Article:

Harrison, G orcid.org/0000-0002-4324-772X, Grant-Muller, SM and Hodgson, FC (2020) New and emerging data forms in transportation planning and policy: Opportunities and challenges for “Track and Trace” data. Transportation Research Part C: Emerging Technologies, 117. 102672. ISSN 0968-090X

https://doi.org/10.1016/j.trc.2020.102672

New and Emerging Data Forms in Transportation Planning and Policy:

Opportunities and Challenges for "Track and Trace" Data

Highlights

- For the first time we define mobile phone generated "Track & Trace" (T&T) datasets within a personal mobility paradigm, illustrating features of spatio-temporal mobility content and socio-demographic (or trip based) individual context.
- We summarise the limitations of conventional data and modelling within transportation planning, conceptualise NEDF and characterise the unique potential (and challenges) of GPS-based T&T data.
- Our meta-analysis of existing reviews and literature related to mobile phone data demonstrates that of New and Emerging Data Forms (NEDF), and T&T data in particular has had little mention to date in the academic literature within an applied policy context.
- We propose that GPS-based T&T data can be integrated into existing methods to create low cost, transferable decision-making tools for holistic and equitable policy.

Abstract

High quality, reliable data and robust models are central to the development and appraisal of transportation planning and policy. Although conventional data may offer good 'content', it is widely observed that it lacks context i.e. who and why people are travelling. Transportation modelling has developed within these boundaries, with implications for the planning, design and management of transportation systems and policy-making. This paper establishes the potential of passively collected GPS-based "Track & Trace" (T&T) datasets of individual mobility profiles towards enhancing transportation modelling and policy-making. T&T is a type of New and Emerging Data Form (NEDF), lying within the broader 'Big Data' paradigm, and is typically collected using mobile phone sensors and related technologies. These capture highly grained *mobility content* and can be linked to the phone owner/user behavioural choices and other *individual context*. Our meta-analysis of existing literature related to spatio-temporal mobile phone data demonstrates that NEDF's, and in particular T&T data, have had little mention to date within an applied transportation planning and policy context. We thus establish there is an opportunity for policy-makers, transportation modellers, researchers and a wide range of stakeholders to collaborate in developing new analytic approaches, revise existing models and build the skills and related capacity needed to lever greatest value from the data, as well as adopt new business models that could revolutionise citizen participation in policy-making. This is of particular importance due to the growing awareness in many countries for a need to develop and monitor efficient cross-sectoral policies to deliver sustainable communities.

Keywords: Transport policy, Track and Trace, mobile phone data, mobility profile, Big Data.

1 Introduction

The primary motivation of this paper is to focus on the potential of Track and Trace (T&T) data of individual mobility profiles towards meeting gaps in the modelling and analysis needs of transportation policy makers. We firstly elaborate the overarching data framework, which illustrates the relative positioning of T&T, as typically collected using mobile device (e.g. phone) sensors, and a type of New and Emerging Data Forms (NEDF) which lie within the broader 'Big Data' paradigm. We argue that in particular GPS-based T&T data currently offers the most potential for revolutionising understanding of travel behaviours, but are yet to be adopted by the majority of practitioner stakeholders in transportation policy and planning.

Local authorities are facing increasing pressure to plan, design and implement schemes that satisfy multiple, and often-conflicting policy targets. From the perspective of the individual, mobility is central to social engagement, health and economic wellbeing (Schwanen et al., 2015). Yet it also impacts on physical activity (Oliver et al., 2016, Saunders et al., 2013, Mueller et al., 2015), pollutant exposure (Xia et al., 2015), safety (Lord and Washington, 2018) and personal security (Beecroft and Pangbourne, 2015). Transportation planners, engineers and policy-makers have long focused on increasing network efficiencies through models based on traditional static surveys and historic data. The potential of Information and Communication Technologies (ICT) for transportation has been recognised for decades (Banister and Stead, 2004). More recently, the challenge of setting transport policy within an expanding cross-sectoral policy framework has emerged in tandem with the growing awareness of new, and particularly digital, data[1], arising from emerging technologies, and an increasingly connected world. These offer

[1] In this paper we use the term 'data' to cover both raw data and information produced from processing or semi-processing of raw data.

the opportunity to augment existing data and analysis methods (Witlox, 2015, Milne and Watling, 2019) and can also influence mobility, accessibility (Cohen-Blankshtain and Rotem-Mindali, 2016) and improve understanding of behaviours. The knowledge and insights gained contribute to the development of better systems and public policy.

1.1 Defining 'Big Data', NEDF, mobile phone technology and Track & Trace

Increasing affordability and access to digital devices with high data processing capabilities allows for the gathering and monitoring of more types and larger volumes of data related to mobility (Ben-Elia and Zhen, 2018) and transport analytics (Ukkusuri and Yang, 2019). Commonly referred to as 'Big Data', these can be characterised in various ways, mainly building on "3 V's": Volume, Velocity and Variety (Laney, 2001), such as Veracity, Visualisation and Visibility to emphasise the potential for an increased quality of geo-spatial data (Li et al., 2016). In a review to identify what may be 'special' about Big Data in transport systems analysis and (medium-term) planning without focussing on any particular type of data, Milne and Watling (2019), identify seven features of Big Data in this context, including monitoring, data ownership, collection rational, acquisition domain, potential connectivity, scale and relationship with other sources. The Transport Systems Catapult[2] (UK) championed development and identified key trends of open transport datasets related to Big Data (TSC, 2015), whereas Cottrill and Derrible (2015) discuss the value of using Big Data for sustainability indicators related to transportation. This paper is distinct from previous papers concerning Big Data and makes an important contribution in that the main focus is on *'Track and Trace'* (T&T) data. This is a particular data type within the category of *'New and Emerging Data Forms'* (NEDF). NEDF are highlighted as a strategic priority in policy agendas and national/international funding

[2] Now incorporated into the new 'Connected Places' catapult: https://cp.catapult.org.uk/

(Chandy et al., 2017, EC, 2019, Hobbs and Hanley, 2014, Lalawat, 2018, NITRD, 2016, NSFC, 2018, Willets, 2013) and form a subset of 'Big Data' more generally. The term "Track and Trace" was first identified in supply chain logistics around the year 2000 (Swartz, 2000), and the technology and practise has been widely adopted (Kros et al., 2019, He et al., 2009). "Tracking" is the determination of state (ie. position) at any point in time, whereas "Tracing" is the ability to determine past states and origin (ie. the trajectory) (Kelepouris et al., 2006). In a personal travel context, the T&T terminology may not have been widely used, but the concepts have (see Section 4). T&T data are generally collected using personal mobile devices such as wearable technology and mobile phones[3], although we are primarily concerned with *mobile phone technology* and GPS-based T&T datasets in particular. As we were unable to identify a clear precise definition in literature, we present our definition of T&T in Figure 1.

At the highest level, Big Data can be approximately divided into two main groups, i.e. traditional data and NEDF (Figure 1). Although the basic form of data or type of collection technology may be 'traditional', technology advances including new pervasive, fixed location sensors such as image and motion sensors satisfy the "3Vs" definition (Teknomo, 2002). NEDF are characterised by the novelty of their source, collection process and potential for multivariate or concurrent observations. There is a need to distinguish the data *collection technology* and the data *form*. Examples of collection technology include mobile phones tablets and a variety of other 'smart' devices[4]. The forms of data, which

[3] In 2017, 85% of the population of Europe were subscribed to mobile phone services, of which 70% connected with smartphones GSMA 2018. The Mobile Economy. Europe 2018. London, UK. Available from: https://www.gsma.com/mobileeconomy/europe/, accessed 06/12/18: GSMA. We found in the literature that 'mobile phone' is normally used for smartphone technologies and will continue that terminology in our work.

[4] An electronic device which collects and shares data with other devices through a common network (eg 3/4G, Bluetooth, RFID) or cloud-based web platform, commonly known as the 'Internet of Things'. Examples other than

may not be originally collected for a transportation application, include individual geo-spatial movements, multi-media and contextual information, and are being widely utilised across many sectors including business (Fan et al., 2015), public policy making (Hilbert, 2016) and academic research. There are a variety of forms of NEDF that include individual potentially high-resolution[5] spatio-temporal context characteristics and which can be used to form T&T datasets, as shown in Figure 1. Lee and Sener (2020) also describe emerging data, specifying both sources and application, though focusing on pedestrian and bicycle monitoring. They identify that (at the time of printing) there has been no clear definition of the terminology used, though also stress the difference between traditional and emerging data.

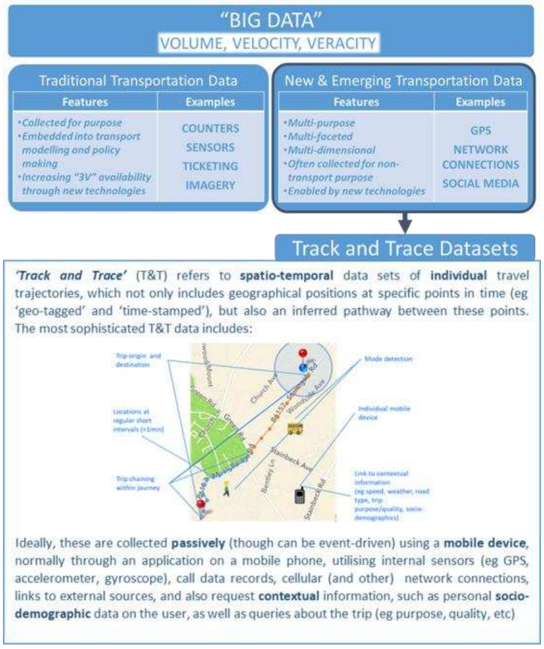


Figure 1: Subsets of 'Big Data' with key examples and showing those which can form T&T datasets (authors' interpretation), and definition of 'Track & Trace' (authors' definition, building on Swartz (2000) and Kelepouris et al. (2006)).

Notes:

The table is not a *comparison* between traditional data and NEDF or intend to suggest that these data serve similar purposes. It is simply illustrating that the types of data that have been traditionally used in transport planning and policy have distinctly different features than NEDF that are applicable to transportation.

---

mobile phones include watches (e.g. 'FitBit'), home control systems (e,g, 'Hive'), and virtual assistants (e.g. 'Amazon Echo Alexa').

[5] By "high-resolution" we mean a spatial and temporal scale of metres and seconds rather kilometres and hours

Traditional ‘ticketing’ (i.e. paper based with limited information) and ‘imagery’ (i.e. static fixed base such as ANPR) are distinct from highly digitised and detailed spatio-temporal ‘smart’ ticketing and mobile media that we view as NEDF in Section 3 – please see further description in this section.
The ‘individual’ travel trajectory is that of the device itself and it can only be used to derive that of the individual who is carrying the device
Although comprehensive T&T datasets consist of multiple forms of mobile phone data (as described), most studies tend to focus on the specific contribution of one specific NEDF (see Section 4) and we categorise NEDFs in this way (see Section 3). In our study we focus on GPS data (explained in Section 4).

1.2 Scope and structure of this paper

As a review paper, this work is aimed at a diverse readership within transportation planning and engineering, anticipating varying degrees of background knowledge and includes examples from a number of countries. We seek to address specific research questions regarding T&T data that have emerged from transportation practitioners, experienced policy makers and students across the transport and public health sectors.[6] This fills a gap identified in the literature (see Section 4), as although there are increasing studies which focus on mobile phone data, they focus on limited types of T&T data and it’s use in transportation modelling and policy applications.

(RQ1) How can NEDF overcome limitations of traditional data and models?

(RQ2) How have different types of spatio-temporal mobile phone data been used to date?

(RQ3) What are the opportunities and challenges in using passive GPS-based T&T data as a specific NEDF, collected through mobile phone technology?

RQ1, which sets out important high-level context, is addressed across Sections 2 and 3. In Section 2 we establish the limitations of traditional data and models, acting as a preface

---

[6] Questions arising from workshops with city stakeholders as part of the ESRC funded HABITs project (habitsdata.org/) and EU funded EMPOWER project (empowerproject.eu), as well as presentations on work from these projects in international conferences POLIS 2018 (www.polisnetwork.eu/2018conference) and ETC 2018 (aetransport.org/en-gb/etc).

to the response to RQ1, which is fully answered in Section 3. In Section 3, NEDF are defined, categorised and a rationale is provided as to how they could overcome the limitations established in Section 2. T&T data are generally collected using mobile personal devices (usually mobile phone technology), hence an overview of state-of-the-art applications of a range of mobile phone data in transportation research is then presented in Section 4. This section provides an answer to RQ2, and includes a focus on T&T data, thereby setting the context to address RQ3 in Section 5. Section 5 provides an in-depth response to RQ3, through discussion of the specific opportunities and challenges of GPS-based T&T data in a transportation planning and policy context. Finally, we draw our conclusions and summarise our response to all research questions in light of our discussions.

2 Background to transportation policy modelling and data needs

In addressing our first research question, *"How can NEDF overcome limitations of traditional data and models?"* we first need to establish the policy context of transportation modelling and data and the limitations of the traditional approaches currently employed.

Partly driven by strategies of international bodies and national governments (DfT, 2017, DoT, 2018, EU, 2011), transportation policy is becoming more 'people-oriented'. This approach recognises the importance of accessibility to services, as well as the range of individual needs and capabilities. Alongside this, the need to address cross-sectoral issues and policies, such as energy, health and environment is becoming increasingly important (Towe et al., 2016) at transnational policy level (Le Blanc, 2015), and at national level (DfT, 2017). Traditionally these sectors have tended to be dealt with separately and there is a recognition that the most efficient decisions need to be based on good data (Zmud et al., 2013).

