Office for National Statistics

OFFICIAL - SENSITIVE

ONS Internal Report

Using mobile phone data within commuting flows

Susan Williams

July 2016

Contents

OFFICIAL - SENSITIVE

OFFICIAL - SENSITIVE

Executive Summary

Around the world, statistical organisations such as the UK's Office for National Statistics (ONS) are actively seeking to assess the potential of using mobile phone data for the production of official statistics on population and mobility. This report provides an overview of the ONS investigation to date on the potential of using mobile phone data with a focus on the application to flows of workers, from home to work location, as customarily produced every ten years in the UK population census. This investigation has involved engagement with the three main Mobile Network Operators (MNOs)[1] currently operating in the UK and public transport bodies who have a similar interest in using the data to estimate transport flows.

Mobile phones generate digital information and send it to the MNO to which the phone is subscribed. These data are used for billing and other operational processing although, more recently, there is additional interest in using these data to develop statistical products. Geo-location data have attracted the most attention as mobile phones offer almost complete coverage of the adult population and, given the custom of carrying mobile devices on the person, can be used to produce information on population densities and mobility. The availability of data in close to real time and for small geographies brings advantages for a number of applications such as in crowd monitoring, retail location planning, tourism and transport flows etc.

Highlights

Privacy and Ethical Issues

Usage of mobile phone data may raise privacy concerns around access to, or use of, personal data.

These concerns may include informed consent as well as potential disclosure risk and identification of individuals arising from unique patterns of movement.

The Data Protection Act 1998 is currently the main legislation covering the processing and handling of personal data although ONS is also subject to obligations set out in the Statistics and Registration Service Act 2007.

Public opinion is important and people must have confidence in the

[1] [redacted]

OFFICIAL - SENSITIVE

safeguards put in place by both ONS and MNOs.

Data Format

Two types of geo-location data are generated from mobile telephony: active events which occur when making or taking calls or text, and passive location updates which take place when a mobile connects to different cell tower or is periodically 'pinged' by the MNO to ascertain its whereabouts.

There are three primary MNOs with a large enough subscriber base to support or enable the production of population statistics, and they have all set up operations to offer such services.

The data are available in various aggregated and anonymised forms, each being suitable for different applications, raising different privacy concerns and requiring different IT infrastructure and processing capabilities.

Data Access

Aside from privacy and ethical concerns, there are data access problems associated with procurement arrangements and the associated costs.

On the other hand, there are examples of data being made freely available for research [redacted]

Use of the Data

Over the past two years public bodies, most notably transport bodies, have conducted a number of pilot studies using mobile phone data to produce estimates of transport flows.

Transport bodies need substantial budgets to carry out roadside surveys to facilitate their work and are seeking alternative data to reduce or even replace such surveys in view of their ever increasing costs and declining response rates.

Procurement

To date, procurement of mobile phone data for use in producing statistical products by public sector organisations has been uncoordinated.

Transport bodies seek similar data for input into their transport models. Different bodies have separately procured transport flows for the same geographic areas and time periods, albeit using

OFFICIAL - SENSITIVE

different MNOs and methods.

The Transport Appraisal and Strategic Modelling (TASM) team in the Department for Transport (DfT) is the main source for guidance on collecting evidence to inform transport initiatives and plans to update this advice to cover the use of mobile phone data during 2016. However, DfT does not have oversight of all public transport bodies operating in the UK and therefore coordination is difficult.

At the same time better co-ordination could facilitate the use of other public sector applications with the potential to support epidemiological work, tourism and the derivation of population densities.

Methodology

Although transport bodies have commissioned pilots using mobile phone data, there is still uncertainty around modelling methods and the quality of the estimates.

This uncertainty arises from a lack of transparency on the methods used, a feature exacerbated by the chain of organisations that may be involved in producing the final modelled estimates.

Different organisations have developed their own modelling techniques and are reluctant to share their knowledge and expertise because of commercial sensitivities. These problems are compounded by statistical issues including questions of bias and differences in geographical coverage.

ONS has set up two work streams to take forward its research into the potential of using mobile phone data for the production of official statistics.

The first will focus on obtaining access to commuting flows derived from mobile phone data to understand the quality issues and to develop recommendations on the use of the data for statistical and research purposes. This work stream is likely to require collaboration with transport bodies that have already purchased modelled mobile phone data.

The second will seek to influence the cross Government approach to using mobile phone data for statistical purposes. This will involve extensive engagement and encouragement of collaboration with government bodies including the GSS, MNOs and other organisations who may wish to acquire such data.

OFFICIAL - SENSITIVE

1. Introduction

This report outlines the work to date on the use of mobile phone data to produce statistics to inform decision-making for public policy and planning. The emphasis is on the use of mobile phone location data within commuting flows, although a number of other applications are also highlighted.

The report starts with a brief background to the work followed by a description of mobile phone data and the statistical applications of most interest within official statistics. The mobile phone market structure is then considered together with the factors that might impact on its stability. There follows a section on the experiences of organisations who have already conducted research using mobile phone data. Finally a discussion of the main methodological questions is made before a conclusion and next steps in this work for ONS.

In addition to the Bibliography there are four appendices provide more detailed information on a number of topics, namely:

Appendix A: Mobile phone data

Appendix B: Formats of mobile phone data with pros and cons

Appendix C: Market composition (as at Aug 2015)

Appendix D: Data access

2. Background

2.1. Official Statistics

The Statistics and Registration Service Act 2007 defines 'official statistics' as all those statistical outputs produced by the UK Statistics Authority's executive office (ONS, by central Government departments and agencies, by the devolved administrations in Northern Ireland, Scotland and Wales, and by other Crown bodies (over 200 in total). Secondary legislation may further allow other statistical outputs to be eligible for official statistics status.

Traditionally, official statistics have been based on data collected through censuses, surveys and, more recently, the operational data collected by government for the administration of services such as welfare, taxation and provision of health services. High quality standards are maintained through the application of the principles and procedures stipulated in the Code of Practice for Official Statistics.

The Government wishes to increase the role of administrative data for the production of official statistics as it considers this will help to create new and more timely information and

OFFICIAL - SENSITIVE

reduce costs Information generated from mobile phones is one of the sources of interest to producers of official statistics.

