or "app", as commonly known, is a type of custom software designed to run on mobile devices of different hardware specifications. Apps can provide users with services, similar to those of desktop computers. Nowadays there are apps for word processing, reading books and pdf documents, or even for learning programming in a desirable language. However, this doesn't mean that smartphone applications can replace desktop computer applications. They are small software units, designed to have relatively limited functionality to be able to run universally on devices of smaller processing power.

The development of applications is a complex and lengthy process where the device specifications, constraints and development requirements must be considered beforehand. The development time of an application can take between 2 – 4 months[10]. This depends on the number of features of the application, the complexity of the design and the user requirements, the operating system the app is developed for, and the team working on the application. Additionally, due to the number of smartphone manufacturers and the variety of devices they produce, developers must be considerate on different hardware specifications and configurations such as screen sizes, sensors, modules etc. Application development requires a Software Development Kit (SDK) which includes some software development tools that assist the developer with the creation of applications. An SDK includes various features such as Integrated Development Environment (IDE), Application Programming Interfaces (APIs), emulators that assist in the visualization of the user interface and the testing phase of the application.

Once the development is finished, applications are placed on an app store ready for users to download them. An app store is a distribution platform, specific for each operating system, where users can search, browse, and download apps that fit their needs. Android's app store is called Google Play. It opened in October 2008 and, as measured in June 2016, it hosted around 2.2 million available apps. Apple's app store is called App Store. It opened in July 2008 and as of June 2016, it had 2 million available apps[11]. Similarly, Windows Phone Store was launched by Microsoft in October 2010 and it currently hosts around 669,000 available apps [11].

---

[9] Context-Aware Personal Navigation Using Embedded Sensor Fusion in Smartphones
[10] http://readwrite.com/2013/01/09/how-long-does-it-take-to-build-a-native-mobile-app-infographic/
[11] http://www.statista.com/statistics/276623/number-of-apps-available-in-leading-app-stores/

2.1.3 Ability to improve the National Travel Survey

Depending on the implementation of an app, smartphones could eliminate some of the respondent burden and improve the quality of data by using some of the above sensors, usually in combination with a prompted recall approach, to take care of under-reported trips and inaccurate trip times. A prompted recall approach is a common practice in travel data collection research. Often, survey implementers develop a web interface, that can be accessed by mobile phones and computers, which displays to the survey participants the data collected from the sensors, and the participants can validate and approve the journeys.

Through the use of an app, smartphones could eliminate some of the respondent burden and help to improve data accuracy. Automating certain processing and analytical needs, allows semantics to be attached to data from sensors and, as a result, infer additional data such as trip duration, travel speed, trip times etc. The increased penetration of smartphones in everyday life makes them suitable to use on the collection of data. Smartphones are so essential nowadays that many people always carry them around and use them at every opportunity they find (Figure 4). This consumer behaviour proposes that survey participants will be consistent at collecting and providing data related to their travel activity. Additionally, smartphones can easily access cloud based services for data storage and information dissemination, minimizing the data loss risk and allowing for faster decision making.

Even though the ability of smartphones as a technology could be useful for NTS purposes, there is still an issue with the ownership of the devices. Unfortunately, there are still many people who don't own a smartphone device for several reasons. For example, elderly people who don't see the need for one and people who face economic or financial issues and therefore can't afford one, or young children who don't have a phone at all. This creates a less representative sample of participants, adds bias to the data and it might impose ethical and coverage issues in the context of tracking young children. Until this situation gets resolved the need for other collection methods that can complement smartphones will be needed. Additionally, the age of smartphones should also be considered. Usually, older devices with older versions of the operating system cannot support newer apps and possible new features they provide.

More specific to the current methods of the survey, the table in chapter 6.2 explains the capabilities of smartphones and the ways they can be used to capture the key data elements of the survey. It is assumed that smartphones have the appropriate sensors needed for the survey. Depending on the implementation of a smartphone app, some of the sensors that can be utilized are: GPS, compass, accelerometer; Wi-Fi. Older devices may not include all the sensors. However, the majority of smartphones will most certainly include accelerometer, GPS and Wi-Fi. The table in chapter 9.1 provides an overview of smartphones and the sensors they include.

2.1.4 Limitations and costs of smartphones

The main costs associated with smartphones are the following:

Device costs

Although the majority of the population owns at least one smartphone, there are still people who don't, for reasons such as cost of the device, usage complexity etc. Furthermore, some will own a version that is too old to support the technology needed for the app. Due to the variety of manufacturers and the operating system the devices come with pre-installed, the typical cost of an average smartphone device varies but it can be estimated at around £250 - £300 with higher end models costing around £500 - £600[12]. The relative cost of these devices makes it possible to provide a few participants with a smartphone device. In the Singaporean Household Interview Travel survey (see chapter 2.1.5), a few participants were provided with devices to use during the study[13]. Similarly, for the Dutch Mobility Panel project, the respondents without a smartphone or a device that is not supported by the iOS or Android operating system were loaned a device (a Samsung Galaxy Gio)[19].

Application development cost

The development cost of an application depends on different factors:

- The mobile operating system
- The technical complexity of the app
- The size of the development team involved in the project
- The timeline to develop the app

Developing an app is not a one-person job: it involves a team collaboration. That's why, there are a lot of software development companies with teams specialized in mobile application development. Usually, such companies have different teams specializing in different mobile operating systems and each team has their own cost.

The table in chapter 9.2 in Appendix B, shows some mobile app development companies in the UK and their pricing rate. As mentioned previously, the cost of an app is proportional to its complexity. Applications with basic functionality and entry level requirements cost around £5 - £15k, medium sized apps and games with requirements such as web services, media uploads, physics engine etc. are in the range of £15k - £50k.

---

[12] http://www.computerworld.com/article/2489944/smartphones/smartphone-prices-are-dropping--and-will-continue-to-dip-through--18.html

[13] The Future Mobility Survey: Experiences in developing a smartphone-based travel survey in Singapore

Large scale apps with web-based dynamic content, usage of various sensors are valued at around £50k - £250k[14].

Data storage and processing

There are some cloud based solutions, where smartphones can easily connect and transfer their data whenever they find an internet connection. The hardware specifications will depend on the size of the study and the processing requirements.

Data transmission costs

The data transmission can be done either via Wi-Fi internet connection or using mobile network data. Most of the time, Wi-Fi connection can be obtained from public access points which are cost free. On the other hand, the mobile network data will depend on the network provider and the type of plan being used.

Limitations

One of the major problems of smartphones is the limited battery life during an average day of use. The number of sensors in combination with background software and services that smartphones provide, along with the constant usage from users, provides a day or half day of battery life for the average user. Figure 5 provides an overview of the services and the amount of battery they are using.

In terms of daily usage, when watching YouTube videos using Wi-Fi, a battery lasts about 3.4 hours, when browsing the web using 3G a battery lasts around 5 hours, and with typical usage a battery lasts around 12 hours. According to Sharkey, the processes that cost the most for the battery are: apps waking up in the background, bulk data transfer, moving between cells/networks, and parsing textual data [15].

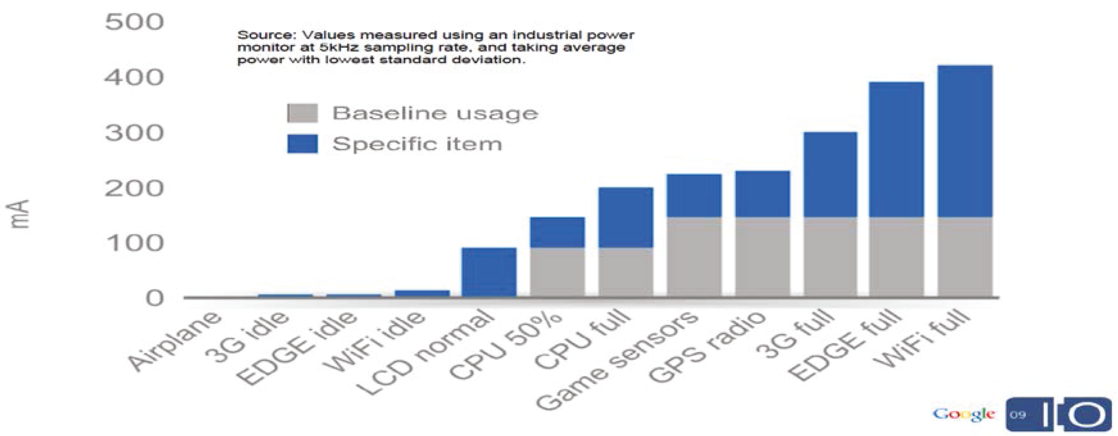


Figure 5: Battery usage from smartphone services and components whose current is measured in milliampere (mA)[15]

[14] http://www.the-app-developers.co.uk/app-dev-news/the-cost-of-an-app/
[15] https://dl.google.com/io/2009/pres/W_0300_CodingforLife-BatteryLifeThatIs.pdf

Designing an app that makes heavy use of the device's sensors and has as a main requirement to gather as much data as possible, becomes challenging considering the limitations mentioned above. Several research efforts have been around developing smartphone apps for travel surveys that have a workaround on the battery life. For example, for the Dutch Mobility Panel project, they used a strategy that was capturing data only when the sensing module detected significant location changes. This way excess battery use from the GPS sensor was prevented.

Another limitation that can affect the data collection process is the accuracy of the GPS sensor. There are a number of factors that can affect the accuracy no matter what GPS module is being used. For more details on that, see section 2.2.2.

2.1.5 Practical issues and used in travel surveys worldwide

The popularity and capabilities of smartphones have attracted the interest of many researchers and academics worldwide on the use of these devices as a data collection tool in travel surveys. However, incorporating smartphones in travel surveys is not a seamless process as there are a lot of challenges and issues to be overcome, to guarantee successful implementation. The main challenges that can affect smartphone-based travel diary surveys are: the participant involvement, battery life of the device, the level of participant's interruption of their daily travel activity, and privacy concerns that vary amongst participants. Additionally, to increase the quality of the data coming from multiple sensors it will be preferable to combine them. This approach, although useful to have, is not a straightforward process. Sensors may provide contradictory data, for example the accelerometer shows movement activity but GPS shows no change in location[16].

There are a lot of studies and experimentation on the implementation of an appropriate system design for dealing with the above concerns. Researchers at the University of Queensland worked on a smartphone based travel survey, for performing personal travel data collection, which was used experimentally as part of the New Zealand National Household Travel survey. The goal of this approach was to test the adequacy of undertaking a household travel survey at a national level using smartphones. The application was designed following a prompted recall approach, meaning the application collected the travel activity of the participants passively, and at the end of each day the participants had to recall and label their activity through the interface of the application. The main challenges of smartphone based surveys were well addressed in the design of this application. Some of the features of the app included: automatic detection of the start and end of the trips and advanced battery optimization algorithms that allowed daily activity monitoring without affecting the user's normal usage[17]. Moreover, to cover privacy

[16] Harnessing smartphone sensors for tracking location to support travel data collection

[17] Design and Implementation of a Smartphone-based System for Personal Travel Survey: Case Study from New Zealand

concerns, the app allowed users to review and delete whole day journeys. During the data analysis, it was noticed that the application recorded some trips which were not uploaded by the participants. This means that the app recorded some trips but the participants deleted them. The conclusions of the study showed a success in providing a pleasant experience for participants which encouraged them to continue their participation after the survey finished. More specifically, 55 participants (71% of the total participants) continued their participation after the end of the survey.

There has been a lot of research and experimentation on the best approaches to use smartphone sensors in a way that doesn't impact battery life and doesn’t place a constraint to daily use. Researchers at the University of Sydney developed a smartphone application for location tracking and stop detection by balancing the needs for data quality and battery life[16]. This study was done to investigate the changes in travelling patterns before and after the construction of specific cycling infrastructure. For the development of the app, the sensors used can determine a location using Wi-Fi and cellular network location services. The accuracy of the two were tested against another version of the app that used GPS. It was found that, when the Wi-Fi radio was on, the accuracy of the data from the Wi-Fi and cellular network were comparable to GPS data. The method they followed to increase battery duration was to request a location once every 5 seconds. If the battery level was decreasing the request time increased; for example, if the battery level dropped below 50 %, the request time increased to once every 15 seconds. This could influence the data collected but for the purposes of this project (looking at travel time, distance travelled, route taken and stops), the researchers found that there wasn’t any impact on the individual points.

