INTELLIGENT MOBILITY AND BIG DATA FOR PLANNING, TRUST, AND PRIVACY

Caitlin D. Cottrill

Introduction

According to the UK’s Transport Systems Catapult,

At its simplest, Intelligent Mobility is about moving people and goods around in an easier, more efficient and more environmentally-friendly way. It uses new ideas and new technologies to look beyond traditional, infrastructure-heavy approaches to transport, and instead to come up with innovative ways to improve mobility and make journeys better and accessible to all.

(Transport Systems Catapult, n.d.)

While much of the focus of intelligent mobility (IM) is on the technologies that enable efficient movement, the collection, integration, and use of data generated by such technologies are of equal importance. In the case of public transport, a shift towards mobile ticketing and fare payments and app-based journey planners has resulted in the development of rich and robust data resources, which have been used for a variety of purposes, ranging from day-to-day operations to more strategic long-term planning activities (Pelletier et al., 2011).

While in aggregate, such data represent a valuable addition to the public transport toolkit, it must also be acknowledged that individual data records possess the ability to directly identify users’ identities and habits as demonstrated through their travel behaviours and the linking of personal financial information. Such potential was demonstrated in 2018, when Public Transport Victoria, Australia, released a dataset containing 1.8 billion travel records covering the period between June 2015 and June 2018. The records, which were generated by 15.1 million users of Victoria’s myki public transport smartcard, contained tap-on and tap-off data (including date, time, and location) from users of the state’s tram, train, and bus network. Researchers from the University of Melbourne were able to use the ‘de-identified’ data (at times combined with other contextual information, such as Twitter posts) to successfully identify a number of travellers, including a Victorian MP (Taylor, 2019; Culnane et al., 2019).

It is evident, then, that the use of emerging data resources in public transport offers great promise; however, the protection and security of such data are also critical considerations. In this chapter, the use of big data resources in the public transport sector, with particular focus
on methods of collection and areas of use, is first explored. Next, the privacy and data protection implications of such use, including considerations of new regulatory requirements and the importance of trust as a mitigating factor, are addressed. Finally, the chapter discusses how together these considerations may impact the use of data resources in the public transport sphere, and issues related to data privacy in the context of MaaS are highlighted.

The use of data in the public transport sector

Effective public transport planning and operations have long relied on an understanding of passenger needs and preferences, routing efficiencies, and origin-destination patterns to determine optimal timing and locations of services. Traditionally, a number of data resources have been used in these planning and operations activities, including household travel surveys, on-board passenger counts, passenger intercept surveys, and, increasingly, forms of automatic vehicle location (AVL) data from vehicles (White, 2016; Behrens & Schalekap, 2010; Richardson et al., 1995; van Oort et al., 2015; Hounsell et al., 2012). While such resources have provided useful and fairly reliable data to be used in ongoing public transport operations, management, and planning, they have often proved insufficient for fully understanding longitudinal aspects of public transport use or linked trips. In addition, these traditional resources have often been temporally deficient in terms of ongoing needs for route modifications due to changing origins and destinations and modification to roadway networks (Bagchi & White, 2005; Chapleau et al., 2018).

A fundamental technology that has had myriad impacts in the public transport realm in recent decades are global positioning system (GPS)-enabled devices, such as mobile phones, GPS receivers, in-vehicle navigation systems, and others. Their decreasing cost and increasing uptake has greatly expanded the availability of location-based services (LBS), defined as “computer applications (especially mobile computing applications) that deliver information tailored to the location and context of the device and the user” (Huang et al., 2018). A variety of types of such services have been developed, including:

- Orientation and localisation
- Navigation
- Search
- Identification
- Event check (Steiniger et al., 2006)

In the public transport realm, the use of GPS data has been frequently cited in a number of areas, including providing insights in travel time variability (Mazloumi et al., 2010), improving public transport journey planners (Allulli et al., 2014), and using collected data to improve policy and planning decisions (Gschwender et al., 2016). In addition, standalone or smartphone-based GPS devices are increasingly used in household travel surveys, which has been seen to improve the robustness of data collected on public transport use, including access between the trip origin and public transport stop and the point of disembarkment and the destination (Bohte & Maat, 2009).