2.2. Mobile Phone data

Statistics collected by the UK communications regulator Ofcom show that at the end of 2013 there were just over 83 million mobile phone subscriptions in the UK with 93% of adults owning such a device. These subscriptions include contracted and pay as you go customers and cover both business and personal mobiles. In many cases people have more than one mobile phone.

When mobile phones are switched on they generate digital information which is transmitted to the mobile network operator (MNO)[2]. This information includes details of any calls or texts made or received and also passive location updates. The growth of smart phones/devices capable of connecting to the Internet has further increased the data generated. This coupled with new generation mobile technologies, large increases in computer power and the availability of analytical software means that MNOs are increasingly interested in using their data holdings to create data products of commercial value.

As the norm is for mobile phones to be carried by their users, there is great interest in the use of mobile phone location data to understand population densities and how people move around. The attraction of mobile phone data is that it is potentially available in close to real time and for small areas. As a result it offers the basis for developing a much richer understanding of population dynamics.

Around the world, statistical organisations such as ONS are interested in the potential of mobile phone location data for the production of official statistics on the population. Such data might be used to help quality assure or even replace official estimates, provide more timely indications of population change or support the development of new indicators on population densities at specific times of day.

Over the past year, ONS has conducted research into the potential of using mobile phone location data to estimate flows of workers from home location to work location as customarily produced in the decennial Census. This research has involved a literature review and extensive engagement with MNOs and transport bodies who are keen to use mobile phone data in a similar way for transport planning.

[2] [redacted]

OFFICIAL-SENSITIVE

Statisticians in the Department for Transport (DfT) have helped to direct this research and will continue to be involved in future statistical work on using mobile phone data for transport applications.

3. Mobile Phone data

Mobile phones connect via radio waves to a nearby cell tower or base station that covers a specific geographic area called a cell[3]. Groups of cells form larger geographies called location areas[4].

Two types of location information are generated from mobile phones:

Call Detail Records (CDR): generated when the phone is active. This occurs when taking or placing a phone call or mobile message or using the Internet on a smartphone. In these situations, the time and the cell id of the cell tower forming the mobile connection is recorded.

Network data: When the phone is switched on but in idle mode there are two ways of generating location data. The first is as a result of periodic location updates, known as 'pinging', sent by the phone to the network; the frequency and location information of these updates is dependent on the generation of mobile phone technology used.

The second form of network location data is mobility-driven: network communication with the mobile phone is transferred to a different cell tower when it moves into that cell tower's cell area. For 4G technology, the new cell id is recorded as the mobile switches connection. For older 2G and 3G technologies, a new location update is recorded only when the mobile moves between location areas.

Appendix A gives more detail on locational mobile phone data and discusses the different formats available.

Another common form of location information is generated via Global Positioning System (GPS) technologies that use satellites to pinpoint highly accurate locations, particularly in outdoor settings where there is no obstruction to the sky. These data are generated on internet accessible devices such as smartphones and tablets, typically when users make a

[3] Cell areas vary in size. In urban areas, where cell towers are densely situated, they may have a range of 300 to 400 metres. In rural areas, cell tower density is very sparse and a cell may have a range of 5km or more.

[4] Location areas can also vary in size. Some can be very large.

OFFICIAL - SENSITIVE

request to know their location via an application requiring GPS capability. The MNO does not have access to this GPS location unless the user is using an app belonging to the MNO. Companies collecting these GPS data include [redacted]

The bias in GPS data is considered to be far greater than cell-tower information as only wealthier individuals are likely to be able to afford internet enabled devices. Moreover, GPS location is only intermittently collected when a call for location is requested and the user has given permission for their location to be recorded. Using GPS uses a lot of battery power which further restricts regular use. Cell tower data are more readily available and cover a larger proportion of the population and so provide more comprehensive location traces.

4. Statistical Applications

Mobile phone location data, as sourced from the network or CDR, has attracted research interest in a number of statistical applications, the main ones being the potential of this data to inform on population density and mobility.

Population density

The ONS has authority to produce accurate population estimates which it does every 10 years through the population census. These estimates are updated every year in-between censuses by adjusting for births, deaths and migration.

As mobile phones are owned by such a high proportion of the population across the country, counting the number of connections to a single cell tower could theoretically have a relationship to the true number of people in the cell area associated with that cell tower. To estimate population totals, adjustments are needed to correct for the market share held by the MNO and for other factors affecting bias.

Research by Deville et al (2014) argues that counts of mobile phones connected to cell towers may provide a credible approach to distributing national population totals to sub regions. This assertion would need to be tested further.

Mobile phone data also present an opportunity to produce timely estimates of populations according to new and flexible definitions. This might include estimates by time of day, day of week or by season.

Mobility analysis

Individual level locational data allows analysis of movement patterns for individuals which when aggregated may provide a basis for understanding population movements. This is vital for transport planning as such information, along with speed, transportation mode and

OFFICIAL - SENSITIVE

journey purpose, is used to build traffic models. The large transport consultancies have carried out a considerable amount of research on the use of mobile phone data and MNOs are actively selling estimates of transport flows to transport planning bodies. ONS is interested in understanding the relationship of such estimates with home-work flows from the Census as mobile phone data might lead to more frequent estimates.

Mobility analysis is applicable in epidemiology, for example, to monitor the movement of people during an outbreak of contagious disease. Understanding how and where people move can help to target services to areas of greatest need. Another mobility application is in tourism statistics where such data may be used to estimate how many people visit attractions and where have they come from.

Other applications

Depending on the format of the data, other applications for mobile phone location data include intelligence on land use and the sphere of connectivity exerted within and across areas. At an individual level, CDR can also inform on social networks.

Applications for other types of mobile phone data

Although the main focus has been on applications using location based data, other data held by the MNO are of relevance for research purposes including identifying cross-border migration. This application is of great interest in countries who do not operate border crossing checks such as in the 26 European countries contained in the Schengen Area. Estonia have been using mobile phone data in this way to estimate various tourism inputs to their National Accounts.

MNOs have access to their customers' data including details on age, gender, home and work information as well as language settings.

5. Market composition and data access (as at Aug 2015)

[redacted]

OFFICIAL-SENSITIVE

5.1. Market Composition

[redacted]

	[redacted]	[redacted]
	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]

[redacted]

[redacted]

OFFICIAL - SENSITIVE

[redacted]

[redacted]

[redacted]

5.2. Market stability and future developments

It is important to note that the market composition will not remain constant over time. MNOs are highly competitive and they strive to gain and retain customers through initiatives offering different services and pricing schemes.