Another novel smartphone based travel activity survey is the Future Mobility Survey (FMS) designed and deployed as part of the Singaporean Household Interview Travel Survey (HITS)[18]. Initially, 74 people registered to participate in the survey, of whom 34 installed the app and collected data, and only 27 of them validated their data. The app was part of a framework designed to function as a prompted recall travel survey. The architecture of the survey used a smartphone app, developed for both Android and iOS platforms, used for data collection, and a server database that stored the data coming from the app and performed machine learning techniques to understand patterns and interpret the data. Finally, there was a web interface that presented the data to the user and allowed them to validate their trips. There were, also, a few stages that took place before and after the data collection process, such as participants being required to register to the system, and having to complete a pre-survey questionnaire that included questions about their household, socio-economic information, vehicle and driving licence ownership in the

[18] The Future Mobility Survey: Experiences in developing a smartphone-based travel survey in Singapore

house. The survey app was scheduled to run for 14 days and, at the end of the collection process, there was an exit survey to collect feedback and comments on how it could be improved. Unfortunately, this pilot showed that the whole participation process, registration, pre-survey, installing the app, collection of data, was very unclear and burdensome for some users. For example, at the validation stage it appeared that only experienced or tech savvy users corrected or added new locations to their journey summary.

A similar approach was used for the Dutch Mobility Panel. A smartphone app, called MoveSmarter, was developed, for both Android and iOS, to automatically detect departure and arrival times, trip origins and destinations, transport modes and travel purposes[19]. The app was in a three-year smartphone-based prompted recall panel survey, where around 600 smartphone and non-smartphone users participated and over 18,000 trips were recorded and validated by the participants during a two-week period. Similar to the Singaporean survey, the application was accompanied by a web-based interface for prompted recall purposes; the users could edit their journeys via an easy to use interface. Participants without a smartphone or one that supported Android or iOS were provided with a loan device. Interestingly, it was noticed that the participants with the loan devices were more prone to forget their phone than the other smartphone owners. The iPhone users were the most consistent in carrying their device with them and, as a result, had fewer missing detections. Although it was concluded that the overall performance of the app was satisfactory, with MoveSmarter able to detect most trips correctly without strong biases, the app struggled to detect trips where the activity time was small or when two successive trips followed each other. Furthermore, short distance car trips were classified as bicycle trips, train trips were undetected and public transport modes (bus, tram, metro) were poorly detected.

2.1.6 Summary

Smartphones have proven to be a very advantageous technology that can be applied to various fields. The increased penetration of smartphones in everyday life along with the ever increasing software and hardware capabilities makes them an important and useful tool for innovation in travel diary surveys. The increasing ownership of smartphones gives the advantage of: lower costs for the survey implementation since there is minimal need to provide equipment to participants; there is minimal risk of data loss; and there is the possibility to process data and obtain insights faster than before. In addition, currently, with the cloud services and the ability of smartphones to access the Internet, the cost of getting access and securing the participants data has also reduced significantly.

[19] Automatic trip and mode detection with MoveSmarter: first results from the Dutch Mobility Panel

The smartphone based studies showed that the challenge of battery life can be controlled, although, depending on the technique followed, it can have an effect the quality of the data. Also, by following a passive data collection approach, most of the information from the data are inferred and a prompted recall approach is essential to ensure the quality of the data. In addition, it was noticed the participation levels, in some of these studies, weren't consistent (people stop validating their trips, forget to install the app etc.) so providing reminders and prompts is very significant for the quality of survey data.

In conclusion, smartphones have many advantages but they also carry disadvantages. The main disadvantage that smartphone users face every day, and one that affects the successful implementation of household travel surveys, is battery life. Given that the average smartphone device lasts for about a day with a single charge, it influences the collection of travel data since it can lead to half completed trips, or even completely missing trips. Another disadvantage that also affects the implementation of travel diary surveys is the lower number of the elderly population that own a smartphone. Although, as seen in figure 3, smartphone ownership in the older generations is increasing year by year, which suggest that smartphone ownership in the elderly population will become less of an issue in years to come.

2.2 Dedicated GPS devices and data

Global Positioning System (GPS) technology is a space navigation system developed by the US military in 1973 and became operational in 1995. GPS provides precise information about time and location of someone with a GPS receiver anywhere on the globe and under any weather conditions. Originally, when the system was deployed, it was using 24 satellites; currently it uses 32, reinforcing the precision accuracy to -5 meters.

There, also, exist two other major space navigation systems: GLONASS and Galileo. GLONASS was developed by the Russian Ministry of Defence, and currently uses 24 operating satellites in the orbit. Galileo, developed by the European Union (EU) and the European Space Agency (ESA) is still under development but is expected to be operational by early 2020 and it will be interoperable with GPS and GLONASS providing more accurate positioning even in environments that tend to obscure the signals such as urban environments. The system promises to have 30 satellites, 24 active in the orbit and 6 active spares, and it will be able to provide two types of positioning services: Basic use, low precision positioning will be around 1 meter and it will be free for public use; high precision positioning will have 1 cm precision and it will be available for paying commercial users.

The capabilities and freedom of use has made GPS a very useful technology with numerous military and civilian applications. For example, known military applications are: navigation, target tracking, missile, and projectile guidance. On the other hand, civilian applications include but are not limited to: astronomy; fleet tracking; robotics; sport; surveying. Since the late 1990s, GPS has been investigated for use in household travel diary surveys all around the world to provide a more in-depth view of travel behaviour and movement. Initially, GPS was used as a validation and correction method of the records that were collected from travel diaries, and now, it has been widely explored as a data collection method for travel diaries. Table 1 provides examples of pilot GPS surveys undertaken by countries around the world [20].

Table 1: Household Travel Surveys around the world

Location	Year	Survey purpose	Device	Sample size	Collection period	Technical details	Processing involved
Four states in Australia	2007–13	Travel behaviour change monitoring	Dedicated GPS device, recording data every second	130 households	15 days (6 waves)	Random sampling; GPS-only survey	TI, MD
Ontario, Canada	2007	Route choice	Smartphone plus a GPS receiver	31 respondents	2 days	Snowball sampling; GPS survey with a pre-interview and a web-based PR survey	TI, PI
France	2007–08	Sub-sample of National Travel Surveys	Dedicated GPS device, recording data every 10 seconds	9% of the main survey	7 days	Random sampling; GPS survey with one-day travel diary	TI, MD, PI
Matsuyama, Japan	2004	Compare GPS records and travel diaries	GPS-equipped mobile phone, recording data every 30 seconds	31 respondents	5 days	Non-random sampling; paper based diary and GPS survey with a web diary	TI

Location	Year	Survey purpose	Device	Sample size	Collection period	Technical details	Processing involved
Jerusalem, Israel	2010	GPS-only household travel survey	Dedicated GPS device, recording data every second	3000 households	1 day	Random sampling; GPS-only with a PR survey	TI, MD, PI
Three cities in the Netherlands	2007	Residential selection	Dedicated GPS device, recording data every six seconds	1104 respondents	7 days	Random sampling; GPS-only survey with a web-based PR survey	TI, MD, PI
Western Cape, South Africa	2008	Assess the reliability of GPS survey	Dedicated GPS device, recording data every second	100 respondents	14 days	Random sampling; GPS survey with two-day travel diary	TI, MD, PI
Borlange, Sweden	1999–2001	Traffic safety	In-vehicle GPS device, recording data every second	310 vehicles	15–243 days	Stratified sampling; in-vehicle GPS survey	TI, PI
Three cities in Switzerland	2008	Explore whether participants pass certain billboards	Dedicated GPS device	4882 respondents	Average 6.6 days	Random sampling; GPS-only survey	TI, MD, PI

Location	Year	Survey purpose	Device	Sample size	Collection period	Technical details	Processing involved
UK	2011	Test the possibility of replacing travel diaries	Accelerometer equipped GPS units, recording data every second	429 households	7 days	Random sampling; pilot survey (GPS only) for National Travel Surveys	TI, MD
Ohio, USA	2009–10	GPS-only household travel survey	Dedicated GPS device, recording data every second	2059 households	3 days	Random sampling; GPS-only survey with a web-based PR survey	TI, MD, PI
Graz and Tullnerfeld, Austria	2009–10	Test an integration of new technologies for a mobility survey	Dedicated GPS device	235 respondents	3 days	Random sampling for four groups (passive GPS-only, active GPS only, GPS with diary, and diary only; pilot GPS survey with PR)	TI, MD, PI
Beijing, China	2010	Sub-sample of Beijing Household Travel Surveys	Dedicated GPS device, recording data every five secs	890 people	1 day	Random sampling; GPS survey with one-day travel diary	TI, MD
Greater Copenhagen Area	2013	Part of the research on travel chain and sustainable mobility	Dedicated GPS device recording data every second	54 households	3–5 days	GPS survey with one-day travel diary	TI, MD

2.2.1 Ability to improve the National Travel Survey

The traditional survey methods are bound by problems such as underreporting of trips, inaccurate locations and times. GPS travel diary surveys have an advantage over the traditional methods since they can potentially reduce respondent burden and can provide more accurate and detailed information on the geographical and time related aspects of participant's travel activities. Typically, GPS devices can provide more data than location and time only. Depending on the devices, the data that can be provided are: date, time, latitude, longitude, altitude, NSAT (the number of satellites that a GPS device used to calculate its position), HDOP (horizontal dilution of precision, measuring how the satellites are arranged in the sky at the time of the record), speed and heading[20].

Although, the data collected from GPS data loggers provide useful insights on certain aspects of the travel activity of people, there are still some data that GPS cannot provide. For example, journey purpose, method of travel, and number of boardings, cannot be collected from GPS and require advanced processing methods to be used. Such methods can be the inclusion of Geographic Information System (GIS) and machine learning algorithms trained to infer details from the data. This leads to surveys following a prompted recall approach where participants, either through an interview or a web interface, can correct errors in their trips, complete and confirm journeys or provide additional information useful to the survey.

The table in chapter 6.2 shows how GPS and certain processing methods can capture certain key data elements contained in the travel diary. Some of these elements can be captured only with GPS and, as explained in the table, some others can be inferred using other data sources and methods.

2.2.2 Limitations and costs of GPS data

The following section outlines the costs associated with GPS technology for travel surveys:

Device costs

The average cost of GPS data loggers starts from £50 and it can reach up to £500. Due to the fact that GPS units are not that common for people to have, the price would be proportional to the sample size of the survey.

Data transfer costs

Depending on the types of GPS devices and the requirements of the survey, there might be a need for data transmission via wireless networks.

[20] http://www.sciencedirect.com/science/article/pii/S1877042814041597

Deployment costs

Given that GPS devices are not common devices for people to have, the process and logistics of distributing GPS loggers to participants and making sure they are comfortable using it, adds to the cost. This cost is relative to the sample size.

Data processing costs

There will be a need for data processing methods to be developed. The costs of this will relate to the labour of a skilled person to do this work and the hardware costs for storing and processing large quantities of data. Currently, with the popularity of cloud computing, renting computer clusters for large processing jobs is reasonably priced.

Limitations

It is well known that GPS can provide highly accurate data. However, there are some problems that affect GPS devices. Such problems are the difficulty of obtaining a signal in certain urban areas due to the infrastructure and the built environment such as skyscrapers, indoor halls, underground or in tunnels. Signal noise and signal loss are two main issues that GPS units have and there will always be situations and areas where GPS will fail to record locations. There are various reasons signal problems occur, cold start or warm start (the time it takes for the device to locate satellites and calculate position) are two of them[21]. These problems usually occur at the beginning of each day (i.e. cold start) or when the GPS device switches from 'sleep mode' to 'working mode' after a person stops for one or two hours (i.e. warm start), and travelling in an urban canyon.

As mentioned before, dedicated GPS devices are not very common among ordinary people, this means that there is a possibility participants will forget to carry the devices with them or keep them charged up ready for use.