Both qualitative and quantitative data are used in policy appraisal processes, with the choice of methodologies driven by the available data. Thus, the quality of policy making has a clear inter-dependence with modelling and data. However, in many countries, municipalities are experiencing increasing resource pressures (e.g. (NAO, 2017)), both monetary and skill-based (TSC, 2015), data governance issues (Abrantes and Linton, 2016), and tend to rely on traditional proven methodologies (DfT, 2014). Despite a growing awareness by policy makers of the potential of NEDF in transportation (TSC, 2015), and the importance of participatory processes (Keseru et al., 2018), there is some sluggishness in the adoption and integration of NEDF into existing processes and models, especially at a local level. The novelty of the data itself and the resource demands of collection and processing demands require researchers and innovators to prove the scalability of the concepts, before it is felt adoption can be justified. This reluctance could be a hindrance to improving policy decision making that currently relies on tools limited to the boundaries of traditional data. Thus, tools and therefore decision making are largely limited to a high-level understanding of behaviours, which neglect to account for inter- or intra-population differences. This is contrary to aspirations to develop policies that are both efficient and effective – providing benefit to all citizens, including the most vulnerable, with limited public funds in an inclusive and transparent process.

In this policy context, we use the next two sub-sections to frame the approaches to modelling and forms of data traditionally used.

2.1 Modelling approaches commonly applied in the transportation sector

Since the 1950s, many modelling approaches have been applied to (and have been central to) transportation research, policy and planning (Hensher and Button, 2008). In Appendix Table A1 we provide an overview summary of modelling approaches, including purpose, underlying concepts, data requirements, weaknesses and further reading. These approaches are complementary, each offering different purposes, scales and contexts, relying on different forms of data and adding value to transportation research and planning. Despite their contribution, each approach has its limitations, for example having been developed with the available data and computer power at the time (e.g. 'HATS' in the UK (Jones et al., 1983)), as well as being framed and contextualised within the transportation research mind-set of the time. Traditionally, transportation modelling has tended to focus on infrastructure/engineering or economics, being related to traffic demand and based around the classic 'four-step model'. Most municipalities use some form of traffic modelling based on this approach in order to plan changes in the network. More recently, modelling approaches from other disciplines are being applied in transportation policy and planning. The most notable of these are system dynamics, agent-based, discrete event, discrete choice and GIS. Both pure and hybridised models are being developed where different modelling approaches are interfaced or combined to overcome limitations of each, or capitalise on the combination of individual strengths (eg combined SD and ABM (Shafiei et al., 2012)). Many of these approaches are being adopted as they allow for a more encompassing recognition of hereto neglected environmental and social impacts, as well as to broaden the scope of transportation modelling beyond an urban and automotive centric approach which has long been the primary concern of transportation practitioners (Kitamura and Kuwahara, 2005). Areas of particular interest include car ownership or passenger demand forecasting and energy or emission accounting. Adoption of contemporary modelling approaches within transportation planning and research has become possible at least partly due to an increase in ICT capabilities, digitalisation and data availability.

2.2 Traditional Forms of Transportation Data

The modelling approaches outlined rely on good quality data for base specification, calibration, validation and scenario design. Wang et al. (2018b) note that early travel behaviour research largely relied on data from manual travel surveys. Worldwide, most developed countries undertake some form of regular travel survey across their population to understand the movements of people and goods, and the needs for infrastructural and policy planning (Cornick et al., 2018, Hoogendoorn-Lanser et al., 2015, Westat, 2018). Such surveys form the "backbone of the transportation data pipeline" (Zmud et al., 2013). These surveys often capture movements plus the background characteristics of the individual and their transport needs, but can be costly and infrequent (Cottrill and Derrible, 2015). Further critique has highlighted the frequency (and duration) of monitoring (Ortúzar et al., 2011), reporting effects (Aschauer et al., 2018), resource intensity (Keseru et al., 2018), response rates and sampling restrictions (Shen and Stopher, 2014), and, understanding the social context, multi-horizon choices and psychological factors (Zmud et al., 2013).

Although travel survey data is perhaps the most broadly used form of data in transport research, there is also an extensive and varied collection of data sets and sources employed (particularly in modelling). These are described in detail in DfT (2014) and summarised in Appendix Table A2**Error! Reference source not found.**.

NEDF have the potential to significantly improve knowledge for the benefit of the transport sector, above that which can be gained from established data (DfT, 2019). There are known limitations of focusing entirely on traditional data (DfT, 2014) and there is great interest in comparing it with new data (such as travel surveys and GPS data – see later discussion in Section **Error! Reference source not found.**). In short, traditional data may be rich in system-oriented content, but generally lacking in context (*"if it moves, count it")*. For example, it is common practice that origin-destination matrices updates are validated based on cordon counts so do not identify the exact route that is taken, nor any detail on the individual and their purpose (and often only assumption on the mode). Furthermore, the focus of many data tends to be on specific problems such as congestion, urbanisation, motorised (and often private) modes, and 'regular' journeys (eg work, school). Although these studies are useful for policy makers who have identified a particular problem, it neglects contextualising the bigger picture, and does not help in recognising what may be unknown or unintended, especially regarding the most vulnerable road users.

2.3 Limitations of traditional modelling and data

To partly address our first research question, in this section we have identified that in general, many of the established models used for transportation planning and research were developed using the data available at the time, e.g., highly aggregated, limited detail on spatio-temporal routes, limited longitudinal, or lacking demographic detail. What was most often lacking were datasets that tracked individuals over time at a fine level of spatio-temporal detail. This limitation gives rise to estimated and extrapolated travel behaviours that may not reliably reflect movement or agents within the system as a whole, restricting the applicability of modelled behaviours and inhibiting the understanding of mobility motivations.

3 New and Emerging Forms of Transportation Data

Having established the limitations of traditional modelling and data for transportation planning and research, the opportunities for NEDF to overcome these limitations must be identified in order to fully address our first research question. In this section we present an overview of NEDF that are particularly relevant to transportation, and in doing so highlight their potential for overcoming the limitations outlined. Furthermore, we contextualise and highlight the specific potential of spatio-temporal mobile phone data and in particular GPS-based T&T data. Whilst the characteristics of NEDF (see Figure 1) are particularly beneficial to transportation planning, it is worth noting that this data is also relevant to transportation-related sectors including for example health care, tourism, and disaster/emergency management.

3.1 NEDF Content, Context and Categorisation

Error! Reference source not found. provides some main examples of the wide range of NEDF related to transportation, their potential technology source and some real world examples.[7] We categorise them into four areas: "GPS tracking", "Smart Ticketing", "Network Connections" and "Multi-media". For an alternative definition of NEDF related specifically to active modes see Lee and Sener (2020), who consider the technologies and applications in terms of if mode is specified or not. Our approach to categorising NEDF in this way is that each type is based around one specific form of data, but we consider the full potential range of data that may be associated with origin data and collection technology. It excludes NEDF that may provide context but are not directly related to mobility, such as smart energy meters and wearable health gadgets (eg "Hive", "Fitbit"). We recognise that some of our examples may not be intuitively 'new and emerging' (indeed, the first "text" was c.3200 BC[8]). However, as characterised in Section

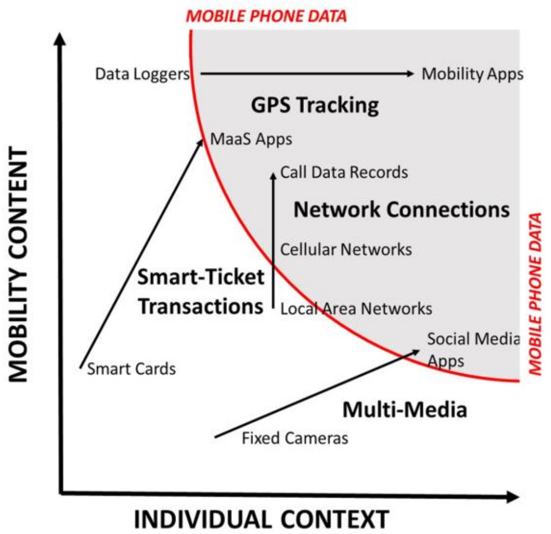
[7] Within NEDF, there exists a degree of interchangeability between the terms for the data itself and the software, technology or device which generates and/or stores the data.
[8] https://sites.utexas.edu/dsb/tokens/the-evolution-of-writing/

1, NEDF is not entirely defined by the source but the data collection process and the volume, velocity and veracity.

Table 1: Examples of NEDF relevant to transport

DATA	TECHNOLOGY / DATA SOURCE	DESCRIPTION
GPS Tracking	GPS loggers Mobile phone applications (eg Strava, Uber, SMART) Wearable technologies (eg FitBit) In-vehicle (eg Tom-Tom)	In public use for over 30 years, GPS technology records exact locations at regular time intervals using the Global Navigation Satellite System (GNSS). Although originally in dedicated devices, GPS is now embedded in many mobile technologies. This data can usually be collected alongside contextual variables (such as weather and personal data) when collected in a dedicated passive mobile phone application, alongside estimated mode.
"Smart" Ticketing Transactions	'Tap-in, Tap-out' (eg Smart-card, mobile phone application, credit card). MaaS mobile phone applications	Smart-ticketing is increasingly utilised by large public transport system operators, allowing seamless, ticketless travel, e.g. the 'Oyster' card in London, UK. Usually the user will 'tap-in' and 'tap-out' at the start and end of a journey, registering location and time, hence 'real-time' trips of individuals can be identified. However only information on public transport journeys are identified, with no knowledge of travel behaviour outside that. Some contextual variables may be associated.
Network Connections	Mobile phone: call data records , cellular towers Local Area Networks: Near Field Communications (eg RFID/Bluetooth), WiFi).	Every time a device (Computer, mobile phone etc) connects to a network, that event is recorded and can be geo-tagged and time-stamped. Individual travel routes may be inferred from these events. These NEDF are limited as they do not provide an exact location or full journey trace and are dependent on the user being active. Although mobile phone network providers may hold limited information (eg name, address) on all or some mobile phone owners, this data may not be available due to privacy issues.
Multi-media (Speech Text Images Videos)	Wearable technology (eg body cameras) Surveillance (eg Dash Cams, CCTV) Satellite and Aerial Mapping (eg google maps) Social Media (eg twitter, facebook, blogs) Other web content (eg google search)	Although there may be some cross-over with traditional forms of Big Data, continuous, geo-tagged and time-stamped multi-media can be analysed to reveal transport behaviours across a population. Longitudinal surveillance and mapping using videos and images are widespread in many countries. People are increasingly engaging online, on social media and other web content, leaving a 'digital footprint' and high volumes of event-driven multi-media. These are highly contextual over a population but may lack detailed content on individual movements.

Our examples of transportation-related NEDF are represented in **Error! Reference source not found.**, which demonstrates (in the authors' view) their relative volume of spatio-temporal **mobility content** (*where* and *when*) and **individual context** (*who, why and what*). These two dimensions were selected as they succinctly demonstrate the advantage of the characteristic "V's" of NEDF in the transportation context. Mobility content includes the geo-spatial location of generated data but also reflects data temporality. The location of fixed location data loggers (eg vehicle counters) is known, whereas mobile sources are assumed to have internal technology that can determine location, for example by GPS or network connections, providing different contextual information. Where spatial data is time-stamped, it could be event-driven or longitudinal. This distinguishes between data generated only when triggered by specific events (eg Ticketing Transactions) and that recorded passively at pre-determined regular intervals (eg GPS tracking). Individual context[9] is characterised here by how much information may be gleaned about the individual related to the mobility data and the reasons for the captured data (eg purpose, preferences, quality). This could be from linked data (eg personal data provided on app registration), inferences from crowd-sourced social-media data, or direct questioning. Information about the trip itself, such as mode, weather or road-types can be captured by internal mobile phone sensors (or other device), or from linked data-sets. It is therefore possible that this type of data may be missing, fragmented or inferred, so may be less reliable than the associated mobility content.



[9] We recognise contextual information is often *derived* from observed patterns and characteristics of the trips particularly when observed over time or obtained through additional surveys, e.g., requests and queries about age, gender, income, etc. What is often highly valuable is the ability to combine the highly detailed mobility content with the individual context for patterns of movement over periods of time.

Figure 2: Content and context of NEDF (see **Error! Reference source not found.**) – authors determination. The arrows show the range of mobility content and individual context that can be expected from each NEDF type, depending on the source origin/collection technology.

When comparing the requirements of T&T datasets (see Figure 1**Error! Reference source not found.**) with the characterisation of NEDF (**Error! Reference source not found.**), those highlighted as having potentially highest resolution mobility content and individual context (primarily associated with mobile phones – i.e. the shaded area in **Error! Reference source not found.**), offer the greatest opportunity to generate T&T datasets.

3.2 Opportunities and Weaknesses of NEDFs

In Section 2, it was established that traditional modelling approaches had been constrained by the limited content and context of the traditional data available. In this section, and in answer to RQ1, we have shown that NEDF offer potentially high resolution alternatives that could substitute or integrate with traditional data and modelling. Furthermore, we have identified that mobile phone technology, and in particular GPS Tracking, present the greatest opportunity for highly grained context and content required to generate T&T datasets.

Although offering great potential, NEDF also have some weaknesses. Two of the biggest challenges are verifying quality and ensuring fitness for use (Li et al., 2016). Data cleaning and structuring is needed before any analysis, requiring significant resources (Fox, 2018), including computing power and working time. For traditional data, well-established algorithms and software have evolved over time for cleaning and structuring (and integrated into processes), but these must be recreated for NEDF. Further complications include extra processing demand from the volume, whilst compatibility with existing data structures, models and tools may be limited. Despite the extra volume of available data, uncertainties will still exist in abundance (Lyons, 2016). Whilst NEDF generate potentially high resolution, individual level data, privacy and security become more important (see Section 5.2).