Mergers and acquisitions between MNOs are also common, leading to the creation of new companies with different strategic objectives and the demise of older companies. This activity is reinforced by the arrival of alternative communication technologies, such as wi-fi, which might provide further competition for existing MNOs.

Factors that may affect the stability of future market composition are discussed in more detail in Appendix C.

5.3. Data Access

International experience shows that accessing mobile phone location data is very difficult in view of the amount of data involved, concerns about privacy and the cost implications. These issues are considered in more detail in Appendix D.

5.4. Privacy

The Data Protection Act 1998 provides the statutory framework for the processing and handling of personal data in the United Kingdom. In addition to safeguarding confidentiality, special attention must be given to issues of consent. More detailed consideration of these and other related ethical issues are outlined in Appendix D.

OFFICIAL-SENSITIVE

ONS presented proposals to source aggregated data relating to mobile phone generated commuting flows to the GDS Privacy and Consumer Advisory Group and the ONS Beyond 2011 Privacy Advisory Group. Despite some reservations both groups recognised the potential benefits associated with making use of such data for statistical purposes

5.5. Chains of data provision

There are three main types of organisation that may be involved in procuring transport flows derived from mobile phone data. The first is the MNO itself, who has primary access to the data. The second may be a company specialising in the processing of large volumes of data who forms a partnership for data access with the MNO. The third may be a transport consultancy who provides specialist advice or guidance on the transport industry leading to the development of appropriate methodology. Further details about the roles of such organisations and the need for appropriate IT infrastructure for the safe and secure processing of any data are set out in Appendix D.

5.6. Commercial Interest

In view of the challenges around data access and privacy it might be preferable for the MNO or partnering data analytics company to undertake the research rather than provide access to the data.

However, the high charge out costs and the amount of time required to process the data to a suitable specification mean that purchasing data may be prohibitively expensive. Very small pilots studies, for areas covering a single Local Authority say, typically command charges of between £50K and £100K.

As transport bodies have traditionally used expensive roadside surveys to collect evidence for transport schemes the costs associated with the purchase of such data could still be cost effective. In other cases, such as work on official statistics and epidemiology, this might not be viable.

5.7. Research data

On a more positive note, there are precedents for making CDR data freely available for research, both at aggregated and anonymised individual level. Academic studies have managed to source this data from individual MNOs. The global telecommunications operator Orange has made millions of anonymised CDRs from Senegal and Ivory Coast available for research for many years, under its Data for Development Initiative.

OFFICIAL - SENSITIVE

6. Experiences of using locational mobile phone data

Various organisations have worked with mobile phone location data, mostly in a research setting. This section outlines some of the work that has been undertaken by public bodies, international statistical organisations and academic institutions.

6.1. UK public bodies

Public transport bodies have pioneered research on the use of mobile phone data for statistical purposes. As response rates for roadside surveys have declined and costs increased alternative ways of generating the evidence required for major transport initiatives have been sought. A number of pilot studies have been conducted.

Transport bodies seek similar data for input into their transport models: different bodies have separately procured mobile phone generated data on transport flows for the same geographic areas and time periods, albeit using different MNOs and methods. There is a growing appreciation that guidance is needed on how to procure these data, how it should be modelled and used for transport planning.

The Transport Appraisal and Strategic Modelling (TASM) team in the Department for Transport (DfT) provides guidance on collecting evidence to inform transport initiatives and plans to update this to cover mobile phone data during 2016. However, DfT does not have oversight of all public transport bodies operating in the UK.

In addition other public bodies could make use of mobile phone data for non-transport related applications (e.g. epidemiology, the derivation of population densities or tourism flows) and these disparate applications might benefit from better coordination.

6.2. International bodies

International statistical organisations have also been considering the potential of mobile phone data within official statistics. Details of some of the work carried out in recent years are outlined below.

The United Nations Economic Council for Europe set up an international project team to consider the opportunities offered by big data for statistical organisations. The team wanted mobile phone data to test big data technologies and methodologies. [redacted] The work substantiated many of the complexities previously outlined (paras 6.3 and 6.4). Unfortunately, initial exploratory analysis revealed that the data were not suitable for big data research.

OFFICIAL-SENSITIVE

Estonia is well advanced in modelling tourism flows using mobile phone data. They use it to estimate the number of unique visitors by country of origin, number of trips/visits, length of stay, area of stay etc. Some of these data are used as input into the official estimates of balance of payments. A study assessing the feasibility of using mobile positioning data for inbound, outbound and domestic tourism statistics including details on other related issues such as access, cost, trust, and technological and methodological challenges, was published in a Eurostat (2014) report.

The Central Statistics Office (CSO) of Ireland has, been conducting research on the use of mobile phone data for tourism statistics. They have had access to a sample of anonymised data from an Irish MNO. Legislation was passed in February 2014, for CSO Ireland to have free access to similar data from all MNO providers for specific statistical use although there are still details to address before this will take place.

[redacted] Procedures to protect privacy, such as mobile ids changing every 24 hours, were put in place so that longitudinal analysis was not possible. Further restrictions meant that an intermediary was appointed to process the data under instruction from CBS and provide aggregate outputs. Investigations undertaken by CBS Netherlands on population densities have highlighted the difficulties associated with assessing quality without access to the raw data.

6.3. Academia

A review of academic studies using mobile phone data was conducted in ONS. With a focus on applications to official statistics the review raised the following observations:

Data access has been granted for some academic studies albeit in highly restrictive forms. In general samples of anonymised CDR, from a single MNO, have been provided. Analysis using CDR is generally held to be challenging due to the widely differing behaviours exhibited by users: for example, some users text or make calls many times a day whilst others do not make calls for days on end.

There is therefore limited scope for mobility analysis with CDR, although some academics such as M. C. Gonzalez et al (2008) have proposed methods. [redacted]

Academic institutions also form collaborations for research purposes as exemplified by the work carried out by Montjoye etc al (2013) and Deville et al (2014). [redacted]

OFFICIAL - SENSITIVE

[redacted]

Within the UK, the Economic and Social Research Council has recently provided funding to set up a Consumer Data Research Centre [redacted] This research centre has a role in accessing consumer data, including mobile phone data, so that it may be used within social and economic research.

7. Methodology

Methodologies to produce worker flows are complex, diverse and not transparent.

