Road traffic open data in Sweden: Availability and commercial exploitation

- A research study on the state of open transportation data in Sweden

Authors:
Sivan Shamoon
Johan Jönsson

Supervisor:
Jordi Olivella

Research Assignment,
Universitat Politècnica de Catalunya,
Barcelona, 2021

Content

1. Introduction

In this chapter, the background of the research study is presented, followed by the purpose and research questions. Finally, the delimitations of the project will be outlined.

1.1 Background

The term "open data" refers to publicly available data that has been released by governments or other organizations and is available for anybody to use for personal or business purposes (Smith, 2015). As a result, this free and open data may be utilized to, among other things, start new businesses or solve difficult strategic problems. National meteorological data, road maintenance, and safety are the most widely utilized examples of open data. However, there are some problems with open data that need to be taken into account in order for it to be used effectively to solve problems in society and for companies (Smith, 2015). First, data must be made easy to use and presented in such a way that different stakeholders can act on it. Second, the amount of open data is immense, however, there are still companies that tend to skip using it simply because they cannot use it appropriately and because open data has a lack of public awareness. The majority of public and private transportation companies today are in need to use data in order to improve several of their operations processes to stay competitive on the market (Smith, 2015). An example of a company that uses open data to improve its customer experience is the Swedish public transportation company Västtrafik. Their main goal is to provide safe, cost-efficient, and sustainable transportation solutions for society (Västtrafik, 2020). However, they aim to benefit the welfare of society by continuously sharing a large amount of their data with the public in order to make urban transportation more efficient. This is due to the fact that the company is a non-profit organization owned by the Swedish government. Furthermore, it is also stated that Västtrafik is dependent on utilizing and sharing data in an efficient way to create value for the consumers (Västtrafik, 2020).

There are a number of different types of data, one of which is open data that is publicly available for free (Open Data Handbook, n.d). By using open data in combination with proprietary data, transportation companies can enhance the efficiency of their operations processes, lower their costs, as well as provide customers with more reliable and sustainable transports (Freeman, 2020). As road transportation is performed on roads funded and maintained by the public sector, transportation companies can benefit from the transparent nature of government data that is typically available to the public (Soriano et.al, 2018). Examples of open data that can be used and that potentially can create value for the companies are weather, geographical, road works, and traffic data (OECD, n.d). Furthermore, the need for open data is crucial to increase the interplay between public and private companies and further make more data available for society. Concludingly, since data is becoming more valuable and important for the success of a company, it is essential for companies to gain insights from all types of data sources (Soriano et.al, 2018). Hence the importance for private transportation companies to understand how public companies today handle and share open data to create value for their stakeholders. In this way, they can take their business to the next level by really understanding their customer needs and making better forecasts using predictive analytics.

1.2 Purpose and Research Question

The purpose of the study is to investigate the state of open transport data in Sweden and to describe how different public and private companies benefit from it. The potential of the available open data will be analyzed by how it can be used independently, or in combination with closed data that companies gather through their own operations. As one crucial component to managing and planning transportation is to perform forecasts for how different activities within companies will be affected by current circumstances, the study will mainly analyze the potential that open data has for improving business processes through predictive analytics.

As data analysis often requires a vast amount of data in order to efficiently scope patterns and draw adequate conclusions from the analysis, organizations such as small companies without the required resources for large scale data collection, start-ups that lack a sufficient amount of historical proprietary data, or transportation companies that operate on a big variety of roads could suffer from the disadvantages they experience due to the limitations of available closed data. Hence, this study will serve to provide alternative solutions to this issue by proposing ways to predict deviations to their transportation operations with openly available data from public transportation agencies. Moreover, data generated by public companies, such as traffic and public transit data, can also enable new business opportunities such as app development, which will also be discussed in this paper in order to highlight the potential that open data has for driving innovation and enabling public and private organizations to thrive from each other. Ultimately, this paper will provide an analysis of the state of open transportation data usage in Sweden, both in regards to the improvement of operations in existing companies as well as for the possibilities of creating new business ideas.

Research questions:

- How can companies within the transportation sector use open data to improve their processes?
- What data-driven business opportunities can be identified through available open transport data?

1.3 Delimitations

This study is limited to only investigating the open data that is collected by public companies within the transportation industry and thus excludes the closed data that different companies possess. This is due to the difficulty of collecting closed data since this type of data is sensitive and can harm the competitiveness of the specific company. The study will focus on the possible improvements to transportation processes, and new transportation business ideas arising from open data in Sweden, as knowledge about the available open data sets and current business applications within the country is needed in order to provide a relevant discussion on the findings. Furthermore, the study will focus on Trafiklab and Trafikverket open data portals as platforms to get access to open data that transportation companies can use due to their huge offerings of transport-related data in Sweden.

2. Method

This chapter includes a description of how the study was conducted. In order to explore the possibilities for private companies to use open data, an extensive literature review was conducted. Furthermore, this helped to get familiarized with the subject of open data and understand how it is utilized today by public companies. While researching different methods for data analytics that are being used in transportation, it was found that predictive analytics was one of the most prominent methods as it can be used in numerous ways in order to improve predictions and planning within organizations. The use of predictive analytics in transportation includes predicting delays and traffic conditions which were found to be appropriate areas of analytics with regards to the types of open data that are commonly available. Hence, these will be the areas of transport analytics that will be focused on in this study.

In order to analyze the full potential of open transport data, both as a means of improving existing businesses as well as to allow for new business opportunities to originate, the methodology had to be considered accordingly. To scope out opportunities for improvement of business activities, research projects were reviewed where a number of types of open transport-related data were used to predict future outcomes of traffic conditions and events in public transportation that could have potential impacts on how daily activities within transportation organizations are performed. The projects were chosen based on the potential accessibility that the data used for the analysis has in Swedish open data sources, in order to make sure that corresponding solutions to the problems are feasible to perform in Sweden. Furthermore, in order to analyze the potential for new businesses to arise from available open data, several existing companies that have gained their success through the use of such data were studied to gain an insight into how value can be extracted from it. To analyze the accessibility of relevant open data in Sweden, Trafiklab, and Trafikverket, two open data sources for transportation-related data have been used. These were chosen in a screening method of the biggest open data sources that offer a large amount of data publicly in Sweden.

3. Theoretical framework

This chapter provides the theoretical background to the discussion of this study. Initially, a definition of data is presented in 3.1, followed by the presentation of open data in 3.2 and predictive analytics (PA) in 3.3. In 3.4, open data communities are presented.

3.1 Data

Datafication is the act of producing data, and it has become increasingly significant in recent years as a result of the digitalization movement, especially because the majority of new technologies demand data as input (Ylijoki & Porras, 2016). Data can be structured, semi-structured, or unstructured and comes from a variety of sources. Since it is frequently arranged in databases or spreadsheets, structured data is often the easiest to deal with (Taulli, 2019). However, according to Taulli (2019), the majority of the data is unstructured. Semi-structured data is a mix of structured and unstructured data and accounts for less than 10% of all data available.

Furthermore, the data quality is a crucial factor to consider. Several scholars contend that high-quality data is critical (Gudivada, Apon & Ding, 2017; Fu & Easton, 2017). The effectiveness of the data application can be harmed if the data is missing or of poor quality (Fu & Easton, 2017). Different dimensions can be used to measure data quality. Juddo, George, Duquenoy, and Windridge (2018), for example, recommend using accuracy, completeness, and consistency. The dimensions of data quality described by Patterson (2017) are illustrated in Figure 1.

