

Apanasevic, Tatjana; Rudmark, Daniel

## Conference Paper

# Crowdsourcing and Public Transportation: Barriers and Opportunities

23rd Biennial Conference of the International Telecommunications Society (ITS): "Digital societies and industrial transformations: Policies, markets, and technologies in a post-Covid world", Online Conference / Gothenburg, Sweden, 21st-23rd June, 2021

### Provided in Cooperation with:

International Telecommunications Society (ITS)

*Suggested Citation:* Apanasevic, Tatjana; Rudmark, Daniel (2021) : Crowdsourcing and Public Transportation: Barriers and Opportunities, 23rd Biennial Conference of the International Telecommunications Society (ITS): "Digital societies and industrial transformations: Policies, markets, and technologies in a post-Covid world", Online Conference / Gothenburg, Sweden, 21st-23rd June, 2021, International Telecommunications Society (ITS), Calgary

This Version is available at:

<https://hdl.handle.net/10419/238005>

### Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

### Terms of use:

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

# Crowdsourcing and Public Transportation: Barriers and Opportunities

Tatjana Apanasevic<sup>a</sup>, Daniel Rudmark<sup>b</sup>

<sup>a</sup>RISE Research Institutes of Sweden,  
Division Digital Systems, Prototyping Society  
Isafjorgatan 22, Kista, Sweden;  
tatjana.apanasevis@ri.se

<sup>b</sup>RISE Research Institutes of Sweden,  
Division Digital Systems, Mobility, and Systems,  
Lindholmspiren 3A, Gothenburg, Sweden;  
daniel.rudmark@ri.se

## Abstract

In the last decade, private companies have successfully used crowdsourcing to revolutionise mobility, while public transport companies are still mostly failing to utilise the benefits of crowdsourcing. The application of crowdsourcing in public transport is a new area of academic research, and research on crowdsourcing *en route* in real-time is missing. This research aims to address this gap, explore opportunities and challenges of this type of crowdsourcing, and conceptualise this phenomenon. The research is based on empirical data collected in five Northern European countries. Our research findings help identify areas where crowdsourcing *en route* can add value to public transport: *new forms of communication, opportunities to communicate with third parties, and improved transit planning and optimisation*. Identified challenges are related to behavioural change for users, a need to develop infrastructure to enable crowdsourcing *en route*, and financial rationalities.

*Keywords: Future of transportation, public transportation, emerging technologies, automated vehicles, crowdsourcing*

## 1 INTRODUCTION

In the last decade, crowdsourcing has played a salient role in revolutionising mobility. This phenomenon involves specific groups of people in decision-making and production processes (Nash, 2018). The perhaps most well-known application is the navigation service Waze fuelled by data from hundreds of millions of car drivers, helping them to navigate congested cities more efficiently. When applied to public transport, several commercial third-party services are aggregating crowdsourced data today. For instance, the public transport app Moovit tracks users as they are switching between transport modes. Google Maps collect data from tens of millions of contributors as a basis for crowd-based predictions for public transit (Li, 2020). Transit.app informs its users about a vehicle's current occupancy level in real-time in response to COVID-19 challenges.

In this mobility landscape, public transport companies are still mostly failing to reap the benefits crowdsourcing entails (Nash, 2018). Hence, there is a risk that the benefits of crowdsourcing will not reach Public Transport Agencies (PTAs) and Public Transport Organisations (PTOs) but will remain with the private sector realm.

To date, the application of crowdsourcing in public transport is a new research area and literature addressing this question is limited (Marzano et al., 2019). Some studies focus on crowdsourcing

applications for modelling passengers' travel patterns (Cairo et al., 2015; Gustarini et al., 2014; Misra et al., 2014) and using crowdsourced data for geospatial mapping (Haklay and Weber, 2008). However, to the authors' knowledge, currently, no studies address the phenomenon of crowdsourcing *en route* in real-time (i.e., crowdsourcing used while riding public transport), and conceptual, theoretical works are thus missing. The present research, therefore, aims to contribute to this topic by understanding and conceptualising this phenomenon. In this study, we address the following research questions:

1. *How can crowdsourcing en route be conceptualised?*
2. *Which opportunities and challenges does such crowdsourcing entail for public transport companies?*

We applied engaged scholarship research and interviews and collected empirical data in five Northern European countries to answer these research questions. Based on collected empirical data, the authors have generalised and identified potential opportunities, challenges, and ways to solve these challenges. Our findings identify areas where crowdsourcing *en route* can add value to public transport. These are new forms of communication, the opportunity to communicate with third parties, and better transit planning and optimisation. Identified challenges are related to behavioural change for users, a need to develop infrastructure for enabling crowdsourcing *en route*, and financial rationalities that need to be solved. Based on collected data, we conceptualised crowdsourcing *en route* in real-time applied in public transport.

The main contributions and novelty of the present study are: (i) bridging the existing gap in the academic research regarding crowdsourcing *en route*; (ii) extending the taxonomy of crowdsourced activities applied for public transport proposed by Nash (2018) and applying the framework of crowdsourcing by Modaresnezhad et al. (2020) to conceptualise crowdsourcing *en route* in real-time public transport; (iii) providing a comprehensive overview and analysis of not only technical and business opportunities but also challenges related to the use of crowdsourcing *en route* in the public transport; and (iv) discussing possible ways to solve difficulties identified. With this study, we also aim to initiate a discussion of crowdsourcing *en route* in real-time public transport and propose topics for further research.

In the next section, we provide a brief overview of the related academic literature, followed by a description of the used methodology. Subsequently, we present our findings, discuss them, and provide our conclusions.

## **2 LITERATURE REVIEW**

### **2.1 Crowdsourcing: definition and use in the public transportation and mobility areas**

The term “crowdsourcing” first appeared in 2006 (Howe, 2006). Crowdsourcing can be referred to as “a type of participative online activity in which an individual, an institution, a non-profit organisation, or company proposes to a group of individuals of varying knowledge, heterogeneity, and number, via a flexible open call, the voluntary undertaking of a task” (Estellés-Arolas and González-Ladrón-De-Guevara, 2012, p.197). As this paper is focused on public transport, we use terms “a public transport user” and “a passenger” interchangeably.

To date, there is limited academic research on crowdsourcing in public transport. From a theory-building perspective, Nash (2018) proposed a tentative taxonomy of crowdsourced activities applied for public transport companies, which are:

1. Reporting activities allow passengers to provide their suggestions and feedback and report problems to a public transport company, such as reporting a pothole.
2. *Collecting and analysing data* refer to the use of today's information technologies and solutions that enable public transport actors ‘to collect, analyse, and interpret data’ (Nash, 2018). One example of such initiatives is crowdsourced real-time status information.

3. *Collaborating* represents activities where public transport companies work together with citizens to solve a problem or jointly plan public transit to find and develop solutions aligned with public transport user preferences.
4. *Acting and providing initiatives* encourage citizens to act on public transport and provide a service or create a product, for example, peer-to-peer services.
5. *Intending activities* are related to collecting crowdsourced information on passengers' intentions, making it possible to offer precisely the services that the customers want. These activities bridge customer qualitative perception of transport service with planned activity, for example, such real-time public transport-like services as Uber Pool or services on demand.

Based on this classification, crowdsourcing can be used in various activities within public transportation, such as planning, service, information, education, etc. However, in this research, we focus on crowdsourcing *en route* in real-time.

Literature addressing questions of crowdsourcing applied in public transport is somewhat limited (Marzano et al., 2019). A few studies have proposed crowdsourcing applications, for example, a mobile service, UnCrowdTPG, allowing users to plan their trips based on the real-time information about the position of vehicles in Geneva (Switzerland) and share their experience of a ride with public transport (Gustarini et al., 2014). Such data could be used to, e.g., better understand and model travel patterns of passengers; individual experience of using public transportation; and assist in designing a more efficient mobility services ecosystem. Another example is using crowdsourced data to identify unscheduled transport networks and plan efficient routes for public transport users based on this data (Cairo et al., 2015). A few studies focus on smartphone sensor data (location and acceleration) in public transport planning (Shin, 2017; Palazzi and Bujari, 2016). These studies open up a discussion about how crowdsourcing could be applied in public transport, which data could be extracted, and which knowledge it could provide.

Articles related to crowdsourcing in other areas discuss the opportunity to collect crowdsourced cycling volume data to predict ridership (for example, Jestico et al., 2016); creating user-generated street maps within the OpenStreetMap project (Haklay and Weber, 2008).

Based on the literature review, we summarise the significant benefits and challenges of crowdsourced data and list them in Table 1.

Table 1. Benefits and challenges of using crowdsourced data.

Benefits	Challenges
<ul style="list-style-type: none"> <li>- Easy to capture and map ridership information and add geospatial and temporal data (Jestico et al., 2016)</li> <li>- Possibility to improve and enrich urban analytics (Jestico et al., 2016; Haklay and Weber, 2008)</li> <li>- Possibility to tailor the service for personal needs and habits and to improve the quality of the service (Misra et al., 2014)</li> <li>- Easy to engage people in solving problems in their communities (Misra et al., 2014)</li> <li>- Possibility to reduce data collection costs (Misra et al., 2014)</li> </ul>	<ul style="list-style-type: none"> <li>- Fitness of crowdsourced data for purposes of data collection (Haklay and Weber, 2008)</li> <li>- Ability to continuously update the information (Haklay and Weber, 2008)</li> <li>- Quality, accuracy, trustworthiness, and reliability of data and ability to aggregate it (Haklay and Weber, 2008; Mirri et al., 2014; Misra et al., 2014)</li> <li>- Recruiting and engaging user participation (Haklay and Weber, 2008; Misra et al., 2014)</li> <li>- Evaluating the contribution of users (Haklay and Weber, 2008; Mirri et al., 2014; Misra et al., 2014)</li> </ul>

## 2.2 Public transport data: types of data, its use, findings, research results

Crowdsourcing is a potential new way to collect data in public transport. Hence, to understand how such data collection could benefit public transport, it is essential to survey how data has been collected

traditionally in public transportation and the advantages and disadvantages of these methods. An overview of traditional data collection methods is provided below.

**Survey data.** For decades, academic research on passenger travel patterns in public transport has focused on identifying dynamics of passengers' travel patterns and their modelling. Studies of transport surveys involve the analysis of many variables and the use of advanced statistical analyses. Traditionally, these studies have focused on passenger travel patterns, including interconnectivity among routes. However, not all transit types or transit passengers can be observed using surveys (Ceder, 2007). It is only possible to collect a fraction of the passenger population's travel habits, making extrapolation of such travel surveys rife with insecurities (Ceder, 2007). Another issue with travel surveys is data collection frequency as transit service is offered in a dynamic environment where passenger demand may change rapidly (Ceder, 2007). Finally, passenger surveys are relatively expensive to carry out (Ceder, 2007; Furth and McCollm, 1987).