Discussions with transport bodies suggest that methodologies are proprietary and that methodologies are constantly being improved, with optimal results involving the merging of mobile phone data with other data from road traffic sensors and surveys, official statistics on population including workplace populations as well as geographic data related to land use. How well, mobile phone data may perform on its own in producing transport flows is unknown.

Some of the key methodological issues are now discussed:

7.1. Population definitions

Official statistics are based on clear definitions and population statistics tend to refer to the usual resident population where a usual resident is a person who has lived, or intends to live, in the UK for 12 months or more. Census commuting flows are based on this definition.

Transport bodies, however, are primarily interested in total population flows, not just those made by the usual resident population. Mobile phone location data may be more suited to the derivation of such 'de facto' population flows as further processing would be needed to infer, with any accuracy, adherence to the usual resident population definition.

Another complexity is that census commuting flows are based on the adult population (those aged 16 and over) only. As children also use mobile phones, they would need to be removed from the data. MNO's indicate that this is difficult as children tend to have contracts taken out by their parents. Transport bodies, with their focus on total flows, would be less concerned about such issues.

The population definition has direct implications for the weighting model, discussed below (para 8.7).

OFFICIAL - SENSITIVE

7.2. Spatial resolution

There is a limit on the spatial resolution possible in terms of identifying the origins and destinations of journeys. With basic cell id data, the maximum resolution possible is a cell varying in size from a few hundred metres (in highly populated areas where cell towers are found in greater densities) to several kilometres (in areas of low population). The mobile phone is typically assigned to the cell centroid although, in reality, it might be at any location within the cell boundary.

Although network and handset techniques are available that can improve locational accuracy these investments have not been made by the MNOs. Cell ID data are therefore the only readily available source of locational data. While the data can be 'enhanced' using supplementary information such as received signal strength from neighbouring cell towers, such additional data are only available when the handset is making calls or texts and are not used by MNOs in their modelling of commercial statistical products.

The spatial resolution means that cell id data is not able to detect journeys that do not move between cell towers. This will be more marked in rural areas where cell towers may be kilometres apart.

Global Positioning System (GPS) positioning uses satellites to determine highly accurate locations and may help with monitoring movements in rural areas, although this is limited by the much smaller number of mobile devices capable of using it, or being turned on regularly to enable the detection of repeated journeys. The MNO will only see GPS location information if one of their apps is being used to generate it.

In areas of dense overlapping cell towers and to manage overload, mobile phone connections can switch tower without the caller having actually moved. In Kung et al (2014), a spatial filtering technique is used so that all movements within a 1km radius of the original cell tower are treated as 'noise'.

A further level of spatial uncertainty is produced when mapping cell areas to the official geographical boundaries commonly used in official statistics such as output areas or administrative boundaries such as wards.

7.3. Home location

Home location is generally assumed to be where the mobile is located during the night or when switched on first thing in the morning. A longer timeseries of location data, spanning multiple weeks, will reveal repeated patterns and give greater confidence about the inference of home location. MNOs may also use the address given on personal contracts to derive home location although this information is considered to be less reliable.

OFFICIAL-SENSITIVE

Modelling difficulties will arise for subscribers with non-traditional patterns of occupancy. These may include shift/night workers and those who work far from home and stay away, or spend a lot of time at an alternate address.

It is not known how accurate the modelling of home location is, although MNOs are comfortable that it is fit for purpose. Once home location is inferred it is set up on the subscriber's data record held in the MNO so that it no longer needs to be modelled. The derivation of home location is updated periodically for the whole subscriber base in at least one MNO.

7.4. Journey purpose

Work location is more difficult to model than home location and tends to be set to the location where mobile phones are found during the day (Mondays to Fridays). Methods are based on the detection of regular and repeated journeys made by the mobile that necessitates trace data spanning a longer period (e.g. varying from two to 6 weeks). Locations where the mobile stops regularly for longer periods of time (3+ hours) during standard work hours on repeated days are possible work locations. Land use data might also be used to indicate whether an area contains mainly residential or commercial properties or public buildings such as hospitals, schools etc.

Clearly, workers who have more flexible arrangements, such as those on part-time or zero hours contracts, who work during the night, , at weekends, in shifts, or at multiple locations etc., may not be easily identified by such a broad approach although in some cases it might still be possible to identify a repeated journey over a period. Furthermore, students travelling to their local college every day might look like a commuter using mobile phone geolocational data.

All MNOs agree that home workers are difficult to detect from location data. It is not possible to distinguish them from workers who do not move far enough to cause a transfer to another cell tower or from non-workers. MNOs have suggested that they might be able to detect subscribers on business contracts or those who make multiple calls to business numbers during the day although there is no basis for assessing how accurate such modelling might be, or whether it is being applied.

7.5. Key demographics

Population flows are sometimes required to be segmented by key demographics such as age and sex. MNOs use the information held on contracted subscribers directly, although concede that children's mobiles may be incorrectly categorised as adults are commonly responsible for the contracts. Contracted subscribers have typically represented around

OFFICIAL - SENSITIVE

50% of the total subscriber base although this has increased to over 60% in recent years due to the increasing take up of smart phones and sim-only deals.

Flows of Pay As You Go (PAYG) subscribers are segmented by age/sex at a later stage of processing. The prevailing method is to use the age/sex distribution from contracted subscribers as a proxy for the distribution of PAYG subscribers. [redacted]

There are an increasing number of academic studies on the inference of demographics from the behaviour exhibited by mobile phone use (Dong et al (2014)). These studies use call event data and make use of relationships between the frequency/length of calls or texts and number of different contacts to infer sex, age or social class. Inference of key demographics is not known to be performed using location data alone. None of the MNOs suggested that they performed this sort of modelling, preferring to use contract information.

Ethnicity is a subjective and contentious concept, however, there are studies (e.g. by Silm et al (2014) in Estonia and by Blumenstock et al (2013) in a South Asian country) on how it may be inferred by using the language setting on the mobile phone as a proxy indicator. Both these studies show a high correlation between the language setting and underlying ethnicity. Other studies consider the regularity of phone calls abroad to associate mobile phone users to a particular ethnicity such as in Douglass et al (2013).

7.6. Mode of transport

The start of a journey is detected when a mobile phone first switches its connection to another cell tower. As the journey continues there will be subsequent switches until the mobile arrives at a final cell tower where it remains for a period, thereby signifying the journey end. The mode of transport used to perform the journey is inferred using a combination of the time taken to switch cell towers and the sequence location of the cell towers.

