Journal of Transport Geography 76 (2019) 235–244

Contents lists available at ScienceDirect

Journal of Transport Geography

journal homepage: www.elsevier.com/locate/jtrangeo

Big data and understanding change in the context of planning transport systems

Dave Milne[*], David Watling

Institute for Transport Studies, University of Leeds, UK

ABSTRACT

This paper considers the implications of so-called ‘big data’ for the analysis, modelling and planning of transport systems. The primary conceptual focus is on the needs of the practical context of medium-term planning and decision-making, from which perspective the paper seeks to achieve three goals: (i) to try to identify what is truly ‘special’ about big data; (ii) to provoke debate on the future relationship between transport planning and big data; and (iii) to try to identify promising themes for research and application. Differences in the information that can be derived from the data compared to more traditional surveys are discussed, and the respects in which they may impact on the role of models in supporting transport planning and decision-making are identified. It is argued that, over time, changes to the nature of data may lead to significant differences in both modelling approaches and in the expectations placed upon them. Furthermore, it is suggested that the potential widespread availability of data to commercial actors and travellers will affect the performance of the transport systems themselves, which might be expected to have knock-on effects for planning functions. We conclude by proposing a series of research challenges that we believe need to be addressed and warn against adaptations based on minimising change from the status quo.

1. Introduction

In recent years there have been enormous technological advances in the capture and storage of data, affecting our potential to monitor both human behaviour and the physical world, and providing possibilities to track and triangulate diverse data sets. A report by the OECD (2013) identified the following types of new data on ‘the human condition’:

- Data from government transactions (e.g. tax, social security)
- Data related to official registration/licensing
- Commercial transactions by individuals and organisations
- Internet data from search and social networking activities
- Tracking data
- Image data (e.g. aerial/satellite images, land-based video)

All of these are potentially relevant to planning transport systems since they either provide insights into the location, timing and frequency of activities that generate travel (such as employment, shopping or social engagement), or they provide direct evidence of the volume, concentration and direction of person movements or vehicular movements. These opportunities have been explored to varying degrees. Saadi et al. (2016) used social security data to infer trip patterns. Chatterton et al. (2015) used annual vehicle test data to understand social variations in vehicle use. De Montjoye et al. (2015) used credit card data to reconstruct individual movements. Smart card data has been widely used by many researchers (e.g. Pelletier et al., 2011; Tao et al., 2014; Tamblay et al., 2016), with a range of applications in public transport planning. Location-based, social media check-in data from services such as Foursquare and Twitter has been used by a range of researchers (e.g. Hasan and Ukkusuri, 2014; Liu et al., 2014; Yang et al., 2014; Abdulazim et al., 2015; Hu and Jin, 2017) to attempt to estimate travel activity patterns, while Jestico et al. (2016) have investigated the potential of the activity tracking app STRAVA for measuring cycling volumes. Social media data have also been used to examine unexpected situations and special events (Pender et al., 2014; Pereira et al., 2015; Gu et al., 2016). Automatic vehicle identification has been used to understand complex travel activity patterns (Ozbay and Ercelebi, 2005; Siripirote et al., 2014). Perhaps most widely exploited for understanding travel/mobility patterns has been mobile phone data (e.g. Calabrese et al., 2013; Blondel et al., 2015; Widhalm et al., 2015) and GPS data (e.g. Frignani et al., 2010; Lin and Hsu, 2014; Montini et al., 2014; Tang et al., 2015; Gong et al., 2016). Returning to the original OECD list of data sources we might add other sources of transportation data not directly related to the ‘human condition’, such

[*] Corresponding author.
E-mail address: d.s.milne@its.leeds.ac.uk (D. Milne).

https://doi.org/10.1016/j.jtrangeo.2017.11.004
Received 9 February 2017; Received in revised form 7 October 2017; Accepted 6 November 2017
Available online 19 April 2018

as spatial mapping of landscapes, human infrastructure and related properties (e.g. rail track condition, emissions levels and noise).

Alongside this data revolution there has been significant social change in how individuals plan movements, network socially, use mobile devices, and accept to provide personal details for commercial, registration and governmental transactions (Chung et al., 2015; Grindrod, 2016; Maréchal, 2016; McCarthy et al., 2016; Fortunati and Taipale, 2017). These changes are leading to rapid growth in both digitised and 'non-purpose-oriented' data. In particular, there is increasing availability of large quantities of data that either: (i) were not specifically captured for transport applications but which have relevance to its understanding; or (ii) have been archived to preserve historical 'pictures' of transport patterns, but without an analytical method in mind. This process completely contradicts classical methods of statistical inference: experimental design, random sampling, even (it might be claimed) 'the scientific method' (Mazzocchi, 2015; Kwan, 2016). It seems clear that something quite fundamental is happening beyond data-sets simply increasing in size and becoming digitised.

Considering these issues from the perspectives of transport system analysis and planning, this paper seeks to achieve three goals:

(i) to try to identify what is truly 'special' about big data in this particular context;
(ii) to provoke debate on the future relationship between transport planning and big data; and
(iii) to try to identify promising themes for research and application.

As authors we approach this topic from a background in transport modelling and have a particular interest in the opportunities, threats and permissive impacts that these emerging data sources may have for travel, and for our ability to understand and predict it in support of planning and decision-making. The focus of our paper is on medium-term transport planning and not on providing real-time system or personalised information, analytics and control, which have been covered elsewhere and which lead to a very different set of considerations (e.g. Zheng et al., 2013; Kitchin, 2014a, 2014b; Mori et al., 2015; Oh et al., 2015).

The paper will also avoid focussing on a single type of big data and its detailed applications, such as mobile phone records which have been covered quite extensively elsewhere (Bohte and Maat, 2009; Calabrese et al., 2014; Steenbruggen et al., 2015; Rojas IV et al., 2016). Instead, the paper will consider any kind of 'big data' (in the senses we shall define) and will discuss: how it differs from more traditional transport data sources; what those differences mean for the information it can provide; how different information has an impact on the analytical techniques that make use of it (from the perspectives of both analysing past data and devising predictive models), and what the implications of all this are for the practice of transport planning. Our ultimate aim is to try to identify what we believe are key questions and areas for debate, which might suggest foci for research regarding the potential impact of big data on the future of transport planning and policy assessment. Our starting point is a simple conceptual understanding of the relationship between data about the transport system and the practical discipline of transport planning. Our proposition is that, in the transport sector, analytical approaches—most of which could be referred to as models—have traditionally played a vital role in linking observable phenomena to decision-making. Therefore, as the volume and nature of observations changes, it is necessary to consider how that may impact on models that are intrinsic to prevailing planning and decision-making processes.

The paper we present is deliberately a mix between a literature review and a think-piece about the implications of such new data opportunities for the theory and practice of transport geography. The content is divided into three major sections. First (in Section 2) we consider how the term 'big data' might be usefully defined in a transport planning context, paying particular regard to what is distinctive about it. In identifying its distinctiveness, we reflect on its relation to more conventional transport data sources and the ways in which this may transform the questions we ask of transportation and travel data. Second (in Section 3) we discuss the potential implications of different types of data for analysing and modelling transport systems. Here we particularly consider the likely future impacts on both those that use models to plan transport systems in real life and those that provide the modelling tools to do so. Third (in Section 4) we attempt to identify the main opportunities and limitations of big data for achieving better transport planning outcomes. In particular, our interest is in the potential transformative aspects of the 'big data revolution', and how it may set off a cycle of impacts for both real-world transport systems and those that aim to understand them. In Section 5 we draw some concluding remarks.

Table 1
Definitions of 'big data' from literature.

"any data that cannot fit into an Excel spreadsheet" (Batty, 2013)
3 V = Volume, Velocity, Variety (Laney, 2001)
5 V + C = 3 V + Variability, Veracity, Complexity (GSR, 2016)
3 V + 2R + E + F = 3 V + Resolution, Relational, Exhaustive, Flexible (Kitchin, 2013)
"Big Data is notable not because of its size, but because of its relationality to other data" (Boyd and Crawford, 2011)
"The end of theory" (Anderson, 2008)

2. Defining 'big data' from a transport planning perspective

There exist many possible general characterisations of big data. Some of those included in previous literature are provided in Table 1, and are explained further below.