Figure 1: Visualization of the six dimensions of data quality based on Patterson (2017).

According to Patterson (2017), data is *complete* when the desired amount of data and the actual stored data are equal. Patterson's requirements of completeness are expressed similarly by Ramsamy and Chowdbury (2020). When data appears only once in a database, it is said to be *unique* (Patterson, 2017). *Timeliness* is an indication of to which extent a set of data represents reality, and *validity* relates to how well data conforms to the format, type, and range specified (Patterson, 2017; Ramasamy & Chowdbury, 2020). As a result, the dimension is concerned with data in the acceptable format and follows the appropriate standards for use in the required context (Patterson, 2017). If the data accurately depicts the world, it is *accurate* (Patterson, 2017). Patterson (2017) pushes this point further by adding that data is accurate when it is similar to what it represents. Data should not differ from other databases' equivalent representations of itself to be regarded as *consistent* (Patterson 2017; Ramasamy & Chowdbury, 2020).

3.2 Open data

Open data is a concept that can be generally described as data that is publicly available and free to use. The Open Definition (n.d) also proposes an additional set of requirements for data to be considered as open which include that the data must be released under an open license, which is required to allow anyone to retrieve, use, modify and share it. The data must also be structured in a machine-readable format in order to allow for efficient usage. The data must additionally be allowed to be used for any purpose and in any field.

As stated by Manyika et. al., (2013, p.4) to clarify the relatedness and differences between open data and big data, “Big data refers to data sets that are voluminous, diverse and timely. Open data is often big data, but “small” data sets can also be open”. Furthermore, open data can be made available from both governmental and private organizations and institutions, although publicly collected data is more commonly available including weather and GPS data. Open data is generally available in the data portals through APIs, which according to IBM (2020) can be defined as a tool that allows an interface and a service to communicate with each other. The way an API works is that a request is made from an application to a web server, after which the API makes a call to the webserver, receives a response from the server, and finally transfers data to the application where the request was initially made.

As stated by Open Data Handbook (n.d), one crucial aspect to having a clearly stated definition of open data is to allow for interoperability, i.e., letting organizations and systems inter-operate. Interoperability is imperative for the ability to combine multiple open data sets in order to build more complex systems, maximizing the potential values that the data is ought to bring in a wider variety of products and services. Manyika et al. (2013, p.7) further highlight the benefits of combining multiple data sets in the form of combining open data with proprietary data, i.e. data that derives from within an organization, to enhance the potential values that data analytics can provide to organizations. An example is given by Deloitte (2012, p.13) for how combining data sets can create business value is for transportation companies to combine proprietary data with open public transport data in order to better tailor efficient route planning within their organization.

According to The World Bank (n.d), sharing open data can provide societal benefits including increased governmental transparency and reducing corruption due to increased insight into governmental activities. It also gives rise to initiatives for improving public services by letting citizens gain access to the data required. As governmental data is the most prominent form of open data, many of the highlighted potentials for open data are in regards to public sector activities. However, a project by the Open Data Institute has underlined the potential benefits of sharing data in the private sector as well. One of these benefits is the rising possibilities of open innovation, which is the concept of boosting innovation by allowing collaboration between different organizations (D'Addario, 2021). Another benefit of data sharing among private companies is the aspect of benchmarking and insights, which helps companies evaluate their performance by being able to compare themselves with other companies operating in the same industry (Maddison & D'addario, 2020). Results of a case study conducted within the project showed that the company HiLo Maritime Risk Management was able to significantly reduce accidents and incidents occurring in the shipping industry by collecting data from various companies within the industry and analyzing it through predictive analytics. By analyzing data from several companies as opposed to the individual companies analyzing their own, more information and insights could be gathered from it, providing both economical, social, and environmental benefits for the companies involved. (Maddison & D'Addario, 2020).

However, the use of open data has a few challenges and drawbacks which need to be addressed by the companies. According to (Charalabidis & Zuiderwijk, 2012) there are three main challenges that open data faces. First, even if there is a large amount of open data available, it does not guarantee it is accessible. Governments and other public companies tend to lose consistency as they improve their way of sharing data. Since data changes continuously, the users must adapt to the fast-changing environment. As a result, users of the data will have a tougher time finding the information and files they want, which is a time-consuming and non-efficient way of working (Charalabidis & Zuiderwijk, 2012). Second, the fact that data is public does not imply that it is 'ready for use'. There may be differences in format and other compatibility concerns. When you compare open data sets across time, you typically find that they contain a varied collection of fields (D'Addario, 2021). This makes it difficult for users to calculate growth over more than three years in a critical. In order to examine trends and make use of the open data, users would require consistent data over the previous five years (D'Addario, 2021). Third, the fact that companies always strive to survive and be the best in their industry makes it difficult to share data publicly with their competitors (Charalabidis & Zuiderwijk, 2012). This can harm their business and eventually lead to impairing their competitiveness in the market.

3.3 Predictive analytics

According to Runkler (2020, p.2), data analytics can be defined as *"the application of computer systems to the analysis of large data sets for the support of decisions"*. One method of data analytics that is being frequently used in transportation is predictive analysis (PA) (Van Bodegom, 2021). Predictive analytics is a type of data analytics that seeks to generate future predictions, including opportunities and hazards, by studying current and historical data (Balali, Nouri, Nasiri & Zhao, 2020). The act of identifying trends or patterns in acquired data can be automated using algorithms, allowing businesses to answer the question "what will happen?" (McCarthy, McCarthy, Ceccucci & Halawi, 2019). PA can lead to a number of benefits, including making it easier to spot new opportunities and providing more accurate market sales projections (Attaran & Attaran, 2019). Several writers propose that PA should be used in conjunction with machine learning algorithms to evaluate larger data sets more quickly, in order to improve its effectiveness (Balali et al., 2020; Shi-Nash & Hardoon, 2016).

Although PA can help businesses develop in a variety of ways, Attaran and Attaran (2019) claim that it also comes with a number of drawbacks, including integration issues, data access issues, and difficulties using the generated results. These issues arise frequently as a result of a company's inadequate expertise and funding for PA adoption.

3.4 Open data sources

Open data can be offered in many different types of sources, however, the most common are open data communities and open data portals. Open data communities are platforms that offer a broad range of open data sets from several open data portals. The biggest benefit of using open data communities is that they gather all open data sets in one platform which are routinely updated to newly released governmental data. Their objectives are to offer fully open, accessible, and re-usable forms of data which can create value for the companies and their stakeholders. The community provides in-depth knowledge, specialized data from different sectors, applications, and real-time datasets. Trafiklab is an example of an open data community that will provide the basis for analyzing the possibilities of open data in this paper together with Trafikverket open data portal. Trafikverket is a nationwide authority and traffic data from all over Sweden is available in their data portal. Trafiklab however is a community that offers data from multiple transportation agencies that operate in several different regions in the country. These open data sources were mainly chosen due to their extensive offering of open data that is related to traffic and transportation.