Train journeys are reported to be easy to identify as, besides speed, multiple phones will switch cell tower at the same time. The sequence of changeovers will also mirror the location of cell towers on rail routes, although it is sometimes the case that major roads follow the same route.

Road vehicles are more difficult to differentiate as buses, vans, lorries, cars, motorbikes and, possibly, cycles might all be moving at identical speed. All MNOs suggest they can produce

OFFICIAL-SENSITIVE

flows made by road vehicles in general although some have suggested that further modelling may be performed to detect HGV vehicles as these have a maximum speed of 60 mph on motorways and have to stop at regular intervals.

It may only be possible to detect journeys by foot in urban areas where cell towers are densely packed. Even here there may be difficulty in distinguishing a pedestrian from a slow moving vehicle.

Air travel could also be derived from the time it takes to travel between two long distance locations. Even if mobiles are switched off during flight, there could well be some activity at the start and end of the journey signifying different airport locations.

As well as the detection of mode of transport it is possible to model likely routes taken. In addition to using the location data from cell towers, [redacted] This type of modelling may be performed within the MNO or alternatively may be outsourced to specialist transport analytics companies.

7.7. Weighting to population totals

The methods used to weight mobile phone data to population totals need careful consideration and are not transparent. [redacted] As greater variability in MNO market shares would be expected for smaller geographies, the suitability of weighting using regional information is unknown.

A more elegant weighting model is used by some MNOs. This method first infers the customer's home location as described above (para 8.3). For a given area, the MNO works out the total number of customers it believes reside there and then uses official population figures to calculate the proportion of the residential population deemed to be customers. The inverse of this proportion is then used as the weight for all customers in that area.

There are other uncertainties associated with the treatment of multiple mobiles by the same individual. It would be expected that unique subscribers would need to be identified first, before weighting to population totals. Further adjustment would need to be applied to account for the population that does not use mobile phones.

As most of the development research has occurred in transport bodies, it is presumed that the goal is to calibrate totals to the number of people travelling, whether resident or not. In

OFFICIAL-SENSITIVE

the specific case of worker flow it may be acceptable to presume that any such flows would belong to a UK resident.

A finer level of weighting might rely on the accuracy of the demographic detail and take account of variations across regions.

7.8. Other modelling challenges

The methodology discussed with MNOs implicitly assumes that mobile phones are switched on all the time and that full location traces are available for all users. One uncertain dimension is how missing data are handled. One transport body told us that some mobile traces had to be removed from their study as there wasn't enough location information, even over 4 weeks, to identify usual journeys with any accuracy. There are a number of potential explanations for this. For example, the user has switched off their phone, leaves the phone uncharged for lengthy periods, is a new subscriber, has changed provider or is away on holiday. The scale of such missing data is unknown, and will have implications for the weighting model. It might be possible to assume that such missing data is spread randomly across the country or within regions and by demographic profile but this would need to be tested. Missing location data will be more problematic if using only call events rather than passive location information as people who text or make calls infrequently will be very hard to model.

The problem of missing data is complicated by the opt-out arrangements for subscribers allowed by some MNOs in order to strengthen their ethical standards. The level of opt-out needs to be known and factored into any weighting. Any assumptions about the level of opt–out would need to be tested.

8. Conclusion

Around the world, statistical organisations such as ONS are actively seeking to identify the potential of using mobile phone data for the production of official statistics on population and mobility. This report has provided an overview of the investigations currently being conducted.

Mobile phones generate digital information that is sent to MNOs for processing and there is an interest in using this to develop statistical products for sale. Location data generated by mobile phones have attracted the most attention as they provide almost complete coverage of the population and so can be used to produce information on population densities and mobility. The availability of data in close to real time and for small geographies brings

OFFICIAL - SENSITIVE

advantages to a number of applications such as in crowd-monitoring, retail location planning, tourism and transport.

Two types of locational data are generated from mobile telephony: from active events such as making or taking calls or text messages, and from passive updates. MNOs will sell these data in various aggregated and anonymised forms for different applications.

The mobile phone market has been shown to contain only three MNOs, with a customer base large enough for the derivation of population estimates.

Privacy, ethical and legislative issues have been highlighted alongside the related issue of public acceptability. Other important factors affecting data access include the complexity of procurement arrangements and associated processing costs.

A brief overview of the experiences of using mobile phone data is provided with a focus on the research projects conducted by UK public bodies, international statistical organisations and academia.

Finally, specific methodological challenges are considered, including difficulties associated with accuracy and transparency.

9. Next steps

ONS will continue its research into the potential of using mobile phone data for the production of official statistics. The work will be organised in two work streams as follows:

The first will seek to access some mobile phone data in order to understand the quality issues, to assess the potential use for official statistics and make recommendations. This work will require collaboration with organisations who have already acquired mobile phone data including public transport bodies, other government bodies, international statistical organisations and academia.

The second will seek to influence the cross Government approach to using mobile phone data for statistical purposes. This will involve extensive engagement and encouragement of collaboration with government bodies including the GSS, MNOs and other organisations who may wish to acquire such data.

10. Bibliography

Blumenstock J, Fratamico L (2013) Social and spatial ethnic segregation: A framework for analyzing segregation with large-scale spatial network data

OFFICIAL - SENSITIVE

Calabrese F (2011) Urban sensing using mobile phone network data. Ubicomp 2011 Tutorial

de Montjoye Y, Hidalgo C A, Herleysen M and Blondel V D (2013): Unique in the Crowd: The privacy bounds of human mobility. Nature.com Article number 1376

Deville P, Linard C, Martin S, Gilbert M, Stevens F R, Gaughn A E, Blondel V D, Tatem A J (2014) Dynamic Population Mapping using mobile phone data

Dong Y, Yang Y, Tang J, Yang,Y, Chawla NV (2014): Inferring User Demographics and Social Strategies in Mobile Social Networks

Douglass RW, Meyer DA, Ram M, Rideout D and Song D (2015) 'High resolution population estimates from telecommunications data

Eurostat (2014) Feasibility Study of the Use of Mobile Positioning Data for Tourism Statistics *Consolidated Report Eurostat Contract No 30501.2012.001-2012.452*

Gonzalez M C, Hidalgo C A, and Barabasi A L (2008): Understanding individual human mobility patterns. Nature, 453(7196):779{782}

Kung K S, Greco K, Sobolevsky S, Ratti C (2014): Exploring Universal Patterns in Human Home-Work Commuting from Mobile Phone Data. PLoS ONE 9(6):e96180. Doi:10.1371/journal.pone.0096180.