The simplest characterisation of big data, typically acknowledged in all definitions, relates to a greatly increased *volume* of information compared to past expectations, and the fact that this can sometimes lead to practical difficulties (or *complexity*) for those who wish to analyse it using existing methods and tools. *Velocity*, on the other hand, refers to an increased temporal frequency of observation. References to *variety*, *variability* and *veracity* are all acknowledgements of the diversity likely to be inherent within non-purpose-oriented data that may mean it does not conform to convenient structures and definitions, or that it may lack accuracy. Other possible features identified include the suggestion that big data may provide complete (*exhaustive*) coverage of a system or population, and that it may be fine grained in *resolution* and detail (a quality potentially related to velocity). The final area on which some definitions focus is the potential ability of big data sources to be used in combination to provide novel insights through the *relational* qualities of what is recorded (e.g. the ability to link observations of different activities via the common fields of time and space) and the associated *flexible* nature of the information. This leads to the suggestion that big data could result in *the end of theory* because it implies a shift towards insight and understanding being driven by empirical evidence rather than starting from theoretical constructs.

Although thought provoking, the lack of agreement between these suggested combinations of features does not provide a sufficiently clear application-oriented definition for our purposes. From a transport planning perspective, the features described in Sections 2.1–2.7 (we propose) provide a more useful characterisation of emerging data sources that fit the 'big data' concept within our chosen context. We are not proposing that these features are necessary conditions that should all be met in order for data to be considered 'big' in a transport planning sense. Indeed, our aim is to move away from hard and fast definitions, because we suspect they are not helpful (especially as technologies change over time), and to focus instead on features that have the potential to change how people think about data and how they use it.

2.1. Continuous monitoring

'Continuous monitoring', as we shall define it, is perhaps one of the most important features in terms of the new opportunities it affords for analysis and insight. In our view, it refers to situations where data, of attributes such as personal and vehicle mobility, possess one or more of the following facets: (i) a high temporal sampling rate (i.e. observations are made so frequently that they are almost/seemingly continuous); (ii) monitoring for which there are no gaps (i.e. as opposed to a high temporal sampling rate covering 08:00–09:00 each working day but with no observations at other times); and (iii) data that are being recorded indefinitely (i.e. with no defined end-time of observation). There are many implications of such facets. For example, gapless historical monitoring means that it is possible to select a sample of past data for analysis as often as is useful, and potentially to go back in time as far as desired (to the start of monitoring, assuming data is retained indefinitely). Over time this creates a historical database that is always growing. Examples of use of this type of data related to transport might include that from mobile devices such as cell phone records and Bluetooth data (Widhalm et al., 2015; Crawford et al., 2017b), electronic ticketing data for public transport (Tamblay et al., 2016), data from fixed sensor, GPS and automatic vehicle identification (AVI) technology such as loop detectors, automatic traffic counters, Trafficmaster and Automatic Number Plate Recognition (ANPR) cameras (Chow, 2016), and location-based social network data such as Foursquare and Strava (Hu and Jin, 2017; Jestico et al., 2016). Even if there are only resources to analyse a limited (but large) sample of observations for any particular study, this sample need not be time-constrained a priori (e.g. all observations in the last year), but might be sampled in some other way that allows temporal changes to be better understood (e.g. automatic traffic counts for a city covering all Januaries in the last ten years). Thus, such data differs from traditional transport-related surveys because there are no beginning and ending points imposing limited temporal windows during which information is available.

2.2. Data may not be owned by the data analyser

Data ownership and use are complex and controversial issues in situations where information related to people's activities and movements may be passively collected, are continuously monitored, could potentially be used to identify individuals, and are retained indefinitely (Mayer-Schönberger, 2010; Cate and Mayer-Schönberger, 2013; Mayer-Schönberger and Cukier, 2013). Traditional surveys of travel activity, based on active data collection, work on the assumption that any organisation which commissions surveys has ownership rights over the information gathered and is also responsible for ensuring that the personal rights of people surveyed are not infringed. As part of this, it is typically the case that data would only be passed to third parties in a form where it could not, for example, be used to identify individuals. In addition, the purposes for which a third party could use such data might be prescribed to exclude issues that could work against the interests of the data owners or fall beyond the consent considered to have been provided by participants during the survey process. This creates a challenge for the use of non-purpose-oriented data for transport planning purposes, as by definition it is likely that data will need to be transferred away from the original owners and used for purposes that were not originally envisaged. For example, it is not unusual for privately owned public transport operators to be unwilling to release ticketing data to transport planners on the grounds that it is commercially sensitive; likewise, mobile phone companies are often only willing to release movement trajectories in an aggregated form to disguise individual identities. For location-based social networks it may not always be clear whether data ownership rights lie with the host organisation or the individual who decided to post their information, and the existence of privacy settings may allow individuals to be selective about which information is available. However, researchers in this area have noted that privacy functions which provide a combination of anonymization, user consent and open access (e.g. the so-called 'Foursquare-Twitter bridge') potentially provide a convenient solution to data ownership problems, albeit at the cost of a possible significant reduction in sample size (Hu and Jin, 2017).

2.3. Data collection may not have been designed for the purpose

In addition to issues of ownership rights, privacy and costs associated with acquiring data for planning purposes, the reuse of information that was not designed for the purpose provides both challenges and opportunities.

Obvious problems related to data reuse are that the information available may not contain all the desired attributes or may not be structured in ways that are easily compatible with established methodologies. In the context of data relating to spatial mobility and travel activity, an example of this is that movement trajectories based on cell phones or Bluetooth devices are unlikely to contain information related to the purposes of journeys or the modes used, both of which would be collected as standard in traditional travel surveys. This has, meant that new techniques have needed to be developed to estimate those elements (Abdulazim et al., 2013; Bhaskar and Chung, 2013; Bwambale et al., 2017; Crawford et al., 2017c). A further issue is that the samples included in external datasets may not be fully or proportionally representative of the populations and activities being considered during spatial and transport planning. For, despite Kitchen's (2013) assertion that big data may have an exhaustive quality, in the current era engagement with mobile technology tends to be skewed towards particular types of (typically younger) people (Yang et al., 2014; Sun and Li, 2015) and, for location-based social media data, technology usage levels vary considerably between individuals, between different times of day and a skewed towards particular types of activity (Sun and Li, 2015; Hu and Jin, 2017).

Opportunities provided by non-purpose-oriented data at the current time might include larger volumes of information and greater levels of spatial and temporal detail than have previously been available from traditional surveys. In many situations, however, these advantages are likely to be temporary, as advances in technological developments over time—and associated reductions in cost—should enable transport planning organisations themselves to conduct purpose-oriented surveys with similar features. In contrast, the opportunity that is unique to big data is the potential to identify patterns that are unlikely to have been observed through traditional methods of investigation. This is an area which is yet to have a major impact on research within spatial and transport planning. As an illustration of the potential which transport studies may aspire to achieve, we may look to the field of medical research. In this field, Mayer-Schönberger, 2016 reports work in which large volumes of routine monitoring data (e.g. heart-rate, pulse, breathing) were used to identify elevated risk of infection in premature babies at a much earlier stage than traditional methods (based on identifying known symptoms) were able to do. A key feature of this study was that it provided sound evidence of a correlation in an area that traditional methods would potentially not have explored, and thereby opened up the possibility of corrective interventions before the causal links were fully understood. In the area of human mobility, there has traditionally been great focus on quantitative predictions of travel activity and on the valuations of benefits resulting from planning interventions. However, these predictions and valuations frequently struggle to capture the complexities of (and temporal changes in) human populations and their behaviour, suggesting that the potential for new insights based on more diverse mobility-related datasets would appear to be immense.