4 State-of-the-art use of spatio-temporal mobile phone data in transportation research.

Our second research question is: "*How have different types of spatio-temporal mobile phone data been used to date?*". Answering this question allows us to contextualise the unique opportunity for T&T datasets and in particular GPS-based data. A number of authors have presented comprehensive critical reviews of mobile phone data used in transport (Chen et al., 2016, Gadzinki, 2018, Steenbruggen et al., 2015, Wang et al., 2018b, Yue et al., 2014, Wismans et al., 2018, Lee and Sener, 2020), which we avoid repeating. Instead, we have carried out a meta-analysis of these reviews alongside identification of a number of additional studies (listed in Table A3) to present research within our context. We specifically draw out the contribution of the different types of NEDF categorised in Section 3 (GPS Tracking, [Smart Ticketing[10]], Networks and Multimedia[11] – see Table 1 for details) in terms of the characteristics of mobility content and individual context are utilised, which are key for T&T data-sets, and the potential for policy making and planning.

[10] Tap-in, tap-out' applications that integrate smart-card technologies are available for mobile phones, but not considered here. In any case, ticketing data is limited in its contribution to T&T as only public transport interactions are recorded. Although ticketing transaction within 'Mobility as a Service' apps are important, and likely to become increasingly so, no relevant studies were identified. They are most likely to be based on the GPS tracking type apps, with an added ticketing function.

[11] Mainly social media.

4.1 Summary of Meta-Analysis of Spatio-Temporal Mobile Phone Data

A number of NEDF can be gathered (either partially or exclusively) through mobile phones, using either dedicated applications for capturing or influencing travel behaviour or data arising through unrelated applications and general mobile phone operation. Such data is revolutionising research into transport behaviours. It has been used to corroborate the importance of socio-demographics, built environment and trip factors identified in previous research, and can lead to greater insights to be used to develop more complex theories (Chen et al., 2016). However, this new form of data faces challenges of representation, data management and analysis (Li et al., 2016). Wismans et al. (2018) recognise that although research into this type of data is still evolving, it is generally used in two ways: "to gain better understandings about (aggregated) mobility patterns and the use of (dedicated) applications for individual travel patterns". Gadzinki (2018) suggests that research focuses on either methodological aspects or analytical purposes (including for public policies) – though examples of the latter were sparse. Wang et al. (2018b) identified various research applications e.g. travel patterns, mode inference, and travel behaviour analysis. They also found studies typically focused on 'old' topics (e.g. reviewing standard questions, justifying new technologies or verifying existing theories), yet failed to offer much insight into underlying mechanisms. This is echoed by Toole et al. (2015), who note that only abstract insights have been offered so far.

The use of spatio-temporal mobile phone data has been widely used in understanding and identifying 'real-world' travel patterns (through the mobility content), of overall populations (in comparison to traditional travel survey data), or understanding specific socio-demographics and individuals (though the individual context). Other mobile phone data studies related to transportation are more concerned with human interaction, such as tourism/leisure (Beeco et al., 2014, Hasnat and Hasan, 2018, Kubo et al., 2020, Saluveer et al., 2020), emergencies, public safety and crowd management (Steenbruggen et al., 2015, Yabe and Ukkusuri, 2020) or public health (Oliver et al., 2015, Liu et al., 2013).

4.1.1 Mobility Content

T&T datasets are increasingly including multiple data sources for a full and rich mobility trace (though Wang et al. (2019) report trip extraction methods are generally absent), with the key contribution being a form of spatio-temporal data. The most accurate are generally recognised as being those that include time-stamped passively tracked GPS datasets.

Network data provides many valuable insights and offer great potential to improve traditional methods of determining O-D matrices, with numerous studies taking this approach, rather than attempting to create full mobility traces. It has also been combined with traditional data to better understand passenger travel patterns. However, many authors recognise that Origin and Destination tend to be inferred, and Network data lacks continuous and contextual information, so requires augmentation with other data sets. Hemmings and Goves (2016) reviewed the use of Network data for transport modelling recognising potential data bias, trade-offs in specifications and making benchmarking recommendations.

Social media data has had limited application in developing full individual mobility tracing due to the reliance on activity and engagement of users, though the spatio-temporal content can provide valuable information on aggregate movements of a population.

4.1.2 Individual Context

In order to achieve a full T&T dataset, the (spatio-temporal) mobility content must be accompanied by, or be possible to link to, highly detailed individual context data. Various forms of mobile phone data has been used to explore impacts on social segregation (Yip et al., 2016, Silm and Ahas, 2014, Barbosa et al., 2018), identifying differences in spatial and temporal activity between ethnic groups.

Studies that are most successful in identifying individual context have used social media, in particular Twitter and Google maps, and Rashidi et al. (2017) suggest that there is great potential for this type of data in transportation engineering. This provides a rich source of contextual information, especially for aggregated populations, but as with Network data, are dependent on user activity and engagement. There are technical challenges to mining relevant data (Grant-Muller et al., 2015) and currently a risk of population bias, given social media is favoured by younger, technology-literate population segments. This data offers potential for specific niche applications, such as understanding specific events. Wu et al. (2020) identified sub-sets of the population due to their mobility patterns but using anonymised data.

GPS tracking data generally is gathered in applications that also hold (or are linked it) socio-demographic information. Other GPS data and most network data may need to be linked with related data-sets (such as Census or social media data) to imply individual context.

4.2 Potential of spatio-temporal mobile phone data for transportation planning and policy

Studies that demonstrate the practical application of mobile phone data to policy tools and planning processes are limited to supplementation of travel survey data or remain experimental. This is partly due to the novelty of the new data. To date, no standards exist for new methodologies or applications (He et al., 2018, Prelipcean et al., 2018), and research continues on the best way to use it (Wismans et al., 2018).

Gadzinki (2018)) recognised that a great limitation in using this data in public policy making is the sheer volume - requiring time and resources for processes that public authorities often lack. The potential of using such data to not only calibrate models and measure impacts, but also to develop transportation sustainability indicators, overcoming limitations of existing data sets, was discussed by Cottrill and Derrible (2015). The great opportunity to combine data from multiple sources, and to use this as spatial and temporal modelling inputs was recognised by these authors and Wismans et al. (2018). A further opportunity for policy makers is to use these large volume of data for visualisations, making public information more accessible and encouraging more participative policy approaches (He et al., 2018).

It is worth noting that mobile phone applications can in themselves be used to shape travel behaviour (Cohen-Blankshtain and Rotem-Mindali, 2016, Brazil and Caulfield, 2013, Sunio and Schmocker, 2017). Andersson et al. (2018) explored the use of mobile phone applications in promoting sustainable travel behaviour, whilst criticising the lack of grounding in explicit behavioural change theory. Furthermore, they identified the need for more research with more extensive data collections and measuring actual travel behaviour change in order to make informed planning decisions.

Recognising the importance of targeting sub-groups of a population in the promotion of sustainable travel, Semanjski and Gautama (2016) explore how attitudinal market segmentation can be based on crowdsourced mobile phone data, which could then be used to facilitate incentivising specific socio-demographics. One study explored incentivisation to modify travel behaviour using gamification (albeit with a small sample size), but found no statistically significant change attributable to the incentives (Castellanos, 2016). Delclòs-Alió et al. (2017) actively analysed mobile data reported perceptions of travel time, finding that women under-perceive more than men and older people are more likely to miscalculate. Maruyama et al. (2015) and Prelipcean et al.

(2018) both found differences in socio-demographic responses in recruitment and response to policy levers, such as younger people preferring rewards.

Even if socio-demographics were not captured, mobile phone data can be used to inform policy and planning. For instance, Guan et al. (2020) applied network data to the understanding of urban park catchment areas, which could have input into transport as part of wider urban development planning. Similarly, Yu et al. (2020) used GPS data to infer potential bike-sharing trips within a population.

Many of these studies conclude that combining data from traditional survey methods with forms of mobile data to enhance conventional modelling methods offers a more accurate understanding of heterogeneous travel behaviours (Nitsche et al., 2014, Prelipcean et al., 2018). This is reflected in government policy in various countries adopting these approaches (Gask and Williams, 2015, Swier et al., 2015, ONS, 2018, Thomas et al., 2018). It is recognised that ease of use and user acceptance are important (Berger and Platzer, 2015), and participation bias also exists with mobile phone methods, so recruitment and incentivisation need to be considered.

4.3 Spatio-Temporal Mobile Phone Data as Track & Trace Data

Whilst there is great interest in the application of spatio-temporal mobile phone data within transportation research, no studies were identified that specifically demonstrated the use of full T&T datasets in policy tools, modelling or planning for transportation systems, although a number identified this opportunity and offered potential approaches for doing so.

We found that few of the identified studies listed in Table A3 (and considered in the previous discussions) that used spatio-temporal mobile phone data explicitly consider full pathway T&T datasets which include individual context. Those that do are highlighted in Table A3. Of the sixty six studies considered in the analysis, only fifteen

(23%) clearly demonstrated both mobility content and individual context. Two out of twenty seven network studies (7%), two of the twelve social media studies (17%) and eleven out of twenty seven GPS studies (41%). The objective and location of these studies varies widely, but we note that the majority rely on GPS and are related to assessment for use in travel surveys (Donaire-Gonzalez et al., 2016, Geurs et al., 2015, Montini et al., 2015, Safi et al., 2017, Thomas et al., 2018, Zegras et al., 2018). Other objectives for papers considering full T&T datasets are the development of methodologies (Calabrese et al., 2013, Thomas et al., 2018, Toole et al., 2015, Zegras et al., 2018), analysis of travel patterns (Gong et al., 2018, Huang and Li, 2019, Toole et al., 2015) and prediction of preferences (Bantis and Haworth, 2017, Gong et al., 2018, Huang and Li, 2019, Semanjski and Gautama, 2016, Xiao et al., 2016), In general, studies using Network data focus on travel demand (mobility content) whereas those analysing Social Media data concentrate on socio-demographics (individual context). This emphasises the limitations of other NEDF types and highlights the need to focus on GPS-based T&T in the next section. It is possible, however, that some conclusions may be transferable to other more forms of T&T and this could be explored in future work.

5 Opportunities and challenges of GPS-based 'track and trace' data for transportation planning, modelling and policy making

In this section we address our final research question, RQ3, "*What are the opportunities and challenges in using GPS-based T&T data?",* building on responses to other research questions in previous sections. The characteristics of GPS-based T&T datasets (based on the T&T definition presented in Figure 1) are set out explicitly in Figure 3, reflecting that not all apps collect the same data, despite having apparently similar features. This impacts on the range of data potentially available to end users and therefore the degree of opportunity the data provides. In addition, the characteristics of the data *collected* may be more detailed than the characteristics of the data *shared* (under secure arrangements), due to user agreements and privacy considerations. Figure 3 contains examples of the features of mobile phone apps in use over the last 2 years, however many apps can be tailored to collect or process specific data for a particular context. The agile development environment for apps of this type presents a spectrum of opportunities, notwithstanding the need for stable and well-proven detection algorithms. These include focused front-end branding, choices of broad spectrum or niche user information functions (with branding, influencing the nature of the engaged demographic) and purpose-specific additional data collection.

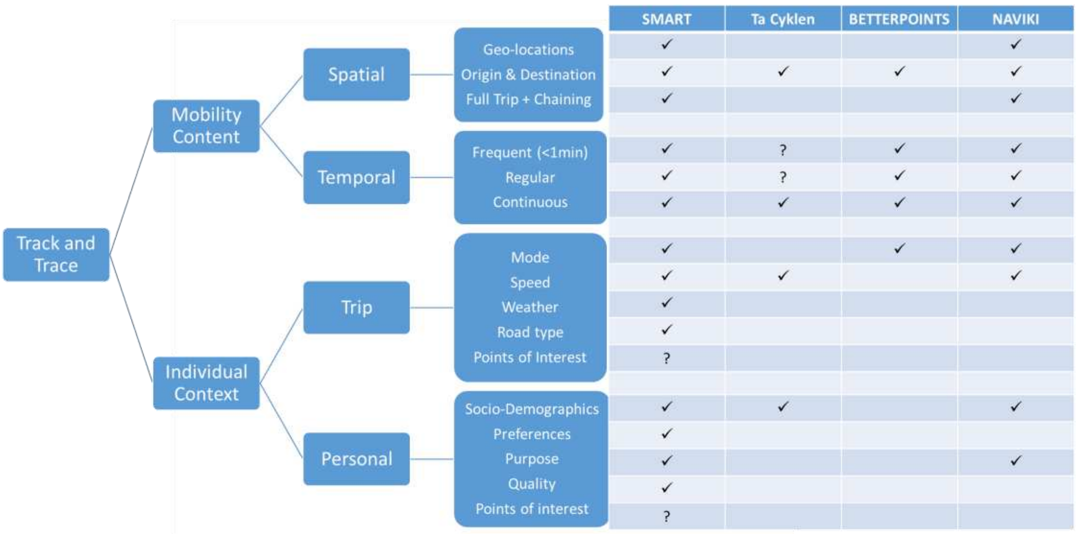


Figure 3: Characteristics of GPS-based T&T data-sets (authors definition), with examples of key mobile phone apps[12] which collect GPS-based T&T data, demonstrating variability in functionality

[12] Ta Cyklen (www.tacyklen.dk/ - in Danish) and Naviki (naviki.org)were apps developed for a specific purpose ie encouraging cycling, SMART (mobidot.nl/en/tools_apps.php)and Betterpoints (betterpoints.ltd/) have more generic application

The UK-based Urban Transport Group (UTG) identified four key challenges for emerging data in transport: sharing and integration; ownership and privacy, quality and standards, and; skills and capabilities (Abrantes and Linton, 2016). These correspond with Milne and Watling's (2018) identified differences in the features and nature of the information from NEDF compared to traditional data forms. They also subsume the three challenges identified in the broader literature: access, privacy and validity (Lazer et al., 2009, Boyd and Crawford, 2012, Kitchin, 2013, Lee and Sener, 2020), so form the basis for the discussion here.