3.4.1 Trafiklab

Trafiklab is a Swedish service that aims to make people choose to travel in a more sustainable way using public transport over car or plane (Trafiklab, n.d.a). The service is developed by Samtrafiken which is an organization that serves to improve the overall intelligence of the Swedish public transport system by collecting and coordinating public transit data within the country (Samtrafiken, n.d). With the vision of boosting innovation and collaboration between public transit authorities and other actors both within and outside of this sector, Trafiklab (n.d.a) has made available open data and APIs for developers and other interested parties to use. By directions of the European Union regarding Intelligent Transport Systems (ITS), all public transit data is required to be released publicly in the correct formats. Samtrafiken has set out to help transportation companies reach these requirements and to make the data available on the Trafiklab service (Trafiklab, n.d.a).

In Trafiklab's open data portal, data from all public transportation agencies in Sweden can be found, as well as for some private ones. There is both static and real-time data available, with static data including points such as scheduled departure and arrival times, and real-time data including points such as roadworks and estimated delays of public transport vehicles. Data sets are regularly updated with some real-time data sets receiving updates as much as every second (Trafiklab, n.d.b). Furthermore, Trafiklab also provides its users with advice and guidelines on how to efficiently use their available data for the user's specific purpose, e.g. to develop route planners that use data sets from the Trafiklab portal (Trafiklab, n.d.c).

3.4.2 Trafikverket Open data portal

Trafikverket is the Swedish transport administration authority and has responsibility for the long-term infrastructure planning within the country, including roads, railroads, sea, and aviation (Trafikverket, n.d.a). Trafikverket has launched an open data portal on their website where numerous kinds of traffic data are available including road data, railway data, traffic information data, and surrounding geodata. The data is available through three different applications in the portal where both companies and interested people can retrieve desired data, with the first being an API that offers real-time information on the traffic situation on roads and railroads. Additionally, a platform for collecting traffic information through the EU standardized Datex II format is available, as well as the Lastkajen system where data on the national roads and railroads can be retrieved (Trafikverket, n.d.b; Trafikverket, n.d.c). As a means of helping users efficiently retrieve and use data from the portal, Trafikverket has, in a similar manner as Trafiklab, included descriptions of the available data and extensive guidelines on how to use it into the portal. Additionally, a console is integrated into the website where the user can test the API by proposing questions to it through pre-designed templates that represent how the use of the API works in practice (Trafikverket, n.d.d).

Different methods can be used to acquire real-time traffic data, however, the most common and reliable method is by using governmental collected traffic data. Travel times and delays in real-time traffic data are collected by using e.g., number plate scanning cameras or road sensors that measure the number of vehicles traveling in a certain period of time (Jansen-Young, (2021). As mentioned earlier, Trafikverket offers several APIs that can be used by private individuals or companies to better understand the current traffic situation (Trafikverket, n.d.d). In figure 2 an illustration of real-time traffic data on Swedish roads is presented. This creates the possibility for the user to choose between two different categories depending on the purpose of the findings: Normal road traffic and commercial road traffic. Furthermore, within those categories, there are different functions and sub-functions which the users can choose based on what type of data they wish to visualize (Trafikverket, n.d.d). For example, the user could choose to get deeper insights into how the traffic situation looks by looking into the planned road maintenance and if there has been a road accident on a specific road in order to minimize the total transportation time. Trafikverket states that the main goal of providing this type of data is to increase traffic awareness among drivers, but also to provide companies and private individuals to plan their routes more efficiently and consequently have less impact on the environment (Trafikverket, n.d.d). Ultimately, the traffic data may be used to detect congestion and increase road safety throughout the road network by comparing transportation times and accident occurrence during free-flow traffic with those during peak hours. It is also feasible to compare bottlenecks with one another and develop rules and activities in the planning process to eliminate or minimize delays.

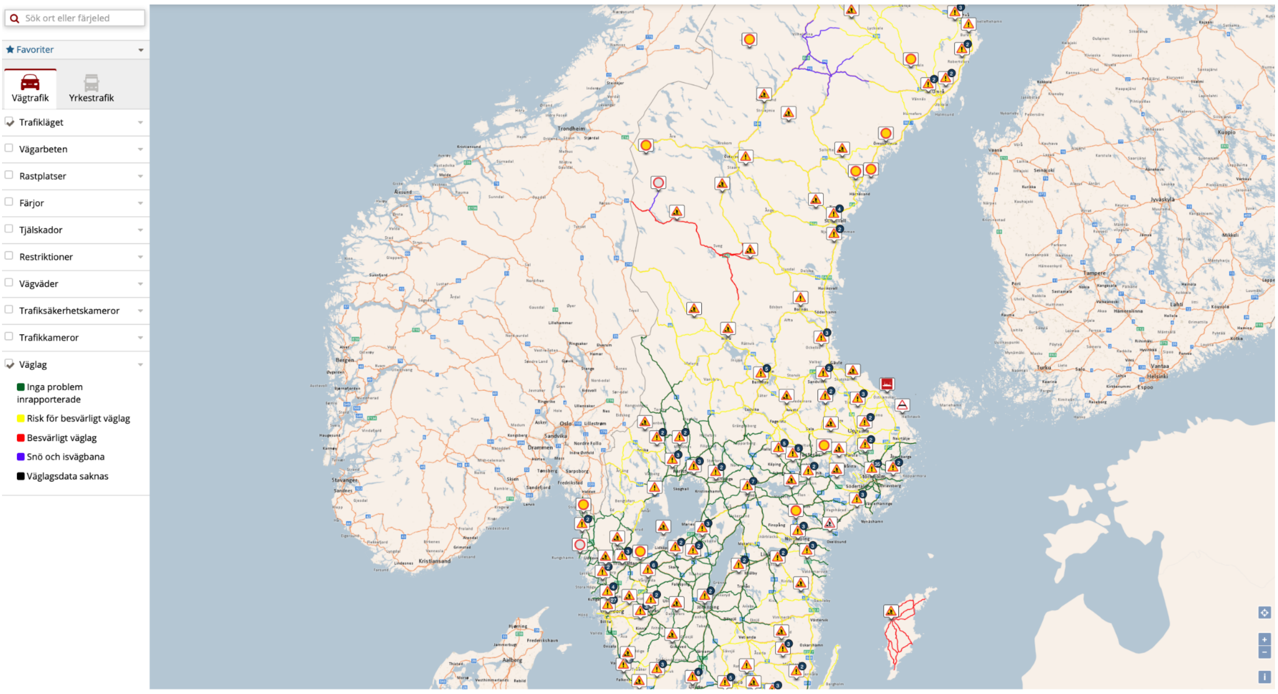

Figure 2: Real-time traffic data visualization in Sweden from Trafikverket. (Trafikverket, 6 December 2021)

4. Current applications

This chapter provides some current applications of open data in the transportation industry. Initially, several studies are presented regarding how open data can be used in combination with data analytics to improve the utilization of public transportation and traffic data. This is then followed by a presentation of several open-data driven businesses to gain insights into how different businesses benefit from open data to create new business opportunities within the private sector.

4.1 Utilization of open data in transportation

To showcase how open data can be used to improve business operations within transportation, previous studies that have implemented predictive analytics in combination with open data are presented here. To underline the significance and potential usage of the open data available at Trafiklab and Trafikverket, the reviewed studies are focused on public transportation and traffic data.