2.4. Data may be acquired from outside transport as currently studied

A natural extension of the opportunities provided by non-purpose-oriented data, discussed above, is the potential to access data beyond the traditional scope of studying spatial mobility and travel activity. The emergence of big data provides the opportunity to find new explanatory variables that are either beyond the currently-received wisdom, potentially requiring understandings and approaches from outside the transport field, or that were not previously quantifiable. In particular it allows for the possibility of wider exploration of data-mining approaches that may reveal new insights about patterns of behaviour and their causes.

For example, involvement of clinical health specialists in analysing information from wearable health monitors might be used to link data about individuals' physical activity and fitness levels to better understand the health impacts of different transport scenarios. Owen et al. (2012) used data from such monitors to explore the relationship between method of travel and mean level of physical activity for children on their journey to school. Similarly, Oliver et al. (2010) considered the potential of GIS, GPS and accelerometry data for studying transport-related physical activity. Subsequently, Carlson et al. (2015) used a similar range of data sources to examine the relation between neighborhood 'walkability' and levels of active travel. In time these approaches may be used to challenge existing theories about travel behaviour, which typically ignore links to physical activity, and hypothesise new ones.

Separately, both De Montjoye et al. (2015) and Lenormand et al. (2015) have used credit card transaction data to investigate spatial and temporal mobility and its relationships to spending patterns. While this type of analysis is always likely to lead to debates about ethical considerations and what can be done to overcome them (Sánchez et al., 2015; De Montjoye and Pentland, 2015), research of this nature has the potential to provide new insights into the ways in which the activities that mobility supports shape and influence travel behaviour.

2.5. There may be an ability to link multiple contemporaneous data sources

The fact that some data sources may be continuously monitored provides the potential for temporal overlap (and thus temporal analysis) of different data sources that may previously have been considered to have only a circumstantial relation. The data may, at one extreme, be related at the level of individuals (e.g. Philips et al., 2017) or may be on a higher level of aggregation (e.g. spatial, demographic). Building on the example of spatially structured data on financial transactions discussed in Section 2.4, such information might be analysed temporally alongside data relating to movements of people and of different types of vehicles and, potentially, even matched with data about weather conditions to lead to better understandings about seasonal variations in spatial activity and travel behaviour.

While this is currently a relatively new research area, a good example can be found in Pereira et al. (2015). This study combined internet data regarding special events with electronic public transport ticket tap-in/tap-out data (for a case study of Singapore), to develop a predictive model of passenger arrivals at event venue locations. They succeeded in improving the quality of transport predictions under special event scenarios, which they claim could lead to a greater ability to plan for and manage such situations in the future.

2.6. Data of sufficient scale to apply statistical inference techniques

In the transport discipline there has traditionally been a paucity of both research and applications using statistical techniques to understand behaviour. One reason has been the previous difficulty in obtaining reasonable amounts of repeated data in a similar environment (e.g. data about origin to destination movements), due to the costs and disruption involved in collecting it. Obtaining sufficient data on the travel behaviour of a city population, when the city and phenomena within it are changing in an uncontrolled way, is a problem to which big data may provide a solution. For example, Crawford et al. (2017a) recently analysed long periods of recorded traffic flows in order to identify systematic sources of variation, such as day-of-week and seasonal effects.

2.7. Synthesis and relationship of big data to traditional information sources

Rather than attempting to define big data in the ways that other authors have done, in this section so far we have presented a series of features which we believe many emerging data sources may possess, and have discussed the main issues they appear to raise for research about spatial mobility and travel activity. Our central argument is that 'bigness' in terms of the number of data elements, while potentially significant from a practical logistical perspective, may be largely irrelevant from a conceptual viewpoint for distinguishing what is special about big data. We believe this view is justified because the spatial and transport planning context we are considering does not rely greatly on speed of analysis. This might be contrasted with an application in which, say, an automated vehicle was being directed through machine vision, during which there is a need to process vary large amounts of image data very quickly to ensure safe operation. By contrast, we are primarily interested in understanding and planning for longer term trends. Actually, we think our discussion has wider applicability: even in a case where the objectives include some form of dynamic management of the transport system in response to real-time information, it is likely that the most effective strategies will involve solutions that are at least partly based on experience of similar past events, rather than those that focus solely on a rapid large-scale analysis of the present.

However, it is important to acknowledge that the features we have identified may not be exclusive to big data. Even within more traditional data sources there are likely to be elements of continuous monitoring (e.g. automated traffic counts, public transport ticket receipts), situations where data is not owned by planning authorities (e.g. private car park arrival and duration data), reuse of data collected for other purposes (e.g. use of vehicle licensing data to estimate traffic-related emissions) and combined use of multiple data sources including data from outside the transport sector (e.g. the use of demographic data from the census and business directory data, together, to estimate spatial patterns of travel demand).

In addition, it is true that not all traditional survey data can truly be considered 'small'. A population census, while only carried out periodically, may collect a wide range of data from every citizen. Likewise, other more focussed data sources related to, for example, land and property purchase, vehicle ownership, or road traffic accidents may be expected to represent a somewhat complete coverage of the phenomenon measured. There has been longstanding use of this sort of 'medium sized' digital data, in conjunction with digital spatial mapping, leading to a considerable body of research that provides analytical insights about human mobility within GIS environments. An approach that has proved particularly useful for investigating travel activity over time has been exploratory data analysis (ESDA) (Buliung and Kanaroglou, 2004), which uses large-scale traditional survey data in combination with an object oriented analysis and design (OOAD) methodology to produce spatial patterns and correlations. While not really fitting the characterisations of big data that we have proposed, this type of work might be considered to have some 'big' qualities, and so is particularly interesting as a comparator against which to discuss emerging data.

In particular, we have in mind the ability of such a methodology to link large (albeit traditionally surveyed) datasets of travel activity and other information, such as demographic, socio-economic and business-related survey data. This research approach continues to be popular in the study of mobility and to have considerable value. For example, Rybarczyk and Wu (2010) used an ESDA approach in conjunction with multi-criteria analysis to propose better ways for planning urban cycling facilities in Milwaukee City, while da Silva et al. (2014) used ESDA approaches in conjunction with population census data and information about road infrastructure to investigate the definition of urban regions in Brazil. An explicit aim of the latter study was to develop methods for use in developing countries where more detailed information (including, by implication, big data) may not be available. More recently, Buckwalter (2017) has used socioeconomic census data in conjunction with ESDA to investigate mode choice for journeys to work in Pittsburgh, while Loidl et al. (2016) used an exploratory spatial and temporal analysis to identify patterns of bicycle accidents in Salzburg.

There are some significant similarities between the ESDA body of research and some of the transport-related work using big data, particularly regarding studies that focus primarily on investigating spatial patterns of activity. However, there are also some key differences concerned with the ways in which research based on big data tends to focus more on high levels of resolution, especially temporal resolution, which is typically missing from traditional datasets. Against that, a critical advantage of medium-sized digital data from traditional surveys is the wealth of explanatory variables that tend to be available, rather than potentially needing to be inferred, meaning that GIS-based ESDA analysis may be more powerful for investigating differences between people and for suggesting more sophisticated aspects of human behaviour. Overall, a big data revolution should certainly not decrease the need for this sort of research, though over time it may impact on the types of data that are available to analyse, if it causes traditional surveys to go out of fashion.

Considering the comparative advantages of different types of data naturally leads to observations about both data collection processes and the nature of information generated. It is reasonably well established that traditional surveys of travel demand and behaviour are not only time consuming and expensive, they can also prove to be both disruptive for the transport system and intrusive for individuals. Yang et al. (2014) identify all these concerns related to the use of household and roadside interview surveys for deriving matrices of origin to destination movements. They argue that this constrains the feasible volume of data to a limited sample at a fixed point in time that is subject to potential sampling bias. Similarly, Abdulazim et al. (2013) consider traditional travel diary surveys to be “non-respondent friendly” because they require participants to put in significant effort to recall and record their activities. They argue that this may affect data accuracy and that it results in the approach being inappropriate for use beyond a few days, meaning that there is little scope for capturing variations. By contrast, approaches which automate data collection through mobile devices offer the potential for the process to be almost invisible, facilitating study over longer periods and aiding data accuracy through both increased sampling and removal of reliance on human responses. Nevertheless a variety of logistical, technical, cost and sampling concerns remain regarding GPS, mobile phone and Bluetooth data. This has led some to view the most promising direction to be open access, location-based, social network data, made available through platforms such as Twitter (Yang et al., 2014).