**Automatic data collection systems (Automatic Fare Collection (AFC), Automatic Counting Systems (ACS), and Automatic Vehicle Location (AVL)).** With the advent of AFC, public transport agencies have been given new travel pattern data collection and processing methods. By the use of such *en route* technology, data quality has increased, data collection costs have dropped, and perhaps most importantly, the turn-around time between a change in travel pattern and public transport agency knowledge of such a change has decreased (Ceder, 2007; Pelletier et al., 2011). This way, ticketing data from smart cards' enables analysis of both day-to-day operation and long-term transit planning (Pelletier et al., 2011) by understanding passenger travel patterns over time (Agard et al., 2006; Bagchi and White, 2005). In the short term, these analyses help design more efficient time schedules (Chu et al., 2009; Park and Kim, 2008) and reduce total passenger waiting time (Sun et al., 2014). For long-term planning, these analyses enable more accurate estimations of the travel demand and service adjustments coping with ridership variations and forecasting overall passenger demand (Trépanier and Morency, 2010; Utsunomiya et al., 2006).

The advantages of data provided by smart cards are the possibility of automatically collecting data on all trips, high quality of collected data, and the opportunity to obtain large volumes of data (Bagchi and White, 2004, 2005; Hanft et al., 2016). Among the drawbacks, it is possible to mention a lack of information on the ultimate origin and destination of the trip, trip purpose, passenger's attitudes towards transport service and its quality, issues identifying individual users, impossibility to identify complex travel chains, and issues related to personal data privacy and security (Bagchi and White, 2004, 2005; Pelletier et al., 2011; Utsunomiya et al., 2006)

With wider penetration of automatic counting systems in vehicles, researchers started using this data in their studies. One example is research by Zhao et al. (2007). The researchers proposed a method for inferring rail passenger trip origin-destination matrices from an origin-only, automatic fare collection system by combining automated counting systems' data and databases of geographic information systems and database management systems. This approach could be used to replace expensive passenger original destination surveys. Integration of AVL and AFC systems can help identify passengers' exact travel patterns and connect data on transfers across multiple unlinked trips (Hanft et al., 2016). Finally, a combination of AVL and Automatic Passenger Counting (APC) data can be used for bus dynamic travel time prediction (Shalaby and Farhan, 2004), and better understanding the transit system performance in terms of efficiency, reliability, and service quality (Pi et al., 2018).

**Other data sources.** Public transport companies in transit planning and for passenger travel pattern analysis additionally integrate different sources of information. One example is data from telecom operators. Integration of this data can enable user-oriented planning solutions (Elias et al., 2016). Mobile phone data can be used for passenger pattern analysis (Tettamanti et al., 2012; Chen et al., 2016), prediction of travel behaviour at an individual level (De Sabbata et al., 2009), detection of complex trip chains (Kim et al., 2010; Feng and Timmermans, 2014) and types of trips (Çolak et al., 2015). This data may enhance traditional data used for public transit planning and allow performing more granular research on the passenger travel pattern change, for example, by estimating the difference in travel patterns depending on the time of the day, time of the week, or year (Wang et al., 2018). However, the

major challenge is adjusting traditional modelling methods to use mobile phone data (Wang et al., 2018). In addition, new theories and methods for mobile phone data are needed to predict individual travel behaviour (Wang et al., 2018).

A summary of the benefits and challenges of surveyed data collection methods in public transportation are presented in Table 2.

Table 2. Benefits and challenges of traditional data collection methods used in public transportation.

Types of data	Advantages	Disadvantages
Passenger surveys	<ul style="list-style-type: none"> <li>- Available information on passenger satisfaction and perception of the transport service, information on non-users of public services or non-smart-card users (Bagchi and White, 2004)</li> <li>- Available information on trip origin and destination (Bagchi and White, 2004)</li> </ul>	<ul style="list-style-type: none"> <li>- Only a limited sample of data (Bagchi and White, 2004; Ceder, 2007)</li> <li>- Low frequency of data collection (Ceder, 2007)</li> <li>- High cost of surveys (Ceder, 2007; Furth and McCollm, 1987)</li> <li>- Manual data collection (Bagchi and White, 2004)</li> <li>- Under-reported data that gets out of date quickly (Shen and Stopher, 2014; Wolf et al., 2001)</li> </ul>
Automatic data collection systems (AFC, AVL, APC)	<ul style="list-style-type: none"> <li>- Data is collected automatically (Bagchi and White, 2004)</li> <li>- Automatic systems provide data on all trips (Bagchi and White, 2004)</li> <li>- Data quality is improved by combining trip data with personal data, and large volumes of data become available (Bagchi and White, 2005)</li> <li>- A combination of data from different automatic systems (AVL and APC, AVL and AFC) enables tracking of passenger journeys across multiple unlinked trips (Hanft et al., 2016)</li> <li>- Automatically collected and more complete large datasets replace manual data (Hanft et al., 2016)</li> <li>- Better understanding of passenger travel patterns and re-design of routes (Hanft et al., 2016)</li> </ul>	<ul style="list-style-type: none"> <li>- No information on trip purpose, passenger's attitudes towards transport service and its quality, no info on non-users of public services or non-smart-card users (Bagchi and White, 2004, 2005)</li> <li>- Data does not identify individual users (Bagchi and White, 2005)</li> <li>- The user's ultimate origin and destination is not provided (Bagchi and White, 2005)</li> <li>- Surveys are needed to verify and confirm analysis based on AFC data and assumptions made (Bagchi and White, 2005)</li> <li>- In the case of AFC, market penetration needs to be sufficient to provide a representative sample of the entire population (Utsunomiya et al., 2006)</li> <li>- Issues related to personal data privacy and security (Pelletier et al., 2011)</li> <li>- A robust transit information system is needed to support the analysis of large volumes of data (Trépanier et al., 2009)</li> </ul>
Other sources of data	<ul style="list-style-type: none"> <li>- Enhancing traditional data used for public transit planning (Wang et al., 2018)</li> <li>- Performing more granular research on the passenger travel pattern change (Wang et al., 2018)</li> </ul>	<ul style="list-style-type: none"> <li>- Difficult to adjust traditional modelling methods to make use of mobile phone data (Wang et al., 2018)</li> <li>- A lack of new theories and methods for using mobile phone data to predict individual travel behaviour (Wang et al., 2018)</li> <li>- Unprecedented amounts of data generated by tracking human movement (Wang et al., 2018)</li> <li>- Technical issues when dealing with multiple standards and unstandardised mobile data (Wang et al., 2018)</li> <li>- Issues related to personal data privacy and security (Wang et al., 2018)</li> </ul>

### 2.3 Research framework

Several researchers (Buechler et al., 2010, Ghezzi et al., 2017; Modaresnezhad et al., 2020) have sought to conceptualise the concept of crowdsourcing. Modaresnezhad et al. (2020) based their conceptual framework for crowdsourcing (see Figure 1) on a structured literature review. The first step in this

framework is to clarify the domain where crowdsourcing activities take place. It might be both business and non-business field. As such, the model caters for a broad range of organisations, including companies, profit-seeking organisations, non-profit organisations and other institutions such as governmental entities and non-business applications (Modaresnezhad et al., 2020).

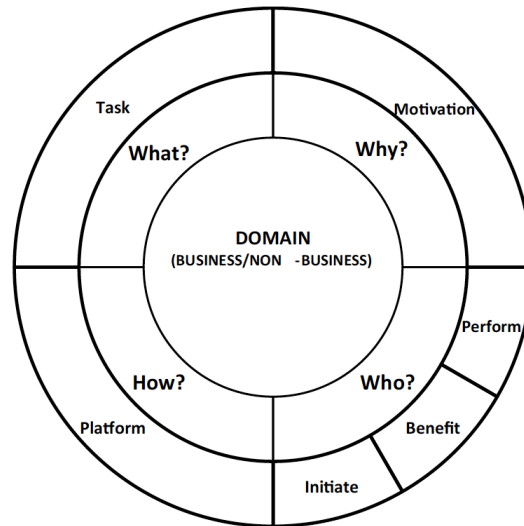


Figure 1. Conceptual framework of crowdsourcing (Modaresnezhad et al., 2020).

### 2.3.1 Who?

The framework further includes activities defining “the strategy, policy and management” (Modaresnezhad et al., 2020), reflected in the model category “Who?”.

*Who initiates the process?* The initiator could be any entity, any type of organisation or institution, or an individual who may use crowdsourcing as a problem-solving mechanism.

*Who benefits from the process?* According to Modaresnezhad et al. (2020), three groups can benefit from crowdsourcing: (i) for-profit companies may use it to meet commercial goals; (ii) the public or communities may benefit from social projects and humanity; (iii) individuals and communities may get mixed benefits ‘(e. g., customers offering suggestions for product improvement can benefit the firm as well as many customers if the idea is valuable to many customers)’ (Modaresnezhad et al., 2020).

*Who performs the task?* For the initiator, it is essential to attract the right crowd to enact the crowdsourcing activity. The crowd, in this case, is “the dynamically formed group of individuals who voluntarily participate in crowdsourcing systems to share their ideas, experiences, knowledge, work, or money” (Zhao and Zhu, 2012). It is also crucial to have good information about the crowd, which is vital for the management and the quality of the process (Modaresnezhad et al., 2020).

One crucial parameter is the source of the crowd. It could be specific communities having specific knowledge and expertise; the general public where anyone can participate; or a combination of both. The diversity and size of the crowd should be considered when planning for tools and resources to store, evaluate, and analyse crowdsourced data. Another key parameter is the level of the crowd’s skill or capabilities needed for problem-solving. Depending on a task to solve, the initiator may need general, specialised, or situational skills.

### 2.3.2 Why?

*Why does the crowd participate in the process?* The question of incentives or motivation for the crowd can be addressed by intrinsic and extrinsic motivation, widely used in psychological theories.

Intrinsic motivation refers to “participants’ inner motives such as natural feelings of competence, satisfaction, or fulfillment” (Modaresnezhad et al., 2020). In this case, an individual may participate in an activity because he finds it enjoyable, fun, engaging, and personally rewarding. Consequently, external motivation to participate is related to external incentives. Such incentives could be a direct or indirect award, monetary compensation, or recognition by others.