4.1.1 Public transportation data analytics

Several studies have proved the efficiencies of using open public transportation data in combination with predictive analytics, e.g. to predict estimated times of arrival and delays in transportation. In a study by Sun et al. (2016) on bus delay prediction for a midsize American city bus system, it was found that through the use of predictive analytics of real-time bus position data and historical arrival and departure data, errors in arrival prediction could be decreased with 25% within a one hour window and with 47% within a 15-minute window compared to the methods that were currently being used by the transit company. In a similar study conducted by Balbin et al. (2020), additional emphasis was put on the occurrences of buses leaving early and on time and the authors were able to map out frequent patterns in bus timeliness on different routes in the Winnipeg bus transit system by studying open data on deviations from bus departure times. The research showed that deviations in departure times varied according to the day and specific route and conclusions were drawn that e.g. weather could be a contributing factor to said deviations.

In another study by Leung et al. (2020), a machine learning algorithm for predictive analytics has been used to analyze the occurrence of delays on streetcars in Canadian city Toronto using open delay data in combination with weather data. In this case, only historical data was used, highlighting the adequateness of PA even in situations where real-time data is not attainable. By categorizing and grading different types of weather phenomena, conclusions on how these affect transportation delays could be drawn. Results showed that some weather conditions, such as snowfall, were significant sources for increased transport delays. Similarly, Alam et al. (2021) constructed a model for predicting irregularities in bus arrival time from public bus transit data and weather data. It was discovered that the inclusion of weather data in the model significantly improved the predicted arrival times in relation to the actual arrival times. These results show that multiple open data types in combination can further elevate the significance of using open data in transportation modeling.

4.1.2 Traffic data analytics

The increasing amount of vehicles appearing on roads due to factors such as population growth and the growth in demand for road transport-based services is causing a significant increase in traffic congestion. Congestion has numerous social, economic, and environmental impacts on society. It increases travel time and causes delays for both people and companies operating on the roads, leading to a waste of time and resources. It also causes road accidents and environmental pollution due to increased fuel consumption (Samal et al., 2020).

Predictive analytics has been proven to be a viable tool for analyzing congestion and allowing better decision-making based on forecasts concerning traffic conditions. Zang, Qu and Fang (2020) developed a deep learning model for traffic congestion prediction on Chinese highways using historical traffic speed data from the same weekday three weeks prior to the forecasted day. The authors found the predicted speed to be virtually the same as the actual speed on the forecasted day and that the congestion state on the roads could be predicted with 93% accuracy. Another way of predicting congestion is by analyzing traffic flow i.e., the number of vehicles passing a road measuring point in a given time period. Li et al. (2021) used recurrent neural network-based models to predict future traffic flow using historical data. Results showed that the best out of the developed models was able to achieve a highly accurate prediction compared to the actual measured traffic flow which can significantly aid congestion management. In a study by Li et al. (2020), traffic congestion prediction was performed using the three parameters average speed, road occupancy and traffic flow density. Here, road occupancy is defined as “the ratio of the actual traffic flow to the maximum capacity in a specific section”, and traffic flow density is defined as “the number of vehicles in particular length of road within a unit time". The three parameters were used in combination to find the congestion level of the studied road segments at different dates and times. Several machine learning algorithms were used for the predictions and the best performing one was found to have a prediction accuracy of 92%.

The evaluation of historical- and real-time traffic data is essential for the operations of a company. In addition to congestion, (Chowdhury et.al., 2017) points out that traffic data is important to consider when measuring the collision risk of the drivers. Intuitively, this means that roads with a considerable amount of traffic are more likely to encounter collisions than roads with little traffic. As a result, crash count modeling frequently requires huge amounts of traffic volume data. This type of traffic data can be collected from several open data sources owned by the government, in combination with the closed data that is collected from their own transportation system to predict and plan the most efficient routes. Hence, the importance of collecting large data sets on a regular basis from numerous open data sources increases the possibility and ability to enable a wide range of novel safety research and applications. However, the process of collecting open data and making it useful is more complicated in practice since most of the data is unstructured (Shi & Abdel-Aty, 2015). In their article *Data applications in real-time traffic operation and safety monitoring* (Shi & Abdel-Aty, 2015), the authors study the possibility of using real-time traffic flow data to calculate the risk of collision in great detail. This risk could then be used by companies to evaluate the safety risk for drivers, but also the risk of collision on a roadway which in turn can lead to logistical delays. The authors also highlight the importance to consider other factors such as the psychological- and physiological presence of the drivers to reassure the safety of the drivers. Parallelly, (Sall et.al, 2016) states that open data has a significant role in creating new business opportunities for companies. The authors urge many states and towns to work towards making traffic data more widely available to encourage third-party developers to create new applications. For example, Trafikverket (2021) has made real-time traffic times, scheduled road events, traffic camera data, and road maintenance amongst others available on their website to make road transportation safer, more efficient, and sustainable. According to (Shi & Abdel-Aty, 2015), this type of data can be utilized to obtain comprehensive traffic flow information which can be applied in data analytics when planning routes and forecasting delivery times, complementing the closed data collected from a company's vehicles.

4.2 Open data-driven businesses

To show how open data can be used to identify and create new businesses within the transportation industry, the companies Citymapper, Moovit, and Waze are presented here. The three companies have created applications that would likely not have been possible without the availability of open data and show that while open data is available to anyone, using it in an innovative way can generate profitable opportunities in the private sector.

4.2.1 Citymapper

One business idea within the private sector that arose from the use of open public transport data is the London-based mobile application Citymapper. The application serves to integrate all different modes of transport for the user to spot the best possible routes for their journey through cities all over the world, focusing mainly on Europe and the United States. The app does not however depend solely on public transport but has additionally added private transportation modes into the service, including taxis and rental services for bicycles and electric scooters (Citymapper, n.d.a).

Citymapper is using data available through Trafiklab in order to integrate the public transportation services available in Stockholm into the application (Trafiklab, 2016). According to Trafiklab (2016), the open data sets that have been used in the application are GTFS Sverige, which contains information about all public transport in the entirety of Sweden, SL Realtidsinformation, which contains real-time information on the public transport in Stockholm, and SL Störningsinformation, which contains information about disturbances in the Stockholm public transportation system.

Furthermore, in order to additionally simplify the use of public transport within cities, Citymapper has launched a service within the application that lets the user sign up for a weekly subscription that gives access to unlimited use of public transport in combination with limited use of other transportation services such as bikes, scooters, and taxis (Citymapper, 2019). The service is currently implemented in London and is meant to include more modes of transport as the development progresses (Citymapper, 2019).

As a way of creating business opportunities from the platform that Citymapper has managed to build to a large degree thanks to the availability of open data, the company also offers software development kits (SDK) and API's for implementation into applications created by external actors. This gives the possibilities for developers to use Citymapper features such as navigation and routing into their own applications through integrating their SDKs, or to build their own user interfaces using the Citymapper API (Citymapper, n.d.b).

Although Citymapper first started out as a timetable application for public transport, it has since become an important player in the development of smart cities and data-driven urban infrastructure. From the open data the application was originally built upon, new insights have been gathered from the collection of user data through the application which tracks the behavior of people, their movement through cities, and the choices they make regarding available forms of transportation. This data is then fed back into the transport systems themselves in order to improve them and make them more efficient through the insights gained from the data (Tavmen, 2020). Apart from the open transportation data that the company retrieves from open data portals such as Transport for London or, in the Swedish case, Trafiklab, the application collects behavioral data and other relevant data from its users in a number of different ways. Firstly, the users can improve the already available data on route information by reporting issues they experience during their transportation on specific routes and additionally users are able to add new data about different transportation modes themselves in order to improve the application and hence giving Citymapper a free source of newly created data. Moreover, the user location data is collected in different ways depending on how users choose to share their data with the company. For users with continuous data sharing, the location data will be gathered in real-time through the entirety of their journeys and thereby the smartphones of active users will provide a network of GPS sensors that track the location and behavior of the users in real-time. For users that do not share their location data, the data points are collected from how they navigate within the application such as from what starting point and end point they use the route planning function (Tavmen, 2020).