5.1 Sharing and Integration

There is increasing interest in openness, transparency and engagement in planning and policy-making, adopting open-source based software, making public data accessible and encouraging participation in the process. Successful examples of sharing information schemes include for example, traffic road conditions in Waze[13], in which travellers share contemporaneous or past fine-grained spatial and temporal travel data to mutual benefit. The co-design and co-creation opportunities using T&T data in modelling are yet to be exploited. However, T&T data offers the opportunity for the creation of new value, and new business models for a range of stakeholders including individual citizens involved in data generation and analysis (Hodgson, Forthcoming). The opportunity of shared data of this type also entails a challenge in terms of data security however, for example mischievous attempts to sabotage the data with spurious entries.

By its nature, GPS-based T&T data is complex, multi-dimensional and not easily accessed in its raw form. The value of this data arises once processed when it opens up possibilities for system integration (such as improved multimodal services through

[13] www.Waze.com accessed 06/08/19

knowledge of end-to-end journeys), influencing travel behaviour (through improved understanding of route choice and departure times, together with direct messaging, either to the collective, the individual or sub-groups) and traffic management (for example new capability to detect and address build-up of demand on the peri-urban and none-instrumented sections of highway infrastructure). As such, data vendors may offer data-sets ready for analysis (mined, cleaned and smoothed), but these can come at significant cost (Lee and Sener, 2020).

GPS-based T&T datasets provide accurate detail on revealed trip choice (see Figure 3). This may overcome some issues of uncertainty concerning the reliability of input data for transportation modelling that exist for traditional data sets. However, new challenges arise in seeking to utilise this data and the format compatibility needed for modelling, raising a fundamental issue on whether the legacy of existing model structures drives future developments, or whether the new capabilities of data form the basis for a *new type* of modelling. Prelipcean et al. (2018), reviewed the relative advantages of using travel diary methods and integration with new approaches using dedicated devices or smart phones. Based on evidence from multi-national trials they concluded that integration improved representation across demographics. However, integration remains problematic without internationally agreed and consistent definitions of components, as exists for the Harmonised European Time Use Survey (Eurostat, 2019), for example.

This raises the broader question of whether existing transportation models can incorporate and make use of more finely-grained detail, and thus produce more insightful outputs, and whether the individuated nature of the data can be used in models based on aggregated transport system approaches and outputs. Where new data would need to be aggregate to interface with models, the advantage of fine granularity is lost and outweighed by some of the challenges of new data, for example representativeness. One approach which has potentially strong interfaces with GPS-based T&T data is Agent-Based Modelling however, where the fine grain, individualised and context rich nature of GPS-based T&T offers the opportunity to simulate more nuanced and representative ‘agents’ in a model which could ‘bridge’ data and existing models. The interfaces and complementarities between Agent-Based Models and more aggregate network or system wide models is already receiving attention in the research community.

5.2 Ownership and Privacy

A number of risks concerning GPS-based T&T data are principally associated with privacy, data anonymization and informed consent. The datasets provide the capability to target specific travel movements, reveal the identities of individuals or vulnerable groups and sensitive locations. In other words, “have quite an incredible capacity to de-pseudo-anonymise and reveal new information about individuals” (Romanillos et al., 2016) through contemporaneous tracking or pattern recognition and analysis. It has been argued that there are problems with ‘informed consent’ and the use of Application Programming Interface (API) data in particular as “customers accept the terms and conditions without fully understanding them” (Schoonmaker, 2016) (reported in (Taeihagh and Lim, 2019)) but also as citizens report heightened concern about their own cybersecurity throughout Europe (Ofcom, 2017b). Even within the context of the new General Data Protection Regulation (EU, 2016) these present new ethical problems for travel data analysts, transportation system operators and transportation sector planners and policy makers. To date, positive approaches to solutions involve desensitising location data, secure data analytic environments, and aggregation of individual patterns but with attendant drawbacks of loss of definition and precision (EMPOWER, 2018). Where location data can be desensitised so privacy is preserved however, this enables new depths in modelling, for example in capturing area-based social inequalities when modelling proposed transport policies (see Grant-Muller et al. (2018) for example). The heightened awareness around privacy and increased transparency on this issue may also have served to increase the rigour of the privacy processes applied, engendered confidence in app-users and ultimately led to more individuals being willing to share data. Engagement by users in understanding the data they generate and how it is used adds further value.

T&T data also offer new opportunities in the governance of data. Principal among these is the opportunity for citizens to identify the value of the datasets they generate through their own travel, (a corollary to the point made by Milne and Watling (2018) with regard to the analyst not 'owning' the data). Further opportunities arise from the ability to realise that value through modelling approaches, and observatories into travel schemes, interventions and applications in collaboration with others (including public) for the benefit of those involved (Keseru et al., 2018).

5.3 Quality and Standards

Standardisation of data collection and processing can contribute to assuring quality. This applies to both the mobility content and individual context contained within GPS-based T&T datasets. The quality of the data will be reliant on the strength of the signal to the mobile phone throughout the journey, as well as links to external sources. This can be a challenge in particular types of urban environments where the types and density of buildings and land-use overall can lead to low data quality. As a result, some apps are marketed towards inter-urban transportation use as the physical environment and network are less challenging to the location algorithms, and data quality tends to be more consistent. Other apps include a data quality indicator in their standard output files, which can inform the data user of persistent quality issues. In principle, it may be possible to weight the data contribution by quality when used for modelling and mitigate poor data by using associated contextual information for data adjustments. Future research may well include learning algorithms and auto-correction to adjust poor quality data along frequent routes and locations where the built environment compromises quality. The accuracy may be subject to the programming and calibration carried out by the app developer and user inputted corrections, such as modal type, trip purpose or trip segment. Furthermore, for a complete data set, the user must carry the mobile with them at all times, with the app switched on and ensure that the battery does not run out. These possible gaps in the data can be overcome if the data set is available for a long period of time and regular trips can be inferred. This issue of data representativeness raise similar challenges to that of scarcity of cell observations in O-D matrix estimation in some existing modelling approaches. However the fine granularity of T&T O-D data means that this can intensify the problem across spatial areas resulting in a need for aggregation, estimation of an O-D matrix and raises the potential of integration with existing O-D matrices.

The demographic component of GPS-based T&T data, like the behaviour part, is not without its short-comings, and is in fact seen by many as one of the major challenges. The datasets are not necessarily fully representative of a population and could create bias when T&T data sets across a population are aggregated. Technology adoption, diffusion and usage patterns (Grant-Muller et al., Forthcoming) are often systematically biased within cultural contexts and the socio-demographic data collected by the API may be patchy and non-uniform. It is common to find that T&T collection technologies are adopted by younger, more affluent groups (Ofcom, 2017a), although the same technologies are also engaged more purposefully in the context of population sub-groups such as older people or those with disabilities, where they may be running unobtrusively on dedicated personal devices. There are further research opportunities to better understand, support and influence the mobility patterns of such sub-groups using T&T

data. Depending on how an app is promoted, it could appeal to certain demographics or strata within a population, and its purpose could also limit adoption to eg, public transport users. In addition, there is variation among the available collection technologies and applications in terms of maturity and the demographic data they collect from the individual user. To be used to its full potential (e.g. by linking to health or environmental burdens), maximising the available contextual information about the individual is an important step.

Potential solutions to overcome patchy data coverage include augmenting with data about the user from other API or with relevant generic demographic data (eg census and household surveys (ONS, 2018, Thomas et al., 2018)) that may infer characteristics. This could be obtained at regional, city or street levels, depending on the available data on the user's location (inferred by travel patterns). Such proxy methods have advantages and disadvantages associated with the eventual model use and in overcoming issues of missing or mixed quality data (Button, 2019), as well as transferring assumptions into other areas/regions where the app was not used at all or different data sets are available.

Further impacts on data quality may also arise from malicious or manipulative activities that could arise from government espionage, organisational intervention or other "hacking". For instance, there have been concerns over the allowance of certain companies in the installation of 5G networks[14], evidence of data tampering in the "fake news" paradigm[15] and in particular for GPS T&T data, direct activism.[16] Those working with data have a duty to be aware of these issues and the subsequent reliability of models it is based on.

[14]https://www.bbc.co.uk/news/business-51497460, accessed 14/02/20
[15] https://www.cbronline.com/opinion/fake-data-decision-making, accessed 14/02/20
[16]https://www.theregister.co.uk/2020/02/03/google_maps_hack_cartful_phones, accessed 14/02/20

5.4 *Skills and capabilities*

New data and methods will require new skills and inter-disciplinarian practice to exploit the opportunities in applying GPS-based T&T data in the field of transportation planning, modelling and policy. These skills are socio-technical in practice, combining technical skills in data analytics, presentation and interpretation for a fundamentally different type of data (generally of high volume and potentially intermittent) with social skills in the data generation (ie through app based schemes, and through interaction with different stakeholders) and governance, (including issues such as privacy, data sensitivity and ethics of the T&T data). These skills go beyond those of the traditional transport planner or modeller. GPS based T&T data offer opportunities for new business models and collaborative practices that develop the necessary skills and partnerships required to analyse and apply the new data. In-house, out-source and co-creation business models for the use of GPS-based T&T data have very different skillset requirements from different stakeholders (e.g., transportation engineers and consultants, API developers, citizens, planning analysts) and very different partnership collaborations (Hodgson and Grant-Muller, 2015). Seter et al. (2019) recognise this knowledge gap particularly between researchers and consultants, requiring policies and guidelines to support upskilling. As part of this there may be a change in focus away from almost solely road traffic modelling towards a recognition of individual mobility profiles (though there certainly remains many opportunities for more accurate traffic forecasting (Ermagun and Levinson, 2018)).

5.5 *Contribution from GPS-based T&T data to transportation modelling*

Expanding on the above discussion, we now review the potential contribution from GPS-based T&T datasets to overcoming the limitations of traditional data and modelling approaches, and answer our final research question. Table 2 is an adapted version of Appendix Table A1 (*Classification of modelling approaches applied to transport policy* *and planning)***Error! Reference source not found.**, with additional columns drawn from our discussions. The table highlights contributions for each modelling approach of the four characteristics attributes of GPS-based T&T data (as presented in Figure 3), as well as providing short explanations of the opportunities and challenges of the datasets for each modelling approach. What can be seen is that all modelling approaches have potential to be improved or augmented by GPS-based T&T data, though some caution is required regarding the limitations of this NEDF.

Table 2: Overcoming weaknesses of traditional data and modelling approaches with GPS-based T&T data

Modelling Approach *(purpose)*	**Traditional Data Inputs**	**Data and Modelling Approach Weaknesses**	**Potential contribution of GPS-based T&T Dataset Characteristics**				**Additional value from GPS-based T&T data**	**Limitations of GPS-based T&T data**
			Mobility		**Individual**			
			Spatial	**Temporal**	**Trip**	**Personal**		
Traditional Four-step *(Traffic Network Analysis)*	Data inputs specific to purpose: network focus, re-using existing survey eg National Travel Survey and Census/population	• Aggregated • Fragmented • Superannuated • Relies on 'rational' behaviours	✓	✓	?	?	Replacement of assumed network behaviour with actual individual mobility traces and linked preference data.	Development of useful algorithms for predictive models from historic data would require the development of new approaches and integration with other modelling types?
System Dynamics *(Policy/ scenario analysis)*	Data inputs specific to purpose plus historical trend data for calibration.	• Aggregated • Fragmented • Superannuated	?	✓	?	✓	More accurate validation and/or calibration of models due to higher granularity of spatio-temporal data.	Data cleaning, anonymization and processing. Direct input of individualised data would require complex subscription of models.

Modelling Approach *(purpose)*	Traditional Data Inputs	Data and Modelling Approach Weaknesses	Potential contribution of GPS-based T&T Dataset Characteristics				Additional value from GPS-based T&T data	Limitations of GPS-based T&T data
			Mobility		Individual			
			Spatial	Temporal	Trip	Personal		
Agent-Based *(Individual movements/ decisions)*	Data specific to all individuals within system.	• Complex • Many assumptions • High processing power required	✓	✓	?	✓	Individual behaviours do not need to be assumed.	Data cleaning, anonymization and processing.
Discrete Event *(Process and network simulation)*	Data inputs specific to purpose.	• Lacks detail • Many assumptions	?	✓	?	×	Individual traces and behaviour within a network.	Cannot account for uncertainties within a network.
Discrete Choice *(Preferences and attitudes)*	Data collected through carefully designed stated preference surveys.	• Not dynamic • Limited transferability	×	?	✓	✓	Insight from individual real-world revealed preferences could replace stated preferences.	Relies on honest answers and representative engagement. Cannot account for hypothetical/future scenarios.
GIS *(Spatial variations)*	Data inputs specific to purpose. Access to spatial data.	High degree of data processing required.	✓	?	✓	?	More accurate and continuous spatial data.	Data cleaning, anonymization and processing.