GPS has provided a solid basis for the collection of location data; however, in recent years, the data landscape for public transport has widened even further, largely as a result of the increasing use of both smart cards and location-sensing mobile apps (see also Chapter 36). Smart cards provide an alternative medium to cash for public transport fare payments, a
function which is particularly important during the COVID-19 pandemic. According to Bagchi and White (2005),

Each smart card can be identified by a unique serial number. The cards can be registered to a given individual, or they can be anonymous. On these cards can be placed electronically a range of fare options such as travelcards or stored value (a monetary amount credited to the card which is debited as and when journeys are made).

(p. 464)

Depending on the system, users may ‘tap-in’ to the system to record a flat fare payment, or ‘tap-in’ and ‘tap-out’ for distance or zone-based fares. For each tap-in or -out, the following data are generated: card number, date, time, validation status, and stop number (Agard et al., 2006). In addition, a determination is made if there are adequate funds available to pay for the trip and, if so, a deduction is made. In some cases, users may set up their cards to allow for direct debit from a linked financial account when funds are running low (Mezghani, 2008). Though many public transport agencies are moving away from bespoke, single-purpose smart cards and toward electronic payment using general debit or credit cards (Keitel, 2011), the expectations remain for public transport operators to have access to the relevant data generated.

Regarding mobile apps, cellular phones are some of the most ubiquitous devices in the world, and smartphones (defined as mobile phones with advanced computing capabilities) have quickly gained a large share of the market. Though adoption rates vary, largely according to national development levels (according to research conducted by the Pew Research Center, “a median of 76% across 18 advanced economies surveyed have smartphones, compared with a median of only 45% in emerging economies” [Tyler & Silver, 2019]), the rapidity of smartphone penetration rates has made them a key component of emerging public transport business models, discussed subsequently.

While there is no comprehensive estimate of the percentage of public transport systems that make use of smartphone apps, a 2017 study from Deloitte reported that of survey participants with a smartphone, 16% of persons in developing and 18% of those in developed countries had paid for public transport with a smartphone (Wigginton et al., 2017), indicating a degree of adoption worldwide. In terms of functionality, Gössling (2018) undertook an extensive review of transport- and mobility-related apps, finding the following represented (amongst others): travel information, planning, and routing; service sharing; payment and price comparisons; and participatory transport systems. More recently, Mobility as a Service (or MaaS) apps have emerged in the marketplace and aim at their simplest level to allow users to ‘bundle’ their mobility choices using a single app for planning, purchase, and delivery of multi-modal tickets and services (Li & Voege, 2017; Wong et al., 2020, see also Chapter 3).

The data collected via use of smart cards and smartphone apps and services have been demonstrated to have great benefits for the public transport sector. Creating more detailed, robust, and timely data on passenger movements, adherence to timetables, and information requests contributes to the overall potential for public transport service providers to more effectively meet the needs of their customers. Smartphone-generated location data can additionally be used for operational purposes (particularly in the absence of AVL data) and to support the provision of real-time passenger information (Corsar et al., 2013). Such applications also enhance the potential for providers to maximise the potential efficacy of public transport services (Schmitz et al., 2016; Edwards et al., 2011; Simonyi et al., 2014). Access to more detailed, accurate, and, potentially, representative data streams may also contribute to efforts to better respond to a
greater variety of user needs (Filippi et al., 2013). Given these benefits, it is evident why public transport agencies worldwide have begun to focus more time, staff, and resources on implementing smart card-based fare payment systems and developing smartphone-based technologies that provide continuing and comprehensive data streams on the actions and events of persons and places.

**Data and privacy**

The benefits of the technologies and the data they generate discussed previously are myriad. However, as noted in the introduction, they are not without concern. A key issue here is privacy – as noted in a report commissioned by the US Federal Highways Administration:

> With many transportation apps tracking and storing sensitive information, such as user location data, concerns are emerging on how to safeguard user privacy . . . and on how this might impact usage of transportation apps and services. . . . Community-based navigation app, Waze; public transport app, Moovit; and community-based running and cycling app, Strava, are transportation apps revolving around the collection and analysis of user data. They raise questions about how these data are being shared, particularly as the three apps partner and share data with cities.

(Shaheen et al., 2016)

Data generated through dedicated GPS receivers, smart cards, and mobile apps are highly identifiable. The presence of unique identifiers (such as a smart card or mobile identification numbers) combined with robust and detailed location data indicates that without adequate data and privacy protection, the potential for users’ private data to be detected or inferred is substantial (Narayanan et al., 2016; Rossi et al., 2015; Riederer et al., 2016).