2.2.3 Practical issues and use in travel surveys worldwide

There have been numerous travel surveys that make use of GPS either as a data collection or data correction method. A relatively recent GPS enhanced Household Travel Survey is the California Household Travel Survey which started in February 2012 and ended in January 2013[22]. This survey was a state-wide survey driven by many state agencies and councils from across the state. This survey followed a dual approach where participants received travel diaries and GPS devices. The sample size of the survey was around 42,500 households, where 5,100 of them were chosen to participate in the survey with the GPS component. In addition, there were three types of GPS devices: a wearable GPS, vehicle GPS and an on-board diagnostic device.

[21] http://usglobalsat.com/forum/index.php?topic=4275.0
[22] 2010-2012 California Household Travel Survey Final Report Appendix

One of the issues that was faced, once the survey was completed, was the retrieval of the equipment that was provided. A retrieval management system was set in place which included phone calls and letters requesting the return of the equipment provided for the survey. Further, additional work needed to be done on the analysis and identification of misreporting or over-reporting trips. The identification was done by comparing the GPS data with the data from the travel diaries. The challenge was to identify the similar trips in both methods, for example, if a respondent made a regular trip from home to work but stopped at a petrol station in the middle, the GPS might count this as two trips but in the travel diary, the respondent will count it as one journey. This analysis created a trip classification of three types: matched trips, the ones that were found in the GPS and the travel diary; underestimated trips, trips found in the GPS but not the diary; and overestimated trips, trips found in the diary but not the GPS. Of all trips made by the 6,241 participants, 3,050 participants (49%) had perfect matches, 11,725 trips were underestimated and 5,757 were overestimated.

Another survey was the Atlanta Regional Travel Survey that was implemented in 2011 with a sample size of 10,278 households with 1,061 of them using GPS devices[23]. The purpose of this survey was to collect detailed information on trips made so the Atlanta Regional Commission could improve their travel demand forecasts in their trip-based and activity-based models. The sub-sample of households using GPS was equipped with both vehicle and wearable GPS equipment to provide an independent measure of travel, more specifically, to estimate levels of trip underreporting in this sub-sample that could be applied to the larger non-GPS sample. The GPS devices were to be used for seven days by the vehicle sample and three days by the wearable sample, with the first day coinciding with the assigned diary/travel day. All household members were asked to keep track of all their trips for a 24-hour period using a paper diary. After the collection period the survey participants reported their travel details by telephone (42% of participants), online (29%) or by mail (29%). Once all the travel details were retrieved, they were processed and subjected to quality control checks. For example, for the GPS household data, a comparison was performed of what was reported by telephone versus the trips detected by GPS data streams. The survey had a 55% completion rate of GPS and diary data collected from the 1,061 households out of the 1,938 households deployed in this study. Additionally, 392 households completed GPS data without reporting diary data. Similar to the California Household Travel Survey, a classification was created of: perfectly matched trips; trips reported by the diary but not captured by the GPS; and trips captured by GPS but not reported by the diary. More specifically, of all trips made by the 649 people, 277 had perfect matches between the diary and the GPS data. For the second classification, 478 diary trips were identified that had no corresponding GPS trips. This is due to technical problems such as respondents placing the device in a position (purse or

[23] http://www.atlantaregional.com/File%20Library/Transportation/Travel%20Demand%20Model/tp_2011regionaltravelsurvey_030712.pdf

backpack) that could receive GPS signal or they forgot to confirm that the device was powered on. For the last classification, there were 798 missed diary trips that reflect a 19.6% missing trip rate across the entire sample.

2.2.4 Summary

GPS is one of the most researched technologies in household travel surveys around the world. Given the existence of another two navigation positioning systems (GLONASS and Galileo) and the ability of GPS based devices to make use of more than two systems (see table in chapter 9.1 in Appendix B) guarantees more accurate readings and additional data.

However, the main problem with GPS technology is that it is not common for people to own a separate GPS recording device. This means that for household travel surveys there will be a need to provide equipment to participants and devise methods to capture the data and collect the equipment back from them. This adds to the cost of implementation in terms of hardware and labour. The requirement to provide equipment doesn't allow one to target a large sample size; this potentially could lead to skewed results and biased views. Additionally, there is the possibility of participants forgetting to carry the devices with them, which could have a negative effect on the data quality unless there is a way of retrospectively adding in missed trips.

The main question in many of the household travel surveys is whether GPS technology can replace to a significant extent the current methods of capturing the data collected in travel diaries. Potentially, with the right framework implementation, GPS could be used as a main mode but currently there are two ways GPS enhanced surveys can replace traditional household surveys: 1) the survey must follow a prompted recall approach where GPS data are gathered and processed and after that, the participants input is required to confirm, correct or even complete the trips; 2) using a GPS-based prompted-recall sub-sample to calibrate models that are then used to impute details on completely passive GPS data collected by the majority of the sample without further input from survey participants. These models could be based on machine learning algorithms that will be trained to infer and understand patterns from GPS data that have been collected passively and not labelled or processed from the survey participants.

2.3 Mobile Network Data

Mobile network data (MND) is a novel technology which has started gaining popularity in the field of transport modelling and intelligent mobility[24]. MND can be acquired from mobile network operators (MNOs) and their properties and types make them suitable for national decision making since they can assist in providing insights into why, how and when people travel, in a timely and comprehensive manner.

A mobile network consists of multiple land areas called cells where each cell is served by at least one fixed location transceiver, known as a cell site or base station (figure 6). The base station provides cells with network coverage which is used for transmission of voice and data information. Each cell operates in different frequencies ensuring there is no interference with neighbouring cells. Combining all the cells together creates a communication network that provides radio coverage over a wide geographic area. This way portable transceivers such as mobile phones can communicate with each other and with fixed receivers and telephones anywhere in the network. As mobile phones communicate with the mobile network they transmit their position and receive data for network operation, billing and regulatory purpose. Each communication event is associated with a distinct, persistent and, for privacy reasons, anonymous user id as well as with a timestamp and a cell identifier. The cell identifier is useful for locating the event within a specific region of the UK by means of a cell map[25].

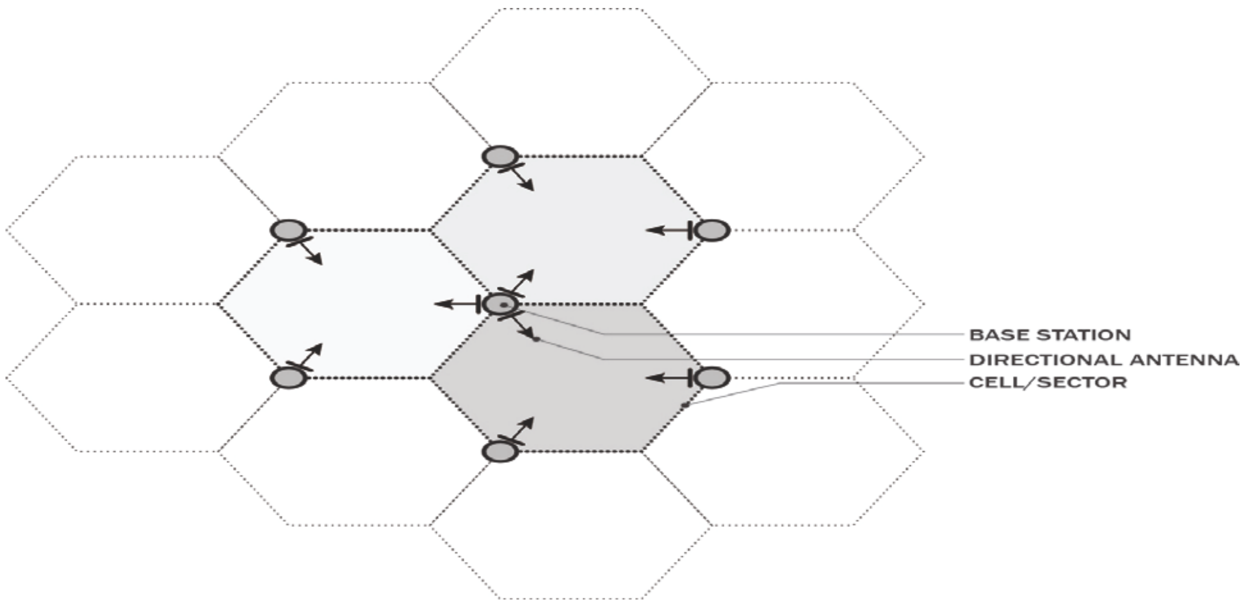


Figure 6: Cellular network organization[26]

[24] http://citilogik.com/sectors/intelligent-mobility/#tab-id-1
[25] Utilizing mobile network data for transport modelling, Internal TSC report: unpublished
[26] http://urbanomnibus.net/2011/07/signal-space/

Mobile network data is a by-product of the interaction of mobile phones with the network and they come in different formats and types. It must be noted that MND are not restricted to smartphones or new generation devices. Any type of mobile phone that can make a phone call, therefore interact with the network, generates MND.

The events described above, can be categorized into active and passive events. Passive events occur when the phone is not used but the user moves from one cell to another and depending on the collection method, the first cell the phone was connected to can be identified. Active events occur when the phone is used to either make a phone call or access the Internet via network data. This way keeps track of the cells used by the user during the phone call. In addition, events can be divided into bands, or generations. Depending on the mobile device, the data plan and the cell coverage, a phone can connect to a 2G, 3G, or 4G network, which then will determine the band type of events generated.

The technology MND are based on, is evolving and expanding in terms of the technological context and analytical processes. For example, the 4G network is growing and more users are switching to it, thereby creating more data with better quality and increased granularity. 4G networks are relatively new and are mainly used for providing mobile broadband. Currently, 2G and 3G technologies are used with 3G being the most popular broadband technology in the UK. 3G provides customers with voice calls, text messages and mobile broadband capabilities, whereas, 2G provides the same services but with very low speed data connections.

In the UK, there are only four major mobile network operators, EE, O2, Three and Vodafone[27]. These companies own and manage the wireless network infrastructure and lease telephone and data services to mobile virtual network operators such as Tesco Mobile, Virgin Mobile and giffgaff[28]. All of them provide 3G and 4G coverage with EE leading the market in 4G coverage at 95% of the UK and 3G coverage at 98% of UK premises[29]. Figure 7, shows a comparison of the 3G coverage amongst the major network operators, the figure shows the distribution of signal from weak (red) to strong (green). As mentioned, MND can be purchased from some mobile network operators, the pricing plan depends on the data specification and the network operator. The developers in need of this data must provide a request to MNOs where they specify the criteria and data they need. Such criteria could be: time, number of days, type of day (weekday, weekend etc.), spatial resolution and cell movements. To define a good data specification, model developers should work together with the MNOs to understand better what type of data

[27] https://www.uswitch.com/mobiles/networks/
[28]https://en.wikipedia.org/wiki/List_of_United_Kingdom_mobile_virtual_network_operators
[29] http://www.4g.co.uk/4g-news/news-features/best-4g-network-networks-compared_300112734.html

they have available, the methods they use to capture data and what processing methods have been applied to the data.

MND are covered by the Data Protection Act 1998, this means data containing personal information cannot be shared, including details that could be identifiable if MND are merged with other data sets. Data needs to be anonymised by the MNO before they are shared with the developers. The anonymization process differs by MNO and it may affect the data by adding bias to them. Developers should consult with MNOs on this process.