6 Conclusions

In this paper we have framed transportation planning and policy in the context of data and modelling needs and the rapidly changing availability of NEDF, in particular GPS-based T&T datasets. We have argued that NEDF overcomes some key limitations of traditional data – mainly in that it can provide high resolution mobility content and individual context, but this in itself brings new challenges of a socio-technical nature. From our discussions, we can draw the following conclusions regarding our original research questions that should provide direction and insight to policy makers, transportation modellers, planners and engineers, and data scientists:

How can NEDF overcome limitations of traditional data and models?

Although traditional data and models remain highly valuable in transportation planning and policy, traditional transportation models have developed subject to the boundaries of traditional forms of data. These are largely aggregate and disparate, lacking individual socio-demographic and spatio-temporal detail. NEDF offer elements of this context, in high volume and at low cost. However, conventional data and models should not become obsolete, but can be augmented by NEDF. There is a general agreement in previous literature that such data require integration with and update of existing methods, though face challenges with capturing representative samples, socio-demographic context and protecting privacy.

How have different forms of spatio-temporal mobile phone data been used to date?

NEDF with relevance to transport can be categorised by the mobility content and individual context it contains, the greatest potential of which are spatio-temporal mobile phone data. Two broad areas of research have been understanding travel behaviours (mainly through comparison to traditional data and surveys), and influencing travel patterns (through targeting and incentivisation). Due to the novelty, to date, most studies have tended to describe the potential data, development of methodologies, descriptive/qualitative results and point to future research, rather than demonstrating any successful policy interventions. There are limited examples of T&T datasets without GPS that offer full tracing pathways for mobility profiles, and less than half of those studies identified in our meta-analysis included individual context within the dataset.

What are the opportunities and challenges in using GPS-based T&T data?

GPS-based T&T datasets are unique not only due to the high-resolution mobility content and individual context they offer compared to other NEDF, but also because of the interaction with users who share that information, that allows for feedback and refinement. This further opens up opportunities for new business models and citizen participation in policy making, as well as creating value for all types of modelling platforms. Similar gaps to traditional data exist, and new socio-technical challenges are raised. It can still lack some individual context and suffer from population bias, induce privacy and consent issues, as well as requiring significant processing prior to analysing, relying on skills that are not traditionally in the realm of transportation modelling and planning. To exploit GPS-based T&T data for transportation planning and policy requires exploration of new value propositions, developing new collaborations and partnerships among stakeholders (local authorities, technology developers, citizens and users, and researchers) to acquire the skillsets required and avoid problems of integration and data availability.

Policy makers draw heavily on data and models from researchers and transport planners in the process of transportation policy decision making. Through NEDF, and particularly GPS-based T&T data, there is an opportunity for citizen engagement that develops a more finely grained picture of up-to-date travel behaviour across the socio-demographics of a population, which is both temporally and spatially reliable, and improves upon limited traditional approaches. This paper has set out the foundations and suggested areas of research and development that are needed to best exploit the opportunities offered. Future research should build on the ideas discussed here and focus on the issues raised for particular modelling types.

REFERENCES

ABRANTES, P. & LINTON, C. 2016. Getting Smart on Data. Challenges and opportunities for transport authorities from emerging data sources. Leeds, UK. Available from: http://www.urbantransportgroup.org/system/files/general-docs/Getting%20Smart%20on%20Data%20-%20Final.pdf: Urban Transport Group.

ALEXANDER, L., JIANG, S., MURGA, M. & GONZÁLEZ, M. C. 2015. Origin–destination trips by purpose and time of day inferred from mobile phone data. *Transportation Research Part C: Emerging Technologies,* 58**,** 240-250.

ALLSTRÖM, A., KRISTOFFERSSON, I. & SUSILOB, Y. 2017. Smartphone based travel diary collection: experiences from a field trial in Stockholm *Transportation Research Procedia* 26**,** 32-38.

ANDERSSON, A., WINSLOTT HISELIUSA, L. & EMELI, A. 2018. Promoting sustainable travel behaviour through the use of smartphone applications: A review and development of a conceptual model. *Travel Behaviour and Society,* 11**,** 52-61.

ASCHAUER, F., HÖSSINGER, R., AXHAUSEN, K. W., SCHMID, B. & GERIKE, R. 2018. Implications of survey methods on travel and non-travel activities. A comparison of the Austrian national travel survey and an innovative mobility-activity-expenditure diary (MAED). *European Journal of Transport and Infrastructure Research,* 18**,** 35.

BACHIR, D., KHODABANDELOU, G., GAUTHIER, V., EL YACOUBI, M. & PUCHINGER, J. 2019. Inferring dynamic origin-destination flows by transport mode using mobile phone data. *Transportation Research Part C: Emerging Technologies,* 101**,** 254-275.

BANISTER, D. & STEAD, D. 2004. Impact of information and communications technology on transport. *Transport Reviews,* 24**,** 611-632.

BANTIS, T. & HAWORTH, J. 2017. Who you are is how you travel: A framework for transportation mode detection using individual and environmental characteristics. *Transportation Research - Part C: Emerging Technologies,* 80**,** 286 - 309.

BARBOSA, H., BARTHELEMY, M., GHOSHAL, G., JAMES, C. R., LENORMAND, M., LOUAIL, T., MENEZES, R., RAMASCO, J. E. J., SIMINI, F. & TOMASINI, M. 2018. Human Mobility: Models and Applications *Physics Reports* 734**,** 1-74.

BEECO, J., HALLO, J. & BROWNLEE, M. 2014. GPS visitor tracking and recreation suitability mapping: Tools for understanding and managing visitor use. *Landscape and Urban Planning,* 127**,** 136-145.

BEECROFT, M. & PANGBOURNE, K. 2015. Personal security in travel by public transport: the role of traveller inforamtion and associated technologies. *IET intelligent transport systems,* 9**,** 167-174.
BEN-ELIA, E. & ZHEN, F. 2018. ICT, activity space–time and mobility: new insights, new models, new methodologies (Guest Editor Introduction). *Transportation* 45**,** 264-272.
BERGER, M. & PLATZER, M. 2015. Field evaluation of the smartphone-based travel behaviour data collection app "SmartMo". *Trasnportstion Research Procedia,* 11.
BLAZY, D., CUILLIER, B., DEROME, C., MOKRANI, K., RENAUD, D., FOCK, S., BOURGEOIS, S. & LEBLOND, V. 2018. To what extent can we utilise mobile phone data? *European Transport Conference 2018.* Dublin, Ireland. 10-12 October 2018: Association for European Transport.
BONABEAU, E. 2002. Agent Based Modeling: Methods and Techniques for Simulating Human Systems. *Proceedings of the National Academy of Sciences,* 99**,** 7280-87.
BOYD, D. & CRAWFORD, K. 2012. Critical Questions for Big Data. *Information, Communication & Society,* 15**,** 662-679.
BRAZIL, W. & CAULFIELD, B. 2013. Does green make a difference: The potential role of smartphone technology in transport behaviour. *Transportation Research Part C: Emerging Technologies,* 37**,** 93-101.
BUTTON, K. 2019. The value and challenges of using meta-analysis in transportation economics. *Transport Reviews,* 39**,** 293-308.
BWAMBLE, A., CHOUDHURY, C. F. & HESS, S. 2017. Modelling trip generation using mobile phone data: A latent demographics approach. *Journal of Transport Geography,* In Press.
CACERES, N., ROMERO, L. M. & BENITEZ, F. G. 2020. Exploring strengths and weaknesses of mobility inference from mobile phone data vs. travel surveys. *Transportmetrica A: Transport Science,* 16**,** 574-601.
CALABRESE, F., DIAO, M., DI LORENZO, G., FERREIRA, J. & RATTI, C. 2013. Understanding individual mobility patterns from urban sensing data: A mobile phone trace example. *Transportation Research Part C: Emerging Technologies,* 26**,** 301-313.
CARREL, A., LAU, P. S. C., MISHALANI, R. G., SENGUPTA, R. & WALKER, J. L. 2015. Quantifying transit travel experiences from the users' perspective with high-resolution smartphone and vehicle location data: Methodologies, validation, and example analyses. *Transportation Research Part C: Emerging Technologies,* 58**,** 224-239.
CASSANDRAS, C. G. & LAFORTUNE, S. 2009. *Introduction to discrete event systems*, Springer Science & Business Media.
CASTELLANOS, S. 2016. Delivering modal-shift incentives by using gamification and smartphones: A field study example in Bogota, Colombia. *Case Studies on Transport Policy* 4**,** 269-278.
CHANDY, R., HASSAN, M. & MUKHERJI, P. 2017. Big Data for Good: Insights from Emerging Markets*. *Journal of Product Innovation Management,* 34**,** 703-713.
CHEN, C., MAB, J., SUSILO, Y. Y. L. & WANGE, M. 2016. The promises of big data and small data for travel behavior (aka human mobility) analysis. *Transportation Research Part C* 68**,** 285-299.

COHEN-BLANKSHTAIN, G. & ROTEM-MINDALI, O. 2016. Key research themes on ICT and sustainable urban mobility. *International Journal of Sustainable Transportation,* 10**,** 9-17.

COLLINS, C., HASSAN, S. & UKKUSURI, S. V. 2013. A novel transit rider satisfaction metric: Rider sentiments measured from online social media data. *Journal of Public Transportation,* 16**,** 21-45.

CORNICK, P., BYRON, C., TEMPLETON, I. & HURN, J. 2018. National Travel Survey 2017 Technical Report. Online: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/729525/nts-technical-report-2017.pdf [accessed 8th January 2019]: NatCen for The Department for Transport.

COTTRILL, C., D.; & DERRIBLE, S. 2015. Leveraging Big Data for teh Development of Transport Sustainability Indicators. *Journal of Urban Technology,* 22**,** 45-64.

CUI, Y., MENG, C., HE, Q. & GAO, J. 2018. Forecasting current and next trip purpose with social media data and Google Places. *Transportation Research Part C: Emerging Technologies,* 97**,** 159-174.

DAVIDSSON, P., HENESEY, L., RAMSTEDT, L., TÖRNQUIST, J. & WERNSTEDT, F. 2005. An analysis of agent-based approaches to transport logistics. *Transportation Research part C: emerging technologies,* 13**,** 255-271.

DAVIES, A., RUTHERFORD, S., BEECH, T., HANSON, P., TOLOUEI, R., WROE, C. & BEFFENDORF, H. 2018. Project EDMOND - Estimating demand from mobile network data in London. *European Transport Conference 2018.* Dublin, Ireland. 10-12 October 2018.: Association of European Transport.

DELCLÒS-ALIÓ, X., MARQUET, O. & MIRALLES-GUASCH, C. 2017. Keeping track of time: A Smartphone-based analysis of travel time perception in a suburban environment. *Travel Behaviour and Society* 9**,** 1-9.

DFT 2014. Tag Unit M 1.2: Data sources and surveys. [Online]. London. Available from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/427119/webtag-tag-unit-m1-2-data-sources-and-surveys.pdf [Accessed 17 November 2018] Department for Trasnport.

DFT 2017. Areas of research interest 2016-2017 [online]. London. Available from: https://www.gov.uk/government/publications/dft-areas-of-research-interest-2016-to-2017 [Accessed 17 November 2018].

DFT 2019. Areas of Research Interest 2019. Research and Evidence. *In:* DEPARTMENT FOR TRANSPORT, H. G., UK (ed.). Available online from: https://www.gov.uk/government/publications/dft-areas-of-research-interest (accessed 03/09/19).

DONAIRE-GONZALEZ, D., VALENTÍN, A., DE NAZELLE;, A., AMBROS, A., CARRASCO-TURIGAS, G., SETO, E., JERRETT, M. & NIEUWENHUIJSEN, M. J. 2016. Benefits of Mobile Phone Technology for Personal Environmental Monitoring *JMIR Mhealth Uhealth* 4**,** 126.

DOT 2018. Strategic Plan for FY 2018-2022. U.S. Department of Transportation.

EC. 2019. *Big Data* [Online]. Available: https://ec.europa.eu/digital-single-market/en/policies/big-data [Accessed 2nd January 2019].

ELIAS, D., NADLER, F., STEHNO, J., KROSCHE, J. & LINDORFER, M. 2016. SOMOBIL – improving public transport planning through mobile phone data analysis. *Tranportation Research Procedia,* 14**,** 4478 - 4485.