Isaacman S, Becker R, Caceres R, Kobourov S, Martonosi M, Rowland J, and Varshavsky A (2011): Identifying important places in people's lives from cellular network data. Article published in proceedings of the 9th international conference on Pervasive Computing, San Francisco, pages 133-151.

Phithakkitnukoon S, Smoreda Z, Olivier P (2012): Socio-Geography of Human Mobility: A Study Using Longitudinal Mobile Phone Data. PLoS ONE 7(6):e39253 Doi:10.1371/journal.pone.0039253.

Ofcom 2014: Children and parents: media use and attitudes report

Silm, S & Ahas, R (2014): Ethnic differences activity spaces: The study of out-of-home non-employment activities with mobile phone data, *Annals of Association of American Geographers* 104(5): 542-559.

Silm, S and Ahas, R (2014): The temporal variation of ethnic segregation in a city: evidence from a mobile phone use dataset, *Social Science Research* 47: 30-43.

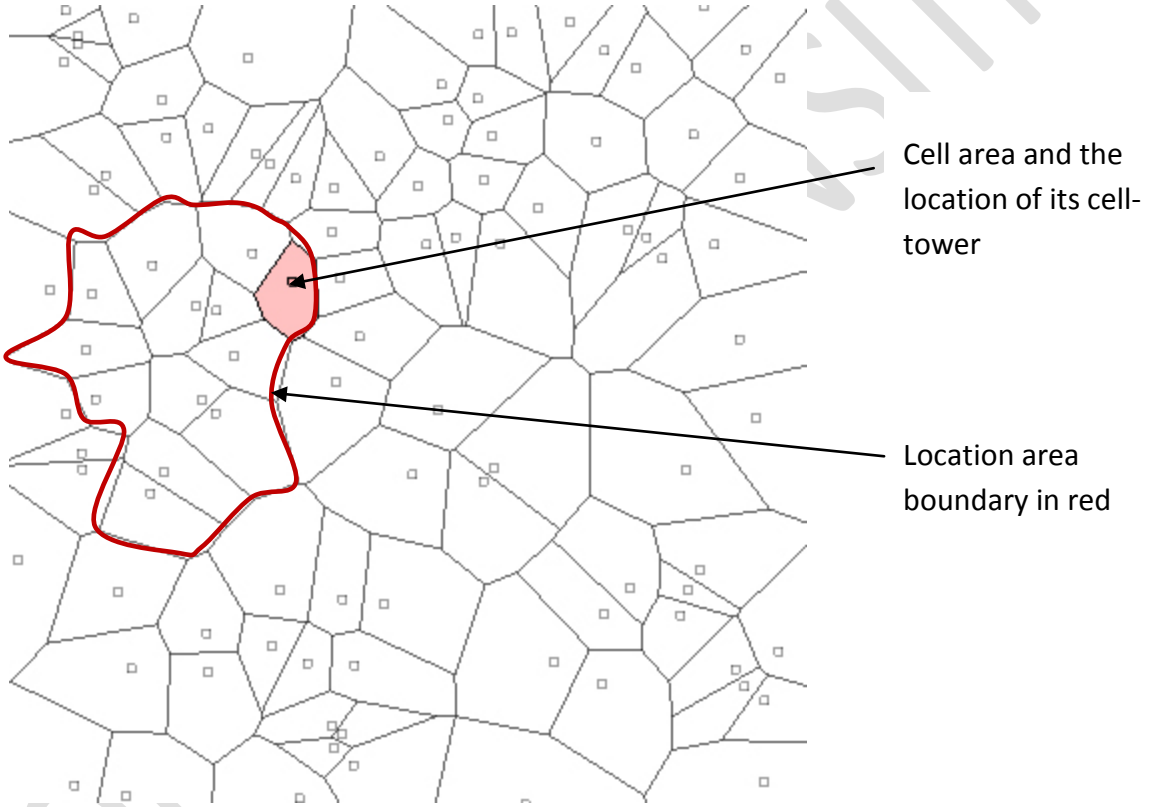
OFFICIAL - SENSITIVE

Appendix A: Mobile phone data

Mobile phones connect via radio waves to a nearby cell tower or base station that covers a specific geographic area called a cell. Groups of cells form larger geographies called location areas. Figure 1 illustrates a small section of such a mobile phone cell grid.

MNOs build networks of cell towers that collectively provide the infrastructure to operate mobile phone telephony. As the mobile phone usually connects to the nearest cell tower, it can be assumed that it is somewhere within that cell. To form a successful connection it is necessary for the mobile to be switched on and for there to be a strong enough signal to a cell tower owned by the relevant network operator.

Figure 1: Illustration of a cell grid with one highlighted location area



Two types of location information are generated from mobile phones:

Call Detail Record (CDR): generated when the phone is active such as taking or placing a phone call or mobile message or using the Internet on a smartphone. In these situations, the time and the cell id of the cell tower forming the mobile connection is recorded.

Network data: When the phone is switched on but idle there are two ways of generating location data. The first is as a result of periodic location updates, known as 'pinging', sent by the phone to the network; the frequency and type of location information produced is

dependent on the generation of mobile phone technology used and Table 2 summarises this.

Table 2 Comparison of frequency and location information according to the generation of mobile phone technology

	Generation of mobile technology (4G enables fastest speed and functionality, 2G the slowest speed and functionality)		
Location Updates (when phone in idle mode)	2G	3 G	4G
Frequency	3 hours	3 hours	30 mins
Location information	Location Area	Location Area	Cell ID

The second form of network location data is mobility-driven: network communication with the mobile phone is transferred to a different cell tower when it moves into that cell tower's cell area. For 4G technology, the new cell id is recorded as the mobile switches connection. For older 2G and 3G technologies, a new location update is recorded only when the mobile moves between location areas.

If the phone is active and the mobile switches cell tower, then the cell-id is recorded at the time of the switch. This is called a 'handover' event.

Illustrations of the different data are shown in Table 3 and Table 4, adapted from Calabrese F (2011).