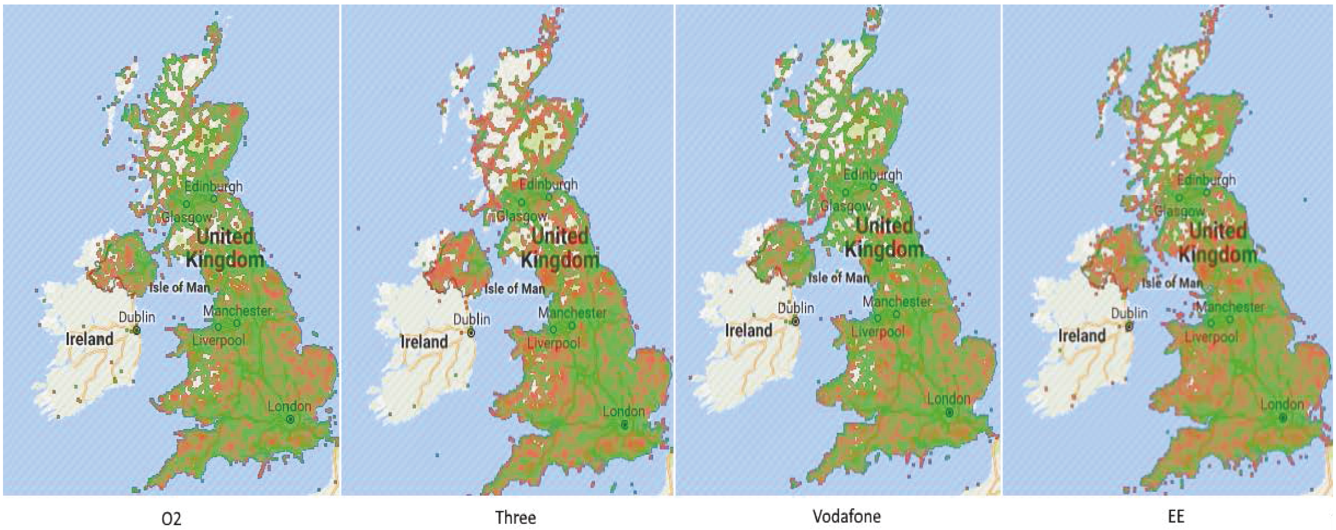


Figure 7: 3G coverage of major operators in UK[30]

2.3.1 Ability to improve the National Travel Survey

Mobile network data can provide information on many fields regarding transportation in the National Travel Survey, however, they lack information on personal characteristics that is essential to the survey. This means that fusion with other data sources is needed, to be able to derive additional information such as census data or public transport usage data.

Chapter 6.2 gives an overview of the information that can be acquired from mobile network data relative to the information needed for the NTS. More specifically, mobile network data are used to create origin and destination matrices. Building on that, the knowledge of origin and destination, along with a consistent user id, a timestamp and a region of space can be used to infer and recreate a user journey. Based on that, travel and movement behaviour of users around the UK can be observed.

30 http://www.broadbandwatchdog.co.uk/story/2009/07/08/3g-coverage-maps-show-3g-and-mobile-broadband-coverage/

Trip purpose and mode of transport, although very hard to obtain, can potentially be inferred from the network events. One way the trip purpose can be inferred is by using a consistent id that spans across multiple days, and by analysing the event's certain locations such as home residence and work can be inferred. This consistent id should be a field in the network data, but is not always the case. The mode of transport is a product of multiple factors such as: speed of travel which can be calculated based on time and location of events, the spatial pattern of events along the journey which can be used to identify the route of the journey, and rail travel can be identified based on a clustering pattern of journey events.

The benefit of using mobile network data is that they require minimal effort of obtaining the data since there is no need to undertake expensive and time consuming surveys. However, there is a level of uncertainty into how mobile network operators process, aggregate and anonymise the data, as well as, the lack of historical precedent and technical processes in using it. The industry is still growing, and processes on how to deal with issues such as data granularity, privacy and quality are being developed.

2.3.2 Limitations and costs of mobile phone data

Costs

The only costs that could potentially be associated with mobile network data is the cost of acquiring the data sources from the network operators. Depending on the volume and variety of data, data storage could be an additional cost but nowadays, with all the different cloud services, the costs are very reasonable.

Limitations

Research from Transport Systems Catapult (TSC) on the use of mobile network data in transport modelling showed there are a few limitations and biases in using mobile network data[25]. These limitations are a by-product of the network coverage of the different operators. It is common knowledge that mobile network operators don't have good coverage in certain areas, for example certain rural areas or occasionally in urban areas where the signals from mobile phones interfere with signals coming from other phones. Furthermore, there are other natural factors that can affect the signal travelling to the cell stations. For example, bad weather such as fog, rain and even the amount of leaves on trees can affect the signals. Another possible limitation are people who don't own a mobile phone and are unable to generate data. However, nowadays this limitation is of very low risk.

Adding to the network coverage, cell resolution is also a source of limitations. Possible limitations are:

- short trips cannot be identified easily, as compared to GPS technology
- attributes such as trip purpose and mode of transport may not be well represented

- multimodal journeys cannot be identified.

The distance of short trips is arbitrary and it depends on the cell resolution and the area studied, for example, spatial resolution in urban areas is usually of better quality than in rural areas. Mobile network operators are comfortable sharing Middle Super Output Area (MSOA) resolution, but a more granular spatial resolution can be achieved with better network coverage that can be achieved with bigger cells in certain areas. In addition, given the data are anonymized and pre-processed for legal issues, all information and insights are inferred and not confirmed. For example, trip purpose such as home to work trips could be inferred with some relative accuracy but there is no way to validate this accuracy.

2.3.3 Practical issues and use in travel surveys worldwide

There is very little practical experience on the use of MND for household travel surveys and at the time this report was written there were no examples of MND in travel surveys around the world. Most of the research on MND has been done around transport modelling and more specifically around traffic and mobility monitoring during natural disasters and emergency situations. The purpose behind that is to monitor population movement and respond with help and assistance more rapidly. A relatively recent example comes from Wesolowski, who made use of call detail records to understand human mobility during the Ebola outbreak in Africa in 2014[31]. From this study, it became clear that MND, with the right processing and analysis, can be used to analyse population movements which is particularly important for improving responses to disasters and for planning outbreak strategies. Furthermore, another field where there has been a lot of research is on the use of mobile network data in urban planning and analysis. Research has focused on: estimating population distribution, estimating types of activities in different parts in the city, estimating mobility patterns, and analysing local events.[32].

2.3.4 Summary

MND have proven to be a very useful data source for transport modelling purposes such as traffic monitoring. For the purposes of the NTS, mobile network data could only play a complementary role since they have some weaknesses that can affect the NTS. The definition of a trip, mode of transport and trip purpose can be inferred from the data but with a certain degree of uncertainty. The ability to obtain such data depends on several factors such as: spatial resolution, the pre-processing from the MNO's side, the ability of the MNO to capture data in greater resolution and so on. Additionally, as seen short trips are very difficult to capture and they also depend on the cell resolution the operator can provide.

---

[31] http://currents.plos.org/outbreaks/article/containing-the-ebola-outbreak-the-potential-and-challenge-of-mobile-network-data/

[32] Urban Sensing Using Mobile Phone Network Data: A Survey of Research

3 Technology Group 2: Web-based travel diary surveys

Web-based travel diaries are not a new concept; travel diaries, such as that used in the NTS, have been used in transportation research for over 60 years and the Web has been a platform for the hosting and publication of surveys since its infancy[33]. Although web-based surveys are not a new technology, the ever increasing access to the Internet along with the development of new technologies and programming languages can advance the development and format of travel diaries to more complex and comprehensive activity diaries.

The Internet has become part of the everyday life and the percentage of the computer literate population has increased. According to the Office for National Statistics (ONS), in 2015 the Internet was accessed every day or almost every day by 78% of adults (as seen in Figure 8) which translates to 39.3 million people. Nowadays, internet access is linked with computer and mobile usage. Last year, 72% of individuals used a computer every day. As seen in Figure 9, the percentage of people aged 55-64 and 65+, since 2006, has doubled and tripled in size respectively. On the other hand, 74% of adults used the Internet 'on the go' on a smartphone or handheld device. Naturally, the adults aged 16-24 constitute the highest percentage (96%) of the population whereas adults aged 65+ were just about 29%[34].

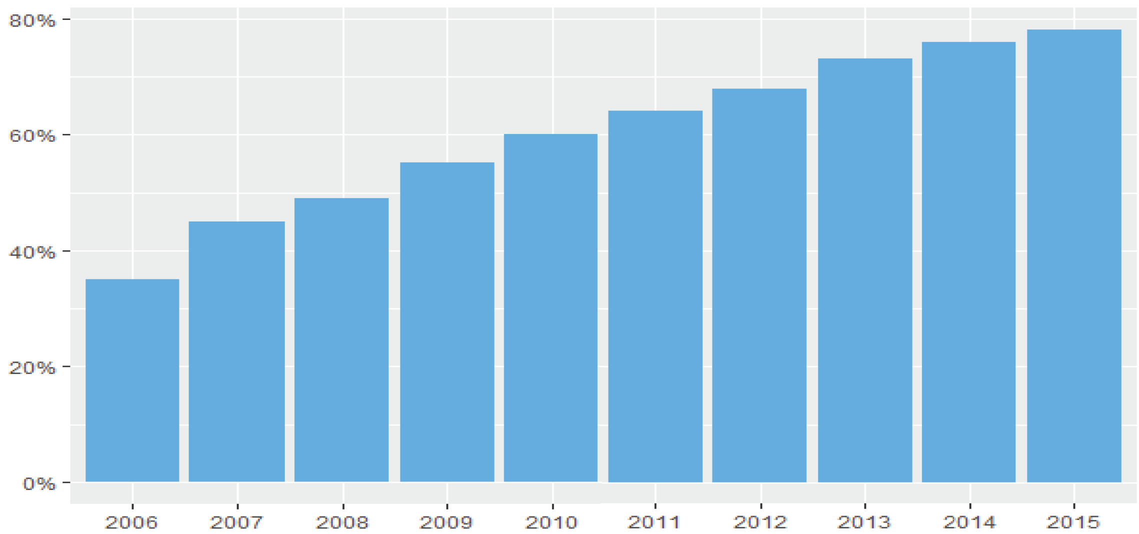


Figure 8: Daily Internet use by adults, 2006 to 2015, in Great Britain

33 Travel Diaries: An Annotated Catalogue

34 http://www.ons.gov.uk/peoplepopulationandcommunity/householdcharacteristics/homeinternetandsocialmediausage/bulletins/internetaccesshouseholdsandindividuals/2015-08-06#mobile-internet-access

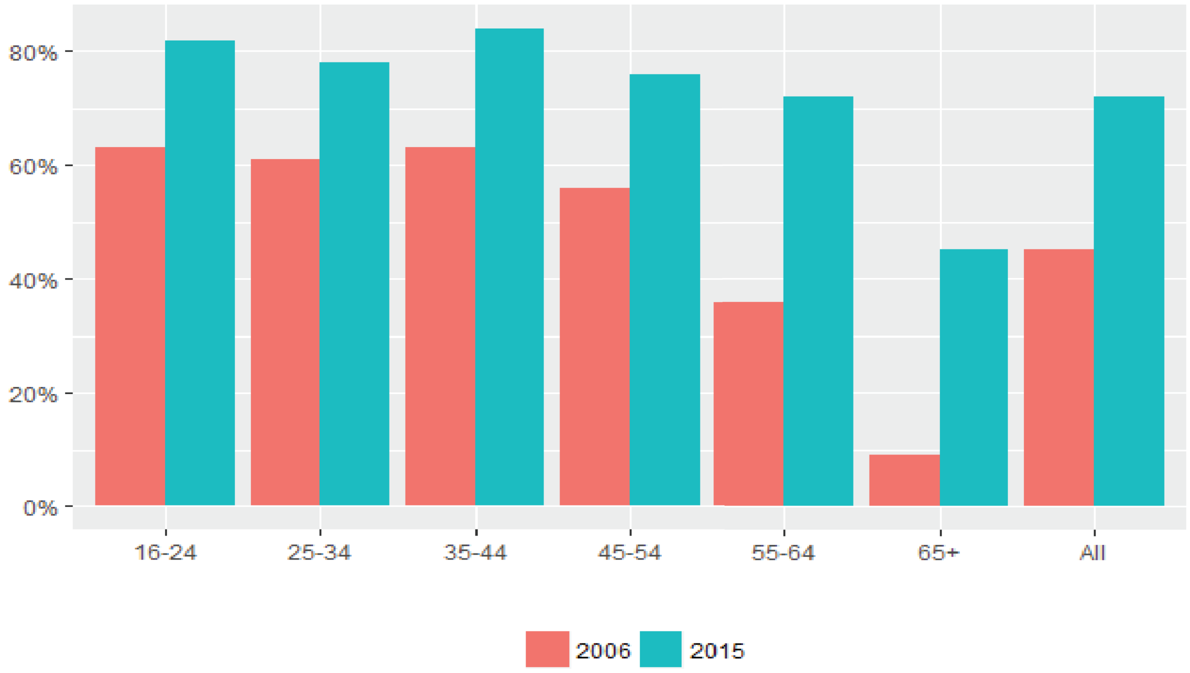


Figure 9: Daily computer use by age group, 2006 and 2015, Great Britain

In Britain 86% of households had an internet connection in 2015 which is an increase from the 57% back in 2006. Notably in 2015, 45% of households with one adult aged 65+ had internet access compared to 9% of households in 2006 (Figure 9). In addition to households, community online centres, schools, public libraries and more workplaces have access to the Internet[35] [36].