With regard to the information generated by different data sources, Hu and Jin (2017) have carried out a particularly thorough audit of the pertinent characteristics. They identify low levels of spatial and temporal resolution as the primary drawback of traditional surveys, while low levels of sampling bias and the potential to collect data for a variety of explanatory variables for travel (e.g. mode, journey purpose and social demographics) are presented as the main advantages. By contrast, they judge most of the big data approaches to offer higher levels of resolution, but with some inevitable sampling bias and a frequent need to infer explanatory characteristics. Yang et al. (2014) argue that location-based social network data have some “unique advantages” that may provide potential to overcome the shortcomings of other sources. These advantages are related to additional activity-related information that is attached to check-in locations (providing some explanations for journeys) and growing levels of penetration (reducing sampling bias).

The limitation of most work to date that uses big data in a transport planning context is that it focuses almost entirely on the ability of the information to replace traditional surveys within existing types of analysis. As a result, the literature contains few examples of big data being used to provide novel insights about mobility in ways that have not previously been considered. Mayer-Schönberger (2016) presents a different conceptual perspective in which the role of big data is seen as “reshaping the scientific method” towards inductive approaches at the expense of deductivism, implying a new study environment that makes use of a broad range of data to reveal the potentially unexpected rather than focusing on the narrower scope of information that fits existing theories and methods. This corresponds to Anderson's (2008) “end of theory” definition, though Mayer-Schönberger is clear that a big data revolution should not require us to “abandon the search for causes”, and that the process of discovery has always been iterative.

3. Using big data for analysis and modelling of transport systems

The nature of the information provided by the big data sources discussed in this paper is likely to be rather different to that from data derived via traditional population and transport-related surveys. This will have implications for data-driven analysis and modelling activities that support and help justify transport planning and policy decisions.

3.1. Traditional transport planning data and its drawbacks

The data that has traditionally informed transport planning has primarily been provided by manual surveys of people and their travel behaviour, for estimating travel demand, by manual mapping, service level and landscape surveys, for estimating transport supply, and by a mixture of manual and automated surveys of movements and transactions, for calibrating flows. In addition to the possibility that useful insights into the performance of transport systems might come from new types of analysis of previously unconsidered external data sources, the main expectation regarding the role of big data in transport planning is that it will replace much of the information that has previously been collected manually. Much of the focus is on digital data that is already being collected, some of it outside the transport sector and for commercial purposes, such data arising as part of the increasing use of computer-based systems for managing human activities and transactions.

Reasons for replacing manual data sources are not simply related to perceptions of the benefits big data might bring and include a number of ‘push’ factors, such as the issues of cost, time, intrusion and disruption already discussed in Section 2.7. Manual data collection has always been both expensive and time consuming, which has often restricted the volume of data that could be obtained to levels below statistically defensible samples. In addition, some types of manual survey (such as on-street origin to destination surveys of traffic movements) are disruptive, making them politically unpopular and, potentially, prone to errors due to the influence on behaviours the data collection may induce (e.g. drivers re-routing to avoid disruptions caused by roadside interview surveys).

3.2. Features of big data in a transport planning context

The fact that big data offers opportunities to resolve problems with traditional data, such as those discussed in Section 3.1, does not necessarily mean that information will be better in all respects. Differences in the *nature* of information that might be expected as a result of a move away from manual data collection include:

- origins and focus of data;
- volume of data collected;
- range and differentiation within data;
- sampling of data observations;
- nature of errors and omissions.

The origins and focus of data might be expected to affect the information provided in respect of its ability to describe phenomena of interest for transport planning. Data acquired from third parties and collected for other purposes may have been defined in ways that limit detail (e.g. constraints on spatial resolution to protect privacy) or in ways that reduce the scope of the information (e.g. by having no link to individuals, so that it impossible to measure the repeatability of activity and mobility patterns from day to day).

It is normally assumed that big data will provide a significant increase in the volume of information available, but that expectation may present challenges particularly in the case of continuous monitoring. Whereas the traditional emphasis in analysis and modelling of transport systems has focussed on attempting to represent long-run average conditions from relatively small amounts of data, it might be expected to change towards attempting to identify short-run stability within far more comprehensive datasets that include variations by hour, by day, by season and related to specific events. It also seems plausible that increasing volumes of data will be accompanied by rising expectations of what the information can be used for. In addition, it may result in demands from decision-makers that models of transport systems should cater for variations to a much greater extent than has been the case before.

By contrast, the range of data collected may actually reduce with a shift away from manual surveys. Probably the most common focus of work to replace manual data with big data in the transport sector is the use of information produced by commercial mobile devices (e.g. smartphone and Bluetooth movement trajectories), in order to replace manual surveys for estimating travel demand (Toole et al., 2015). This involves the acquisition of data collected independent of a transport planning context which provides potentially very detailed information about the movements of people in space and time. We have already touched on the nature of the information generate by the new data sources in Section 2.7. In this regard, although the *volume* of data might be expected to be much greater than from (say) traditional roadside origin to destination surveys, the *range* of information directly available from passive sources is likely to be significantly reduced, with no ability to differentiate features that add meaning to our understanding of travel such as mode, vehicle type, vehicle occupancy, journey purpose and various demographic features. In addition, the spatial range of the information may be compromised by difficulties in identifying the precise start and end points of journeys, which can only be inferred from movement patterns. Some significant progress has been made over a period of time towards developing new analytical approaches to address these issues (Bohte and Maat, 2009; Diao et al., 2015; Çolak et al., 2015), but it is acknowledged that significant issues remain especially for widespread practical application, including within more detailed settings (Rojas IV et al., 2016).

It is also to be expected that, in many digital data scenarios, sampling will not be random and the implications of that, both for individuals and their activities, will need to be dealt with. It seems very likely that—in the current era at least—observations based on transaction and tracking data are likely to be skewed towards the most economically active, most technologically equipped and, potentially, younger members of society. For example, in studies using mobile phone data it has been acknowledged that there are problems associated with variations across the population in levels of phone ownership and use (Rojas IV et al., 2016).

Finally, the nature of errors and omissions should be expected to differ significantly between digital and manual datasets. In traditional manual transport surveys, errors are most likely to occur due to incorrect recording of observations, a problem that may be difficult to quantify and control for without duplication of effort (Watling et al., 2012). Omissions, on the other hand, are typically the result of constraints on time and resources or of contextual problems related to the feasibility of manual surveys. Automatically recorded digital datasets might be expected to eliminate recording errors if the data collection procedures are well designed, but they may lack an ability to question apparently illogical observations (which, for example, a human interviewer carrying out a roadside origin to destination survey can do). They are also likely to be much more prone to errors that might be introduced during the handling and transfer of large volumes of raw data into a usable format for analysis. Omissions are most likely to relate to technological failure and can lead to a complete loss of usable observations for the duration of the problem, though that may be compensated for by the greater volumes of data available overall. However, in studies using GPS and mobile phone data, loss of signal has sometimes been found to be a significant constraint on data quality (Rojas IV et al., 2016).

3.3. Implications for planning and modelling transport systems

Current trends suggest that it is inevitable that automatically recorded, digital data will come into mainstream use both for academic study and for the practical planning of transport systems. However, alongside this trend, what also seems likely is that inputs from more traditional ‘small data’ sources will still be necessary, in order to make up for the lack in many ‘big data’ sources of demographic information and other unobserved elements related to individuals and activities, all of which add important meaning, context and motivation to the information. At the same time, it is possible that features of digital data may lead to a greater focus on generic and *transferable* understanding of travel, across different contexts, scenarios, cities, and over time (see, for example, the plethora of works inspired by complexity science on the search for ‘universal laws’, such as the city scaling laws studied by Cebrat and Sobczyński, 2016). Such a greater pooling of information from different situations would challenge the traditional of transport planning on particular case studies, and the understanding of phenomena for specific locations and times.