EMPOWER 2018. Ethics eBook Ethical Guidance for New and Emerging Data Forms. *In:* PROJECT, E. W. (ed.). http://2pa49x1fw4cm3s4d2s3tqlse-wpengine.netdna-

ssl.com/wp-content/uploads/2018/03/Interactive-Ethics-eBook-1.pdf. Last accessed 04032019.
ERMAGUN, A., FAN, Y., WOLFSON, J., ADOMAVICIUS, G. & DAS, K. 2017. Real-time trip purpose prediction using online location-based search and discovery services. *Transportation Research Part C: Emerging Technologies,* 77**,** 96-112.
ERMAGUN, A. & LEVINSON, D. 2018. Spatiotemporal traffic forecasting: review and proposed directions. *Transport Reviews,* 38**,** 786-814.
EU 2011. White Paper 2011: Roadmap to a Single European Transport Area - Towards a competitive and resource efficient transport system.: European Commission.
EU 2016. REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). *In:* COMMISSION, E. (ed.). Available online from: https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32016R0679 (Accessed 03/09/19).
EUROSTAT 2019. Harmonised European Time Use Surveys (HETUS) 2018 Guidelines, 2019 Edition. Luxembourg: Publications Office of the European Union.
FAN, F., LAU, R. Y. & ZHAO, J. L. 2015. Demistyfying big data analytics for business intelligence through the lens of marketing mix. *Big Data Research,* 2**,** 28-32.
FOX, C. 2018. *Data Science for Transport. A self-study guide with computer exercises.*, Springer.
FRISO, K., JACOB HENCKEL, STEFAN DE GRAAF, PETER VAN DER MEDE, KORF, J., COFFENG, G. & KOOIJ, J. V. D. 2018. The use of mobile phone data in transport planning in the Netherlands: Experiences and Vision. *European Transport Conference 2018.* Dublin, Ireland: AET.
GADZINKI, J. 2018. Perspectives of the use of smartphones in travel behaviour studies: Findings from a literature review and a pilot study. *Transportation Research Part C* 88**,** 74-86.
GASK, K. & WILLIAMS, S. 2015. Comparing travel flows between 2011 Census and Oyster card data. *ONS Methodology Working Paper Series No 5.* Office of National Statistics.
GE, Q. & FUKUDA, D. 2016. Updating origin–destination matrices with aggregated data of GPS traces. *Transportation Research Part C: Emerging Technologies,* 69**,** 291-312.
GEURS, K. T., THOMAS, T., BIJLSMA, M. & DOUHOU, S. 2015. Automatic trip and mode detection with MoveSmarter: first results from the Dutch Mobile Mobility Panel. *Transportation Research Procedia* 11**,** 247-262.
GKIOTSALITIS, K. & STATHOPOULOS, A. 2015. A utility-maximization model for retrieving users' willingness to travel for participating in activities from big-data. *Transportation Research Part C: Emerging Technologies,* 58**,** 265-277.
GONG, L., KANAMORI, R. & YAMAMOTO, T. 2018. Data selection in machine learning for identifying trip purposes and travel modes from longitudinal GPS data collection lasting for seasons. *Travel Behaviour and Society,* 11**,** 131-140.
GRANT-MULLER, S. M., F.C.HODGSON; & MEEUWISSEN., M. Forthcoming. Software apps and public engagement in sustainable transport Living Labs - diffusion and the influence of marketing instruments.
GRANT-MULLER, S. M., GAL-TZUR, A., MINKOV, E., NOCERA, S., KUFLIK, T. & SHOOR, I. 2015. Enhancing transport data collection through social media

sources: methods, challenges and opportunities for textual data. *IET Intelligent Transport Systems* [Online], 9. Available: https://digital-library.theiet.org/content/journals/10.1049/iet-its.2013.0214.

GRANT-MULLER, S. M., GILLIAN HARRISON, FRANCES C. HODGSON, NICK MALLESON & REDFERN, T. 2018. Transport-health equity outcomes from ICT enabled incentives schemes: the Newcastle case study. *hEART 2018: 7th Symposium of the European Association for Research in Transportation.* Athens, Greece.

GSMA 2018. The Mobile Economy. Europe 2018. London, UK. Available from: https://www.gsma.com/mobileeconomy/europe/, accessed 06/12/18: GSMA.

GU, Y., QIAN, Z. S. & CHEN, F. 2016. From Twitter to detector: Real-time traffic incident detection using social media data. *Transportation Research Part C: Emerging Technologies,* 67**,** 321-342.

GUAN, C., SONG, J., KEITH, M., AKIYAMA, Y., SHIBASAKI, R. & SATO, T. 2020. Delineating urban park catchment areas using mobile phone data: A case study of Tokyo. *Computers, Environment and Urban Systems,* 81**,** 101474.

HASAN, S. & UKKUSURI, S. V. 2014. Urban activity pattern classification using topic models from online geo-location data. *Transportation Research - Part C: Emerging Technologies,* 44**,** 363 - 381.

HASNAT, M. M. & HASAN, S. 2018. Identifying tourists and analyzing spatial patterns of their destinations from location-based social media data. *Transportation Research - Part C: Emerging Technologies,* 96**,** 38 - 54.

HE , S. Y., MILLER & SCOTT 2018. Guest editorial: Big data and travel behaviour. *Travel Behaviour and Society* 11**,** 119-120.

HE, W., TAN, E. L., LEE, E. W. & LI, T. Y. A solution for integrated track and trace in supply chain based on RFID & GPS. 2009 IEEE Conference on Emerging Technologies & Factory Automation, 22-25 Sept. 2009 2009. 1-6.

HEMMINGS, T. & GOVES, C. 2016. Utilising Mobile Network Data for Transport Modelling. *In:* PRODUCED FOR UK DEPARTMENT FOR TRANSPORT, D. R. M. R. (ed.).

HENSHER, D. A. & BUTTON, K. J. (eds.) 2008. *Handbook of Transport Modelling*: Elsevier.

HENSHER, D. A. & JOHNSON, L. W. 2018. *Applied discrete-choice modelling*, Routledge.

HILBERT, M. 2016. Big data for development: A review of promises and challenges. *Development Policy Review,* 34**,** 135-174.

HOBBS, A., ; & HANLEY, S. 2014. POSTnote. Big and Open Data in Transport. POST-PN-472. *In:* TECHNOLOGY, P. O. F. S. A. (ed.). Online: https://researchbriefings.parliament.uk/ResearchBriefing/Summary/POST-PN-472 [Accessed 22nd December 2018]: House of Parliament.

HODGSON, F. C. Forthcoming. Social Media Services and Business models. *In:* STEVENS, A. & BURTON, P. (eds.) *Traffic and Travel Information: Collection, Delivery and Development.* London, UK: IET Academic Books Ltd.

HODGSON, F. C. & GRANT-MULLER, S. 2015. Business models for social media data, a report of the CEDR Understanding New and Improving Existing Traffic Data projec. Available from the UNIETD project. .

HOOGENDOORN-LANSER, S., SCHAAP, N. & OLDE KALER, M.-J. 2015. The Netherlands Mobility Panel: An innovative design approach for web-based longitudinal travel data collection. *Transportation Research Procedia (10th International Conference on Transport Survey Methods),* 22**,** 311-329.

HUANG, H., CHENG, Y. & WEIBEL, R. 2019. Transport mode detection based on mobile phone network data: A systematic review. *Transportation Research Part C: Emerging Technologies,* 101**,** 297-312.
HUANG, W. & LI, S. 2019. An approach for understanding human activity patterms with the motivations behind. *Interantional Journal of Geopgraphical Information Science,* 33**,** 385-407.
HUANG, Z., LING, X., WANG, P., ZHANG, F., MAO, Y., LIN, T. & WANG, F.-Y. 2018. Modeling real-time human mobility based on mobile phone and transportation data fusion. *Transportation Research Part C: Emerging Technologies,* 96**,** 251-269.
JÄRV, O., AHAS, R. & WITLOX, F. 2014. Understanding monthly variability in human activity spaces: A twelve-month study using mobile phone call detail records. *Transportation Research Part C: Emerging Technologies,* 38**,** 122-135.
JONES, P. M., M. C. DIX, M. I. CLARKE & HEGGIE, I. G. 1983. *Understanding Travel Behaviour,* Brookfield, USA, Gower.
KELEPOURIS, T., THEODOROUS, L., MCFARLANE, D., THORNE, A. & HARRISON, M. 2006. Track and Trace Requirements Scoping. Cambridge: Available online: http://aero-id.org/research_reports/AEROID-CAM-004-TrackTrace.pdf [Accessed 23 January 2019]: Aero-ID.
KESERU, I., WUYTENS, N. & MACHARIS, C. 2018. Citizen observatory for mobility: a conceptual framework. *Transport Reviews*.
KITAMURA, R. & KUWAHARA, M. 2005. *Simulation approachs in transportation analysis: Recent advances and challenges.,* New York, United States of America, Springer Science + Business Media Inc.
KITCHIN, R. 2013. Big data and human geography: Opportunities, challenges and risks. *Dialogues in Human Geography,* 3**,** 262-267.
KROS, J. F., LIAO, Y., KIRCHOFF, J. F. & ZEMANEK, J., JAMES, E. 2019. Traceability in the Supply Chain. *International Journal of Applied Logistics (IJAL),* 9**,** 1-22.
KUBO, T., URYU, S., YAMANO, H., TSUGE, T., YAMAKITA, T. & SHIRAYAMA, Y. 2020. Mobile phone network data reveal nationwide economic value of coastal tourism under climate change. *Tourism Management,* 77**,** 104010.
LALAWAT, P. 2018. *How Indian Government is using Big Data analytics to improve economy and public policy* [Online]. https://www.analyticsinsight.net/how-indian-government-is-using-big-data-analytics-to-improve-economy-and-public-policy/ [Accessed 23rd Janaury 2019]: Analytics Insight. [Accessed].
LANEY, D. 2001. 3D data management: Controlling data volume, velocity and variety. *META group research note,* 6**,** 1.
LAZER, D., PENTLAND, A., ADAMIC, L., ARAL, S., BARABASI, A.-L., BREWER, D., CHRISTAKIS, N., CONTRACTOR, N., FOWLER, J., GUTMANN, M., JEBARA, T., KING, G., MACY, M., ROY, D. & VAN ALSTYNE, M. 2009. Social science. Computational social science. *Science (New York, N.Y.),* 323**,** 721-723.
LE BLANC, D. 2015. Towards integration at last? The sustainable development goals as a network of targets. *Sustainable Development,* 23**,** 176-187.
LEE, K.-S., YOU, S. Y., EOM, J. K., SONG, J. & MIN, J. H. 2018. Urban spatiotemporal analysis using mobile phone data: Case study of medium- and large-sized Korean cities. *Habitat International,* 73**,** 6-15.

LEE, K. & SENER, I. N. 2020. Emerging data for pedestrian and bicycle monitoring: Sources and applications. *Transportation Research Interdisciplinary Perspectives*, 100095.

LI, S., DRAGICEVIC, S., CASTRO, F. A., SESTER, M., WINTER, S., COLTEKIN, A., PETTIT, C., JIANG, B., HAWORTH, J., STEIN, A. & CHENG, T. 2016. Geospatial big data handling theory and methods: A review and research challenges. *ISPRS Journal of Photogrammetry and Remote Sensing* 115, 119-133.

LINTON, C., GRANT-MULLER, S. & GALE, W. F. 2015. Approaches and Techniques for Modelling CO2 Emissions from Road Transport. *Transport Reviews,* 35, 533-553.

LIU, F., JANSSENS, D., CUI, J., WANG, Y., WETS, G. & COOLS, M. 2014. Building a validation measure for activity-based transportation models based on mobile phone data. *Expert Systems with Applications* 41, 6174-6189.

LIU, H.-Y., SKJETNE, E. & KOBERNUS, M. 2013. Mobile phone tracking: in support of modelling traffic-related air pollution contribution to individual exposure and its implications for public health impact assessment. *Environmental Health* 2013, 12-93.

LORD, D. & WASHINGTON, S. 2018. Chapter 1. Introduction. *Safe Mobility: Challenges, Methodology and Solutions.*

LOVELACE, R., NOWOSAD, J. & MUENCHOW, J. 2019. *Geocomputation with R,* Online: https://geocompr.robinlovelace.net/ [Accessed 7th January 2019], Chapman and Hall/CRC.

LYONS, G. 2016. Transport analysis in an uncertain world. *Transport Reviews,* 36, 553-557.

MA, J., LI, H., YUAN, F. & BAUER, T. 2013. Deriving Operational Origin-Destination Matrices From Large Scale Mobile Phone Data. *International Journal of Transportation Science and Technology,* 2, 183-204.

MARUYAMA, T., SATO, Y., NOHARA, K. & IMURA, S. 2015. Increasing Smartphone-based Travel Survey Participants. *Transportation Research Procedia* 11, 280-288.

MILNE, D. & WATLING, D. 2019. Big data and understanding change in the context of planning transport systems. *Journal of Transport Geography,* 76, 235-244.

MONTINI, L., PROST, S., SCHRAMMEL, J., RIESER-SCHÜSSLER, N. & AXHAUSEN, K. W. 2015. Comparison of travel diaries generated from smartphone data and dedicated GPS devices. *Transportation Research Procedia* 11, 227-241.

MUELLER, N., ROJAS-RUEDA, D., COLE-HUNTER, T., DE NAZELLE, A., DONS, E., GERIKE, R. & NIEUWENHUIJSEN, M. 2015. Health impact assessment of active transportations: a systematic review. *Preventive Medicine,* 76, 103-114.

MURCIO, R., SOUNDARARAJ, B. & LUGOMER, K. 2018. Movements in Cities: Footfall and its spatio-temporal distribution. *In:* PAUL LONGLEY, J. C. A. A. S. (ed.) *Consumer Data Research.* London, UK: UCL Press.

NAO 2017. A short guide to local authorities. [online]. https://www.nao.org.uk/wp-content/uploads/2017/09/A-Short-Guide-to-Local-Authorities.pdf [Accessed 15th January 2019]: National Audit Office.

NI, L., WANG, X. & CHEN, X. 2018. A spatial econometric model for travel flow analysis and real-world applications with massive mobile phone data. *Transportation Research Part C: Emerging Technologies,* 86, 510-526.

NITRD 2016. The Federal Big Data Research and Development Strategic Plan. *In:* EXECUTIVE OFFICE OF THE PRESIDENT, N. S. A. T. C. (ed.). Online: https://www.nitrd.gov/PUBS/bigdatardstrategicplan.pdf [Accessed 7th January 2019]: The Networking and Information Technology Research and Development Program.

NITSCHE, P., WIDHALM, P., BREUSS, S., BRÄNDLE, N. & MAURER, P. 2014. Supporting large-scale travel surveys with smartphones – A practical approach. *Transportation Research Part C* 43**,** 212-221.

NORTH, M. J. 2010. Tutorial on agent-based modelling and simulation AU - Macal, C M. *Journal of Simulation,* 4**,** 151-162.

NSFC 2018. National Science Fund Guide to Programs 2018. Online: http://www.nsfc.gov.cn/english/site_1/pdf/NationalNaturalScienceFundGuidetoPrograms2018.pdf [Accessed 2nd January 2019]: National National Science Foundation of China.