The framework incorporates the following categories of motivation: (i) external motivation related to external rewards and monetary compensations; (ii) introjected motivation is about gaining recognition of others; (iii) identified motivation is related to the feeling of greater freedom due to one’s the behaviour to a greater degree is adjusted to one’s personal goals and identity; (iii) integrated motivation is related to “a sense of virtual community where the activities are considered as meaningful and significant” (Modaresnezhad et al., 2020), and (v) intrinsic motivation, such as perceived enjoyment, fun, interests, etc.

When applied to crowdsourcing, examples of external motivation could be direct compensations, such as monetary benefits, a free product, or a small cash prize. Examples of intrinsic motivation are learning, gaining additional knowledge, competence, skill, networking and building personal and professional relationships, enjoyment, etc.

### **2.3.3 What?**

*What is the task?* A clear task definition is essential to identify tasks’ functions, participation modes, and complexity. The crowdsourced need to have precise functions or objectives. It is possible to specify four broad crowdsourcing functions which can be used in different combinations in one crowdsourcing initiative:

- (i) Crowd creation, related to “a contribution via a new design, product, concept, or solution” (Modaresnezhad et al., 2020);
- (ii) Crowd wisdom, related to “the projects that utilise individuals’ cognition, coordination, and cooperation through internet-mediated technologies” (Modaresnezhad et al., 2020);
- (iii) Crowd labour, related to “denotes a contribution via activities that range from simple to specialised tasks” (Modaresnezhad et al., 2020), for example, to test and evaluate, to filter content, translate, rate products, etc.; and
- (iv) Crowdfunding, which refers to asking the crowd to invest in different initiatives of individuals or organisations.

A participation mode can be integrative or selective. Integrative participation means that individual contributions need to be combined to be valuable. In contrast, selective crowdsourcing is focused on the set of provided options.

Based on complexity, crowdsourcing tasks can represent three categories: simple, moderate, and complex. Simple tasks are structured tasks “that can be broken into a series of steps” (Modaresnezhad et al., 2020) and have a single acceptable answer, such as tagging images. Moderate tasks have a higher level of complexity, for example, designing a logo or an advertisement. Complex tasks are less structured, require in-depth knowledge and experience, and are hard to evaluate. Examples of such tasks are generating a project idea and forecasting market trends.

### **2.3.4 How?**

*How does the crowd perform the task?* Initiations of crowdsourcing initiative can require a dedicated platform, communication tools (for example, Skype) or cloud-enabled collaborative solutions (for example, Dropbox and Google Drive). The ubiquitous reach of communication and internet-based solutions reduce the threshold to participate in crowdsourcing initiatives around the globe.



### 3 METHODOLOGY

The findings presented in this paper have been generated from a two-year qualitative approach following engaged scholarship principles (Van de Ven, 2007). More specifically, we worked with two PTAs, one PTO, the Swedish Transport Administration, and a systems supplier. The data generated from this collaboration adhered from full-day workshops, individual interviews with participants, and online meetings. Additionally, we used a complimentary interview study (Ritchie et al., 2013) to triangulate our findings regarding the value of crowdsourced data. The used approach was in-depth semi-structured interviews with appropriate representatives of five public transport companies located in Northern Europe. Public transport in this region is well-developed with dense public transport networks, meaning that it will be easier to make generalisations and comparisons across them regarding the role of crowdsourcing.

The authors have combined the knowledge gained during both engaged scholarship and interview study to identify the benefits, opportunities, and challenges of crowdsourced data for PTAs/PTOs.

#### 3.1 Engaged scholarship

Engaged scholarship is defined as “a participative form of research for obtaining the different perspectives of key stakeholders (researchers, users, clients, sponsors, and practitioners) in studying complex problems” (Van de Ven, 2007, p. 9). While engaged scholarship may be practised in four distinct ways, we opted for *collaborative basic research*. This strand can be seen as a middle ground between a detached outside view and more interventionist action research. Here, knowledge is co-produced, which enabled us to explore not only how things are but also how they can be (Argyris et al., 1985) leading to richer insights than traditional interviewing typically do (Adler et al., 2009).

The nexus of our engaged scholarship was five organisations within the Swedish public transport industry and concerned the subject of this paper, the use of crowdsourcing en route. Participating organisations included two PTAs (Skånetrafiken and Värmlandstrafik), Keolis, the Swedish Transport Administration (Trafikverket), and Veridict, a systems supplier. The two-year cooperation emerged as results of these actor’s interest in exploring how crowdsourcing en route could be used as a means to improve public transport. The data generated from this collaboration adhered from full-day workshops (where new ideas were explored), individual interviews with participants as well as shorter online meetings every three to four weeks (where particular matters were discussed). Using this data, we were able to map the benefits and barriers to crowdsourcing.

#### 3.2 Interview study

Under the course of our engaged scholarship study, our findings pointed towards travel patterns being the primary benefit of crowdsourcing. However, while our participating organisations provided their view on this matter, we were eager to analyse and quantify the direct and indirect benefits public transport companies would get from more precise information about passenger travel patterns. To this end, we contacted and interviewed public transport companies in Norway (Ruter and Kolumbus), Denmark (Movia), Finland (Helsinki Regional Transport – HSL), and Germany (Public Transport Authority Berlin-Brandenburg – VBB). These PTAs were chain-sampled on the criteria from peers in the industry of being trailblazers in using travel patterns. In addition, and following the recommendations of the interviewed PTAs, we also interviewed a representative from a transportation planning software provider. The list of interviewed companies is provided in Table 3.

Table 3. List of interviewed companies.

Country	Organisation	Type of actor	Contacted expert	Interview date
Norway	Ruter	Public transport company	Mobility as a Service (MaaS) specialist	18 May 2020
	Kolumbus	Public transport company	Head of strategy and development Strategic planner Project manager	2 June 2020
Denmark	Movia	Public transport company	Expert in passenger data analysis Expert working with new mobility modes	3 June 2020
Finland	HSL	Public transport company	Research specialist, an expert in passenger data collection and analysis Public transit planner	17 June 2020
Germany	Public Transport Authority Berlin-Brandenburg (VBB)	Public transport company	Expert A from Customer Information Department Expert B from Customer Information Department	9 July 2020
The Netherlands / the USA	Remix	Transportation planning software provider	Business developer	15 June 2020

The approach to data collection was in-depth semi-structured interviews with appropriate representatives of these public transport organisations, who are working with public transport planning, data collection, and analysis. In total, we conducted seven (7) interviews, comprising 12 interviewees. The interviews followed a pre-crafted protocol, lasted for approximately 60 minutes, and were recorded and transcribed.

The questions of the interview protocol were formulated based on the research purpose and objectives. Matters discussed during the semi-structured interviews were related to:

- Specifics of transit planning at specific public transport company;
- Possible monetary gains for the public transport company if the passenger location-based information is known and can be used for transit planning;
- Potential indirect benefits for the public transport company if the passenger location-based information is known and can be used for transit planning.

A sample interview protocol is presented in Appendix A.

### 3.3 Data analysis

To conceptualise crowdsourcing *en route* in real-time in public transport, we apply the crowdsourcing process framework put forward by Modaresnezhad et al. (2020) and describe Who – Why – What – How activities. The conceptual framework is summarised in Table 4.

To describe the benefits and challenges of crowdsourcing *en route* in real-time in public transport, we summarise information gained during semi-structured interviews with public transport companies. When analysing interview texts, we sought to identify similarities and differences across responses of different companies. In the next section, we present our findings.

Table 4. Conceptual model (Source: Modaresnezhad et al., 2020).

<b>Who?</b>	Perform	Crowd-source	General public	Specific group	Combination	
		Crowd-skill	General	Specialised	Situational	
	Initiate and control	For-profit organisation	Non-profit organisation	Government institute	Individual	
		Benefit	Community	Individual	Combination	
<b>Why?</b>	Motivation	External	Introjected	Identified	Integrated	Intrinsic
<b>What?</b>	Task	Function	Creation	Wisdom	Labor	Funding
		Complexity	Low	Moderate	High	
		Participation mode	Integrative	Selective		
<b>How?</b>	Platform	Self-designed	Third-party platform	Communication/ collaboration tools		

## 4 FINDINGS

We extended our findings gained during the engaged scholarship research with the help of exploratory interview study to probe: (i) which PTAs/PTOs use data for transit planning; (ii) if PTAs/PTOs have data about a travel chain between points A, B, and C, and what value has such information; (iii) what are benefits of crowdsourced data for public transport companies; and (iv) what are the challenges.

### 4.1 Data used for transit planning

As a starting point, we sought to understand which data PTAs/PTOs are currently using for transit planning purposes. Our findings are presented below.

**Survey data.** A passenger **survey** allows a more precise gathering of information about passenger patterns. Among our respondents, HSL arranges an extensive travel survey, a *travel diary survey*, once in five years on average. This survey provides information on trips between different areas and about the market share of different transport modes. Germany has quite strict regulations regarding personal data protection. For this reason, passenger surveys are one of the essential information sources on passenger patterns for VBB. The survey is arranged every 2-3 years and provides information about the entire travel chain: how people got to the first stop, from where they are departing, and where they are going.

Some companies (Kolumbus and HSL) mentioned that they consider **customer feedback** when planning routes. For example, HSL uses different methods to gain information and feedback on the plans, such as a map-based questionnaire, blogs, etc.

**Automatic data collection systems.** Ticketing data from the AFC system is one of the core sources of information used for transit planning by all interviewed companies. For most companies (Ruter, Kolumbus, HSL, and VBB), this data only provides statistics about the starting point of a journey, where the passengers check-in (i.e. pay for their trip or board on the vehicle). Movia uses a system where passengers are required to both check-in and check out. This ticketing validation process means that Movia can follow the customer's entire journey from A to B and C. VBB implemented a similar check-in/check-out system; however, due to strict privacy regulations, they could not make use of this valuable data.

However, **travel smart cards** are just one type of ticket, such as mobile ticketing apps and single paper tickets. Availability of alternative tickets affects the quality of data provided by smart travel cards. For example, HSL has mentioned that in recent years, it has deteriorated, mainly because more passengers switch to mobile ticketing solutions. Currently, only between 40 and 50 per cent of Movia passengers use smart travel cards (*Rejsekort*), while the remaining passengers use other types of tickets (paper-based, mobile ticketing, etc.).

“We have perhaps better statistics on single trips with the *Rejsekort*. <...> Outfashioned monthly passes on paper, single tickets, app-based tickets are not check-in/check-out” (Movia).