For the discipline of modelling transport systems, the greatest likelihood would appear to be that models will become more empirically-based as a result of such a data revolution. On the other hand, it may be the case that more data will facilitate a greater number of opportunities to test *theories* against real-world evidence. Modelling idealisations, such as equilibrium, economic-man, gravity and value-of-time, may begin to be seen as less important. That may, in turn, add fuel to debates about economic evaluation and decision-making. It may as a result open up new ways of understanding impacts of infrastructure changes, beyond conceptually limited calculations of travel time savings for existing patterns of journeys over fixed and relatively short time horizons. Indeed, new opportunities to examine longitudinal effects could result in more focus on transient properties, periods of change and drivers of change. In time it should become possible to understand the influence of more factors within transport systems, including longer-term issues and factors that have previously been very difficult to quantify, such as the effect of political cycles if data spans several parliamentary periods. Overall, it would be no surprise to see modelling become more data-driven with an influx of pattern-matching approaches, potentially at the expense of more subjective approximations. It also seems likely that there would be increasing cross-uses of data, even for conventional modelling (such as use of engineering data to calibrate behavioural models and vice versa) as part of previously ignored relationships between variables being identified. Related to this, greater use of techniques such as machine learning and data analytics would be expected, to gain new insights into critical elements.

However, changes are unlikely to be limited to analytical and modelling practices, with transport systems themselves likely to evolve in response to new information. This is something that models and the policies they are used to justify will need to account for. Certainly, it seems inevitable that providers of transport systems (e.g. public transport companies and authorities responsible for road network management) will increasingly use real-time information as part of their operations, possibly from competing information providers. This will affect the real ways in which we travel. For example, information that makes predictions of incident impacts will make travellers more aware of unreliability and, thus, it should be expected that they may increasingly factor it into their typical behaviour, which this behavioural adaptation in turn we will then need to understand for modelling and planning. Alongside the use of data by providers, as individuals receive information that is more personally-tailored they will find it easier to make choices to satisfy their requirements. However, the greater the number of information providers, the more difficult it may become to *coordinate* that information and control policies based upon it.

A policy-related research area that builds on the ideas of data technology in transport is 'Mobility as a Service' (MaaS), which relies on the understanding that digital information can be used to coordinate inter-modal transport alternatives for individuals. The aim of MaaS is to provide them with door-to-door service options for journeys that would previously have involved a series of different information sources and, potentially, financial transactions (Ambrosino et al., 2016). The basic dimensions of the MaaS concept are still being developed, but a common assumption of all of them is the use of personal mobile devices in real time for travel information and associated transactions. The original underlying idea behind MaaS (Heikkilä, 2014) was to reduce reliance on the private car in urban areas by matching the door-to-door service it provides with other modes. There was also, potentially, an implicit aim to increase the coordinating power of public agencies with responsibility for transport, towards achieving greater integration of services and payment across all the different providers. However, as the MaaS idea is being disseminated through different environments, multiple interpretations are currently emerging, including the possibility that a technology-based private provider of transport, such as Uber, might become a major driving force of integration in some situations. These institutional and political choices may have a profound impact on the effects of data on the transport system, as well as on its availability and use for planning purposes.

4. The opportunity potential of big data for transport planning

4.1. Opportunities with big data

A major feature of big data that has played a significant part in its adoption in other fields is that it allows analysis at a more 'raw' level, free of assumptions sometimes made in converting raw data to a manageable form (e.g. 'mechanisms' to convert inductive loop data to vehicle counts). Continuous monitoring allows the study of new kinds of variation (time-of-day, day-to-day, time-of-year, scenario-specific) to correlate with data on events/weather, and to monitor unexpected events or disasters (e.g. the bridge collapse studied by Zhu et al., 2010; or the earthquake/tsunami studied by Hara and Kuwahara, 2015).

More widespread monitoring may also allow finer disaggregation of effects and more opportunity to study small and/or disadvantaged groups. As we have mentioned earlier, there are concerns about the representativeness of some of the new data sources (e.g. skewed towards younger people, or biased away from certain groups), but what might at first seem contradictory is that such sources could still open up possibilities for studying minority groups, even if such groups are under-represented in the data, due to the sheer scale of the overall data set. As an example, suppose that a minority group forms 0.1% (1 in 1000) of the population of a large city. Through a traditional travel survey, 1000 individuals are randomly sampled (i.e. an unbiased sample), meaning that on average we will only observe one individual (and perhaps in a particular case no individual) from the minority group. Now a new form of passive data provides information on 100,000 individuals, but in this data set it is known that the minority group is under-represented, and so only makes up 0.05% (1 in 2000) of the sample. In spite of this, 0.05% of 100,000 means that on average 50 individuals of the minority group will be observed, which seemingly gives sufficient data for some kind of focus on that group specifically. It may be that the sampled individuals of that group are in some sense atypical of all members of it, in which case we would need to take care in making any inference, but if not atypical we would also have a basis for making implications about the minority population of the city.

Furthermore, big data provides an opportunity to become less reliant on stated preference approaches because there is more chance of obtaining revealed data of the same individuals in a variety of contexts, or of obtaining/inferring perceptions of non-chosen options. It also provides potential to develop transferable behavioural models with more explanatory factors, due to much larger sample sizes which may be applicable to a wider range of policy contexts, socio-political backdrops and locales, rather than just marginal changes from the present, as is often the case in current studies.

There is also the possibility that new symbiotic relationships could emerge between data owners and planning agencies, under which information could be provided for mutual benefit. A longstanding example of symbiosis potentially similar to the big data context is that, since the beginning of flight, aircraft have made observations about the weather for their own safety reasons, but since World War I they have also provided information to aid wider understanding of meteorological processes. Automated reports of aircraft observations have been available since 1979 and the subsequent growth in commercial airline activity means that they now play a vital role in improving the performance of weather prediction models across the globe (Moninger et al., 2003). This has provided benefits to meteorological agencies serving wider populations, but it has also clearly been a benefit to the airline industry too as both aircraft performance and safety have been improved through better weather predictions. Initially this type of scenario may constitute 'data reuse' to provide additional benefits. However, over time, such data may evolve with the explicit intention that it can serve multiple purposes.

Changes within the data and modelling spheres are bound to have knock-on impacts for practical transport planners, in particular related to the expectations to which they are subject. Intuitively, greater availability of detailed data from continuous monitoring seems likely to lead to greater expectations of focussed planning for more specific situations than is the norm at present. For example, planners may be expected to be capable of creating policies that deal with differences by day of week, season and weather condition.

In parallel, new data and analytical approaches should empower transport planners. The ability to detect unexpected trends and changes may give planning the opportunity to become more contingent. In addition, rather than needing to carry out all analysis of the potential impacts of policy ideas in advance, potentially leading to considerable resource costs and delays before anything is implemented, continuous monitoring may provide opportunities for more trial and error approaches to policymaking, with data giving continuous feedback. This may help facilitate policies that encourage gradual changes based on a series of 'nudges' rather than sudden step-changes. The types of data anticipated will also be highly suited to visualisation, which would fit well with moves towards more public participation in planning processes. For example, mobile phone data has been used to produce both temporal density maps (Ahas et al., 2015) and spatio-temporal trajectories (Gao, 2015).