OFCOM 2017a. The Communications Market:UK. https://www.ofcom.org.uk/research-and-data/multi-sector-research/cmr/cmr-2017/uk. Last accessed 04032019. .

OFCOM 2017b. International Communications Market Report. https://www.ofcom.org.uk/research-and-data/multi-sector-research/cmr/cmr-2017/uk. Last accessed 04032019. .

OLIVER, M., PARKER, K., WITTEN, K., MAVOA, S., BADLAND, H., DONOVAN, P., CHAUDHURY, M. & KEARNS, R. A. 2016. Children's out-of-school independently mobile trips, active travel, and physical activity: A cross-sectional examination from the kids in the City study. *Journal of Physical Activity and Health,* 13**,** 318-324.

OLIVER, N., MATIC, A. & FRIAS-MARTINEZ, E. 2015. Mobile network data for public health: opportunities and challenges *Frontiers in Public Health Technology Report.* 10.3389/fpubh.2015.00189.

ONS 2018. Research Outputs: Using mobile phone data to estimate commuting flows. https://www.ons.gov.uk/census/censustransformationprogramme/administrativedatacensusproject/administrativedatacensusresearchoutputs/populationcharacteristics/researchoutputsusingmobilephonedatatoestimatecommutingflows.

ORTUZAR, J. D. & WILLUMSEN, L. G. 2011. *Modelling Transport: Fourth Edition*, John Wiley and Sons.

ORTÚZAR, J. D. D., ARMOOGUM, J., MADRE, J.-L. & POTIER, F. 2011. *Continuous Mobility Surveys: The State of Practice*.

PAULE, J. D. G., SUN, Y. & THAKURIAH, P. 2019. Beyond Geotagged Tweets: Exploring the Geolocalisation of Tweets for Transportation Applications. *In:* UKKUSURI, S. & YANG, C. (eds.) *Transportation Analytsics in the Era of Big Data.* . Cham: Springer.

PRELIPCEAN, A. C., SUSILO, Y. O. & GIDÓFALVI, G. 2018. Collecting travel diaries: Current state of the art, best practices, and future research directions. *Transportation Research Procedia,* 32**,** 155-166.

PUURA, A., SILM, S. & AHAS, R. 2018. The Relationship between Social Networks and Spatial Mobility: A Mobile-Phone-Based Study in Estonia. *Journal of Urban Technology,* 25**,** 7-25.

RASHIDI, T., H.;, ABBASI, A., MAGHREBI, M., HASAN, S. & WALLER, T. S. 2017. Exploring the capacity of social media data for modelling travel behaviour: Opportunities and challenges. *transportation Research - Part C: Emerging Technologies,* 75**,** 197 - 211.

ROMANILLOS, G., AUSTWICK, M. Z., ETTEMA, D. & DE KRUIF, J. 2016. Big Data and Cycling. *Transport Reviews,* 36**,** 114-133.
RUIZ, T., MARS, L., ARROYO, R. & SERNA, A. 2016. Social Networks, Big Data and Transport Planning. *Transportation Research Procedia* 18**,** 446-452.
SAFI, H., ASSEMI, B., MESBAH, M. & FERREIRA, L. 2017. An empirical comparison of four technology-mediated travel survey methods. *Journal of traffic and transportationation engineering* 4**,** 80-87.
SALUVEER, E., RAUN, J., TIRU, M., ALTIN, L., KROON, J., SNITSARENKO, T., AASA, A. & SILM, S. 2020. Methodological framework for producing national tourism statistics from mobile positioning data. *Annals of Tourism Research,* 81**,** 102895.
SAUNDERS, L. E., GREEN, J. M., PETTICREW, M. P., STEINBACH, R. & ROBERTS, H. 2013. What are the health benefits of active travel? A systematic review of trials and cohort studies. *PLOS ONE,* 8**,** e69912.
SCHOONMAKER, J. 2016. Proactive privacy for a driverless age. *Information & Communications Technology Law,* 25**,** 96-128.
SCHWANEN, T., LUCAS, K., AKYELKEN, N., SOLSONA, D. C., CARRASCO, J. A. & NEUTRANS, T. 2015. Rethinking the links between social exclusion and transport disadvantage through the lens of social capital. *Transportation Research Part A: Policy and Practice,* 74**,** 123-135.
SEMANJSKI, I. & GAUTAMA, S. 2016. Crowdsourcing mobility insights – Reflection of attitude based segments on high resolution mobility behaviour data. *Transportation Research Part C* 71**,** 434-446.
SETER, H., ARNESEN, P. & HJELKREM, O. A. 2019. The data driven transport research train is leaving the station. Consultants all aboard? *Transport Policy,* 80**,** 59-69.
SHAFIEI, E., STEFANSSON, H., ASGEIRSSON, E. I., DAVIDSDOTTIR, B. & RABERTO, M. 2012. Integrated Agent-based and System Dynamics Modelling for simulation of sustainable mobility. *Transport Reviews,* 33**,** 44-70.
SHAW, S.-L. & RODRIGUE, J.-P. 2017. Chapter 10: Geographic Information Systems for Transportation (GIS-T) *In:* SLACK, J.-P. R. W. C. C. A. B. (ed.) *The Geography of Transport Systems (Fourth Edition).* New York: Routledge.
SHEN, L. & STOPHER, P. R. 2014. Review of GPS travel survey and GPS data-processing methods. *Transport Reviews,* 34**,** 316-334.
SHEPHERD, S. P. 2014. A review of system dynamics models applied in transportation. *Transportmetrica B: Transport Dynamics,* 2**,** 83-105.
SILM, S. & 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.
STEENBRUGGEN, J., TRANOS, E. & NIJKAMP, P. 2015. Data from mobile phone operators: A tool for smarter cities? . *Telecommunications Policy* 39**,** 335-346.
STERMAN, J. D. 2000. *Business Dynamics: Systems Thinking and Modelling for a Complex World,* Boston, McGraw-Hill Higher Education.
STILLWELL, J. C. H. & CLARKE, G. 2004. *Applied GID & Spatial Analysis,* Chichester, Wiley & Sons.
SUNIO, V. & SCHMOCKER, J.-D. 2017. Can we promote sustainable travel behavior through mobile apps? Evaluation and review of evidence. *International Journal of Sustainable Transportation,* 11**,** 553-566.
SWARTZ, J. 2000. Changing retail trends, new technologies, and the supply chain. *Technology in Society,* 22**,** 123-132.

SWIER, N., KOMARNICZKY, B. & CLAPPERTON, B. 2015. Using geolocated Twitter traces to infer residence and mobility. *GSS Methodology Series No 41.* Office of National Statistics.
TAEIHAGH, A. & LIM, H. S. M. 2019. Governing autonomous vehicles: emerging responses for safety, liability, privacy, cybersecurity, and industry risks. *Transport Reviews,* 39.
TAKO, A. A. & ROBINSON, S. 2012. The application of discrete event simulation and system dynamics in the logistics and supply chain context. *Decision support systems,* 52**,** 802-815.
TEKNOMO, K. 2002. *Microscopic pedestrian flow characteristics: Development of an image processing data collection and simulation model.*, Tohoku University Japan.
THOMAS, T., GEURS, K. T., KOOLWAAIJ, J. & BIJLSMAB, M. 2018. Automatic Trip Detection with the Dutch Mobile Mobility Panel: Towards Reliable Multiple-Week Trip Registration for Large Samples. *JOURNAL OF URBAN TECHNOLOGY,* 25**,** 143-161.
TOOLE, J. L., COLAK, S., STURT, B., ALEXANDER, L. P., EVSUKOFF, A. & GONZÁLEZ, M. C. 2015. The path most traveled: Travel demand estimation using big data resources. *Transportation Research Part C: Emerging Technologies,* 58**,** 162-177.
TOWE, V. L., LEVITON, L., C.;, C., SLOAN, C., TAIT, M. & ORLEANS, T. 2016. Cross-sector collaborations and partnerships: essential ingredients to help shape health and well-being. *Health Affairs,* 35**,** 1964-1969.
TRAIN, K. E. 2009. *Discrete Choice Methods with Simulation* Cambridge University Press.
TSC 2015. The Transport Data Revolution. Investigation into the data required to support and drive intelligent mobility. Milton Keynes, UK.: Transport Systems Catapult.
UKKUSURI, S., V; & YANG, C. (eds.) 2019. *Transportation Analytics in the Era of Big Data,* 10.1007/978-3-319-75862-6: Springer International Publishing.
VAN DIJK, J. & KRYGSMAN, S. 2018. Analyzing Travel Behavior by Using GPS-Based Activity Spaces and Opportunity Indicators. *Journal of Urban Technology,* 25**,** 105-124.
VLASSENROOT, S., GILLIS, D., BELLENS, R. & GAUTAMA, S. 2015. The Use of Smartphone Applications in the Collection of Travel Behaviour Data. *Int. J. ITS Res,* 13**,** 17-27.
WANG, F., WANG, J., CAO, J., CHEN, C. & BAN, X. 2019. Extracting trips from multi-sourced data for mobility pattern analysis: An app-based data example. *Transportation Research Part C: Emerging Technologies,* 105**,** 183-202.
WANG, Y., CORREIA, G. H. D. A., VAN AREM, B. & TIMMERMANS, H. J. P. 2018a. Understanding travellers' preferences for different types of trip destination based on mobile internet usage data. *Transportation Research Part C: Emerging Technologies,* 90**,** 247-259.
WANG, Z., HE , S. Y. & LEUNG, Y. 2018b. Applying mobile phone data to travel behaviour research: A literature review. *Travel Behaviour and Society* 11.
WESTAT 2018. 2017 NHTS Data User Guide. Online: https://nhts.ornl.gov/assets/2017UsersGuide.pdf [Accessed 8th January 2019]: Prepared for the Federal Highway Administration, US Government.

WILLETS, R. 2013. *Eight great technologies [online]* [Online]. Online: https://www.gov.uk/government/speeches/eight-great-technologies Department for Business, Innovation & Skills. [Accessed 2nd January 2019].
WISMANS, L. J. J., AHAS, R. & GEURSA;, K. T. 2018. From the Guest Editors: Mobile Phones, Travel, and Transportation. *JOURNAL OF URBAN TECHNOLOGY,* 25**,** 3-5.
WITLOX, F. 2015. Beyond the Data Smog? *Transport Reviews,* 35**,** 245-249.
WOODARD, D., NOGIN, G., KOCH, P., RACZ, D., GOLDSZMIDT, M. & HORVITZ, E. 2017. Predicting travel time reliability using mobile phone GPS data. *Transportation Research Part C: Emerging Technologies,* 75**,** 30-44.
WU, Y., WANG, L., FAN, L., YANG, M., ZHANG, Y. & FENG, Y. 2020. Comparison of the spatiotemporal mobility patterns among typical subgroups of the actual population with mobile phone data: A case study of Beijing. *Cities,* 100**,** 102670.
XIA, T., NITSCHKE, M., ZHANG, Y., SHAH, P., CRABB, S. & HANSAN, A. ADDIN EN.REFLIST 2015. Traffic-related air pollution and health co-benefits of alterative transport in Adelaide, South Australia. *Environment International,* 74**,** 281-290.
XIAO, G., JUAN, Z. & ZHANG, C. 2016. Detecting trip purposes from smartphone-based travel surveys with artificial neural networks and particle swarm optimization. *Transportation Research Part C: Emerging Technologies,* 71**,** 447-463.
YABE, T. & UKKUSURI, S. V. 2020. Effects of income inequality on evacuation, reentry and segregation after disasters. *Transportation Research Part D: Transport and Environment***,** 102260.
YANG, C., ZHANG, Y., UKKUSURI, S. V. & ZHU, R. 2019. Mobility Pattern Indentification Based on Mobile Phone Data. *In:* UKKUSARI, S. & YANG, C. (eds.) *Transportation Analytics in the Era of Big Data.* . Cham: Springer.
YANG, F., JIN, P., J.;, CHENG, Y., JIAN;, Z. & RAN, B. 2015. Origin-Destination Estimation for Non-Commuting Trips Using Location-Based Social Networking Data. *International Journal of Sustainable Transportation,* 9**,** 551-564.
YIP, N. M., FORREST, R. & XIAN, S. 2016. Exploring segregation and mobilities: Application of an activity tracking app on mobile phone. *Cities,* 59**,** 156-163.
YU, Q., ZHANG, H., LI, W., SUI, Y., SONG, X., YANG, D., SHIBASAKI, R. & JIANG, W. 2020. Mobile phone data in urban bicycle-sharing: Market-oriented sub-area division and spatial analysis on emission reduction potentials. *Journal of Cleaner Production,* 254**,** 119974.
YUE, Y., LAN, T., YEH, A. G. O. & LI, Q.-Q. 2014. Zooming into individuals to understand the collective: A review of trajectory-based travel behaviour studies. *Travel Behaviour and Society* 1**,** 69-78.
ZEGRAS, P. C., LI, M., KILIC, T., LOZANO-GARCIA, N., GHORPADE, A., TIBERTI, M., AGUILERA, A. I. & ZHAO, F. 2018. Assessing the representativeness of a smartphone-based household travel survey in Dar es Salaam, Tanzania. *Transportation,* 2018**,** 335-363.

ZMUD, J., LEE-GOSSELINE, M., MUNIZAGA, M. & CARRASCO, J. A. 2013.

Transport Surveys: Considerations for Decision Makers and Decisions Making. *In:*

ZMUD, J., LEE-GOSSELINE, M., MUNIZAGA, M. & CARRASCO, J. A. (eds.)