As opposed to other data-driven companies, Citymapper is not using their collected data to target advertising to users or selling it to third-party companies. Instead, the company claims to use the data collected as a means of improving the development of smart cities instead (Tavmen, 2020). This, however, has not yet made the company profitable as it experienced losses of 31 million GBP in 2019. As for now, the revenue streams in the company come from paid user subscriptions, collaborations with other companies through the use of Citymapper SDKs and APIs, as well as fees collected by Citymapper when users buy services such as taxi rides from private transportation companies through the application (Taylor, 2021).

4.2.2 Moovit

Moovit is a MaaS (Mobility as a Service) provider and developer of the urban transportation application under the same name. The application is targeted at people who use public transportation as a means of commuting and, much like Citymapper, is focused on multimodal transportation including services such as public bus and metro transit, taxis, bikes, and scooters. The source of data used in the application is a combination of data collected from public transit companies and other authorities, and proprietary data collected from the users of the application (Moovit, n.d.a). Moovit is available in the majority of major cities around the world and as a means to allow implementation in additional cities where the application is yet to function, the company has incorporated a crowdsourcing solution to make the expansion more efficient. The way crowdsourcing works is that users can sign up as editors and contribute to the data collection in the application by registering and maintaining transit data such as public transportation stops and timetables (Moovit, n.d.a; Moovit, n.d.b; Moovit, 2020).

In addition to developing the mobile application, the company has also been involved in a number of collaborations with external parties to aid them in the development of their services. One example of collaboration on a national level is with the urban transportation provider Arriva and the development of their mobile application Glimble, which is an application with similar features as Moovit, combining multiple modes of urban transportation within the Netherlands into one service. In the project, Moovit helped Arriva to incorporate solutions for payment of mobility services into the application, as opposed to the physical payment card that was needed to pay for fares formerly. Additional features were also incorporated, including optimal route calculation and sorting possible modes of transport by the amount of CO2 emitted (Moovit, 2021.a).

Another collaborative project is one between Moovit and the Aberdeenshire Council with the launch of their Ready2Go application. The application is directed to an on-demand bus service that operates in the Scottish town Inverurie and offers the ability to book bus fares through the application whereas this had formerly been done by phone. In addition to the ability to perform bookings, the application users are also able to follow direct tracking of the buses and to view timetables of other scheduled bus services operating in the area. Furthermore, the user is also given suggestions on how to efficiently reach the pickup points for the bus from the position where they are currently situated (Moovit, 2021.b).

However, looking from another perspective, the pandemic has shown how vulnerable our society and the companies are to bigger changes (Bus-news, 2021.b). Many companies have struggled to get their operation working and are most certainly still finding new ways to fight the pandemic. However, the urge to survive creates new business ideas for the companies and at the same time stay innovative. For example, social distancing is a recommendation that everyone should follow. However, our society was not built on that. According to Santos and Nikolaev (2021), this is a problem that Moovit really quickly addressed and started to look into how they can benefit from the open data that is valuable and accessible both from the pandemic but also the governmental open data such as real-time traffic information in Trafiklab. The company ended up developing real-time crowding information as an API integrated in their platform that will help people to provide the relevant information to keep their social distance and hopefully feeling more safe during their journey (Bus-news, 2021.b). The solution is based on real-time information on how empty or full a transportation vehicle is, which allows the traveler to make more educated and conscious decisions. Furthermore, Santos and Nikolaev highlight that this opportunity is fundamental for elder people who are in the risk zone but still want to use public transport (Santos & Nikolaev, 2021). Thanks to this, the app users have not fluctuated as previously expected and thus COVID-19 had a minor effect on the company's performance (Bus-news, 2021.b). The company also highlights the importance of using real-time open data since it helps the user to get a clearer picture of the transportation, which also makes the planning a lot easier (Bus-news, 2021.b). Arabghalizi and Labrinidis (2019) emphasizes the idea of making public transportation smarter and more predictable which will make it more reliable, safer and more effective. To this solution, Moovit also added an API that shows if there is place free or if the transportation vehicle is adapted for wheelchairs i.e. wheelchair accessible buses (Bus-news, 2021.b). Furthermore, this simplifies the journey planning for disabled people and empower them to use public transportation (Bus-news, 2021.b). In simple words, the solution would help to plan a complete journey for disabled persons with wheelchairs which includes vehicle options, stations, routes and appropriate lines. In addition to this, Darsena et.al (2021) presents the benefits and business applications of having real-time crowding data. The authors explain that operators and developers can use this type of data to adapt some service features such as: increasing service frequency, reallocating vehicles to different lines to avoid traffic jams, planning alternative routes etc.). The main problem these features solves are dealing with spatially and/or temporally localized crowding situations that are not addressed by traditional statistical tools used today in the transportation system design (Darsena et.al, 2021).

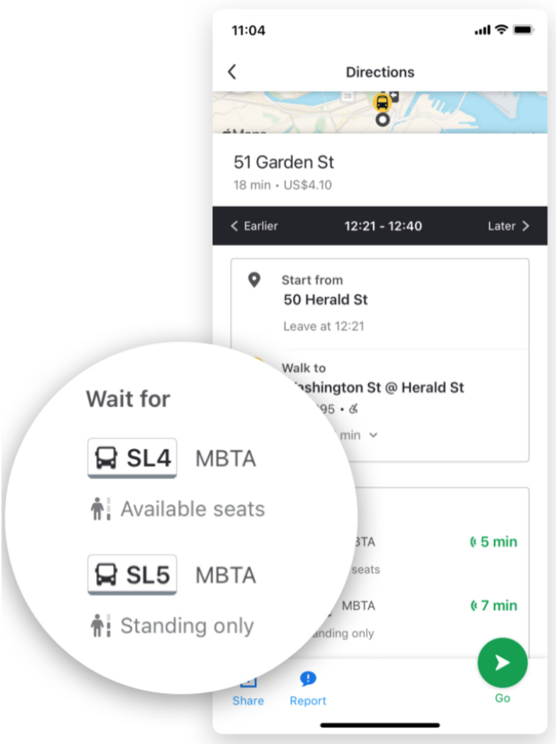

Figure 3: Real-time crowding information using Moovit App - showing available seats and standing area available (Bus-news, 2021.b)

4.2.3 Waze

Waze is a mobile navigation application founded in 2009 in Israel and is focused on improving mobility by letting its users stay alerted on updates regarding ongoing traffic conditions. By providing real-time traffic information, users can plan their trips accordingly in order to find the best possible routes based on how the traffic situation is at the specific moment (Waze, n.d). According to data.europa.eu (2019) the application uses open data from municipalities and governments on a local and national level, but what makes it different from other navigation applications is that it is to a large degree driven by crowdsourced data from its user community. While in use, the application collects data about how users are moving which can be analyzed to get insight into how the current traffic conditions are affecting mobility in certain areas. Aside from that, the users are also able to report traffic conditions such as road works, accidents, police controls, and current speeds, which can help other drivers find the best routes for their journeys. Additionally, users can also report current prices at local gas stations which can further affect how drivers choose the preferred route to their destination (Appel, 2013).