Finally, a major failing of transport planning in the past has been a paucity of ex-post studies to check how forecast outcomes of the impacts of decisions compare to reality. A key reason for this has been a lack of sufficient appropriate data (Nicolaisen and Driscoll, 2014) which is often related to lack of funding, despite significant evidence that such work improves the quality of predictions (ITF, 2017). One reason ex-post studies may sometimes have been avoided is the threat they pose to political capital give the inevitable risk that impacts of decisions may not appear as good as expected. Big data should provide the potential to address this by providing much more evidence on which such studies could be based. Indeed, sufficient data may be openly available to allow well informed independent studies to be conducted. The use of independent organisations to audit transport planning decisions is one of the central recommendations of the most comprehensive international review to date about ex-post assessment in the transport sector (ITF, 2017).

4.2. Limitations and difficulties associated with big data

Clearly, there are—and will continue to be—many technical, analytical and computational issues associated with using the types of data discussed in this paper. Great advances are already being made in machine learning, data science, data analytics, and the ease with which we may interface with and use device data. However, a particular question that remains to be resolved is how analytical methods can be 'future-proofed', if there is no guarantee that the same data will continue to be available indefinitely in the same form. A key risk of non-purpose-oriented data is that the external providers may undergo a change of priorities in which data and how much data they collect, and in their willingness to allow it to be used for other purposes (let alone any access charges). Even if those problems do not arise, technological changes are bound to lead to a multitude of compatibility issues over time that could reduce the temporal power of data considerably, unless appropriate common standards can be adopted.

For situations where data is not open access another key question is how data owners will respond over time to the understanding that their information has value. It is already clear that mobile phone operators, and others who acquire data through technology-based operations, are interested in charging considerable sums to public planning agencies for access to their information as a substitute for traditional travel activity surveys. It is possible that data owners will go further and explicitly seek out new markets to sell information for commercial gain. Though in some cases the retention of value over time may be dependent on the ability of data owners to secure widespread public engagement in ways that are useful for planning purposes, it nevertheless seems likely that potential sources of data will increase in number and decrease in cost.

The inherently invasive nature of many data sources, especially those involving continuous monitoring, also leads to potential for considerable issues related to privacy and its trade-off with data fidelity. Mobile sensors, such as cellular phones and other portable devices with GPS and Bluetooth capability, offer potential to provide a wealth of information about human mobility behaviour in time and space. However, the information poses a risk that individuals could be identified and their detailed movements tracked, leading many data owners to filter information and reduce spatial and/or temporal accuracy before allowing it to be reused. This typically reduces its potential for providing insights, prompting research to identify approaches that can protect privacy without losing the fine-grained qualities of the original data (Sun et al., 2013).

Beyond basic concerns about individual privacy, the collection and use of significant volumes of data brings with it potential for significant ethical issues. For example, data may be used for (and may even lead to) the targeting of policies at different population sub-groups. For example, if information were to become available to demonstrate that certain people have a particular genetic make-up that causes them to be more pre-disposed to have road traffic accidents, then that information could be used both to aid technological advances to make them safer and to discriminate against them in the commercial insurance market. This leads to the question of whether there should be controls on the types of uses of new data and the findings based upon it. In the case of insurance, the traditional underlying concept of 'human solidarity based on ignorance' (Mayer-Schönberger and Cukier, 2013) may be put at risk if appropriate safeguards are not adopted; a similar concept of solidarity based on a 'veil of ignorance' has already been argued for healthcare (ter Meulen, 2016).

Finally, there may be environmental implications of assuming unconstrained data opportunities in the sense that the capacity for data generation, storage and transmission may not always be infinite. Evidence already exists related to the uptake of streaming and cloud storage/retrieval that the energy implications are far from insignificant and could potentially grow over time to become a comparable problem to the energy used for physical movement (Mills, 2013), though there is also potentially evidence of a trade-off between ICT and physical movement (Gelenbe and Caseau, 2015). It may be useful to compare this situation with the historic uptake of other technologies, such as the motor car, where initial perceptions of a wholly positive future without any constraints on capacity have proved wide of the mark. Certainly, if data capacity does need to be constrained in future that may be expected to put pressure on all actors to ensure multi-functionality.

5. Concluding remarks

This paper has attempted to open up thinking on what the emergence of new digital big data sources may mean for transport planning and for the analytical research and modelling that supports it. In the process it has been able to reach few firm conclusions. However, from our discussion, two points do emerge strongly.

First, research regarding the potential of big data for transport planning needs to think about more than how big data can make it easier to pursue existing approaches (e.g. to derive a demand pattern in the form of an origin to destination matrix, or to correlate trip rates to trip lengths). Rather, it needs to engage in a fundamental reassessment of what data can tell us about transport systems that help us better understand how they function and what we can do to influence them in positive ways.

Second, the areas of data, predictive models and planning are a triple that must be considered together. If, for whatever reason, one of them undergoes a significant change, then they must all adapt. In the case of big data, it is that adaptation process which is likely to prove critical if we are to unlock the greatest potential for improvements in transport planning and policy-making that increased volumes of information may allow, while addressing issues that may limit its range and quality.

From our discussion we believe there are a number of 'big challenges for big data' that need to be addressed in order to understand and gain most benefit from the transitions that lie ahead. These might be briefly summarised as:

1. developing a clear understanding of how data related to transport systems is likely to change compared to traditional 'small data' sources;
2. tackling limitations to the information that emerging data sources can provide to attempt both to avoid loss of useful detail and to maximise benefits from new features;
3. opening up transport planning to new opportunities for using data, analytical approaches and specialist understandings outside the traditional scope of the discipline;
4. identifying and measuring the impacts that new data sources have on real transport systems, through the private and commercial use of information and its impacts on travel patterns and related behaviours;
5. re-specifying analytical and predictive modelling approaches in response to the modified data landscape and the new insights it facilitates; and
6. reconsidering the relationships that data analysis and predictive modelling have with transport planning, policy formulation and decision making, to try to ensure that interventions are based on the best understanding and information available.

We have discussed some issues relevant to these challenges, but much more remains to be done. Perhaps the most important message may be that, in our response to new data opportunities, we should avoid attempting to build new analytical approaches, models and planning

Fig. 1. Elwood Haynes in his first automobile, the Pioneer, c 1910 (Wikimedia Commons).

practices only in the image of what has gone before, but should seek instead to embrace change and look to the future. If we insist on repeating the precedent from the early days of the combustion engine, by innovating based on the most minimal adaptations (see Fig. 1), then we can be sure that it will not be long before our designs are swept away and replaced by new structures that are barely recognisable.

Acknowledgements

Would we like to thank the three anonymous referees who provided valuable constructive criticism on an earlier version of this paper. The research was partially supported by a grant from the Engineering & Physical Sciences Research Council project STEP-CHANGE, ref. EP/I00212X/1, EP/I00212X/2.

References

Abdulazim, T., Abdelgawad, H., Habib, K.M.N., Abdulhai, B., 2013. Using smartphones and sensor technologies to automate collection of travel data. Transp. Res. Rec. 2383, 44–52.

Abdulazim, T., Abdelgawad, H., Habib, K.M.N., Abdulhai, B., 2015. Framework for automating travel activity inference using land use data. Transp. Res. Rec. 2526, 136–142.

Ahas, R., Aasa, A., Yuan, Y., Raubal, M., Smoreda, Z., Liu, Y., Ziemlicki, C., Tiru, M., Zook, M., 2015. Everyday space–time geographies: using mobile phone-based sensor data to monitor urban activity in Harbin, Paris, and Tallinn. Int. J. Geogr. Inf. Sci. 29 (11), 2017–2039.

Ambrosino, G., Nelson, J., Boero, M., Pettinelli, I., 2016. Enabling intermodal urban transport through complementary services: from flexible mobility services to the shared use mobility agency. Workshop 4. Developing inter-modal transport systems. Res. Transp. Econ. 59, 179–184.

Anderson, C., 2008. The end of theory: the data deluge makes the scientific method obsolete. Wired Magazine 16 (7).

Batty, M., 2013. Big data, smart cities and city planning. Dialogues Hum. Geogr. 3 (3), 274–279.

Bhaskar, A., Chung, E., 2013. Fundamental understanding on the use of Bluetooth scanner as a complementary transport data. Transp. Res. C: Emerg. Technol. 37, 42–72.