As mentioned before, there is a growing number of people who accessed the Internet 'on the go' via a smartphone or tablet device. This is due to the availability of free wireless hotspots in places such as pubs, cafes, and nowadays, even in buses, trams and trains. In Leeds and Bradford, for example, over 24 trains and buses have been equipped with free Wi-Fi and more than 600,000 unique users use the service. In Edinburgh, 713 buses and 27 trams have 2.6 million unique users using the free service[37].

This increase in 'on the go' internet access has brought changes on the way the internet activities are conducted. Now, more and more people do online shopping, participate in social networks and complete online surveys from their smartphones. Nowadays, even if a web survey is not developed for completion from mobile devices, participants are still

35 http://www.peoplesnetwork.gov.uk/about.html
36 http://www.tinderfoundation.org/what-we-do/uk-online-centres
37 https://www.gov.uk/government/news/millions-of-commuters-now-using-governments-free-wi-fi-on-public-transport

using their smartphones to respond to the surveys. Although there are travel diaries that abide with the new behavioural theories, few have incorporated recordings with the use of new technologies and standards.

A new standard that came into play around 2010 and changed the way information can be accessed by smartphones, is the mobile Web. The mobile Web is a concept whereby websites and related media that can be accessed by the browser of a smartphone or tablet are re-designed for mobile screens following standards tailored for the specifications of mobile devices[38]. Such standards aim to improve the interoperability, usability, and accessibility of mobile web usage. Considerations regarding the technical development on smartphones usually include two options, using the device's in-built web browser or using an application that runs within the operating system of the device. There are advantages and disadvantages for both options. These are summarised at the table in chapters 9.3 and 9.4 in appendix B[39].

As with every technology there are advantages and disadvantages regarding the quality of the data collected, the respondent burden and cost. Particularly, when implementing mixed mode surveys there are certain limitations that can affect the quality of data collected. All the benefits and limitations will be explained further in the sections following.

3.1.1 Ability to improve the National Travel Survey

Considering a proper and careful survey design, web-based travel diary surveys as a technology have the capability to improve the NTS because they allow the collection of data on the full context of travel including the purpose of travel, the scheduling of the trips and the sociability of journeys. Depending on the questions asked, a web travel diary can collect the travel diary data used in the NTS.

An additional benefit to using this technology is the ability to keep a detailed track of completion rates while the study is in progress. This offers the ability to send targeted reminders to respondents who haven't completed the diary or parts of the diary for a particular day.

Given that respondent burden is a big challenge in such surveys, web-based technology can provide various features that can help reduce respondent burden and improve completion times and data quality. Autocompletion functionality and some visual effects, such as a map to display the user's journey, for example, can reduce respondent burden. Additionally, automated checks such as extreme travel speeds or impossible combinations

[38] https://en.wikipedia.org/wiki/Mobile_Web#Standards
[39] Making it fit: how survey technology providers are responding to the challenges of handling web surveys on mobile devices

of answers such as being a passenger without having a travel companion[40] can be implemented to ensure the quality of the data. There has been a significant amount of research on the use of machine learning algorithms for pattern recognition, in particular, machine learning classifiers learn as the user interacts with the travel diary and provide answers for the user to speed up the process and reducing the burden of completing the diary.

Although, some of these reminders and checks may prove problematic in the duration of the survey, their inclusion could improve the data quality and reduce respondent burden in a manner in a less irritating and intrusive manner than repeated phone calls.

3.1.2 Limitations and costs of Web-based travel diaries

Costs associated with Web-based travel diaries:

Survey development

This cost relates to the labour required to develop the travel diary. The skillset needed for this job revolves around web development and web design. There might also be the need for a survey designer with experience in transport related surveys.

Hosting

There are web hosting services where with a reasonable price they can host a website on their servers and make it accessible via the Internet. Additionally, a domain name should be purchased and registered. The domain name will be the name or the address of the website; the participants of the survey will have to use this name to access the survey.

Data storage

Depending on the sample size of the survey and the amount of information that needs to be collected, there might be a need for additional data storage. Currently, there are cloud-based services with reasonable prices that allow integration with websites and other technologies.

Limitations

The main limitation for a web-based travel diary is the need for access to an internet connection. The number of households that have an internet connection keeps increasing, in 2015 86% of households in Great Britain had an internet connection, this was up from 84% in 2014 and 57% in 2006, and people of all ages are becoming more knowledgeable on that matter; however, there is still a percentage of the population that has yet to reach and adapt to the technologies and associated behavioural changes. Although, as mentioned before, there are places such as community centres and public libraries that provide Internet and computer access, the potential of using them could ease the conduct of the survey but it could also add extra to the cost and labour required. The more realistic

[40] http://www.sciencedirect.com/science/article/pii/S235214651500318X

alternative is to run the survey as a mixed mode and retain the paper diary along with the web-based diary. Additionally, there might be some issues over the willingness of the respondents to conduct such activities online. Particularly, people who are occasional Internet users might be concerned with security and would find it hard to give personal details online.

3.1.3 Practical issues and use in travel surveys worldwide

Web travel diaries have long been tested and used in household travel surveys around the world. Another study on the development of an online travel diary for the longitudinal investigation of travel activities in the inner-city of Sydney, Australia was conducted by the University of Sydney[41]. The diary was developed with the intent to be intuitive to use, quick to complete and introduce less burden to the participants. More specifically, the diary made use of several features to simplify the process of data entry, and improve participant recall and completeness of travel. Such features were: auto-fills, prompts, trip editing capabilities, favourite trips, and a drag-and-drop technique for capturing travel mode information. Another important feature was the ability to view a GPS-based Google map of daily travel while completing the diary to assist with recall. The goal of the GPS component was to assist participants with their recall of travel and to verify data collected by the diary.

Particular interest for this study was how to improve the reporting of short walking and cycling trips as individual trips or part of multi-modal trips. One option that was implemented in the diary was to mimic the prompts used in some telephone interviews, where participants are asked further questions if certain types of trips were reported. The survey had 37 participants, where 30 were elected to use the diary with the GPS and 7 elected to use only the diary. Of the total, 33 participants completed all seven days of the diary which accounts for an 89% completion rate. The reaction of the participants to the diary was assessed via the email and telephone enquiries during the survey, the exit surveys and the usage statistics obtained by querying the survey data itself. In terms of the enquiries, 26 were received which makes an average of less than one per participant, regarding access issues, questions on how to complete and GPS-related issues. The exit surveys showed that most of the participants enjoyed completing the survey with only 10% stating that they would not want to do the survey again in 12 months. The negative reaction was due to functionality issues such as browser version, speed and the device used to access the diary. The diary was developed for completion by a PC but usage analysis showed that participants used tablets and smartphones as well.

[41] https://trid.trb.org/view.aspx?id=1289621

3.1.4 Summary

Web-based technologies, in the context of online diaries, provide many advantages compared to paper-based diaries and GPS-based devices for the collection of travel data. Online diaries improve the quality of the data collected, since the data are reported directly from the survey participants and are not inferred as happens with the GPS-based devices. However, the quality of the data is related to the motivation and drive of the participants to provide actual, truthful data. In addition, the task and costs related to survey implementation and distribution is simplified since more and more people own a device that can access the Internet. Identification of non-completed diaries is easier and measures to ensure completion rates can be established.

This technological innovation became possible due to the increased number of households with internet access and the increased number of people with computer knowledge. In addition, the technological advancements with mobile phones and their ability to access the Internet introduced many improvements in the development of web-based diary surveys. As more people use the Internet and as the age profile of users becomes less biased, internet surveys are becoming a more viable data collection option for large and nationally representative samples. Notably, there are still some people, and more specifically the elderly, that don't own a mobile device, do not have access to the Internet or are not confident internet users. For those people, there is a need for alternative or mitigating solutions. Such solutions could be: either retain the paper-based diary or set up phone-desk support, face-to-face support set up at community centres and public libraries or other public places where people feel comfortable going to. However, as seen from Figure 9, nowadays, there is a large percentage of people 65+ that are using a computer daily, so it can be presumed that in the immediate future these numbers will continue to increase.

4 Technology Group 3: Social media/ crowdsourced data

4.1 Twitter platform

Twitter is a social media and microblogging platform that allows users to share thoughts that are in the form of a 140-character message called tweets. Twitter registered users can read and post tweets that can be viewed by 'followers', which is a group of people, normally with similar interests, who choose to receive tweets from another person.

Twitter was created in around 2006 and it immediately gained millions of users using the platform daily. At the end of the first quarter of 2016, the platform averaged at 310 million active users. Figure 10, shows the steady increase of monthly active users worldwide from 2012 to 2018.

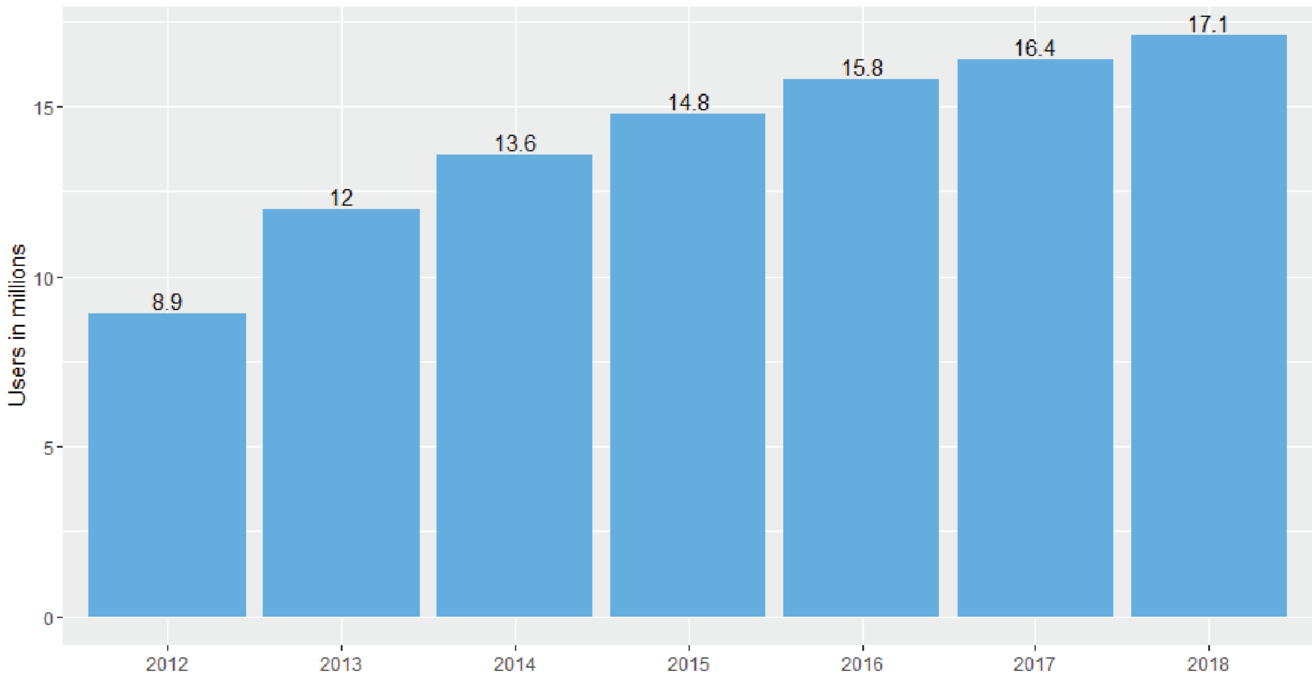


Figure 10: Number of Twitter users in UK from 2012 with a forecast to 2018[42]

[42] http://www.emarketer.com/Article/More-than-One-Fifth-of-UK-Consumers-Use-Twitter/1010623

Twitter has over 15 million active users in the UK and over 65% of these users are under the age of 34. Over 80% of the 15 million active users access the social network from their mobile with a further 29% checking their Twitter feed multiple times during the day[43].