To further elevate the application, Waze has been collaborating with event organizers in order to gather data on how different events such as marathons are affecting driving conditions due to e.g. road closures adjacent to the occurring events (Sawers, 2016). According to Sawers (2016), Waze is offering the interested parties "custom access to its traffic management and communications channels" in exchange for receiving information from the event organizers on how said events are contributing to traffic conditions as well as real-time traffic information in proximity to the event locations. Hence, the collaborations will benefit both of the involved parties. The resulting changes in traffic conditions that the events give rise to are then integrated into the Waze application to let users avoid areas that may cause interruptions in their driving.

In order to let more people apart from the users of the Waze application benefit from the large amounts of data that are gathered through the platform, the company has partnered up with cities and public agencies by initiating a service called Waze for Cities. Partners can access data traffic data collected from Waze for free through Google Cloud and are also provided with a set of tools for data analytics to help gather useful information from the available data, although some more advanced analytics tools will come at a cost (Edelstein, 2019). The aim of the service is to let partners gain value from the data provided by Waze in order to help improve the infrastructure planning and other projects leading to decreased traffic impact in cities. In exchange for providing its partners with data, Waze receives data from the partners regarding traffic conditions in the cities, thus leading to a collaboration where both citizens and businesses benefit from each other (Sawers, 2019).

While Waze themselves gain revenues from selling ads to companies that are made visible through the application, the company evidently creates additional societal value from their crowdsourced data which is used in traffic planning within the cities the application operates. While sensors and other forms of installed equipment provided by national traffic agencies already collect information about accidents and incidents occurring in city traffic, the data gathered by Waze may bring additional information about occurring events to complement the governmentally collected data. In order to analyze the relevance of the crowdsourced data, studies have been made addressing the importance and reliability of it with regards to traffic management. According to Lenkei (2018) the traffic sensors in cities have limited coverage which proposes that additional data from other sources such as smartphone users could be a viable alternative. Lenkei (2018) highlights possible benefits of the use of crowdsourced Waze data including more detailed information than what can be obtained from automated sensors, reporting of accidents that cannot be obtained through sensors such as missing road signs and potholes, and a faster reporting of accidents. Results of the study show that both the relevance and reliability of the crowdsourced accident reports from Waze are highly dependent on the number of users in the affected area. The amount of Waze reports in areas with a high user rate is significantly higher than the number of reports from Trafikverket. Hence, the value that crowdsourced data can provide is dependent on the amount of active users in the location of accidents and is thus more likely to be of high relevance on large roads near major cities than on smaller roads. It is moreover difficult to determine in some cases the relevance of the accidents reported through Waze as more accidents with less significance are reported through the application than from Trafikverket (Lenkei, 2018). Concerning the reliability of Waze reports, the application evaluates it from the rating of the reporter and feedback from other users. This results in that the reliability of a reported accident cannot be determined instantly but rather has to wait until feedback on the report has been given. From this fact, it seems important to come up with a solution for accurately and quickly verifying reports in order to take the appropriate traffic management actions in the case of accidents (Eriksson, 2019).

4.2.4 Sygic

Sygic is a company and application developer mainly focused on providing navigation services for commercial use. The company offers services tailored towards a number of different branches of the transportation sector such as long haul transportation, last-mile delivery, taxi, and smart city mobility. Included features are, among others, real-time navigation, route planning, and driver behavior evaluation, and the majority of features are available both online and offline hence working regardless of current cellular reception status. In order to simplify route planning for trucking fleets, attributes such as vehicle dimensions, weight, fuel type and cargo (including hazardous materials) can be specified to ensure that the calculated routes are suitable with regards to vehicle characteristics as well as goods carried (Sygic, n.d.a). Additionally, to ensure driver safety, the application will calculate suitable rest stops during longer transportations which will be included in the arrival-time estimation. In the case of last-mile delivery, the application has the possibility of displaying exact house numbers in order to aid door-to-door delivery and upcoming delivery stops will automatically be updated according to the planned delivery route once a delivery has been made (Sygic, n.d.b). Furthermore, to avoid additional expenses such as fees or fines, the application will help drivers avoid low emission zones where the vehicle in use may not be allowed to enter. In that case, the driver will be routed around the affected areas or be alerted if entering (Prekopová, 2019).

Sygic is using multiple kinds of data in their applications, some of which are available in the form of open data such as GPS and traffic data. According to data.europa.eu (n.d.a), the company is using transport data to create visualizations for users through 3D and regular maps. Furthermore, traffic data is used for the ETA (estimated time of arrival) function where historical traffic intensity data for different times in combination with real-time traffic data is used to make future predictions for how travel times will be affected by traffic conditions. Additionally, in cities that provide data collected from e.g. road sensors, cameras, and public transport, the Sygic applications are able to integrate these data in order to make traffic predictions even more accurate (Sygic, n.d.a; Sygic, n.d.c). Aside from externally collected data, Sygic uses proprietary data collected through user devices to analyze and improve driving behavior. The users are given in-app feedback on their driving through analysis of different parameters such as speed, acceleration, and breaking. This data can be used to give drivers a score based on their driving performances and can serve as a tool for companies to ensure that transports are performed in a safe and reliable manner based on company directives (Sygic, n.d.d). The data is saved and can be further analyzed to spot patterns and gain knowledge of how driving behavior affects different outcomes in order to improve decision-making. Additionally, as in the case of Citymapper, Sygic offers SDKs to be used in third-party applications where desired functions developed by Sygic can be integrated according to the needs of the user (Sygic, n.d.a).

To showcase how data generated by Sygic can be used in other areas other than the application itself, a study by Scherrer et al. (2018) was conducted where Sygic data was analyzed to map out the behavior of how people are moving around in countries based on if they belong one of the two subclasses travelers or locals. The study was made in Australia and showed that insights on human mobility could be drawn from analyzing such data to be used for planning and tracing purposes. From the user data, travelers were determined as the class of people that moved around in large areas and stayed in the same location for limited time periods, whereas locals were determined as those that move around mostly in the same areas and traveling shorter distances. Scherrer et al. (2018) highlighted the value of the insights gathered from the analysis in for example tracing the movement of people and the interaction with tourists and locals during the ongoing covid-19 pandemic. This way useful information can be withdrawn from the virus spread based on the way people are moving around within and between countries. Additional insights could aid the planning of tourism related issues which can have substantial impacts on the economy, as well as insights on how urban land within countries is utilized. Moreover, such data could also be used in predicting the demand of carpooling services in and between specific geographical areas. The challenges with analyses like the one in this study proposed by the authors are that the reliability of the gathered insights is difficult to determine. Hence, in order to use this kind of data analysis for governmental decision making, a suitable approach for validating the insights would be necessary (Scherrer et al., 2018).