Blondel, V.D., Decuyper, A., and Krings, G. (2015). A survey of results on mobile phone datasets analysis. arXiv:1502.03406 [physics.soc-ph].

Bohte, W., Maat, K., 2009. Deriving and validating trip purposes and travel modes for multi-day GPS-based travel surveys: a large-scale application in the Netherlands. Transp. Res. C 17, 285–297.

Boyd, D., Crawford, K., 2011. Six provocations for big data. In: A Decade in Internet Time: Symposium on the Dynamics of the Internet and Society. Sept 21, 2011.

Buckwalter, D., 2017. Modal choice in the Pittsburgh metropolitan statistical area: an exploratory data analysis. Prof. Geogr. 69 (1), 94–106.

Buliung, R.N., Kanaroglou, P.S., 2004. An Exploratory Spatial Data Analysis (ESDA) toolkit for the analysis of activity/travel data. In: Laganá, A., Gavrilova, M.L., Kumar, V., Mun, Y., Tan, C.J.K., Gervasi, O. (Eds.), Computational Science and its Applications – ICCSA 2004. ICCSA 2004. Lecture Notes in Computer Science: 3044. Springer, Berlin, Heidelberg.

Bwambale, A., Choudhury, C., Hess, S., 2017. Modelling trip generation using mobile phone data: a latent demographics approach. (Submitted for Publication).

Calabrese, F., Diao, M., Di Lorenzo, G., Ferreira Jr., J., Rattia, C., 2013. Understanding individual mobility patterns from urban sensing data: a mobile phone trace example. Transp. Res. C: Emerg. Technol. 26, 301–313.

Calabrese, F., Ferrari, L., Blondel, V., 2014. Urban sensing using mobile phone network data: a survey of research. ACM Comput. Surv. 1–23.

Carlson, J.A., Saelens, B.E., Kerr, J., Schipperijn, J., Conway, T.L., Frank, L.D., Chapman, J.E., Glanz, K., Cain, K.L., Sallis, J.F., 2015. Association between neighborhood walkability and GPS-measured walking, bicycling and vehicle time in adolescents. Health Place 32, 1–7.

Cate, F.H., Mayer-Schönberger, V., 2013. Notice and consent in a world of Big Data. Int. Data Priv. Law 3 (2), 67–73.

Cebrat, K., Sobczyński, M., 2016. Scaling laws in city growth: setting limitations with self-organizing maps. PLoS One 11 (12).

Chatterton, T., Barnes, J., Wilson, R.E., Anable, J., Cairns, S., 2015. Use of a novel dataset to explore spatial and social variations in car type, size, usage and emissions. Transp. Res. Part D: Transp. Environ. 39, 151–164.

Chow, A., 2016. Heterogeneous urban traffic data and their integration through kernel-based interpolation. J. Facil. Manag. 14 (2), 165–178.

Chung, N., Han, H., Koo, C., 2015. Adoption of travel information in user-generated content on social media: the moderating effect of social presence. Behav. Inf. Technol. 34 (9), 902–919.

Çolak, S., Alexander, L.P., Alvim, B.G., Mehndiretta, S.R., González, M.C., 2015. Analyzing cell phone location data for urban travel: current methods, limitations and opportunities. Transp. Res. Rec. 2526, 126–135.

Crawford, F., Watling, D.P., Connors, R.D., 2017a. A statistical method for estimating predictable differences between daily traffic flow profiles. Transp. Res. B Methodol. 95, 196–213.

Crawford, F., Watling, D.P., Connors, R.D., 2017b. Identifying road user classes based on repeated trip behaviour using Bluetooth data. (Submitted for Publication).

Crawford, F., Watling, D.P., Connors, R.D., 2017c. Assessing the feasibility of using Bluetooth data to examine the repeated travel behaviour of road users. (Submitted for Publication).

De Montjoye, Y.-A., Pentland, A., 2015. Response to comment on “Unique in the shopping mall: on the reidentifiability of credit card metadata”. Science 347 (1274-b).

De Montjoye, Y.-A., Radaelli, L., Singh, V.K., Pentland, A., 2015. Unique in the shopping mall: on the reidentifiability of credit card metadata. Science 347, 536–539.

Diao, M., Zhu, Y., Ferreira Jr., J., Ratti, C., 2015. Inferring individual daily activities from mobile phone traces: a Boston example. Environ. Plann. B 43 (5), 920–940.

Fortunati, L., Taipale, S., 2017. Mobilities and the network of personal technologies: refining the understanding of mobility structure. Telematics Inform. 34, 560–568.

Frignani, M.Z., Auld, J., Mohammadian, A.K., Williams, C., 2010. Urban travel route and activity choice survey: internet-based prompted-recall activity travel survey using global positioning system data. Transp. Res. Rec. 2183, 19–28.

Gao, S., 2015. Spatio-temporal analytics for exploring human mobility patterns and urban dynamics in the mobile age. Spat. Cogn. Comput. 15, 86–114.

Gelenbe, E., Caseau, Y., 2015. The impact of information technology on energy consumption and carbon emissions. Ubiquity (June edition, journal of the Association of Computing Machinery).

Gong, L., Liu, X., Wu, L., Liu, Y., 2016. Inferring trip purposes and uncovering travel patterns from taxi trajectory data. Cartogr. Geogr. Inf. Sci. 43, 103–114.

Grindrod, P., 2016. Beyond privacy and exposure: ethical issues within citizen-facing analytics. Phil. Trans. R. Soc. A 374 (2083).

GSR, 2016. GSR Technologies Inc. http://www.gsrti.com (visited 9/3/16).

Gu, Y., Qian, Z.S., Chen, F., 2016. From twitter to detector: real-time traffic incident detection using social media data. Transp. Res. C Emerg. Technol. 67, 321–342.

Hara, Y., Kuwahara, M., 2015. Traffic monitoring immediately after a major natural disaster as revealed by probe data – a case in Ishinomaki after the Great East Japan Earthquake. Transp. Res. A Policy Pract. 75, 1–15.

Hasan, S., Ukkusuri, S.V., 2014. Urban activity pattern classification using topic models from online geo-location data. Transp. Res. C Emerg. Technol. 44, 363–381.

Heikkilä, S., 2014. Mobility as a Service: A Proposal for Action for the Public Administration, Case Helsinki. MS thesis. Aalto University, Aalto, Finland (unpublished).

Hu, W., Jin, P., 2017. An adaptive hawkes process formulation for estimating time-of-day zonal trip arrivals with location-based social networking check-in data. Transp. Res. C 79, 136–155.

ITF, 2017. Ex-post assessment of transport investments and policy interventions. In: ITF Roundtable Reports. OECD Publishing, Paris.

Jestico, B., Nelson, T., Winters, M., 2016. Mapping ridership using crowdsourced cycling data. J. Transp. Geogr. 52, 90–97.

Kitchin, R., 2013. Big data and human geography: opportunities, challenges, risks. Dialogues Hum. Geogr. 3 (3), 262–267.

Kitchin, R., 2014a. The real-time city? Big data and smart urbanism. GeoJournal 79, 1–14.

Kitchin, R., 2014b. The Data Revolution: Big Data, Open Data, Data Infrastructures and their Consequences. Sage, London.

Kwan, M.-P., 2016. Algorithmic geographies: big data, algorithmic uncertainty, and the production of geographic knowledge. Ann. Am. Assoc. Geogr. 106, 274–282.

Laney, D., 2001. 3D Data Management: Controlling Data Volume, Velocity and Variety. META Group Research Report.

Lenormand, M., Louail, T., Cantu-Ros, O.G., Picornell, M., Herranz, R., Arias, J.M., Barthelemy, M., Miguel, M.S., Ramasco, J.J., 2015. Influence of sociodemographic characteristics on human mobility. Nat. Sci. Rep. 5, 10075.

Lin, M., Hsu, W.-J., 2014. Mining GPS data for mobility patterns: a survey. Pervasive Mob. Comput. 12, 1–16.