As mentioned above, the main activity of the users is to read and post tweets. These tweets normally consist of text, media such as photos or videos, hash-tags and/or check-in data. Twitter, the company, provides different methods with different policies for accessing the data. They provide two APIs: the Search API which searches against a sample of recent tweets published in the past 7 days[44]; a Streaming API which provides access to Twitter's global stream of data. The latter, however, doesn't provide historical data[45]. Furthermore, if there is a need for a larger volume of data, there is a company called GNIP which is an official Twitter enterprise API platform providing historical and real-time data as well as deep insights into audiences and content[46]. Specifically, they provide live streams which means that when a stream opens, it starts collecting data from that moment on, and, they also provide, historical data from the very beginning of Twitter dating back to 2006. This data is part of a charging service; an estimated pricing plan can be acquired by providing GNIP with a detailed description of intended use of the data. As a point of estimate, their entry level datasets include up to 1 million Tweets over a 40-day period.

Social networking services have changed the way people share, produce and access information. It becomes evident that there is a wealth of information available from social media and with sensible analysis, a lot more information can be uncovered. Social media usage is also proving valuable for disaster or emergency management, aiding faster response and relief operations.

4.1.1 Ability to improve the National Travel Survey

The Twitter platform is a significant source of data and has the potential to provide a large amount of travel-related data needed by the NTS. However, it is better suited as a complementary data source to the current methods of the NTS since the data (geolocated tweets) are quite varied in their user base.

Notably, geolocated tweets are not very popular among users due to privacy reasons but still, the global Twitter traffic generates a considerable amount of data. For the purposes of travel survey methods, the most useful data to work with are the check-in and hash-tag data as they are associated with an event and/or a location. Check-in data which include a location with each tweet can be used to infer an activity, for example tweets linked to some sort of event are often associated with the activity of that event. Hash-tags are

---

43 http://www.thelasthurdle.co.uk/demographics-of-uk-social-media-users/
44 https://dev.twitter.com/rest/public
45 https://dev.twitter.com/streaming/overview
46 https://gnip.com/

associated with activity, event, location, or group. Additionally, activity time and duration can be easily extracted from the above information.

The table in chapter 6.2 explains the ways and capabilities smartphones offer, and how they can be used to capture the key data elements of the survey.

4.1.2 Limitations and costs of Twitter data

The costs that can be associated with this technology are:

Data costs

Depends on different parameters such as the budget available for the data, the time data is needed from, and any rules to filter the data. GNIP, the company that sells Twitter data, provides as entry level datasets up to 1 million tweets for a 40-day period.

Data storage

Depending on the size of the data there might be needed additional storage to process and manage the data. There are some cloud-based services with very reasonable costs.

Technical labour

To derive insights and understanding from Twitter data, a person with skills and knowledge to deal with the data noise is required. To extract information that can be related to mobility and human travel, knowledge of advanced text mining, linguistic and data mining techniques is required.

Big data technologies

Twitter data is often associated with big data because, very often, the volume of the data and more importantly, the variety of the data is too large for common computers to work on. Usually, technologies such as Apache Hadoop, Apache Hive or some other big data processing framework is used to process and analyse data.

Limitations

It's been established that Twitter data can provide large volumes of data for certain time periods, however, the availability of geolocated data, needed for the purposes of a household travel survey, is less than 2% of all tweets[47]. Even though the proportion is too little, the volume is still considerably big enough, considering the amount of such tweets being sent every day in Great Britain. The reason not many tweets share their geolocation is because users are cautious of their privacy. Additionally, there is a big difference in age ranges, it is very likely that the majority of active users will be under 35 years old. This could increase the sample bias and skew the results.

[47] Using geolocated Twitter traces to infer residence and mobility

4.1.3 Practical issues and use in travel surveys worldwide

There have been numerous studies and experiments around the world on the use of social network data to understand mobility and travel patterns. Researchers at the University of New South Wales in Australia worked on a proof of concept on the application of social media to complement or even replace travel diary data. After dealing with the noise and processing the data, the authors recognized the need for three main applications that Twitter data can be used for. These are: in-home activity data which is a challenge to obtain, tour formation which require collecting information about trips forming a tour of activities starting from home and ending at home and forecast potential activities to happen in the future.

In 2015, the Office for National Statistics did research on how to use geolocated Twitter data to infer a user's residence and analyse mobility patterns. The goal of the project was to understand how population and different groups move around the country[47]. To achieve that, they developed methods to cluster individual geolocated Twitter data and then they combined these clusters with additional address data to get information about the location of clusters. Additional analyses were then done using comparable data from the 2011 Census. Some of their interesting findings were: identification of user's residence by clustering geolocated activity traces from Twitter and identification of mobility pattern between local authorities.

Although, there are significant benefits of social media, there are some challenges that need to be overcome, for example, social media trends can be affected by unexpected technological trends and behavioural changes. Interestingly, during the research, it was found that the release of Apple's iOS8 operating system in September 2014 added more features for privacy control which had an impact on the volume of the data collected as well as the socio-demographic characteristics of the users.

4.1.4 Summary

Social media is a rich source of data and information used in many fields of academics and the industry. Twitter data have proven to have potential to be used for analysing travel behaviour and mobility patterns of users. The only caveat is that to build a complete picture of the trips, the user must continuously document their journeys on Twitter which is not always the case. With the right framework, Twitter data can be used for more advanced applications in transport modelling, management, and operational purposes. Although the amount and type of information that can be acquired from Twitter data is very useful and relevant to travel surveys, it's not an indication of whether they can be used as a standalone method of conducting household travel surveys. Literature has shown that Twitter data can be used to complement travel diary data to a greater effect if they are combined with other data sources such as census data.

5 Technology Group 4: Smartcard payments

5.1 Transport smartcard data

Smartcard payment systems have been introduced in transportation systems all around the world and help shape the goal of the multi-modal and seamless travelling experience of users across transport modes. Smartcards look and feel like credit cards, they contain a computer chip that can store and process information. Currently, most of the widely-used smartcard systems consist of a contactless architecture whereby each mode of transport has a card reader where the travellers have to touch in and touch out at the start and/or the end of their journey. The cards have a stored monetary value and by touching the card to the reader, a monetary amount, relative to the trip, is debited from the card.

There are various smartcard schemes around the world. Hong Kong has the Octopus card, which is the second oldest contactless smartcard system in the world, launched in September 1997[48]. Ontario, Canada has the Presto card, Singapore has the EZ-Link card and London has the Oyster card. In the UK, various cities such as Birmingham, Bedford, Cambridge and Manchester are making use of contactless smartcard systems for bus, tram and metro transport networks.

As a by-product of their usage, smartcards collect large volumes of travel data and data about traveller's personal characteristics[49]. This is due to the fact that, every time the card is used to make a trip or renew a stored value, details of the use are recorded. This data can be used to analyse trip rates and derive travel behaviours, both useful information to have when taking decisions on how to improve travelling experience.

5.1.1 Ability to improve the National Travel Survey

Smartcards can provide a variety of data related to spatial, product and cardholder information. It is widely perceived that smartcard data (SCD) can replace sample survey data, but it is popular opinion that their role is better as a complementary data source to household travel surveys. Depending on the way the smartcard systems and the databases are setup, the data attributes that can be captured by smartcards are: cardholder information, such as name; age; gender; trip origin and destination; time and date of the trip, and other information relevant to the product such as date; time of purchase as well as price and location[49].

---

[48] https://en.wikipedia.org/wiki/Octopus_card

[49] Use of public transport smartcard data for understanding travel behavior

5.1.2 Limitations and costs of smartcard data

The costs that can be associated with this type of technology are:

Data acquisition

The acquisition of data from the existing smartcards around UK might be difficult since these smartcard schemes belong to either private operators or city councils. The Swift smartcard in Birmingham, for example, belongs to National Express West Midlands; the OneCard scheme in Leicester is part of the Leicester Council.

Data storage

Depending on the requirements of the project, the need for large volumes of data is quite likely to occur. This creates the need for databases and hardware to store this data for further processing. Cloud-based solutions can again be used here.

Limitations

Important data attributes such as trip purpose cannot be captured by smartcards although, it could potentially be inferred given the time of the day and other variables such as location and frequency of trips. Also, there is a geographical coverage limitation, meaning that mostly, cities with a large public transport network make use of a smartcard system. Finally, trips outside the public transport system, such as walks, cannot be captured either. In addition, a recent trend to use contactless debit or credit cards on public transport has arisen and it has the potential to replace smartcards.

5.1.3 Practical issues and use in travel surveys worldwide

The analysis of trip rates and understanding of travel behaviour depends on the data; the better quality the data is, the better the analysis will be. The quality of data coming from smartcards is often debatable and, as mentioned before, it depends on the architecture of the smartcard system and the scheme implementation they follow. The analysis, for consistency purposes, assumes that all trips made by a user were made with a particular smartcard. This consistency can easily be broken if the card is not used regularly when making a trip (for example, the user tends to forget the card), or if the user alternates between buying paper tickets and using the card. In addition, for a journey to be completely tracked there should be a boarding and departing record. However, buses charge a flat fare for their trips so there is no need to register the exit (touch out) from a bus.

According to the literature, the technology of smartcard data has not been used to conduct household travel surveys. It has only been used to either enhance or complement travel surveys around the world. Spurr et al. have done a research on how smartcard data can be used to evaluate the accuracy of the data collection methods of the Montreal household travel survey[50]. The goal of their research was twofold: firstly, they wanted to determine if the respondent's smartcards could be identified in the transaction database based on their travel behaviour – it appeared that the identification is possible for half of the cases; secondly, they wanted to use the matched smartcards to assess the accuracy of the travel survey responses. Due to the limited number of available survey observations, the second goal couldn't be completed in full, so they were only able to propose a respondent classification. Three types of respondents were recognized: those who respond almost perfectly, those who underreport their travel, and those who report typical, rather than actual travel.

In 2015, ONS did research on the comparison of Oyster card data from tube travel in London against travel flows from the 2011 Census[51]. The goal of the research was to understand the similarities between the data sources and the limitations in using the Oyster data for travel analysis. For the purposes of the analysis they had to assume that journeys from the Oyster card that start at peak time, around 7am – 10am on weekdays, are journeys made by workers. The analysis concentrates on the distribution flows between an origin local authority (LA) and a destination local authority. Furthermore, they also analysed patterns for Middle Super Output Areas (MSOAs) which are smaller geographies than LAs. From this analysis, several factors that affect the usefulness of Oyster card data were uncovered, these factors are: distorted flows in areas with mainline train stations, there is also some distortion in cases where other transport modes such as cars and over-ground trains are available. Additionally, another factor that concerns the usefulness of Oyster card data is tourists and visitors. These factors cause Oyster card data to be more skewed than census data with a lower number of small flows and a higher number of large flows.

5.1.4 Summary

As seen SCD have great potential as a technology for collecting travel activity data. The main problems that the current methods of the NTS and the other technologies such as smartphones and GPS are trying to solve are: respondent burden, accuracy of data, data access. Smartcards have numerous advantages on all the above. They reduce respondent burden since they require no additional effort from the respondent other than validating the fare and they impose no restrictions on equipment or battery life. Additionally, the cost of conducting the survey and getting access to this data is considerably lower than the current methods. Getting access to the data depends on the architecture of the system, but it could easily be a matter of obtaining the card ID and use it to access the data from a database, or even using a card reader to scan the cards and store the data in a portable storage.