In recent years, the public have become more aware of the potential for privacy violations to occur in the mobility sector, largely as a result of incidents such as the myki data release discussed previously; a 2017 reveal that the ride-hailing company Uber had been tracking its customers for up to five minutes following completion of a ride (Wamsley, 2017); and a 2018 investigative report by the Associated Press that revealed that many Google services on Android devices and iPhones store users’ location data, even if the user has enabled a privacy setting that is intended to prevent Google from doing so (Nakashima, 2018). Such violations, and their reporting in the popular media, have begun to impact user awareness and perceptions of privacy, with a 2019 report undertaken by the location data technology company Blis finding that, “Nearly two in three [American] consumers are more aware of how their personal information is being used today than they were just a year ago, and 83% of people are aware that their location is tracked” (Blis, 2019). Given such awareness, it is unsurprising that previous research has indicated that privacy concerns can be a barrier to the intent to download and use mobile apps (Gu et al., 2017) and that privacy expectations may be violated by mobile app practices (Lin et al., 2012). More specifically, with reference to travel apps, Dastjerdi et al. (2019) found that information privacy concerns contributed negatively to the likelihood of Danish consumers creating a registered account that would utilise personal information and location tracking to provide incentives and motivations to engage in ‘green’ travel.

In this climate, a number of new policies, regulations, and directives have emerged that directly address data privacy. While perhaps the most notable is the EU’s General Data Protection Regulation [GDPR—(EU) 2016/679], which became enforceable in May 2018 and
superseded the EU’s Data Protection Directive (95/46/EC), a number of others have also been enacted worldwide. Determann (2020) notes,

In June 2018, California enacted the California Consumer Privacy Act (CCPA), a novel and extremely broad and prescriptive law against data selling . . . Maine and Nevada already followed with their own state laws against data selling. In August 2018, India introduced a new bill and Brazil enacted its first data protection law, both modelled after the GDPR.

According to the United Nations Conference on Trade and Development (UNCTAD), as of 2020, “107 countries (of which 66 were developing or transition economies) have put in place legislation to secure the protection of data and privacy” (UNCTAD, 2020). Of particular note in these developments is that location data are increasingly considered to be ‘personal’ data. For example, under Article 4 of the GDPR:

“Personal data” means any information relating to an identified or identifiable natural person (‘data subject’); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person.

(General Data Protection Regulation, 2016)

The inclusion of both location data and online identifiers in the definition of ‘personal’ data indicates recognition of the ability of these data resources to identify individuals, and is generating increasing awareness in the public transport industry (Cottrill, 2020; Sørensen & Kosta, 2019). The importance of this was also acknowledged in a 2018 report from the Open Data Institute, which stated,

journey data is transforming relationships between companies and customers, drastically increasing the importance of trust, and raising critical questions of ethics, equity and engagement which cannot go unanswered. The impact of these trends will only increase as people grow more aware of these issues and as they gain more rights over data about them

(p. 4)

Under such circumstances, issues of transparency and communication of data practices in public transport, as well as trust in the system and providers, are critical to address.

Transparency, communication, and trust

The concept of transparency is critical in terms of data privacy and protection, with the GDPR defining it as follows: “The principle of transparency requires that any information addressed to the public or to the data subject be concise, easily accessible and easy to understand, and that clear and plain language and, additionally, where appropriate, visualisation be used” (General Data Protection Regulation, 2016). Of note is that methods of communication are also critical in terms of ensuring transparency, as many data privacy policies and practices have
been criticised for being overly technical, unclear in application, or time-consuming to read (McDonald & Cranor, 2008; Cottrill & Thakuriah, 2014; Lederman et al., 2016). In terms of both transparency and communication, Kitchin (2016) uses Transport for London (TfL), which oversees the provision of public transport in London, as an example of good practice, stating that, “TfL has adopted a transparent approach to data privacy and data protection policies, which are published on their website. These policies are short, clear and unambiguous, written in plain English that avoids dense legal language” (p. 57).