Table 3: Sample of call event data for an anonymous mobile id

Origin Mobile ID	Origin cell id	Terminating Mobile ID	Terminating cell id	Date ID	Duration
24393943	10121	17007171	10121	24031517	29
24393943	5621	17007171	2721	25141136	38
24393943	17221	39134589	9411	34160370	111

OFFICIAL - SENSITIVE

Table 4: Sample of network data for an anonymous mobile id

Origin Mobile ID	Loc. Area	Cell id	Time
4ba232e4d96f47dc	558	207	1246689869
4ba232e4d96f47dc	559	218	1246690869
4ba232e4d96f47dc	562	225	1246691869
4ba232e4d96f47dc	565	235	1246692869
4ba232e4d96f47dc	568	250	1246798694
4ba232e4d96f47dc	568	253	1246889869
4ba232e4d96f47dc	568	253	1246989869
4ba232e4d96f47dc	568	253	1246999869
4ba232e4d96f47dc	565	235	1247090869
4ba232e4d96f47dc	562	225	1247091869
4ba232e4d96f47dc	559	218	1247092869
4ba232e4d96f47dc	558	207	1247098694
4ba232e4d96f47dc	558	207	1247189869
4ba232e4d96f47dc	558	207	1247289869

Different Location Areas in short timescale – indicating movement

Period updates have same location area and cell id that suggests mobile has stopped moving

OFFICIAL-SENSITIVE

Aggregation

The tables above illustrate the individual level location data generated by mobile phones. MNOs also aggregate this type of data for different spatial areas and time periods. For example, at a cell tower aggregation, aggregates might represent the total number of calls or location updates in a specific time of day. The format of the data, aggregated vs individual level, has implications for data access and potential statistical applications that are discussed in other sections of this report.

OFFICIAL - SENSITIVE

Appendix B: Formats of mobile phone data with pros and cons

A summary of the different applications for different formats of location data is presented together with the pros and cons in Table 5 sourced from Calabrese F (2011).

Table 5: Applications for different formats of mobile phone location data, with pros and cons

Format	Pros	Cons	Application
Aggregated cell tower statistics	Easy to manage, Possibly in real time	Unable to model user mobility	Population density, Land use
Aggregated CDR with cell tower location information	Easy to manage	No individual interaction information	Connection between places, Regional partitioning
Individual CDR	Individual communication patterns	Large dataset, mostly not real time	Social network analysis
Individual CDR with cell tower location	Individual communication and mobility patterns	Large dataset, mostly not real time	Mobility analysis between large areas
Individual event-driven triangulated location	Individual mobility patterns, possibly in real time	Large dataset, Possibly need for special hardware to access data	Origin destination, transportation mode
Individual Network driven data	Individual mobility patterns, possibly in real time	Large dataset, Possibly need for special hardware to access data	Mobility analysis between large areas

OFFICIAL-SENSITIVE

Appendix C: Market composition considerations

This section discusses/considers the representativeness of the data as well as issues of market continuity.

Non-residential UK

For the development of UK population statistics, mobile phone users who subscribe to networks in other countries should be excluded from consideration. These may include tourists, business travellers or other short term visitors to the UK. MNOs in the UK and abroad have roaming agreements so that mobile network services can be provided to their customers through another network provider when they go abroad. The customer usually has to pay an enhanced tariff for using their phone abroad. Such arrangements should encourage longer term migrants to the UK to subscribe to UK networks (although this price differentiation will soon cease to exist within EU countries).

Sub-regional market share

From discussions with MNOs it is understood that at smaller geographies market share can vary wildly.

As the same pricing models prevail throughout the UK, factors influencing take up in different regions could include the coverage and quality of service. Moreover, the make-up of the population may affect user need and affect the choice of service provider. These factors together with targeted marketing initiatives, may influence take up.

Demographics

It is useful to compare the gender and age distributions of the subscriber base with those of the population to understand the representativeness of the subscriber base. Ofcom estimates for the first quarter of 2014 show that people in older age groups are less likely to own a mobile phone than people in younger age groups: for example, mobile phones are owned by 99% of 25-34 year olds but only 68% of those over 75 years.

Ofcom research also suggests that the distribution of socio-economic classes is important as higher socio-economic classes have adopted smartphones in greater numbers than those in lower socio-economic groups. In Quarter 1 2014, Ofcom estimated that 97% of people in the ABC1 socio-economic classes had a mobile phone compared to only 91% of people in classes D or E.

Amongst children, ownership of a mobile phone increases with age. Ofcom shows that mobile phone usage ranges from 0% at age 3 to 90% at age 15.

OFFICIAL-SENSITIVE

More than half of children aged 5-15 use pay-as-you-go tariffs although there is increasing use of monthly contracts, thought to be attributable to the growth of smartphones. As children are not allowed to enter into contracts with MNOs it will be difficult to differentiate between mobiles used by children and by adults.

Market stability and future developments

Table 1 underlines the fluctuations in market share in the UK during the period 2014-15. The main factors driving these changes are now considered.

Pricing models

Until fairly recently, around 50% of subscribers were on yearly contracts and 50% on pay as you go tariffs. As contracted customers tend to be more profitable, there has been a drive to make two year contracts the norm and as a result the proportion of pay as you go subscribers has fallen. The increase in longer term contracts will tend to stabilise market share.

In contrast, there has been a growth of sim only deals that allow customers to use their existing handsets yet take advantage of competitive pricing from competitor networks. Such deals mean that some individuals may change providers more frequently.

Mergers and Acquisition

Mergers, acquisitions and new entrants to the market are also common. [redacted]

Alternative communication technologies

Another feature to impact on mobile phones is the increased use of alternative communications technologies such as Voice over Internet Protocol (VOIP) services that allow individuals to place calls and other multimedia sessions over the Internet, at much cheaper rates or at no cost. Skype is an example of such a service. Traditionally implemented by placing calls via computers, the functionality is now available on any device capable of connecting to the Internet such as tablets or smartphones. Increasingly, the connection to the Internet from smartphones involves going through local wi-fi networks rather than through cell towers as wi-fi might have a stronger/better signal.

OFFICIAL - SENSITIVE

The increased use of the Internet and wi-fi for messaging will reduce the number of call events detected at cell towers from mobile phones although passive location update data will still be collected in the normal way. This reduction in call events is already evidenced in Ofcom's 2014 communications market report.

[redacted]

3G and 4G

[redacted]

The growth of 4G telephony will lead to increased location data from more frequent pinging and the recording of movement between cell towers.