---

[50] A smartcard transaction "travel diary" to assess the accuracy of the Montréal household travel survey

[51] Comparing travel flow between 2011 Census and Oyster card data

Admittedly, a restriction of using SCD is that it only applies to public transport infrastructure and it doesn't provide more granular information about other types of travel such as cars, bicycles and taxis. Adding to that, SCD doesn't provide the trip purpose either. Although it can be approximately inferred, it misses small trips such as the walk to the bus stop and people forgetting the smartcard and have got to buy a paper ticket. Additionally, there is a growing trend, mostly in London, where commuters can use the public transport with their contactless bank card. These sort of restrictions indicate that SCD can only play a complementary role in travel surveys and cannot stand alone in the implementation of household travel surveys.

6 Summary of findings

6.1 Individual technology highlights

The following tables provide some main advantages and disadvantages of each technology as they emerged from the desk-based research.

Smartphones	
Advantages	**Disadvantages**
• The majority of the population aged 16+ own a device • Smartphones are widely used in everyday life • Can target a large sample size of the population for the survey • Additional data can be inferred if smartphone collected data combined with other data sources and technologies • Data can be collected passively without introducing respondent burden	• There is still a significant section of the population, mostly among the elderly who don't own a device • Can't provide some important data fields such as trip purpose, methods of travel, number of boardings etc. • There might be a need to provide smartphones on loan to survey participants • Battery life is an issue that needs careful consideration • Network coverage • Privacy concerns (some people won't want their location tracked) • Age of smartphone – can it support the app?

GPS Devices	
Advantages	**Disadvantages**
• Reduce respondent burden by collecting data passively • Can provide more data than just location and time only. These types of data are: date; time; latitude; longitude; altitude; speed and heading	• Survey managers have to provide the equipment (GPS devices) • Survey participants have to make sure their devices are always charged and they may be required to carry a power supply with them • GPS can't provide information regarding journey purpose, methods of travel, number of boardings etc. Some of this information can be inferred relatively accurately with the use of GIS database or additional data sources • GPS data are in some cases questionable (urban environments, "cold start" of the device) so there might be a need for an approach where participants have to use travel diaries in parallel with GPS or they have to validate the data collected from GPS in a prompted recall fashion

Mobile Network data	
Advantages	**Disadvantages**
• Large volume and variety of data about national or regional requirements • Can be used to create origin and destination matrices. This technique can be used to recreate user journeys • Effort to acquire the data is minimal since there is no need for recruiting participants or running expensive and time consuming surveys	• Have to be acquired by mobile network operators which requires for a contract to be signed between the involved parties • Fusion with other data sources might be needed • Trip purpose and mode of transport are not given but could be obtained if mobile network operators are able to infer it and are able to share it • Short trips cannot be identified very well • Information such as trip purpose, mode of transport and short trips depend on the cell resolution MNOs can offer. Due to network coverage issues and privacy reasons MNOs are comfortable sharing information at the MSOA resolution • Personal information such as age and gender will be missing

Web-based travel diary surveys	
Advantages	**Disadvantages**
• Web-based surveys are not limited to computers anymore; smartphones and tablets can be used for the completion of the survey • Nowadays, more than half of the population has an internet connection and owns an internet enabled device which means that a larger sample size can be targeted • All the NTS travel diary data can be captured • Features such as autocomplete functionality, visual effects e.g. maps, machine learning algorithms for pattern recognition and classification can all aid respondents when completing the survey • Reminders can be sent via email and text message to participants to prompt them to fill the survey	• The quality of the data is based on the respondent's memory, on their ability to either be consistent in tracking every trip they make or recall trips they have made. • Depending on the amount of days the survey is running for, there is a risk of respondent burden which might affect the completeness of the survey • There is still a significant portion of the population, such as the elderly, who don't have an internet connection or internet access

Twitter data	
Advantages	**Disadvantages**
• The Twitter platform can provide a large volume and variety of geolocated data (tweets) • The cost of obtaining this • data is very minimal since the Twitter provide APIs that can help do that job • There are no restricting privacy issues on the data • Can be combined with other data to infer additional information	• Twitter data can only be used to complement the current methods of the NTS • Further information such as trip purpose, mode of transport etc. are not given but potentially could be extracted • An advanced skillset and technologies are needed to process and analyse the data • There might be a bias affecting the sample size since Twitter is not used by all ages and geolocated tweets are not the norm

Smartcard data	
Advantages	**Disadvantages**
• They provide a large volume of actual travel related data • No major effort is required to collect the data, there is no need to recruit participants or conduct surveys • Respondent burden is reduced, cardholders/participants only have to validate the fare • There is no need to supply equipment • There are no technical restrictions e.g. battery life	• Data are limited to journeys done using public transport • Short trips (walking), car and taxi trips are not identified • Can be a complementary data source to the NTS. They cannot replace the current methods • Acquiring the data is not an easy process since most of the smartcard systems belong to private companies • Trip purpose and other metadata are not provided by smartcard data but can be inferred with the right process

6.2 Comparison of all technologies to provide NTS elements

The following table summarizes the ways and techniques each technology can provide the data needed for the NTS. Web-based surveys are not mentioned in the table since the participant is providing the data needed.

Travel Data Elements (as presented in the current methods)	Smartphones	GPS	Mobile Network Data	Twitter data	Smartcard data (limited to public transport)
Purpose of journey	Can be inferred but is imperfect. Specifically, it can be approximated with the use of predefined deterministic procedures or rules. For example, combining the data with a GIS database and applying heuristic rules to trip start and end times, speed, duration of stay etc. can potentially uncover home or work location. Other research has shown that this approach could infer journeys with purpose of physical activity like cycling or running session[52].	Not directly but it could be inferred, if certain highly visited locations are known, in combination with a GIS database. The locations could be identified following a heuristic approach	No. Home to work trips could potentially be inferred but cannot be validated	It can be inferred from the tweet text (if it is mentioned)	No. Home to work trips could potentially be inferred but cannot be validated

[52] UbiActive: A Smartphone-Based Tool for Trip Detection and Travel-Related Physical Activity Assessment

Travel Data Elements (as presented in the current methods)	Smartphones	GPS	Mobile Network Data	Twitter data	Smartcard data (limited to public transport)
Time left	Can be inferred using location sensors and accelerometer sensors to monitor movement and significant change in location	It can be inferred with the right pre-processing. For example, from the gaps in the time feeds or the density of the GPS locations	Yes, it can be inferred from the timestamp of the communication of the phone with the cell tower	Yes, it can be found from the time tag of each tweet	Cannot know time left from a place e.g. time left home to walk to bus stop
Time arrived	Can be inferred using location sensors and accelerometer sensors to monitor movement and significant change in location	It can be inferred with the right pre-processing. For example, from the gaps in the time feeds or the density of the GPS locations	Yes, it can be inferred from the timestamp of the communication of the phone with the cell tower	Not directly, unless it is specified in a tweet	Cannot know actual time arrived to a place e.g. no record time got off bus to walk to work
Origin – where the journey started (from village/ town/local area)	Can be calculated using the location based sensors such as Wi-Fi signals, GPS and GSM networks	Yes, can be extracted very easily	Needs processing to derive it	If people tweet frequently and share their locations. The tweets can act as a GPS tracker[53]	Large areas or areas with popular or central stations can be acquired

[53] Utilizing Location Based Social Media in Travel Survey Methods: bringing Twitter data into the play

Travel Data Elements (as presented in the current methods)	Smartphones	GPS	Mobile Network Data	Twitter data	Smartcard data (limited to public transport)
Destination – where the journey ended (from village/town/ local area)	Can be calculated using the location based sensors such as Wi-Fi signals, GPS and GSM networks	Yes, can be extracted very easily	Needs processing to derive it	If people tweet frequently and share their locations. The tweets can act as a GPS tracker	Large areas or areas with popular or central stations can be acquired
Method of travel (car, train, bus, bike) (only walks that were more than one mile, or took more than 20 minutes are included)	Can be inferred from GPS and accelerometer data in combination with rules that consider average, minimum and maximum speeds, along with additional information such as car or bike availability in the household. These methods can also be complemented with GIS data to distinguish between modes that have similar movement characteristics[54]	Not directly but it could be inferred by speed, heading and time	Depends on the area resolution. With data, such as speed, spatial pattern of events and clustering pattern of events mode could be classified as motorized or non-motorized.	If, specified in the text of tweets or can be inferred by the location	Yes, if it's a public transport mode and use a smartcard system

54 Transportation Mode Detection using Mobile Phones and GIS Information

Travel Data Elements (as presented in the current methods)	**Smartphones**	**GPS**	**Mobile Network Data**	**Twitter data**	**Smartcard data (limited to public transport)**
Distance travelled (miles or meters)	Can be calculated based on location changes fused with GIS information to get accurate network distance.	Can be calculated based on location changes	Yes, can be calculated based on location changes	Can be inferred based on time travelled and traffic conditions mentioned in the tweet text	Only distance travelled with public transport
Time spent travelling (minutes)	Can be calculated and inferred based on location and accurate mode detection	Can be calculated and inferred based on location, speed and heading	Yes, it can be inferred speed, spatial pattern and distance	Yes, could be found from the tweets and their time stamp	Only time spend travelling with public transport
Number in party (split into adults and children)	No	No	No	No, unless specified in the tweet text or combined with other data sources	No
Which car, motorcycle etc. used (only if journey was made not by public transport, but by car, motorcycle etc.)	No	No	No	No, unless specified in the tweet text or combined with other data sources	No

Travel Data Elements (as presented in the current methods)	Smartphones	GPS	Mobile Network Data	Twitter data	Smartcard data (limited to public transport)
Driver or passenger?	No	No	No	No, unless specified in the tweet text or combined with other data sources	No
Parking cost	No	No	No	No, unless specified in the tweet text or combined with other data sources	No
Ticket type (single/ return/ travel card etc.)	No	No	No	No, unless specified in the tweet text or combined with other data sources	Yes, if the smartcard architecture captures this type of data
Ticket cost	No	No	No	No, unless specified in the tweet text or combined with other data sources	Yes
Number of boardings (the number of trains, buses etc. used to reach journey destination)	Could be inferred indirectly using GPS and accelerometer data.	No	It could possibly be inferred from speed, spatial pattern and dwell time in between trips	No, unless specified in the tweet text	Yes

Keys

- Represents the fields that can be acquired without pre-processing
- Represents fields that need pre-processing
- Represents fields that cannot be acquired or are hard to get

7 Conclusion and recommendations for future work/next steps

Household travel surveys are firmly founded in the early stages of urban transport planning since they provide insights into travel behaviour and reinforce decision making. The purpose of these surveys is to provide accurate and representative data of people in a city, region or nation. To achieve this, it is important that the survey remains 'future proof', and the potential of technologies to improve, supplement, or replace current data collection methods is examined in order to ensure that a travel survey uses methods which are modern, efficient, and acceptable to respondents.

Paper and pencil interview surveys are still the pillar of household travel surveys around the world. However, there is a slow wave of change on existing and emerging technologies that can be used as alternative survey methods that have the possibility to improve the accuracy and completeness of the data required. This report has researched technologies and data sources that have potential to collect and enhance data from the National Travel Survey that is conducted every year in England. More specifically, the technologies examined were: GPS-based devices such as smartphones and dedicated GPS receivers, web-based travel diary surveys, and data sources such as mobile network data, social media data and smartcard data.

A literature review on GPS based devices showed that GPS receivers have already established their ability to collect good quality data but, nowadays, smartphone assisted travel surveys are slowly increasing in importance in collecting data on people's activities. Some technology highlights and comparisons that occurred from the desk based research are as follows:

- Many smartphone devices have integrated GPS technology similar to GPS receivers.
- Smartphones are a ubiquitous device and there is no immediate need for the survey managers to provide equipment to participants. Contrary to GPS receivers which is not a very popular device to own.
- Smartphones can support the development of applications for the collection of travel activity data. Also, smartphones have an appropriate sized screen and an interface which can facilitate prompted recall surveys on the device speeding up the collection of insights and decision making.
- In order to get data from GPS receivers, the devices will have to be collected and data collection software used.
- Smartphones, depending on the model and manufacturer, come with a variety of sensors such as accelerometers, gyroscopes and Wi-Fi sensors that can provide additional data.