“We have a number of tickets bought, but we do not necessarily know where they have been bought and on which line they have been used. We get this information from the HSL cards. But more and more people are changing to mobile tickets, and the quality of passenger boarding data is getting worse and worse” (HSL).

“Approximately about 85% of all tickets that are sold are monthly or yearly subscriptions. Most of them, in recent times, were sold as paper tickets with no chance to track them at all. We have been changing that and migrating monthly and yearly tickets to smart cards” (VBB).

Moreover, some companies (Kolumbus and HSL) mentioned combining travel card data with **account-based ticketing (passenger account)** data for transit planning. However, it needs to be said that the popularity of mobile ticketing solutions negatively affects the quality of ticketing data provided by smart travel cards.

All interviewed companies are using **ACSs** on vehicles. These systems allow PTAs and PTOs to know the numbers of passengers travelling on a specific vehicle.

“We know where people go on and off. We know huge clusters. Today, the most are from our automatic counting from our vehicles” (Ruter).

“Today, all buses driving for Kolumbus have a passenger counting system. <...> This gives us general but quite good information about how full the bus is, how many are entering at each stop, and how many are exiting at each stop. But we do not know who is travelling between the stops nor the travel patterns” (Kolumbus).

[With the help of an automatic counting system] “we have been very good at knowing how many people used a given line from a given bus stop at a certain time” (Movia).

“The VBB already uses sensors of automatic passenger counting systems on almost all vehicles. <...> However, this data is not evaluated in real-time, but rather post-processed” (VBB).

Representatives of HSL have mentioned that they use **previous year run-time data** of buses for each route to plan timetables for the coming year.

**Other data sources.** Several companies use available **open statistical data**. For example, planners at Movia use existing e-metrics borrowed from a National Transport Model as underlying assumptions. These e-metrics describe in a comprehensive way how people travel between different zones. Census data is used at Ruter and Kolumbus. Also, planners at Kolumbus use statistical data on the concentration of jobs and population. Planners at HSL use information and plans of the travel network.

Some companies (Ruter, Kolumbus, and HSL) are starting trials to use the data from **their mobile apps** to track their passengers’ travel patterns. For example:

- Passengers of Kolumbus can remain anonymous and share their position with the company’s mobile app. This solution enables both a better trip planning service and context information. Currently, about 50 per cent of app users do it. Passengers can also select and share their locations with Kolumbus to improve service. Due to GDPR, the company cannot see the exact location but can see the area. Less than 15 per cent of app users agree to share their position with Kolumbus.
- HSL is developing a tracking system in its mobile app to track their customers if they give their permission. The systems will allow following the customer and recognise the transport mode on each part of the journey. At the moment of the study, this system is in testing mode.

Some companies (Ruter, Kolumbus, and HSL) mentioned that they tried using **mobile network operator data** for planning. Ruter has been running projects related to mobile network operator data for years. Kolumbus collaborated with Telenor and Telia, mobile operators in Norway, to see if their data can provide helpful travel patterns. HSL has a new pilot project with Telia focused on using the data of this mobile operator for analyses on (i) the total amount of trips from different areas; (ii) the trip distribution between different areas; (iii) the difference between an average working day (Monday to Thursday), Friday, Saturday, and Sunday; and (iv) the trip variation around an average working day.

VBB was interested in using mobile network operator data for planning, but mobile operators are not offering it.

“In our route planning project, this [Telia mobile operator] data was used already last year as background information. It was used not as a direct way to make decisions but, for example when there was a single comment from a customer that there was a need for a certain connection between two areas. Telia origin-destination data can be used to check if there are more people than just this one or just a handful. And there was an actual need in connection, and the planning was changed accordingly based on that data. It was a single case but a good example of possibilities of these data sources” (HSL).

Due to GDPR, mobile network operators provide information about the flow of passengers in clusters, for example, with no less than 25 people in the cluster for the urban area and even larger sample sizes for rural areas. As Kolumbus representatives commented, this gives some trends, but the technology is not yet good enough to provide useful travel patterns.

HSL is actively looking for new ways to involve customers in the planning process. One example is **workshops** with a group of passengers, and another planned activity is creating a **research panel**. It will be a digital crowdsourcing application for surveys and discussions.

## 4.2 Data on travel chain (tracking the passenger trip between points A, B, and C)

We tried to understand if the data sources used by our interviewed PTAs/PTOs were sufficient in getting information on a whole travel chain. In other words, if PTAs/PTOs can track the passenger trip between points A, B, and C, and what is the value of such data. In this sub-section, we present a summary of this discussion.

A few companies confirmed that they know the passenger travel patterns when moving from A to B, and C. VBB gets this information through surveys.

“We are asking: where do you come from? And where do you go? How do you get to the first stop: on foot, by car, by bike? Here we are asking about the whole travel chain. But as manual data, that is a lot of work. This survey is done every 2 or 3 years” (VBB).

Movia happens to be the only public transport company out of those in this study that can track the passenger travel chain with the help of its ticketing system. Previously, the planning depended on data from automatic counting systems in some of Movia’s buses or on manual counting systems. Now, the passengers are required to check-in and check-out within the public transport system. This way, Movia gets the data on how passengers change from one bus to another, from bus to train, and vice versa. Also, there is an internally developed system, which can be used for the analysis of all data from travel cards in combination with passenger account data. Such a combination allowed for:

- The possibility of investigating more complex network effects of making changes at a specific line in a particular area.
- The possibility to better understand how people change between busses, trains, metro, or between different lines at some of the more significant hubs. With the help of the system, Movia has got excellent knowledge about how people move around at the Central station in Copenhagen.
- Also, this data has much potential when used to model different travel scenarios. In the future, Movia plans to investigate on a pretty detailed level how such changes will influence the travel pattern in the transport system.

Analysis of data on how passengers change from one route to another was considered the most valuable for the public transport company. This dataset allowed Movia to optimise lines marked by a high degree of interchanges by maximising the correlation between these lines. The major problem is that only 40-50 per cent of passengers are currently using the check-in/check-out system. To approximate behaviour for all passengers, the data is hence scaled up to the whole population.

Representatives of Kolumbus commented that the travel chain data “*will be very valuable data for many reasons: (i) for the service, (ii) for new lines, (iii) for work capacity, and (iv) for travel capacity. For*

example, if you see a travel pattern with changes of buses, maybe it would be an idea to have a direct line.”

Representatives of HSL noted that “Until now, we have not got any information on the whole trip. We also have not got information from the transfer. Because there are more and more transfers, it would be good information to gain; the knowledge about **what happens during the transfers and how the transfer is actually working**. <...> And the planning is done accordingly so that the timetables would coincide and the transfer would be as good as possible. But even if we plan it perfectly, it does not necessarily work in reality. Then this would give us the information on how it works in real life. For example, is there any sort of tail light effect when a passenger always sees taillights of the bus and always is 1 minute late.”

### **4.3 Benefits of crowdsourced data**

Crowdsourced data (for example, location-based data and passenger feedback) may have several benefits for PTAs/PTOs. Benefits and opportunities of crowdsourced data *en route* in real-time were in the focus of engaged scholarship research. A discussion of such benefits and opportunities became a central part of interviews with PTAs/PTOs. A summary of these findings is presented below.

#### **4.3.1 Enabling new forms of communication between public transport operators and authorities and their customers**

Offering a more attractive service involves increased opportunities to communicate with passengers. As Movia and HSL noticed, the knowledge of passenger travel patterns in real-time would lead to an accurate and more targeted passenger information service in the case of breakdown of the public transport system, or delay of the buses, or construction works in real-time. For example, public transport companies could send push messages to their passengers’ seating in a specific bus, inform about a delay of the bus operating an inter-connected route and suggest alternative ways to continue their trip. This will positively affect passengers’ attitudes towards the company if they are better informed about the travel situation and the reasons for delays.

“Today, when there is road construction or there is a breakdown in the system. We have a lot of difficulties in informing passengers about alternatives. <...> If we have this data, we could send push messages to people seating in that specific bus and suggest changes for them” (Movia).

“Our information department would like to give our passengers information according to their movements, on where they exactly are, at a particular time. For example, if something happens in the transport system, they could inform passengers, and maybe they could change the route or attitudes could be more positive if they could know what is going on and what the delays are” (HSL).

Another example (tested during the project) was communicating the *experienced* level of crowdedness in a public transport vehicle. While onboard equipment such as cameras, built-in vehicle weighing instruments, and ticketing machines can be used to assess the overall current passenger count of a vehicle, this measure does not necessarily convey how passengers experience the current level of crowding. To address this shortage, Värmlandstrafik installed QR codes throughout a subset of their vehicles that passengers could scan to land on a mobile web page where such more experiential cues could be entered. This way, more objective passenger count data could be augmented by subjective experiences.

Consequently, crowdsourcing has the potential to make the public transport service more interactive due to the possibilities of passengers reporting back to PTAs/PTOs about unexpected delays or issues. This information could result in better management and planning of such problems and enable a quicker response to the changing travel patterns. For example, travel surveys happen once in 3 to 5 years. But if a public transport company gets a daily or weekly update on passenger travel patterns through its mobile app, this would allow a much more flexible response based on their customers’ needs.

#### **4.3.2 Enabling communication with third parties**

One more benefit of crowdsourced data *en route* is the potential to enable communication with third parties. This benefit is possible to realise when the first and the last miles are known. For example,

crowdsourcing allows an opportunity to inform a medical institution that a passenger will be late due to an unexpected transport delay.

Crowdsourced information *en route* would be crucial for MaaS providers and would allow them to tailor different transportation modes for the needs of individual passengers. For example, if a passenger pattern shows a need for a car, a MaaS company could suggest a shared car suiting that travel pattern. Such usage of MaaS will result in the broader penetration of more sustainable transport modalities, such as shared cars, city bikes, e-scooters, etc.

“We will gain from knowing much more about the passengers. [We will be able] to tell them: ‘With your pattern sometimes you need a car. But you do not have to buy a car.’ <...> Then, if someone needs a car, we [being a MaaS company] will provide him with a shared car or a rental car for that moment” (Ruter).

“First and last mile data is of extreme interest for us. <...> we would like to look into using data from people’s smartphones and how we could share knowledge and data with some of the companies that are mobility providers, for example, scooters or electric cars, and build networks with them” (Movia).

“To have enough data and see where people are travelling would make the service more for the people’s needs, where we could provide different modes of transportation. <...> In the future, we would like to suggest different transport modes that are more suited for each individual” (Kolumbus).

“More efficient services can be created because the transport companies will have more detailed information to provide a service to match that” (Remix).