4.2.5 Vélib' Métropole

One trend within smart city mobility that has been present for a number of years is bicycle-sharing services. These are available in the majority of major cities around the world from a variety of different companies. One such company is Vélib' Métropole which provides bikes for use in and around Paris with docking stations placed out every 300 meters around the area. Users can choose between a few different subscription types depending on the amount of biking time desired every month, as well as the type of bicycle since the company provides both electric and normal bikes for use. In addition to the natural environmentally beneficial aspects of biking as opposed to e.g. driving, the company has other policies including fuel types for company vehicles and the recycling of unusable bike parts (Vélib' Métropole, n.d.a).

According to data.Europa.EU, the company uses open data in the form of geospatial and transportation data. Geospatial data is used to provide users with maps through the application in order to find available bikes as well as routes within the city area. Furthermore, the transportation data is used to identify busy roads within the operating area and by doing so, the company can make better decisions on where to place docking points based on the demand in the specific area (data.Europa.EU, n.d.b). Aside from the open data used by Vélib', the company also publishes open data themselves for other interested parties to use e.g. for analysis. Vélib has APIs available that include dynamic data that allows the user to in real-time follow events occurring within the service. Examples of data available through the APIs are the amount of electric or normal bikes available at each docking point and the number of free docking point spots available (Vélib' Métropole, n.d.b).

4.2.6 Västtrafik

Vässtrafik is a public transportation company in Sweden that helps travelers to plan their journey from point A to point B (Västtrafik, 2020). What makes this solution so unique is that the company uses several types of transportation vehicles, depending on the preferences. The most common ones are bus and tram but there are other alternatives such as walking or taking a cab. Obviously, the planned routes that are suggested are the most efficient and the fastest routes for the traveler. In addition to this, the company also provides a cost-proposition for the journey and thus, the journey can be bought directly when planning your journey (Västtrafik, 2020). In this way, the company has managed to collect all of the important aspects of a journey and gather them into a common platform. The company is publicly owned which means that additional investments in technology development and building infrastructure are provided by the government. Hence, the app is free for everyone to use without additional costs (Västtrafik, 2020).

Since traffic information changes very rapidly, it is very important for the company to stay the traveler updated on the traffic information (Västtrafik, 2020). One of the most important aspects for Västtrafik is customer experience and customer satisfaction. Furthermore, an essential factor that can affect these aspects is travel times and delays. Consequently, any travel delays will have a negative impact on the customer experience and customer satisfaction. In order to prevent this from happening, or at least minimize the impact of delays on the traveler, the company provides real-time traffic updates. This type of data is gathered from governmental open data in combination with other types of open data that may affect public transportation traffic. This type of data also helps to improve the API and modularization in the app which further helps to improve the customer experience (Västtrafik, 2020).

More specifically, the open data that is collected by västtrafik is processed in different ways to be more competitive on the market by creating new business ideas and staying in line with the digital development (Lubian & Morin, 2021). First, the authors explain that open data is collected in a cloud system that is available for all the developers. This cloud system can be accessed by everyone that is interested in contributing by developing their own ideas of new APIs'. Vässtrafik acts as a common forum for developers to communicate and network by discussing and developing new APIs of the open data that has been collected by the company (Lubian & Morin, 2021). One of the most important APIs for Västtrafik was developed this way in 2017. The open traffic data that had been collected was used by developers to create a real-time analysis of the traffic API called WSO2 IDP (Lubian & Morin, 2021). The new API was built on previous internal- and external traffic management APIs that were put together. The new API was connected to the real-time traffic situation present on Trafikverket to find the fastest travel way (Lubian & Morin, 2021). Rudmark (2013) discusses such example, every time a customer would search for a journey, the API would first look for the fastest way to get from point A to point B and check if this route is affected by the current traffic situation in Trafikverket. If the route was affected, it would give notice to the traveler that there is a risk for the delay and also provide real-time explanation data from Trafikverket (Rudmark, 2013).

In figure 3, it is evident that Västtrafik has their own specific way of working with open data and swaggers (Open-data APIs). The company also states how important it is that they create their own way of processing data and have their own site of developers to continuously keep their platform updated (Västtrafik, 2020). In the future, they hope to be able to use already developed open APIs such as google maps open APIs to calculate and analyze residual time ratios. The example brought up earlier by Rudmark (2013) can be addressed to this vision by combining the real-time crowding information with already developed APIs. Sabo (2018) studied how a company can build a platform that will be built on several external APIs and that will help to calculate the difference between using a car or traveling collectively based on factors as time, safety, accident risk, and environmental impact. Companies like Västtrafik already uses the Google open APIs to create incentives for the traveler to skip the car and instead travel collectively (Sabo, 2018). With this as a basis, a system for collecting travel time data using the Google Maps Directions API and calculating residual time ratios and other metadata based on this collected data had been created (Sabo, 2018). Furthermore, the system is used to make daily runes and stores collected data in a database at Västtrafik. The system also has a dashboard that can be used to perform searches in the database that has been built up and in this way retrieve, visualize and analyze the desired data. Travel time data for public transport obtained from Google has been randomly compared with travel time data from Västtrafik even though there are some deviations (Sabo, 2018).

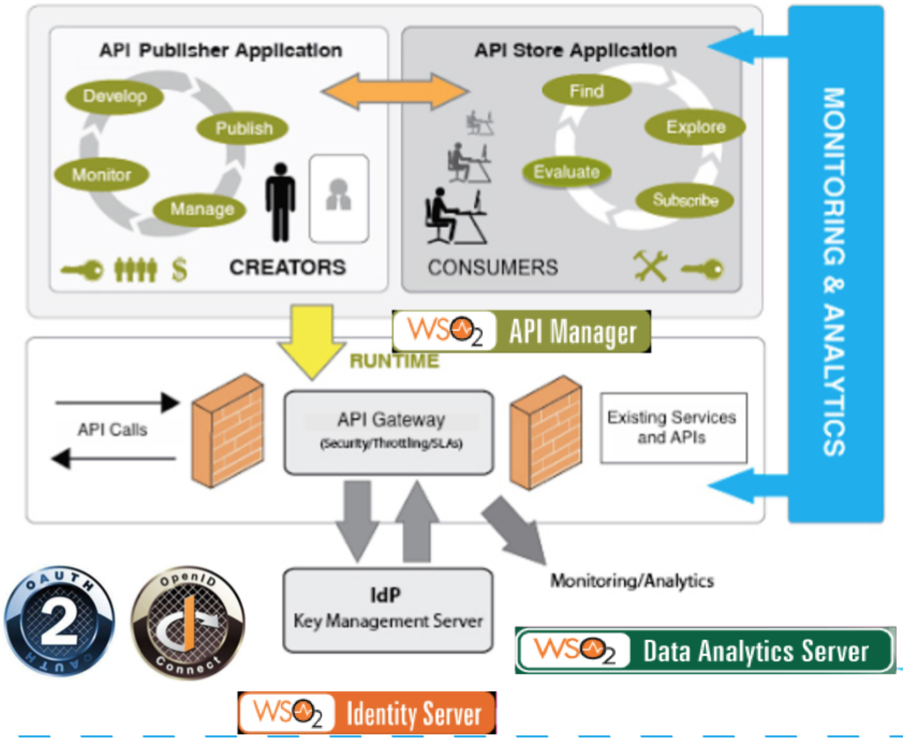

Figure 4: The process of API development and monitoring. (Västtrafik, 14 January 2022)

4.2.7 Gaia Public Transport

Gaia Public Transportation is a real-time, cloud-based system that makes travel easier for passengers, drivers, and operators (Gaia, 2021). It is a central system that is not reliant on information from individual vehicles but instead relies on open national data, allowing companies to have a better understanding and management of public transportation information flow in real-time (Gaia, 2021). The solution primarily aids in the control of public transportation vehicles, as well as their status, signs, and exclamations. The information system enables the corporation to give real-time information to passengers, drivers, and traffic control centers that are continually updated and relevant, independent of unit or location. Gaia Public Transport is cost-effective, easy to deploy and update since it is cloud-based and not connected to specific hardware or restricted by space. Furthermore, it is designed to be versatile and simple to upgrade with the functionalities required in public transportation, both today and in the future (Gaia, 2021). The adaptable software system integrates the whole information flow and serves as the foundation for future public transportation. The passenger experience is improved, and travel becomes more accessible, thanks to increased functionality and pertinent information.