Technological Continuity

Technology may also affect the future stability of these data. For example, current mobile phones use SIM cards to hold information such as key contacts and phone numbers. From their inception in the early 1990, SIM cards offered a convenient way of storing such information whilst facilitating portability without compromising security.. However SIM cards have been replaced by new technologies. For example, consumers may use highly secure cloud environments to preserve their data by simply logging on in a similar way to online banking. Using a SIM card is an obstacle to switching providers as consumers are effectively locked into a specific network. If future technologies remove the need for SIM cards then it is likely that consumers would switch more frequently, to secure the cheapest tariffs. This would have the effect of distorting market shares more frequently.

Appendix D: Data access

In addition to the technical challenges and, in the UK, the high cost of data access, ONS must consider and take account of its obligations to protect privacy and safeguard the confidentiality of all the information that it accesses or uses.

These considerations are introduced in turn:

Privacy and Ethics

ONS can only access and use data where it has the legal authority to do so. In all cases full account will be taken of our legal obligations (e.g. s.39 Statistics and Registration Service Act 2007) to safeguard confidentiality.

ONS has presented a proposal for its research on the potential of using mobile phone data within official statistics to the GDS Privacy and Consumer Advisory Group and the ONS Beyond 2011 Privacy Advisory Group. Both groups, although wary of the acquisition of individual level data, were supportive of the use of aggregated data for research purposes, especially if it is to be aggregated within the mobile phone company.

The main issues are:

Identification of individuals

The Data Protection Act (1998) (DPA) is the main legislation governing the processing and handling of personal data. It stipulates that data identifying an individual should not be passed onto third parties without consent.

ONS is only proposing to source aggregated data for research and as a result it will not be possible to identify individuals.

MNOs are known to pass anonymised, individual level location data to third parties, however, a study in Scientific Reports concluded that it only takes as little as four locational data points to identify unique movement patterns for 95% of users. With appropriate linking to other data, it may be possible to identify a user. ONS is not seeking this form of data.

Table 6 illustrates how the MNOs are currently safeguarding privacy in the data they sell.

OFFICIAL - SENSITIVE

Table 6: Measures to safeguard privacy across the MNOs

	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]

Informed consent

The DPA also states that individuals need to give consent to the collection of their personal information and to its legitimate use.

Commercial companies require that customers agree to their standard terms and conditions (T&Cs) in order to be provided with their service. It is argued that the detail over the use of personal data is 'hidden' within these T&Cs and individuals are therefore compelled to give consent even though they may prefer not to. It is further argued that individuals do not always understand the use that their data may be put to and that there should be specific opt-in to new uses.

[redacted]

OFFICIAL - SENSITIVE

Table 7 illustrates how the MNOs have approached this.

Table 7: Options to opt-out across the MNOs

	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]

ONS complies fully with the terms and conditions specified by individual companies. In all cases the information will be used for statistical or analytical purposes only.

Inferred personal variables

Another concern is the modelling of data to infer personal information, even if it has not been directly given. For example, home and work locations are inferred using mobile phone location data.

ONS is only interested in aggregated data. It will not be possible for ONS to identify the home and work location of individuals. The modelling to infer home and work location requires a longitudinal trace of individual level data and will have been performed within organisations with legal access to the data.

Public Acceptability

In addition to legal or ethical issues account must be taken of public acceptability. This is as important for the end-user of the data as it is for the MNO.

In view of the need for, and importance of, maintaining trust ONS pays great attention to any issues of public concern.

Other ethical considerations

OFFICIAL - SENSITIVE

ONS is only sourcing aggregated data for statistical use. However, there are still ethical considerations for organisations with legal access to individual level data:

- Data may be collected for a specific use. One of the features of big data is that it could be mined to reveal patterns that were not known before. This might lead to situations where observations in the data could reveal insights that are different to the specific use for which they were collected
- Mobility analysis ideally requires a complete longitudinal trace of location data. However, it may be seen as unethical to promote the use of mobile phones in vehicles. A police campaign in North Wales has appealed to motorists to 'switch off before you drive off' in an attempt to reduce the number of accidents.

Chains of data provision

[redacted]

OFFICIAL-SENSITIVE

	[redacted]		
[redacted]	[redacted]	[redacted]	[redacted]

Many public transport organisations procure the services of a major transport consultancy using their existing supplier frameworks. It is the responsibility of the transport consultancy to ensure that appropriate arrangements are made for both access to, and disclosure of, any data.

It has been noted that MNOs are increasingly forming exclusive agreements with data analytics companies who in turn have established relationships with specific transport consultancies. [redacted]

OFFICIAL - SENSITIVE

[redacted]

IT infrastructure

Aggregated CDR are the easiest data to process and present the fewest privacy concerns whilst individual level networked data present such large technological and privacy issues that processing needs to remain either within the MNO or in specialist big data analytics companies with appropriate confidentiality safeguards. Transferring individual networked data to another organisation, aside from demanding the technical capacity, will require strict controls around data security which may involve effective anonymisation procedures such as changing the id every 24 hours.

Legislation

Access to, and use of, data is governed by legal requirements, and other regulatory and ethical standards. In the UK the Information Commissioner's Office has the role of upholding information rights.

At the present time there are no provisions to authorise ONS to seek access to mobile phone data.

Regulators in countries across the world also have an important role in managing data access controls around personal data. The EU Data Protection Directive was first written in 1995 but, given the technological changes of the past 20 years and the increasing challenges associate with data access and use, a new regulation is being drawn up.

Given the UK's referendum decision to leave the EU, it is unclear on the implication of the new EU regulation for UK access to mobile phone location data in the near future.

Research data

There are precedents in making CDR data available for research. Academic studies have managed to source data from individual MNOs operating in specific countries. [redacted]

Commercial Interest

The challenges around data access mean that it might be preferable for the MNO or a data analytics company to undertake the research rather than provide access to individual level data. In such cases there would be financial implications both in terms of processing and analytical charges.

Transport bodies have traditionally used roadside surveys to collect evidence for transport schemes and so have large budgets for collecting data. This level of funding makes it viable to purchase data from MNOs for quite high cost as this still may be more cost effective than roadside surveys. For other mobility type applications, such as in official statistics and epidemiology, the cost of data purchase may be too expensive to justify.

OFFICIAL - SENSITIVE