Liu, Y., Sui, Z., Kang, C., Gao, Y., 2014. Uncovering patterns of inter-urban trip and spatial interaction from social media check-in data. PLoS One 9 (1), e86026.

Loidl, M., Traun, C., Wallentin, G., 2016. Spatial patterns and temporal dynamics of urban bicycle crashes: a case study from Salzburg (Austria). J. Transp. Geogr. 52, 38–50.

Maréchal, S., 2016. Modelling the acquisition and use of information sources during travel disruption. In: Paper presented at the 48th UTSG Annual Conference. University of the West of England and University of Bristol.

Mayer-Schönberger, V., 2010. Beyond privacy, beyond rights-toward a "systems" theory of information governance. Calif. Law Rev. 98 (6), 1853–1885.

Mayer-Schönberger, V., 2016. Big data for cardiology: novel discovery? Eur. Heart J. 37, 996–1001.

Mayer-Schönberger, V., Cukier, K., 2013. Big Data: A Revolution that will Transform how we Live, Work, and Think. John Murray, London.

Mazzocchi, F., 2015. Could big data be the end of theory in science? EMBO Rep. 16, 1250–1255. http://dx.doi.org/10.15252/embr.201541001.

McCarthy, O.T., Caulfield, B., O'Mahoney, M., 2016. Technology engagement and privacy: a cluster analysis of reported social network use among transport survey respondents. Transp. Res. C 63, 195–206.

ter Meulen, R., 2016. Solidarity, justice, and recognition of the other. Theor. Med. Bioeth. 37 (6), 517–529.

Mills, M., 2013. The Cloud Begins with Coal: Big Data, Big Networks, Big Infrastructure and Big Power. (Report produced by the Digital Power Group).

Moninger, W.R., Mamrosh, R.D., Pauley, P.M., 2003. Automated meteorological reports from commercial aircraft. Bull. Am. Meteorol. Soc. 84, 203–216.

Montini, L., Rieser-Schüssler, N., Horni, A., Axhausen, K., 2014. Trip purpose identification from GPS tracks. Transp. Res. Rec 2405, 16–23.

Mori, U., Mendiburu, A., Álvarez, M., Lozano, J.A., 2015. A review of travel time estimation and forecasting for advanced traveller information systems. Transportmetrica A Transp.Sci. 11, 119–157.

Nicolaisen, S.M., Driscoll, P.A., 2014. Ex-post evaluations of demand forecast accuracy: a literature review. Transp. Rev. 34 (4), 540–557.

OECD, (2013). New data for understanding the human condition: International perspectives. OECD Global Science Forum Report, February 2013.

Oh, S., Byon, Y.-J., Jang, K., Yeo, H., 2015. Short-term travel-time prediction on highway: a review of the data-driven approach. Transp. Rev. 35, 4–32.

Oliver, M., Badland, H., Mavoa, S., Duncan, M.J., Duncan, S., 2010. Combining GPS, GIS, and Accelerometry: methodological issues in the assessment of location and intensity of travel behaviors. J. Phys. Act. Health 7, 102–108.

Owen, C.G., Nightingale, C.M., Rudnicka, A.R., van Sluijs, E.M.F., Ekelund, U., Cook, D.G., Whincup, P.H., 2012. Travel to school and physical activity levels in 9–10 year-old UK children of different ethnic origin; child heart and health study in England (CHASE). PLoS One 7 (2), e30932.

Ozbay, S., Ercelebi, E., 2005. Automatic vehicle identification by plate recognition. World Acad. Sci. Eng. Technol. 9, 222–225.

Pelletier, M.-P., Trépanier, M., Morency, C., 2011. Smart card data use in public transit: A literature review. Transp. Res. C Emerg. Technol. 19, 557–568.

Pender, B., Currie, G., Delbosc, A., Shiwakoti, N., 2014. Social media use during unplanned transit network disruptions: a review of literature. Transp. Rev. 34, 501–521.

Pereira, F.C., Rodrigues, F., Ben-Akiva, M., 2015. Using data from the web to predict public transport arrivals under special events scenarios. J. Intell. Transp. Syst. 19 (3), 273–288.

Philips, I., Clarke, G., Watling, D., 2017. A fine grained hybrid spatial microsimulation technique for generating detailed synthetic individuals from multiple data sources: an application to walking and cycling. Int. J. Microsimul. 10 (1), 167–200.

Rojas IV, M.B., Sadeghvaziri, E., Jin, X., 2016. Comprehensive review of travel behavior and mobility pattern studies that used mobile phone data. Transp. Res. Rec. 2563, 71–79.

Rybarczyk, G., Wu, C., 2010. Bicycle facility planning using GIS and multi-criteria decision analysis. Appl. Geogr. 30, 282–293.

Saadi, I., Boussauw, K., Teller, J., Cools, M., 2016. Trends in regional jobs-housing proximity based on the minimum commute: the case of Belgium. J. Transp. Geogr. 57, 171–183.

Sánchez, D., Martínez, S., Domingo-Ferrer, J., 2015. Comment on “Unijque in the shopping mall: on the reidentifiability of credit card metadata”. Science 351 (1274-a).

da Silva, A.N.R., Manzato, G.G., Pereira, H.T.S., 2014. Defining functional urban regions in Bahia, Brazil, using roadway coverage and population density variables. J. Transp. Geogr. 36, 79–88.

Siripirote, T., Sumalee, A., Watling, D.P., Shao, H., 2014. Updating of travel behavior parameters and estimation of vehicle trip-chain data based on plate scanning. J. Intell. Transp. Syst. 18, 393–409.

Steenbruggen, J., Tranos, E., Nijkamp, P., 2015. Data from mobile phone operators: a tool for smarter cities? Telecommun. Policy 39, 335–346.

Sun, Y., Li, M., 2015. Investigation of travel and activity patterns using location-based social network data: a case study of active mobile social media users. ISPRS Int. J. Geo Inform. 4, 1512–1529.

Sun, Z., Zan, B., Can, X., Gruteser, M., 2013. Privacy protection method for fine-grained urban traffic modeling using mobile sensors. Transp. Res. B 56, 50–69.

Tamblay, S., Galilea, P., Iglesias, P., Raveau, S., Carlos, J., 2016. A zonal inference model based on observed smart-card transactions for Santiago de Chile. Transp. Res. A Policy Pract. 84, 44–54.

Tang, J., Liu, F., Wang, Y., Wang, H., 2015. Uncovering urban human mobility from large scale taxi GPS data. Phys. A Stat. Mech. Appl. 438, 140–153.

Tao, S., Rohde, D., Corcoran, J., 2014. Examining the spatial–temporal dynamics of bus passenger travel behaviour using smart card data and the flow-comap. J. Transp. Geogr. 41, 21–36.

Toole, J., Çolak, S., Sturt, B., Alexander, L.P., Evsukoff, A., González, M.C., 2015. The path most traveled: travel demand estimation using big data resources. Transp. Res. C 58, 162–177.

Watling, D.P., Milne, D.S., Clark, S., 2012. Network impacts of a road capacity reduction: empirical analysis and model predictions. Transp. Res. A 46, 167–189.

Widhalm, P., Yang, Y., Ulm, M., Athavale, S., González, M.C., 2015. Discovering urban activity patterns in cell phone data. Transportation 42, 597–623.

Yang, F., Jin, P.J., Cheng, Y., Zhang, J., Ran, B., 2014. Origin-destination estimation for non-commuting trips using location-based social networking data. Int. J. Sustain. Transp. 9 (8), 551–564.

Zheng, Y., Liu, F., Hsieh, H.-P., 2013. U-Air: when urban air quality inference meets big data. In: Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining, New York, http://dx.doi.org/10.1145/2487575.2488188.

Zhu, S., Levinson, D., Liu, H.X., Harder, K., 2010. The traffic and behavioral effects of the I-35W Mississippi River bridge collapse. Transp. Res. A Policy Pract. 44, 771–784.