Transport Survey Methods. Best Practice for Decision Making. Bingley, UK: Emerald Group dPublishing Limited.

LEE, K. & SENER, I. N. 2020. Emerging data for pedestrian and bicycle monitoring: Sources and applicat

Appendix

Table A1: Classification of modelling approaches applied to transport policy and planning (based on Linton et al. (2015) and expanded by authors)

Modelling Approach	Primary Purpose	Underlying Concepts	Data Requirements	Limitations	Further Reading
Traditional Four-step 1. Trip generation 2. Trip distribution 3. Mode split 4. Route assignment	Traffic Network Analysis	Estimates relationships between trips and characteristics of zones and/or households. Regression or category analysis Gravity model, entropy maximising Utility model eg logit User equilibrium, stochastic methods, microsimulation, activity-based	Data inputs specific to purpose: network focus, re-using existing survey eg National Travel Survey and Census/population	• Aggregated (not reflecting individual characteristics). • Could be based on fragmented and/or superannuated data • Relies on 'rational' behaviours	(Hensher and Button, 2008, Ortuzar and Willumsen, 2011)
System Dynamics	Policy/scenario analysis eg: • Alternative fuel vehicles • Airlines and Airports • Supply chain management	Links qualitative causal link diagrams with quantitative stocks and flow to capture dynamics system feedbacks.	Data inputs specific to purpose plus historical trend data for calibration.	• Could be based on fragmented and/or superannuated data. • Aggregated.	(Shepherd, 2014, Sterman, 2000)
Agent-Based	• Traffic simulation • Urban planning	Bottom-up model of system behaviour through individual and autonomous agents.	Data specific to all individuals within system.	• Complex • High processing power required	(North, 2010, Bonabeau, 2002, Davidsson et al., 2005)
Discrete Event Simulation	• Logistics and supply chain	Based on entity flow and resource sharing at specific events in time.	Data inputs specific to purpose.	• Lacks detail	(Cassandras and Lafortune, 2009, Tako and Robinson, 2012)
Discrete Choice Modelling	• Car ownership preferences • Attitudes • Willingness to Pay	Determines preferences and attitudes of demographics within populations.	Data collected through carefully designed stated preference surveys.	• Not dynamic • Limited transferability	(Hensher and Johnson, 2018, Train, 2009)
Spatial/GIS modelling	Spatial variation, eg: • Car Ownership • Accessibility	Spatial visualisation of data sets.	Data inputs specific to purpose. Access to spatial data.	High degree of data processing required.	(Lovelace et al., 2019, Stillwell and Clarke, 2004, Shaw and Rodrigue, 2017)

Table A2: Examples of sources of traditional transportation data (other than travel surveys) We note that data availability seems to be biased to the Global North, and the selection of examples may be biased to our own experience.

CATEGORY	DATA TYPES	INTERNATIONAL EXAMPLES
Government statistics	Vehicle and driver licensing, vehicle ownership, taxation, emissions, health, census information, national road traffic forecasts	UK: Statistics at DfT US: Bureau of Transportation Statistics NL: Netherlands Institute for Transport Policy Analysis CHINA: Statistical Year Book INDIA: Open Government Data Platform SOUTH ASIA: Transport Data & Statistics EUROPE: Eurostat
Non-Governmental Transport Interest Groups	Ad hoc studies, data repositories	UK: RAC Foundation UK: Better Transport USA: Transportation for America EUROPE: Transport & Environment EUROPE: European Cyclists Federation GLOBAL: Federation Internationale de l'Automobile
Road and highway maintenance	Traffic counts/volumes, (though counters, sensors, Automatic Number Plate Recognition (ANPR) etc), Infrastructure lengths/ condition, investment	UK: Highways England AUSTRALIA: Dept of Planning, Transport and Infrastructure USA: Federal Highway Administration GLOBAL: International Road Federation
Public transport	Passenger demand, routes, pricing	UK: Rail Delivery Group USA: American Public Transport Association GLOBAL: International Association of Public Transport
Road Safety	Road traffic incidents, casualties and fatalities	AUSTRALIA: Safety Statistics KOREA: Traffic Accident Information EUROPEAN: European Transport Safety Council
Vehicle manufacturers and traders	Vehicle sales and predictions, Type approvals and emissions, Ad hoc reports	UK: Society of Motor Manufacturers and Traders EUROPEAN: European Automobile Manufacturers Association GLOBAL: International Council on Clean Transportation

Table A3: Studies that collect and/or analyse spatio-temporal mobile phone data "T&T" column indicates (Yes/No) if the study explicitly utilises a Track & Trace dataset (as defined in this paper). "IC" column indicates (Yes/No) if the T&T dataset includes individual context (as defined in the paper) .Highlighted rows indicate studies that explicitly use a T&T dataset including individual context.

NEDF CATEGORY	ARTICLE	OBJECTIVE	LOCATION	T&T	IC
NETWORKS	Calabrese et al. (2013)	Methodology for developing mobility traces using CDR and cellular network connections	USA (Boston)	Y	Y
	Ma et al. (2013)	Used network data with traditional data to estimate OD matrices but recognised limitations arising from unknown modes and capturing short trips, as well as low individual context.	USA (Sacremento)	Y	N
	Järv et al. (2014)	Uses network data to analyse seasonal variation in individual activity spaces.	Estonia (Tallinn)	N	Y
	Liu et al. (2014)	Constructs OD matrix using CDRs and compares to traditional data.	Ivory Coast	N	N
	Silm and Ahas (2014)	Temporal difference in ethnic segregation using CDRs. Residential areas more segregated.	Estonia (Talin)	N	Y
	Alexander et al. (2015)	Estimation of OD matrices from CDR validated against travel surveys	USA (Boston)	N	N
	Oliver et al. (2015)	Reviews call and network data for mobility related public health.	N/A	N	N
	Toole et al. (2015)	Integrates CDR, census and road network surveys, created web visualisations to explore patterns.	USA (Boston, San Francisco), Brazil (Rio de Janerio), Portugal (Lisbon)	Y	Y
	Chen et al. (2016)	Traces from call and network data. Compares characteristics / opportunities against survey data.	N/A	N	N
	Elias et al. (2016)	Development of method to improve traffic modelling for public transport planning	Austria	Y	N
	Bwamble et al. (2017)	Proposed OD trip generation model where demographic groups treated as unobserved.	Switzerland (Lausanne)	N	Y
	Blazy et al. (2018)	Formulation of real-time mobility statistics - validating OD matrices from CDR and networks connections with traditional data (travel surveys and ticketing).	France (Paris)	N	Y
	Davies et al. (2018)	Traffic demand estimation (OD matrix) from CDR, cellular and wifi network connections, data fusion with London travel survey.	UK (London)	N	Y
	Huang et al. (2018)	Combines network data with traditional data to predict traffic demand and severe congestion.	China (Shenzhen)	N	N
	Friso et al. (2018)	Developing OD matrices for traffic flows from network data.	Netherlands	N	N
	Lee et al. (2018)	Uses network data to compare urban activities between cities	South Korea	N	N
	Murcio et al. (2018)	Estimation of retail footfall from wifi "SmartStreetSensors"	UK	N	N
	Ni et al. (2018)	Used CDRs for OD flow modelling and discussed policy implications for facilities provisions and transit accessibility.	China (Hangzhou)	N	N

	Puura et al. (2018)	Used network data to establish relationships between activity spaces and social networks, finding a strong gender influence.	Estonia	N	N
	Wang et al. (2018a)	Used network data to create mobility traces with google maps poi data and mobile internet usage to segment population by travel preferences.	China (Shanghai)	N	N
	Huang et al. (2019)	Approaches to mode detection using mobile phone data were reviewed finding most relied on network data and favouring spatial proximity to network, though few are fully validated.	N/A	N	N
	Yang et al. (2019)	Model location sequences from CDR and clusters users mobility patterns	China (Shenzhen)	N	N
	Bachir et al. (2019)	Creates OD flows for specific transport modes	France (Paris)	N	N
	Caceres et al. (2020)	Used passive and active network data to develop OD matrices and compare to traditional travel survey data	Spain (Malaga)	Y	N
	Guan et al. (2020)	Derived urban park catchment areas	Japan (Tokyo)	N	N
	Wu et al. (2020)	Identified population subsets through mobility patterns	China (Beijing)	N	N
SOCIAL MEDIA	Collins et al. (2013)	Evaluation of transit ride satisfaction from Twitter data.	USA (Chicago)	N	Y
	Hasan and Ukkusuri (2014)	Uses Twitter/Foursquare geo-tags to identify individual travel activities and patterns.	USA (New York)	N	Y
	Gkiotsalitis and Stathopoulos (2015)	A method for using crowd-sourced twitter data to assess willingness to travel for certain activities.	UK (London)	N	Y
	Swier et al. (2015)	Used twitter for aggregate inferring of residence and mobility patterns for consideration by the ONS, finding some key instabilities in the reliance of social media data.	UK	N	N
	Yang et al. (2015)	OD Matrices from 'Foursquare' log in data	USA (Chicago)	N	Y
	Gu et al. (2016)	Identifying traffic incidents using Twitter data	USA (Philadelphia)	N	N
	Ruiz et al. (2016)	Analysis of sentiments of travel, uses sensors and CDR to infer trip	Spain	N	Y
	Ermagun et al. (2017)	Combined traditional travel surveys and google places to predict trip purposes	USA Minneapolis-St. Paul	N	Y
	Cui et al. (2018)	Combined crowd-sourced Twitter data and Google Places to determine and predict trip purposes.	USA (California)	N	Y
	Hasnat and Hasan (2018)	Analyses spatio-temporal patterns from Twitter data, clustering tourists and residents.	USA (Florida)	Y	Y
	Huang and Li (2019)	Analyses spatio-temporal patterns and infer mobility motivations from geo-tagged tweets	Canada (Toronto)	Y	Y
	Paule et al. (2019)	Assessment of the differences in spatial and content coverage between non-geotagged tweets geolocalised using different approaches .	N/A	N	Y
GPS	Liu et al. (2013)	Proposed methodology that infers vehicle trajectories using GPS for health / pollution exposure	N/A	N	N
	Nitsche et al. (2014)	Combining GPS smartphone data with travel survey.	Austria (Vienna)	N	N

	Berger and Platzer (2015)	Evaluation of app for travel survey. Potential privacy issues.	Austria	N	Y
	Carrel et al. (2015)	Used GPS T&T data to assess public transit network performance.	USA (San Francisco)	N	Y
	Donaire-Gonzalez et al. (2016)	Compared to tracker, app was better	Spain (Barcelona)	Y	Y
	Ge and Fukuda (2016)	Uses aggregated GPS traces to estimate OD matrices.	Japan (Tokyo)	Y	N
	Geurs et al. (2015)	Evaluation of app performance, being used for Dutch mobility panel	Netherlands	Y	Y
	Maruyama et al. (2015)	Methods to increase participation – young preferred rewards.	Japan (Kumamoto)	N	Y
	Montini et al. (2015)	Compared to GPS tracker – GPS better	Switzerland (Zurich)	Y	Y
	Vlassenroot et al. (2015)	Comparison of different apps	Belgium (Flanders)	N	N
	Castellanos (2016)	Explores incentives to modify travel behaviour – monetary, non-monetary, gamification – no statistical difference.	Columbia (Bogota)	N	Y
	Semanjski and Gautama (2016)	Attitude segmentation combined within GPS tracking app for suggestion responsiveness to policy options	Belgium (Flanders)	Y	Y
	Xiao et al. (2016)	Combined GPS data with travel survey responses to predict trip purposes with relatively high accuracy.	China (Shanghai)	Y	Y
	Yip et al. (2016)	Tracks mobility and explores impact on social segregation	Hong Kong	N	Y
	Allström et al. (2017)	Evaluates app for collecting travel diary – found feasible.	Norway (Stockholm)	N	N
	Bantis and Haworth (2017)	Combines GPS/GIS data with environmental and socio-demographic data to better understand individual mobilities	UK	Y	Y
	Delclòs-Alió et al. (2017)	Differences between perceived and real time perception. Women under-perceive, older more likely to miscalculate.	Spain (Barcelona)	N	Y
	Safi et al. (2017)	Evaluates performances of tracker, smartphones apps and web–based. Smartphone lowest quality, older preferred web.	New Zealand	Y	Y
	Woodard et al. (2017)	Used aggregated GPS to assess congestion and predict travel times.	USA (Seattle)	Y	N
	Gadzinki (2018)	Literature review and pilot. Collected traces through an app and individual data through a survey.	Poland (Poznan)	Y	Y
	Gong et al. (2018)	Explores seasonal data reliability for identifying trip mode and purpose.	Japan (Hakodate)	Y	Y
	Thomas et al. (2018)	Use of an app in the Dutch mobility panel.	Netherlands	Y	Y
	Prelipcean et al. (2018)	Integrating GPS smart phone data with national travel surveys	Sweden	N	Y
	van Dijk and Krygsman (2018)	Data from a GPS tracking app to explore individual activity spaces in conjunction with travel survey data, mapping data, and sociodemographic accessibility opportunity and willingness to change travel behaviours.	South Africa (Stellenbosch)	N	Y
	Zegras et al. (2018)	Uses GPS based tracking ("future mobility sensing") for travel survey data, noting the usual limitations of	Tanzania (Dar es Salaam)	Y	Y

		bias and verification and also provides an overview of how the technology works.			
	Wang et al. (2019)	Propose a framework for trip extraction from multi-sourced data, based on an app.	USA (Puget Sound)	Y	N
	Yu et al. (2020)	Infer potential bike-sharing trips within the population using GPS OD	Japan (Tokyo)	N	N