One more example that was discussed during the project was using crowdsourcing mechanisms in connection to school buses. In more rural areas, such public school buses are typically the only way for children to travel to school. However, many parents of students in the lower grades often call the bus driver to make sure that their children have indeed boarded the bus. Since this consumes both time and attention for the driver, we discussed how techniques rooted in crowdsourcing could be used instead. Here, children would, for example, scan a QR code during on- and offboarding that would trigger a notification for the parent regarding the child’s travel status.

### 4.3.3 Utilising crowdsourced data for planning and optimisation purposes

Crowdsourcing is a potential new way to collect data for public transport companies, providing data on the entire passenger travel chain. During interviews, we have discussed the potential value of having crowdsourced passenger location-based data for PTAs/PTOs. In this case, such information provides exact data about passenger travel patterns. Our interviewed PTAs/PTOs commented that this would lead to many economic gains because of **transit optimisation and better planning**. Representatives of Movia additionally commented that this data would help avoid spending money on unnecessary replacement buses during substantial construction works on the rail network.

“<...> if we could only optimise the running of our buses by a very, very small number, way less than 1 per cent, we would have easily paid for this system many times.” [If in the future, passenger location-based data would be available, it would help to] “have more agile planning of buses by use of this data” (Movia).

Improved planning of public transport would lead to a more smoothly operating system. Consequently, this would result in more **efficient services**.

“If we can make a system that runs a little more smoothly and efficiently, with better changes between buses, trains, and so on. Then we would have to some extent more passengers, or passengers will travel a little more than that would be otherwise. That is one side of it. Another side, the way we optimise our system, and the way we use the money for running our system can be done in a more precise way” (Movia).

“If we get the same service with fewer vehicles, this is direct economic gains. Or the same money can be used for extra service for a congested line” (HSL).

“We have to see that we are using the available infrastructure and budget in the best possible way. [This is efficiency] but not in terms of minimising costs. It is about using the available budget and getting most of the service for it” (VBB).

The majority of interviewees (Ruter, Kolumbus, Movia, HSL, Remix) stated that knowledge about exact passenger patterns would allow **offering a more attractive service for passengers**. As one side of

improved service, it could be better management and planning of unexpected delays. Another side would be a quicker response to the changing travel patterns (Remix).

“We are trying to get as much information as we can about our current passengers and possible future passengers so that we could provide the best service” (Kolumbus).

“Second biggest concern and complaint in public transport is the unexpected delay. And this kind of data can help us both to estimate the effect of delays and to plan delays. Can we do some work around this in a way that it would affect fewer people?” (Movia).

“This more detailed information gives us more possibilities to plan a better service” (HSL).

“It allows them [PTAs] to respond more quickly. <...> But if it is an app where they can see a weekly summary and where they [PTAs] have an update on travel patterns, this just allows them to be much more flexible in responding to the needs of the community” (Remix).

## 4.4 Identified challenges

During the engaged scholarship research, we have identified several challenges related to introducing and implementing crowdsourcing *en route* in real-time. A summary of these findings is presented below.

### 4.4.1 User challenges

The first type of challenge concerns *users* and their adoption of the service. While achieving public transport user adoption is essential for all types of applications described in Section 4.3, it is paramount to gain the primary identified benefit, i.e., utilising crowdsourced data for planning and optimisation purposes. A core task to enable the benefits of *en route* crowdsourcing described above is to pervasively connect the users to a technical infrastructure where crowdsourcing interactions and data collection can occur. At the very heart of such a connection lies that the user needs to disclose their positions. While such disclosure may seem trivial, it fundamentally changes the relationship between passengers and public transport. Public transportation has a long history of offering private ways of travelling where the passenger may, for example, pay in cash to travel anonymously throughout the transit system. This anonymity contrasts with other, more contemporary transport modalities where disclosing the user’s current position is designed deep into the service. Thus, such disclosure is a non-negotiable requirement to use, for example, hail-riding or e-scooters.

For instance, Kolumbus noticed that their position-sharing app functionality were used by just some 15% of its users. In addition, these numbers were for beta version app usage, targeting early adopters (that assumingly are more willing to share their position). As such, this points to the necessity of creating user incentives to share their positions.

### 4.4.2 IT System Challenges

All applications described in Section 4.3 will require that IT systems are re-architected to some degree. For instance, enabling more personalised traffic information to passengers is beyond the capabilities of many current systems that offer more general traffic information.

However, our study points towards that the immediate primary benefit for agencies to engage in crowdsourcing is to pervasively collect data about the movement in the public transport system. Such data collection requires that the application know the context in which the user is currently located. This context can be a particular vehicle operating a particular route at a specific time or a certain bus stop. During our work, we saw the thus need to establish a “digital link” between the rider’s phone and the infrastructural component that the user is currently utilising (for example, a vehicle or a bus stop). This link is typically designed deep into these services in newer modes of transport, such as hail-riding, car-sharing, and e-scooters. For instance, when a passenger boards an e-scooter, the user must be identified and acknowledge what vehicle s/he is using.

However, during our meetings and workshops, PTAs were often hesitant or outright negative towards installing specific on-board equipment to enable pervasive crowdsourcing. This was primarily due to the associated costs to both install and maintain on large fleets of vehicles.



### 4.4.3 Financial Rationales

Finally, public transport agencies operate under strained financial circumstances, which have been severely augmented due to the ongoing pandemic. Consequently, there are several possible competing investments into which the scarce resources may be invested at any given time. Since the applications of crowdsourcing in this research do not generate direct revenues, our data suggests that the benefits crowdsourcing do entail, very well can be converted into financial gains. A prime example of this was brought forwards by Movia as they were able to reschedule their operations using travel exchange data only by a small fraction, this would justify the investments, as explained by a representative at Movia:

“[I]f we could only optimise the running of our buses by a very very small number, a way less than 1 per cent, we would have easily paid for this system for a number of times” (Movia).

## 5 DISCUSSION

### 5.1 Data used for transit planning vs crowdsourced data

The findings from our research show that the interviewed PTAs/PTOs in transit planning mainly rely on data collected from surveys, automatic data collection systems, and historical data. Automated data collection systems, such as AFC and ACS, used together with passenger account data, give a picture of larger clusters and help to create efficient time schedules, which is in line with findings of the previous research (Agard et al., 2006; Bagchi and White, 2005; Chu et al., 2009; Park and Kim, 2008; Trépanier and Morency, 2010). However, this data does not disclose individual passenger travelling (Bagchi and White, 2005) and, in all cases except Movia, does not provide information on the full passenger travelling chain. On the contrary, passenger surveys offer much more detailed information on the full passenger travelling chain. However, these surveys are carried out manually once in 3-5 years and require substantial resources, both in conducting the survey and in the further processing and analysis of data, which is in line with findings of the previous research (Ceder, 2007; Furth and McCollm, 1987). To summarise, certain limitations of these data sources are:

- Data from both surveys and automatic data collection systems are limited to people using public transport (Bagchi and White, 2004, 2005). However, it is also interesting for public transport companies to know “*the travelling information for those who are not using public transportation*” (Kolumbus).
- Automatic data collection systems do not provide data on passengers’ preferences and attitudes towards public transport services (Bagchi and White, 2004, 2005). Representatives of VBB mentioned that by using all data collection methods, except the questionnaires, “*we only know if a certain vehicle is full or not. And what we do not know is if this is their [passengers’] preferred way or do they take it because the planning is bad. You can find out this only if you know the actual starting destination of the journey. That is why we need to know more about passengers: from where and to where are they actually travelling? Why are they taking a certain route?*”
- PTAs/PTOs offer a wide range of available ticketing services and tickets, including paper tickets, mobile tickets, and travel smart cards. However, the share of passengers using smart travel cards is declining, affecting the quality of this data. Previous research corroborates this finding, stating that a sufficient penetration of this type of ticketing is necessary (Utsunomiya et al., 2006).

Our findings also show that PTAs/PTOs are actively using and integrating new sources of information in transit planning to optimise it and provide a better service. Many PTAs/PTOs use open statistical data, giving background information about the population. Mobile network operator data and data from mobile apps of public transport companies are of interest. However, using these data may inflict on personal data privacy and security (Wang et al., 2018). In addition, PTAs/PTOs need to find out how this data could be used and how it could be combined with the existing data.

Due to the check-in/check-out ticketing system, Movia knows the exact travel patterns or the exact travel chain of its passengers. The highest value of this data is in providing knowledge about what happens during the transfers and how transfers are currently working. This information can significantly improve and optimise transit planning. And this is the identified area where crowdsourcing *en route* in real-time could benefit public transport companies the most.

### 5.2 A conceptualisation of crowdsourcing *en route* in real-time

To conceptualise crowdsourcing *en route* in real-time in the public transportation area, we use the framework proposed by Modaresnezhad et al. (2020) (see Table 5). Description of all elements is provided below.

Table 5. Conceptual model of crowdsourcing *en route* in real-time.

<b>Who?</b>	Initiate and control	PTA / PTO
	Benefit	<b>Combination:</b> Initiator, individual, community, third party provider
	Perform	<b>Crowd-source:</b> by a specific group (public transport users)
<b>Why?</b>	Motivation	External and Intrinsic
<b>What?</b>	Task	Function: Creation
		Complexity: Low and Moderate
		Participation mode: Integrative or Selective
<b>How?</b>	Platform	Self-designed or Third-party platform

*Who initiates the process?* In this case, the actor that starts the process is a PTA/PTO.

*Who benefits from the process?* In the case of crowdsourcing *en route* in real-time, it is possible to specify a combination of four groups that can benefit from this. The core parties are an initiator (a PTA/PTO) and an individual (a public transport user or a passenger).

a. Benefits for a PTA/PTO. By gaining crowdsourced information on the location of their passengers in real-time, public transport companies gain knowledge on the whole travel chain of their passengers. This information could help optimise transit planning and offer a more efficient, attractive, and smooth service, which can flexibly adapt to changing passenger patterns. In addition, PTAs/PTOs can enable new forms of communication with their passengers by informing them about sudden delays and traffic changes in real-time. Finally, crowdsourced information shared with third parties could help to bridge different MaaS services with public transportation. Benefits for a PTA/PTO are summarised in Figure 2.