Given the complexities of technologies and related data described previously, such approaches are becoming increasingly important, particularly given the influence of trust in the willingness of consumers to utilise public transport technologies such as smart cards and mobile apps. Trust has been indicated as a mediating factor in privacy concerns by researchers such as Joinson et al. (2010), Miltgen and Smith (2015), and Kehr et al. (2015). In general, researchers have found that as trust in the collecting agency grows, so too does the willingness of consumers to provide data. This indicates a need for public transport agencies to ensure that they continue to be trusted to obtain and use private data through approaches such as privacy-by-design (Cavoukian, 2009; Avoine et al., 2014) in development of data-collecting technologies and transparency in their data practices. Finally, it is critical that public transport agencies be transparent not only about their collection of data but also in how they are used. Research has indicated that consumers are more likely to feel comfortable with the idea of sharing personal or potentially identifiable data when they see a clear nexus between these data and the intent of use, particularly if this may confer direct benefits to them (Cottrill & Thakuriah, 2015). In the case of public transport, providing users with clear, easily understandable information regarding what collected data may be used for (e.g. route planning, service provision, or modification of schedules) may provide adequate impetus for sharing to take place.

### Mobility as a Service in the spotlight

Issues of data privacy are particularly relevant in the case of Mobility as a Service, as demonstrated in Cottrill (2020). Given the multiplicity of actors involved in MaaS systems (including transport operators, data providers, technology and platform providers, insurance companies, and regulatory organisations [Kamargianni & Matyas, 2017]), providing transparency regarding how data resources are collected and shared is of critical importance. Data required for MaaS systems to function are likewise wide ranging, including not only those related to the transport system and public transport options (including where customers can access/egress transport assets and services, pricing information, asset characteristics, routes, and timetables [Datson, 2016; Kamargianni & Goulding, 2018]) but also user data, including origin and destination locations, acceptable transport options, and financial data for payments. Such a complex ecosystem of data and actors results in increased potential for privacy violations to occur, making data and privacy protection key considerations.

When developing MaaS systems, utilising a privacy-by-design approach will be helpful in ensuring trust of both customers and involved operators, as it will provide for considerations of how the system is enacted with respect to data sharing and use across involved actors. Utilising a trusted third party for data storage may be considered here, as this will provide protections in terms of access and data linkage, thus minimising the potential for privacy violations to occur. Ensuring that relevant data protections are built into the system and clearly and transparently presented to users will further enhance trust and respond to applicable regulation (such as GDPR in European contexts).
Conclusion

As public transport increasingly moves towards the use of digital technologies for providing passenger information, the purchase of tickets and fares, and collecting information for planning purposes, it is important to consider the implications of the data generated on consumer privacy and trust. While these technologies have the potential to generate significant benefits in terms of improving service provision and the ease of customer travel, it is critical to ensure that the data practices of providers are clearly communicated to the user in plain, non-technical language; that data protection is built in to the development of systems; and that data uses are aligned with customer expectations. By taking such precautions, it is anticipated that consumers will continue to view public transport agencies with trust and be willing participants in the intelligent mobility ecosystem. As the use of technology increases in the public transport sector, further research will be required on how best to build in privacy protections that consider both regulatory requirements and the enhancement of trust by travellers.

In the wake of the COVID-19 pandemic, issues of data privacy and protection have become only more critical. With an increasing number of countries and governments at all levels looking to contact-tracing smartphone applications to help manage the spread of the virus, considerations of how data will be treated and protected following collection have not always been fully addressed. There are clear parallels seen in privacy and data protection within the public transport realm, particularly given the spatio-temporal richness of data collected.

In addition to generalised contact-tracing apps, some countries have also implemented measures for specifically tracking public transport journeys, given concerns of the close proximity of passengers and drivers in enclosed spaces. In China, for example, passengers are required to scan a QR code for entrance onto public transport vehicles, which both indicates their ability to travel under China’s ‘traffic light’ system (with green, amber, and red codes indicating likely health status) and allows tracking of public transport trips (Law, 2020). In many locations, the use of cash for fare payment is also being discouraged, and passengers are encouraged to book travel in advance to limit crowding on vehicles (UK Department for Transport, 2020; Australian Government Department of Health, 2020). While such measures are appropriate in the context of the virus, they also have the potential to introduce additional personal data into the public transport ecosystem, and how these data are treated and protected requires additional consideration.

References


Intelligent mobility and big data


Intelligent mobility and big data