- Smartphones can minimise the risk of data loss and speed up the collection of insights since they can upload data to online databases whenever they detect a Wi-Fi signal.
- Representativeness issues are eminent to both technologies. People who are slow adapters of a technology or have no need for technology (a large number of the elderly population, for example) induce the risk of lower data quality and possible sample biases.

The data sources studied can be categorized as big data sources since the data comes in large volume and variety and would require advanced technologies and processing. Some conclusions from these chapters are:

- The acquisition of this data is an elaborate process which involves close collaboration with private companies and organizations.
- To acquire mobile network data, the interested party needs to communicate with mobile network operators and develop a data specification of datasets and information that can be obtained.
- A similar approach must be followed for the smartcard data. Interested parties need to communicate with local councils and private transport operators in Great Britain. Twitter data can be acquired, with minimal cost, either by using the APIs the platform provides or they can be purchased from enterprising companies (GNIP).
- There is no standard cost to obtain such data, a relative estimate can be obtained only by talking with the providers to scope the availability and range of data.
- These data sources cannot replace the current methods of the NTS, they can be used to validate and potentially enhance some of the assumptions and findings that occur.

Finally, web-based travel diary surveys continue to remain a prominent tool in transport related studies. Although, they still require the participant's engagement, which could impose completeness burden and non-response bias, if measures are taken to counter that, they have high potential to collect good quality data. Additionally, the cost of conducting the survey is minimal since all the equipment that is needed is either a computer or a smartphone which currently are both ubiquitous devices.

The recommendations that arise from the overall desk based research:

- GPS-based devices (smartphones and dedicated GPS devices) can replace the current paper travel diary of the NTS but only to an extent. There is still the need for the participant's interaction to correct and validate their journeys.
- Discussions with researchers suggest that smartphones in 5 years' time will be suitable to replace the current paper travel diary.
- Given the different population types (elderly people, non-tech savvy people), a mixed mode data collection approach might be of help.
- A helpline or help desk should exist which participants could then contact for technical problems or survey related issues. The data sources that were researched in this report can only be used to validate assumptions and findings from the survey

8 Appendix A

8.1 Glossary

Term	Meaning
TSC	Transport Systems Catapult
DfT	Department for Transport
NTS	National Travel Survey
NatCen	NatCen Social Research
HTS	Household Travel Survey
FMS	Future Mobility Survey
GIS	Geographic Information System
MND	Mobile Network Data
HITS	Household Interview Travel Survey
SCD	Smart Card Data

9 Appendix B

9.1 Samsung devices with sensors and GPS modules

Phone	Sensors	GPS
Samsung Galaxy On7 Pro	Accelerometer, proximity	Yes, with A-GPS, GLONASS/ BDS (region dependent)
Samsung Galaxy On5 Pro	Accelerometer, proximity	Yes, with A-GPS, GLONASS
Samsung Galaxy J Max	Accelerometer	Yes, with A-GPS, GLONASS
Samsung Galaxy J2 (2016)	Accelerometer, proximity	Yes, with A-GPS, GLONASS
Samsung Z3 Corporate Edition	Accelerometer, proximity	Yes, with A-GPS, GLONASS
Samsung Galaxy S7 active	Fingerprint, accelerometer, gyro, proximity, compass, barometer, heart rate, SpO2	Yes, with A-GPS, GLONASS
Samsung Galaxy J3 Pro	Accelerometer, proximity	Yes, with A-GPS, GLONASS/ BDS (region dependent)
Samsung Galaxy C5	Fingerprint, accelerometer, proximity, compass	Yes, with A-GPS, GLONASS/ BDS (region dependent)
Samsung Galaxy A9 Pro (2016)	Fingerprint, accelerometer, gyro, proximity, compass	Yes, with A-GPS, GLONASS/ BDS (market dependent)
Samsung Galaxy J7 (2016)	Accelerometer, proximity	Yes, with A-GPS, GLONASS/ BDS (region dependent)
Samsung Galaxy J5 (2016)	Accelerometer, proximity	Yes, with A-GPS, GLONASS/ BDS (market dependent)
Samsung Galaxy Tab A 7.0 (2016)	Accelerometer	Yes, with A-GPS, GLONASS
Samsung Galaxy S7	Fingerprint, accelerometer, gyro, proximity, compass, barometer, heart rate, SpO2	Yes, with A-GPS, GLONASS, BDS
Samsung Galaxy J1 Nxt	Accelerometer	Yes, with A-GPS, GLONASS
Samsung Gear S2 classic 3G	Accelerometer, gyro, heart rate, barometer	No

Samsung Galaxy J1 (2016)	Accelerometer, proximity	Yes, with A-GPS, GLONASS
Samsung Galaxy A9 (2016)	Fingerprint, accelerometer, gyro, proximity, compass	Yes, with A-GPS, GLONASS/ BDS (market dependent)
Samsung Galaxy A7 (2016)	Fingerprint, accelerometer, proximity, compass	Yes, with A-GPS, GLONASS/ BDS (market dependent)
Samsung Galaxy A5 (2016)	Fingerprint, accelerometer, proximity, compass	Yes, with A-GPS, GLONASS/ BDS (market dependent)

9.2 Smartphone App development companies in UK

Company	Pricing
Intellectsoft	$40 - $75/hr
hedgehog lab	$78 - $112/hr
Ready4S	$19 - $37/hr
nomtek	$38 - $75/hr
Appinstitute	$78 - $112/hr
The Distance | App Developers	$38 - $75/hr

9.3 Characteristics of App based mobile surveys[39]

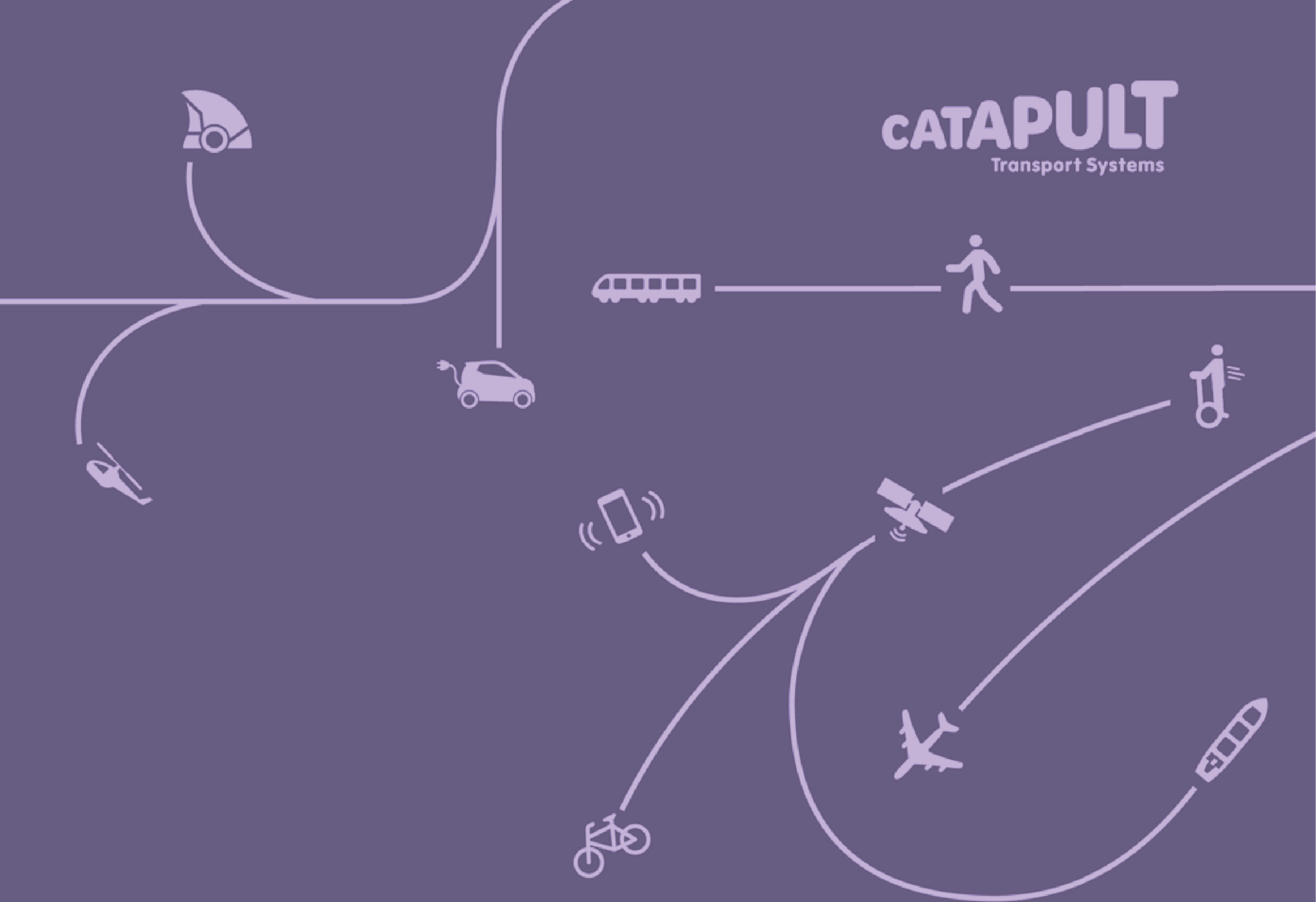
Advantages	Disadvantages
• The survey does not need a permanent stable internet connection at all times and the app can be designed to allow data to be collected offline until communications are resumed. • An app can be designed to exploit all the programmable capabilities of the device, which includes alarms, capture and upload of pictures and video, GPS or cell-based geolocation tagging and recognizing bar-codes through the devices built-in camera • The survey designer can have greater certainty about how the survey will be delivered and appear on the range of devices supported by the app.	• The App must be downloaded to the device prior to the start of the survey. • Apps are specific to the device, or the operating system on the device. It is difficult to develop apps that cover the entire range of devices, and these are constantly changing. • It may be more difficult to deploy the same survey to both mobile and standard browsers, unless the survey platform supports both HTML and App modes.

9.4 Characteristics of browser based mobile surveys[39]

Advantages	Disadvantages
• Uses standard HTML to achieve high coverage by not being restricted to a subset of devices. • HTML also facilitates multiple formats within one survey instrument so that a mixed mode approach can be taken, with the same survey instrument working on desktop and laptop PCs, tablet PCs as well as mobile devices. • Participant can engage immediately without having to locate and download a specific application on to their device. • Avoids having to exclude mobile devices from conventional web surveys.	• If using a standard or mixed-mode web survey tool, a stringent, lowest common denominator approach to survey design must be followed. • Web survey tools may, in some places, rely on flash, or java, which the device may not support, or on the use of mouse buttons (e.g. to achieve a shift-click), which the device does not have. Hard to take advantage of other useful capabilities of the mobile device, such as alarms, geolocation finding, taking photographs or scanning barcodes. • A stable internet connection is needed throughout the period of the survey.

9.5 Smartphone based travel surveys around the world

Location	Survey purpose	Sample size	Collection period	Technical details	Financial incentive	Costs
Singaporean Household Travel Survey (Future Mobility Survey)	Collect spatial and temporal travel data for transport modelling purposes	74 people initially	14 days	Prompted recall approach followed; smartphone devices were provided	SG$30 (US$25)	Device cost, web hosting
Sydney, Australia	Investigate changes in travel and health outcomes before and after the construction of cycling infrastructure	63 people	7 days	Prompted recall approach followed;	$50/year for participation plus 15/year if they used the app	Web hosting account; labour hours
New Zealand	Smartphone based travel survey system was developed and tested as part of the national household travel survey of New Zealand	77 participants	3 days	Prompted recall approach followed; three servers for data storage, data management and visualization and presentation were used	Not known	Not known
Netherlands	Smartphone based travel survey developed for the Dutch Mobile Mobility Panel	600 smartphone and non-smartphone users	2 weeks	Prompted recall approach followed; smartphone devices were provided	Not known	Device cost, web hosting



Transport Systems Catapult

The Pinnacle, 170 Midsummer Boulevard, Milton Keynes, MK9 1BP

Tel: 01908 359 999

www.ts.catapult.org.uk
linkedin.com/company/transport-systems-catapult
Twitter: @TSCatapult