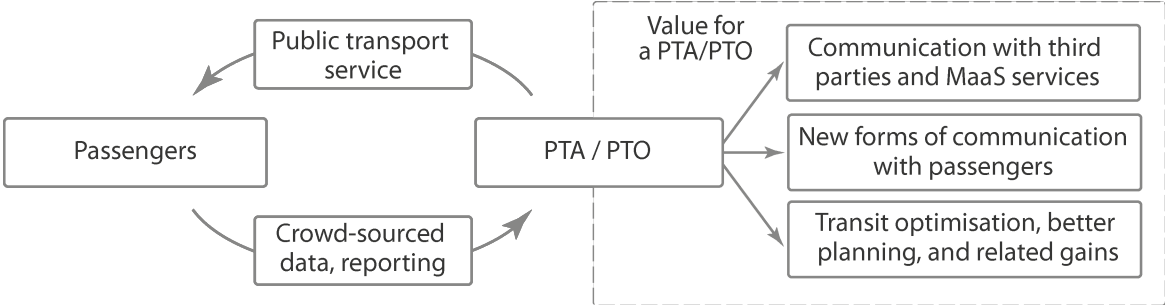


Figure 2. Benefits of crowdsourcing *en route* in real-time for a PTA/PTO.

b. Benefits for an individual (a public transport user). There are several benefits that public transport users can gain by crowdsourcing their data to PTOs/PTAs. The benefits for passengers are a better and more efficient service, which can be personalised and targeted depending on the passenger’s need and location. Crowdsourcing can make this service more interactive by allowing passengers to report back to PTAs/PTOs about unexpected delays, trip experience, or noticed issues. Finally, the opportunity of connection to MaaS would result in a smooth passenger experience of a whole journey, including the first and last mile. Benefits for a passenger are summarised in Figure 3.

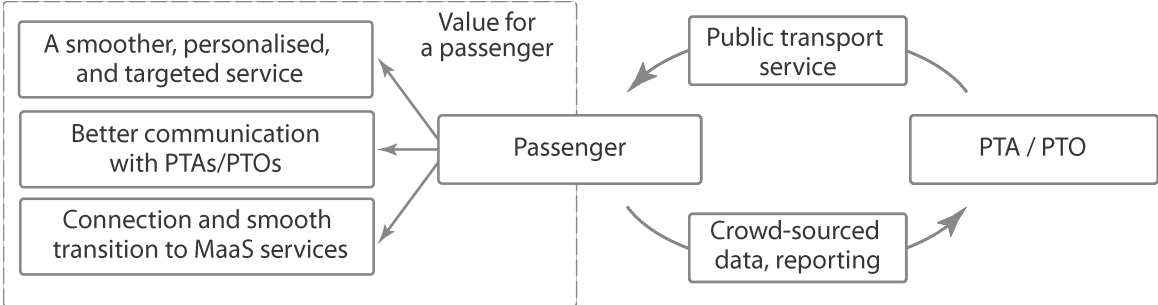


Figure 3. Benefits of crowdsourcing *en route* in real-time for a passenger.

c. Benefits for a community. Exact information on travel patterns might also benefit communities. PTAs/PTOs can adjust and offer a better and more flexible public transport service through the ability to react to changing travel demands quickly. For example, suppose a PTA/PTO company based on passenger travel pattern notices a need to introduce a new line. In that case, this will be beneficial not only for single passengers but for the entire community.

d. Benefits for third parties. Crowdsourcing *en route* makes it possible to share the information on the first and last mile with MaaS providers and make an interchange between different transport modalities smoother. Sharing this information in different contexts may also benefit other involved parties depending on the context; for example, parents get informed about their children trip to school, or a doctor gets notified about a patient being late.

*Who performs the task?* Public transport users will typically initiate crowdsourcing *en route*. They are expected to share their location-based information, inform public transport companies about their trip experience (for example, crowdedness in a vehicle), etc.

*Why does the crowd participate in the process?* In crowdsourcing *en route* real-time, the expected categories of motivation for the crowd would be external motivation and intrinsic motivation.

External motivation to share location-based information is related to a better targeted and personalised public transport service and a smooth experience of public transport and MaaS services. In addition, PTAs/PTOs might consider using specific service discounts to people who crowdsource. This way, by crowdsourcing and sharing their data, passengers gain particular value back (see Figure 4). In contrast, intrinsic motivation might be connected to the enjoyment of contributing to a co-creation of a better service.

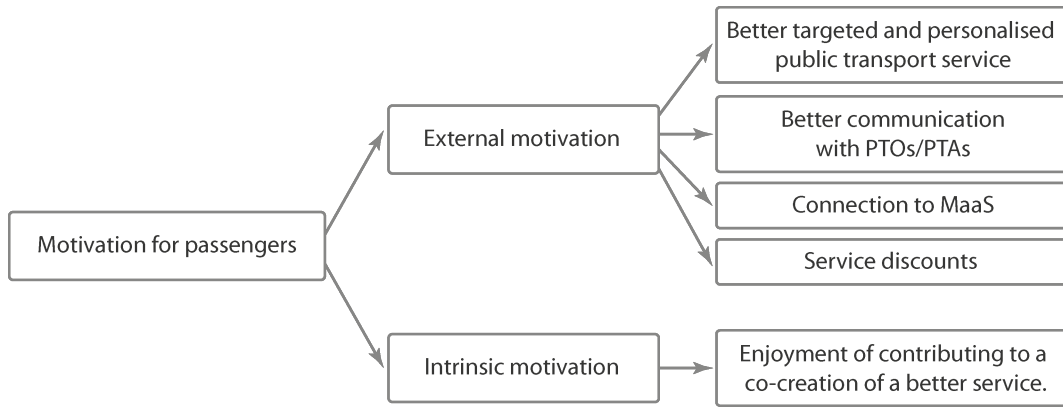


Figure 4. Motivation for the crowd to take part in crowdsourcing *en route* in real-time.

However, it is essential to highlight that crowdsourcing *en route* in real-time is a way to maintain relationships between a PTO/PTA and a public transport user during a trip and is an instrument to create added value of the public transport service. This type of crowdsourcing is only sustainable over time if both sides, PTOs/PTAs and public transport users, gain value from it. There should be interest and motivation coming from both sides, meaning that an (inter)dependence exists between them (see Figure 5). Crowdsourcing *en route* will not happen if a PTO/PTA is willing to accept this data and use it for service improvement, but passengers do not share their data. Likewise, crowdsourcing *en route* will not happen if passengers report about their trip experience and share their data, but a PTO/PTA does not react.

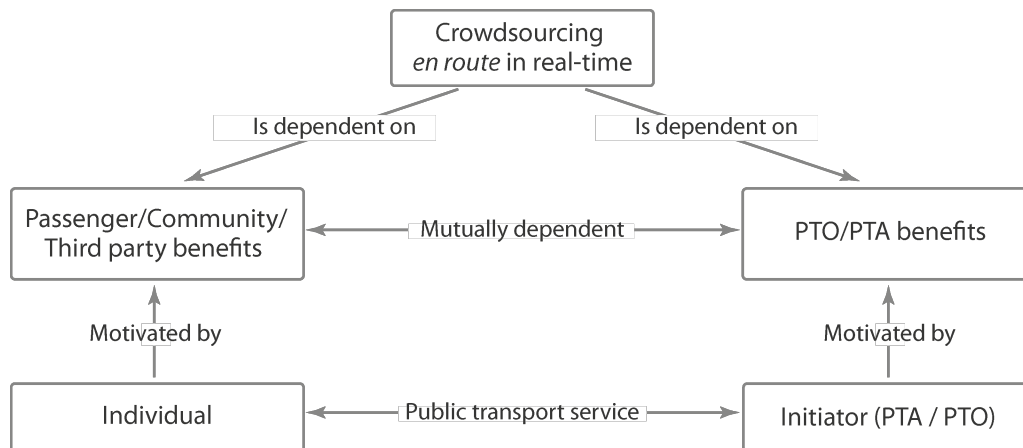


Figure 5. Interdependence between a PTA/PTO and a passenger in the case of crowdsourcing *en route* in real-time.

*What is the task?* In the case of crowdsourcing *en route* in real-time, the task would be related to crowd creation. Applying Nash (2018) taxonomy to our findings, the most relevant activities would be associated with the first and the second activities: (1) reporting and providing feedback and suggestions; and (2) providing real-time location-based information for collecting and analyses.

The fifth activity specified by Nash (2018) – intending activities – would be relevant in the context of crowdsourcing *en route* in real-time. It could be related to using passenger feedback as a base for

providing services on passengers' demand and according to their needs. PTAs/PTOs are interested in these activities, but this is a task with higher complexity for passengers.

In addition, we extend Nash (2018) taxonomy by adding a situation of communication between a passenger, a driver, and a third party actor.

The complexity of task 1 and 2 is low. Passengers need to share their location data automatically and provide their feedback, which is a structured task. The task 5 is relevant but has a higher complexity.

The crowdsourcing initiators may select whether they focus on integrative or selective participation.

*How does the crowd perform the task?* To fully benefit from crowdsourcing *en route* in real-time, PTAs/PTOs might need to design a specific platform or use a platform created by third-party providers. A core task of such a platform is to establish a link between the passenger and the vehicle that the user is currently riding. This is important since many tasks to be executed by the passenger require knowledge on what vehicle that the passenger has boarded. As mentioned above, these links can be established by onboard equipment such as NFC scanners or QR codes. However, more recent AI developments also allows for platforms that matches a phone's GPS signature towards real-time vehicle movements (or even timetable and road data) to automatically identify the current vehicle that passenger has boarded.

### 5.3 Proposed ways to address identified challenges

Our research helped to identify a few difficulties: disclosing anonymity, connecting passengers' and public transport's physical infrastructure, and financial rationales/motivation (see Section 4.4). Below we discuss alternative solutions to them.

The first identified challenge is related to *disclosing user anonymity*. Public transportation historically is an anonymous way to travel. Disclosing information *en route*, passively or actively, encompasses *user adoption and behaviour change* meaning that passengers may need to give up some anonymity to provide data. We have identified three primary ways of addressing user position disclosure during the project:

- The first strategy was to *provide sufficient incentives* for the user to disclose their current position voluntarily. Here, services that were "too good to miss out of" were discussed in detail, and examples were, for instance, APC data complemented by subjective crowdsourced perceptions of crowdedness, as well as communications with third parties.
- A second identified consent-based user strategy entailed asking the user for location-sharing when an app is first launched. This approach's benefit is that it does not require profound service redesigns as in the previous example, but on the other hand, it hinges on users' goodwill.
- A third identified strategy for pure data collection purposes is to scan buses for devices using Bluetooth and Wi-Fi protocols. While our investigations showed that this consent-absent approach is possible both technically and probably legally (considering legal privacy frameworks such as GDPR (Thomas, 2018)), it entails several drawbacks. First, it requires installing scanning equipment on all infrastructural components and thus substantial financial investments. Second, the use of non-consensual data risks creating a negative public momentum around such data collection.

However, this study offers little empirical evidence on this subject, and we thus argue that further research on a public transport user incentives are necessary to address this challenge.

Another challenge is related to the fact that crowdsourcing applications require a "*digital link*" between the rider's phone and the infrastructural component that the user is currently utilising (i.e., a vehicle). This link is fundamental to know where the passenger is located and to be able to yield value from crowdsourced data. We found the following potential ways of establishing such a digital link in public transport:

- The first principal way is that the user establishes such a link. This link can be created by the user manually entering what line and departure s/he is currently on or bus stop that the passenger is at presently. Another user-initiated link is to scan onboard equipment or QR codes located within the public transport network to determine the exact location. If the purpose is data collection exclusively, the chances are slim that sufficient user penetration will be achieved. This link needs to be established in an automated fashion in these cases.
- The second way of establishing such a link is thus without user involvement, and where the app instead calculates what public transport network component the public transport user is using. If GPS is installed on the vehicle and its movements are recorded by the public transport agency, it is feasible to match the phone's GPS signature with the most similar vehicle's GPS signature. Moreover, when such a real-time GPS signature does not exist, we experimented with machine learning to determine the vehicle's most probable location by drawing timetables, historical data, and road geometries.

Finally, the question of *financial rationales* must be solved. In our ongoing interactions with PTAs/PTOs, we found that financial motivations centred around increased user satisfaction, as this is an aspect on which these organisations are evaluated. Our work identified two different factors of increasing user satisfaction through crowdsourcing *en route*.

- First, designing a more *demand-responsive public transport offering* was brought up by all PTAs. The lack of data points disables sufficient analyses on how passengers are using the public transport system. For instance, one of the PTAs had a current overall customer satisfaction of 48% yet a target of 85%. Thus, the financial rationale for this PTA to engage in crowdsourcing was to design planned traffic that was better aligned with customer demand. Also, the possibility of responding to real-time disturbances in a better way, such as knowing how many passengers are stuck at a particular stop or station during, for example, railway breakdowns, would accommodate a more suitable response action increased user satisfaction.
- Second, increasing user satisfaction through the services that are enabled through crowdsourcing was also brought up. This strategy included communication with third parties (other mobility providers, parents, school bus passengers) and tailored traffic information.

## 6 CONCLUSIONS

In this study, we focused on crowdsourcing *en route* in real-time and its application in public transportation. Our aims were (i) to conceptualise this phenomenon and (ii) to understand the opportunities and challenges that this type of crowdsourcing entails for public transport companies. We base our findings on results of engaged scholarship research (with a focus on Sweden) and semi-structured interviews with public transport companies located in Norway, Denmark, Finland, and Germany.

### 6.1 Opportunities and challenges of crowdsourcing *en route* in real-time

Today, public transport companies use several different data sources for transit planning. These data sources encompass passenger surveys, automatic data collection systems, and open statistical data. To enrich these data, public transport companies additionally use open statistical data, mobile network operation data, and data from mobile apps. Triangulation of these data sources helps to optimise transit planning. However, these data sources have certain limitations; for example, public transport companies know if a specific vehicle is complete or not; however, there is no clear answer on how passengers' full travel from A to B and C and transfers between lines looks like. And this information is the most interesting for public transport companies.

Our research points to several areas where crowdsourcing can add value to public transport.

First of all, crowdsourcing enables *new forms of communication* between public transport operators and authorities and their customers. Its value is related to:

- Targeted and more accurate traffic information, for instance, about traffic changes and delays in real-time and offering alternative routes.
- Enabled opportunity for passengers to report back to PTAs/PTOs, for instance, about the experienced level of crowdedness.
- Benefit from more interactive services, for instance, reporting *en route* violence and other disturbances with a PTO/PTA operator.

Secondly, crowdsourcing *en route* enables public transport actors to *communicate with third parties*, such as medical institutions or last-mile transportation (for example, MaaS providers).

Finally, it is possible to *utilise crowdsourced data for transit planning and optimisation purposes*. More specifically, crowdsourced data could fill the gap and provide information on the entire passenger's travel chain. In particular, data about an interchange describing what happens during the transfers and how the transfers are working will be the most valuable. Such interchange data is mostly missing today. However, it has a high potential to model or simulate different travel scenarios and provide a very detailed analysis of how changes in the transport system would influence passengers' travel patterns. All this would lead to better timetables and better inter-connections between lines, and, as a result, a better passenger experience.

Our research helped to identify a few challenges. First of all, the biggest challenge is related to *disclosing anonymity*. Public transportation historically is an anonymous way to travel. Disclosing information *en route*, passively or actively, encompasses *user adoption and behaviour change* meaning that passengers may need to give up some anonymity to provide data. We have identified three primary ways of addressing user position disclosure during the project:

- *Providing sufficient incentives* for the user to disclose their current position voluntarily;
- *Implementing a consent-based user strategy* entailed asking the user for location-sharing when an app is first launched; and
- *Scanning buses for devices using Bluetooth and Wi-Fi protocols*.

The second challenge is to establish a “*digital link*” between the rider's phone and the infrastructural component that the user is currently utilising (i.e., a vehicle). In public transport, we found the following potential ways of establishing such a digital link: (i) the user establishes such a link, or (ii) such a link is established without user involvement.

Finally, the question of *financial rationales* could be solved by increased user satisfaction. With the help of crowdsourcing *en route*, this can be reached by: (i) designing a more *demand-responsive public transport offering*, and (ii) enabling communication with third parties (other mobility providers, parents, to school bus passengers) and tailored traffic information.

## **6.2 Theoretical contribution: Conceptualisation of crowdsourcing *en route* in real-time**

To date, crowdsourcing *en route* in real-time is a novel phenomenon in academic research, and with this article, we seek to open a discussion in this area. As such, we both extend previous important work and provide a novel theorisation of crowdsourcing *en route* in real-time in public transportation.

We contribute by conceptualising crowdsourcing *en route* in real-time for public transport. For that purpose, we use a framework proposed by Modaresnezhad et al. (2020) and describe Who – Why – What – How activities. More specifically, the initiator of crowdsourcing *en route* in real-time is a PTA/PTO. We identify four groups that would benefit from crowdsourcing *en route* in real-time and specifically focus on values and benefits for a PTA/PTO and a public transport user. We highlight that crowdsourcing *en route* in real-time may only happen if both sides (PTOs/PTAs and public transport users) are motivated to do it and if both sides gain value from it.

In addition, our findings extend the taxonomy of crowdsourced activities applied for public transport companies (Nash, 2018) by identifying the additional activity of communication between a driver and a third party.

This paper is the first to start the discussion in crowdsourcing *en route* in real-time in public transport. Further research is needed to understand this phenomenon. One especially interesting avenue for such future research concerns investigating how to design services that simultaneously creates benefits for both PTAs/PTOs and public transport users. For instance, while pervasive travel patterns are the most beneficial crowdsourcing outcome for PTAs/PTOs, many of the services beneficial for public transport users do not necessarily produce such detailed data. We also argue that further research on public transport user incentives is necessary to address *user adoption* of crowdsourcing *en route* in real-time public transport.

### 6.3 Practical implications and limitations

Our research provides essential guidelines for practitioners. We offer a comprehensive overview and analysis of technical and business opportunities related to crowdsourcing *en route* in real-time public transport. We also discuss challenges associated with *crowdsourcing en route* in public transport and suggest possible ways to solve difficulties identified. In this paper, we argue that sufficient passenger adoption of crowdsourcing *en route* would entail adequate financial rationales.

The major limitation of this research is that it has involved just a limited number of PTOs/PTAs. Widening the scope of research could provide more reliable results. This could be the area for future research.

## 7 Acknowledgement

This work was supported by the Swedish Innovation Agency (Vinnova) through the DriveSweden project Welcome Aboard (Välkommen Ombord). We are very thankful to all interviewees who shared their experience and expectations on open accounts payable data.

## References

- Adler, N., M. Elmquist, and F. Norrgren. 2009. "The challenge of managing boundary-spanning research activities: Experiences from the Swedish context." *Research Policy*, 38, no. 7: 1136–1149. DOI: 10.1016/j.respol.2009.05.001.
- Agard, B., M. Trépanier, and C. Morency. 2006. "Mining public transport user behaviour from smart card data." *IFAC Proceedings Volumes* 39, no. 3: 399–404. DOI: 10.3182/20060517-3-FR-2903.00211.
- Argyris, C., R. Putnam, and D. Smith. 1985. *Action science*. 1st ed Edition. San Francisco, Jossey-Bass.
- Bagchi, M. and P. R. White. 2004. "What role for smart-card data from bus system?" *Municipal Engineer* 157, no. 1: 39–46.
- Bagchi, M. and P. R. White. 2005. "The potential of public transport smart card data." *Transport Policy* 12, no. 5: 464–474. DOI: 10.1016/j.tranpol.2005.06.008.
- Buechler, T., J. H. Sieg, R. M. Fuchslin, and R. Pfiefer. 2010. "Crowdsourcing, open innovation and collective intelligence in the scientific method: A research agenda and operational framework." In: *Proceedings of the Alife Twelfth Conference*, Odense, Denmark, August 19-23.
- Cairo, O., J. Sendra Salcedo, and J. O. Gutierrez-Garcia. 2015. "Crowdsourcing information for knowledge-based design of routes for unscheduled public transport trips." *Journal of Knowledge Management* 19, no. 3: 626–640. DOI: 10.1108/JKM-02-2015-0053.
- Ceder, A. 2007. *Public Transit Planning and Operation: Theory, Modelling and Practice*. 2<sup>nd</sup> edition. Oxford, UK, Elsevier.



- Chen, C., J. Ma, Y. Susilo, Y. Liu, and M. Wang. 2016. "The promises of big data and small data for travel behavior (aka human mobility) analysis." *Transportation Research Part C: Emerging Technologies* 68, (July): 285–299. DOI: 10.1016/j.trc.2016.04.005.
- Chu, K.K., R. Chapleau, and M. Trépanier. 2009. "Driver-Assisted Bus Interview (DABI): Passive transit travel survey using smart card automatic fare collection system and its applications." In: *88th Annual Meeting of the Transportation Research Board*, Washington.
- Çolak, S., L. P. Alexander, B. G. Alvim, S. R. Mehndiratta, and M. C. González. 2015. "Analysing cell phone location data for urban travel: current methods, limitations, and opportunities." *Transportation Research Record: Journal of the Transportation Research Board* 2526, no. 1: 126–135. <https://doi.org/10.3141/2526-14>.
- De Sabbata, S., S. Mizzaro, and L. Vassena. 2009. "Where do you roll today? Trajectory prediction by SpaceRank and physics models." In *Gartner G., Rehrl K. (eds) Location Based Services and TeleCartography II. Lecture Notes in Geoinformation and Cartography*. Springer, Berlin: 63–78. DOI: 10.1007/978-3-540-87393-8\_5.
- Elias, D., F. Nadler, J. Stehno, J. Krösche, and M. Lindorfer. 2016. "SOMOBIL – improving transport planning through mobile phone data analysis." *Transportation Research Procedia* 14: 4478–4485. DOI:10.1016/j.trpro.2016.05.370.
- Estellés-Arolas, E., and F. González-Ladrón-De-Guevara. 2012. "Towards an integrated crowdsourcing definition." *Journal of Information Science* 38, no. 2: 189–200. DOI: 10.1177/0165551512437638.
- Feng, T. and H. J. Timmermans. 2014. "Extracting activity-travel diaries from GPS data: towards integrated semi-automatic imputation." *Procedia Environmental Science* 22: 178–185.
- Furth, P.G. and B. McCollom. 1987. "Using conversion factors to lower transit data collection costs." *Transport Research Record*, 1144: 1-6. Retrieved 12 May 2021, from <http://onlinepubs.trb.org/Onlinepubs/trr/1987/1144/1144-001.pdf>.
- Ghezzi, A., D. Gabelloni, A. Martini, and A. Natalicchio. 2017. "Crowdsourcing: A Review and Suggestions for Future Research." *International Journal of Management Reviews* 20, no. 2: 343–363. DOI: 10.1111/ijmr.12135.
- Gustarini, M., J. Marchanoff, M. Fanourakis, C. Tsiourti, and K. Wac. 2014. "UnCrowdTPG: Assuring the experience of public transportation users." In: *Proceedings of the 2014 IEEE 10th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*, Larnaca, Cyprus, October 8-10: 1-7. DOI: 10.1109/WiMOB.2014.6962141.
- Haklay, M. and P. Weber. 2008. "OpenStreetMap: User-generated street maps." *IEEE Pervasive Computing* 7, no. 4: 12–18.
- Hanft, J., S. Iyer, B. Levine, and A. Reddy. 2016. "Transforming bus service planning using integrated electronic data sources at NYC transit." *Journal of Public Transportation* 19, no. 2: 89–108.
- Howe, J. 2006. "The rise of Crowdsourcing". *Wired Magazine*. Retrieved 24 December 2020, from <https://www.wired.com/2006/06/crowds/>.
- Jestico, B., T. Nelson, and M. Winters. 2016. "Mapping ridership using crowdsourced cycling data." *Journal of Transport Geography* 52: 90-97. DOI: 10.1016/j.jtrangeo.2016.03.006.
- Li, A., 2020. Google Maps using crowdsourced data to help you avoid crowds on public transit. <https://9to5google.com/2020/06/08/google-maps-transit-crowds/> [Accessed 23 October 2020]
- Kim, D., H.,Y. Kim, D. Estrin, and M. B. Srivastava. 2010. "Sensloc: sensing everyday places and paths using less energy." In: *Proceedings of the 8th ACM Conference on Embedded Networked Sensor Systems (ACM)*, November: 43–56. DOI: 10.1145/1869983.1869989.
- Marzano, G., J. Lizut, and L. O. Siguencia. 2019. "Crowdsourcing solutions for supporting urban mobility." *Procedia Computer Science* 149: 542–547. DOI: 10.1016/j.procs.2019.01.174.

- Mirri, S., C. Prandi, P. Salomoni, F. Callegati, and A. Campi. 2014. "On Combining Crowdsourcing, Sensing and Open Data for an Accessible Smart City." In *Proceedings of the 2014 Eighth International Conference on Next Generation Mobile Apps, Services and Technologies*, Oxford, UK: 294-299. DOI: 10.1109/NGMAST.2014.59.
- Misra, A., A. Gooze, K. Watkins, M. Asad, and C. A. Le Dantec. 2014. "Crowdsourcing and its application to transportation data collection and management." *Transportation Research Record: Journal of the Transportation Research Board* 2414, no.1: 1–8. doi: 10.3141/2414-01.
- Modaresnezhad, M., L. Iyer, P., Palvia, and V. Taras. 2020. "Information Technology (IT) enabled crowdsourcing: A conceptual framework." *Information Processing and Management* 57, no. 2: 102135.
- Nash, A., 2018. "Crowdsourcing techniques for transport planning and operations." In *Proceedings of 7th Transport research Arena (TRA)*, Vienna, Austria, April 16-19. doi: 10.5281/zenodo.1435672
- Palazzi, C. E., and A. Bujari. 2016. "Fostering accessible urban mobility through smart mobile applications." In *Proceedings of the 13th IEEE Annual Consumer Communications & Networking Conference (CCNC)*, Las Vegas, NV: 1141–1145. doi: 10.1109/CCNC.2016.7444950.
- Park, J. Y. and D. J. Kim. 2008. "The potential of using the smart card data to define the use of public transit in Seoul." *Transportation Research Record: Journal of the Transportation Research Board* 2063, no.1: 3–9. DOI:10.3141/2063-01.
- Pelletier, M. P., M. Trépanier, and C., Morency. 2011. "Smart cards data use in public transit: A literature review." *Transportation Research Part C* 19: 557–568.
- Pi, X., M. Egge, J. Whitmore, Z.Qian, and A. Silbermann. 2018. "Understanding transit system performance using AVL-APC data: An analytics platform with case studies for the Pittsburgh region." *Journal of Public Transportation* 21, no. 2: 19–40.
- Ritchie, J., J. Lewis, C. M. Nicholls, and R. Ormston (eds). 2013. *Qualitative Research Practice: A Guide for Social Science Students and Researchers*. Thousand Oaks: Sage.
- Shalaby, A. and A. Farhan. 2004. "Prediction model of bus arrival and departure times using AVL and APC data." *Journal of Public Transportation* 7, no. 1: 41–61.
- Shen, L., and P. R. Stopher. 2014. "Review of GPS travel survey and GPS data-processing methods." *Transport Reviews* 34, no. 3: 316–334. DOI: 10.1080/01441647.2014.903530.
- Shin, D. 2017. "Urban sensing by crowdsourcing: analysing urban trip behaviour in Zurich." *International Journal of Urban and Regional Research* 40, no. 5: 1044–1060.
- Sun, L., J. G. Jin, D. H. Lee, K. W. Axhausen, and A. Erath. 2014. "Demand-driven timetables design for metro services." *Transportation Research Part C: Emerging Technologies* 46: 284–299.
- Tettamanti, T., H. Demeter, and I. Varga. 2012. "Route choice estimation based on cellular signaling data. *Acta Polytech. Hungarica* 9: 207–220.
- Thomas, R., 2018. GDPR: how will the new data protection law affect the transport sector? Intelligent Transport. 25 May. Available at: <<https://www.intelligenttransport.com/transport-articles/68331/gdpr-new-data-protection-law-transport-sector/>> [Accessed 7 June 2021]
- Trépanier, M. and C. Morency, C., 2010. Assessing Transit Loyalty with Smart Card Data, In: *Proceedings of 12<sup>th</sup> World Conference on Transport Research (WCTR)*, Lisbon, Portugal, 11-15 July 2010.
- Trépanier, M., C. Morency, and B. Agard. 2009. Calculation of Transit Performance Measures Using Smartcard Data. *Journal of Public Transportation*, Vol. 12, No. 1, 2009, pp. 79-96.
- Utsunomiya, M., Attanucci, J., and Wilson, N., 2006. Potential uses of transit smart card, registration and transaction data to improve transit planning. *Transportation, Research Record: Journal of the Transportation Research Board*, No. 1971, 119–126.

- Van de Ven, A. 2007. *Engaged Scholarship: A Guide for Organizational and Social Research*. Oxford: Oxford University Press.
- Wang Z., S.Y. He, and Y. Leung. 2018. "Applying mobile phone data to travel behaviour research: A literature review." *Travel Behaviour and Society* 11 (April):141–155. DOI:10.1016/j.tbs.2017.02.005.
- Wolf, J., Guensler, R., Bachman, W., 2001. Elimination of the travel diary: experiment to derive trip purpose from global positioning system travel data. *Transp. Res. Rec.: J. Transp. Res. Board* 1768, 125–134.
- Zhao, J., Rahbee, A., Wilson, N.H. (2007). Estimating a rail passenger trip origin-destination matrix using automatic data collection systems. *Computer-Aided Civil and Infrastructure Engineering*, 22(5), 376–387. DOI: 10.1111/j.1467-8667.2007.00494.x
- Zhao, Y., and Zhu, Q. (2012). A conceptual model for participant's motivation in crowdsourcing contest. In: *Proceedings of the Eleventh Wuhan International Conference on E-Business*, pp. 429–439. <https://aisel.aisnet.org/whiceb2011/92>

## Appendix A. Sample interview protocol

### 1. About transit planning

- How is public transport transit optimisation and planning organised at your company?
- How do you collect needed information about travel patterns of passengers?
- Which are the data sources of information? (data from mobile tickets, travel cards, mobile operator data, etc.) And is it enough to get a full picture?
- Can you track full trip of a passenger from A to B and to C? and if **Yes**:
  - o How is this reflected in public transport route planning?
  - o How are privacy concerns dealt with?
  - o How has the general public received this?
- Do you think that use of such sources of information as crowdsourcing (real-time passenger comments, questions) in your own app, social and other media would be useful?
  - o And in which way could this be used?
  - o Does anyone use this information for public transport planning today?

### 2. Direct (economic/monetary) gains and benefits

Imagining the situation that data on travel patterns of passengers is available and based on it you have identified an opportunity for a new route/service.

- How do you make these estimations? Is there a specific method that you use?
- Which direct economic/monetary gains/benefits this would mean? (for example, number of new jobs, time saved per route, Return On Investment, etc.)
- How is this figure related to the proportion of travelers?
  - o For instance, if only 10 % of the movement were captured, would this still be of value for you?
  - o Are there any “threshold values”?
- Which costs would this imply? (for example, infrastructure development, etc.)

### 3. Indirect benefits

Imagining the situation that data on travel patterns of passengers is available and based on it you have identified an opportunity for a new route/service.

- Which indirect benefits would this mean? (for example, reduced demand for taxi, improved customer experience, car drivers switching to public transport, etc.)