Gaia Public Transportation can create digital twins of vehicles, passengers, and travel in real-time using open data. This enables them to observe exactly what is going on at any single time and to share that information immediately (Gaia, 2021). Every each piece of data may be kept in the cloud at the same time, making actions and occurrences verifiable and allowing massive volumes of data to be evaluated to improve travel and the passenger experience. It also allows for more complex analyses in the future, such as the ability to foresee and automate mistake avoidance. As a result, future estimates can be based on real-time information (Gaia, 2021).

Open data can be used in many ways to create new business opportunities. However, the fact that different public transportation companies are nisched in different areas implicates that that the type of APIs they use varies. In the case of Gaia Public Transport their main focus is to provide support to othe public transportation companies that uses open data by gathering all important data in a collective cloud system (Gaia, 2021). However, they are today working with several projects that helps to improve this cloud system. The goal with these projects is to provide a wide range of new solutions for companies to use directly in their coding, offering a complete external platform that can be integrated directly (Gaia, 2021). One project that is under development for the moment is GTFS Gaia Public Transport Dev (Trafiklab - developer, 2021). This project uses trafiklabs datasets to gather complete APIs developed by Trafiklab to gather information regarding the speed of public buses (Trafiklab - developer, 2021). Examples of the APIs that are used in this developing project is GTFS Regional Realtime, GTFS Regional Static data, and NeTEx Regional Static data (Trafiklab - developer, 2021). The idea is to create a tool for companies where the goal is to visualize the speed of several bus routes - and further generate a live-map of a specific region’s bus speed. In addition, this will help the public transportation companies to plan the routes more efficiently by combining the live-map of bus speed with traffic weather open data. For example, according to Gerstle (2012) a public transportation company can easily adjust the calculated time of travel by bus based on weather. This is clearly a problem for many companies since the winter in Sweden is well-known to be very though for the traffic. The maintanence work on roads increases to make roads safe to drive on which naturally affects the speed on which the bus can drive on the roads (Gerstle, 2012). From a safety point of view, this solution will help to plan with delays more accurately and provide bus drivers with accurate data on speed to minimize risk of accidents and help increase safety on the roads (Gaia, 2021). Gerstle also adds that other business opportunities can arise with this solution i.e. the government applies this platform in their system when planning maintanence on the roads and developing the roads to be more effective and safe. In this way, many stakeholders can plan the roads, bus routes and the transportation time with greater precision by predicting how the traffic is affected based if it is winter or summer.

Another business opportunity that is under development today for Gaia Public Transport is the safety of travelers (Trafiklab - developer, 2021). These type of services are mostly focused on reaching small niched groups e.g. people that are scared of going home alone. Gaia Public Transport are developing an API that collects real-time traffic data of the traveler and shows exact position using a map (Trafiklab - developer, 2021). This type of API can be used in e.g. police initiative called “Nattknappen”. People in Sweden can call between 11 pm and 4 am if the traveler feels insecure on the way home (Rudmark, 2013). The feeling of security increases when you talk to someone on the phone and the person you talk to on the “Nattknappen” can also act quickly if something should happen after all. Furthermore, Rudmark (2013) explains that in the systems used by Natknappen's volunteers, there is a connection to the public transport APIs developed by public transport companies, which enables the police to guide the caller over the telephone in the public transport system.

5. Discussion

Open data can be used to stay innovative by creating new business ideas but also by improving several processes such as customer experience or decreasing delay times. Those factors are crucial within the transportation industry and have a big influence on the business. In recent years, digitalisation has gained an increasing influence on society and today plays a central role in all types of business and the fact that our society is being more digitalized makes the use of data so important. Not least when the governments in several countries today strive for developing smart cities, one way to do so is by utilizing the open data available. Using open data is for free and can help private- and public companies to offer creative job offerings for developers. A clear example of that is Västtrafik, a public transportation company that has created a developer website accessed by everyone and anyone. This website helps curious developers to practice on coding and creating their own APIs using open data. The platform is an ingenious idea since it increases the awareness of open data and its potential, but it also benefits the society by providing a platform where developers can be imaginative, creative and innovative in their way of thinking. In addition, it helps create a network of developers where sharing knowledge is essential for personal development. A big downfall is that these developers lack incentives to use the platform since no salary is being offered, however, there is a chance for employment.

Important to mention is that all types of data is continuously gathered, updated and uploaded which means that new business opportunities are being created. Public transportation companies carry a huge responsibility to different stakeholders such as their customers/users, communities and employees in terms of accuracy, safety and user interface. This is obviously due to the large size of the public companies' customer base. Based on this, it is important for private transportation companies to learn from public companies in how they tackle several problems using open data. For example, Moovit made their operations more safe and secure for their users but also for their employees by gathering real-time open traffic data on how many persons that are traveling to inform users whether there is a free sit-place on the bus to minimize the risk of infection due to COVID-19. Another improvement they did was to indicate for their users if the vehicle is capable of handling disable people. This type of data could be beneficial for taxis by knowing exactly what type of taxi to send to the customer based on their needs. In addition to this, a vast amount of open data generated by public transportation companies, such as traffic and public transit data can be used to develop existing and new applications. For example, using open data based APIs from dataset companies such as trafiklab or trafikverket can help to create more reliable and user-friendly application. A clear example of this is gather data on what a user spends time on and how much time is being spent when using the app which indicates on what is important to fulfill the customer needs and improve user experience by e.g., providing better design which is easier to handle.

Business opportunities also include service offerings for the users. As seen by examples such as Moovit, open data does really facilitate the interchange of services between commercial and public entities. App developers and other similar firms may leverage open transit data to build their apps and businesses, and then use services generated by private actors to assist municipalities and public transportation companies improve their applications and processes. Apps that integrate a variety of modes of transportation have the opportunity to collect a greater quantity of data on how people travel around cities than public transportation companies. As a result, they may be able to provide services that are more complex than those provided by bus companies.

A huge obstacle which many private transportation companies have struggled to overcome is being effective with their deliveries, especially product-focused transportation companies such as Budbee or DHL. These companies often deliver packages inside the city centers which indicates that they may have a difficult time predicting delays. This is mainly due to the fact that the used routes are not always the same, which means they may not have the proper historical data to monitor that precise delivery cycle. Thus, it is necessary to get an understanding of how the delivery will be affected by the traffic to estimate the delivery time of a package in order for people to be home at the appropriate time for delivery. Buses, for example, drive on almost all roadways within cities and their data is freely available to study. As a result, private transportation companies should be allowed to participate in the open data that the government and public transportation companies provide to examine historical bus-route data where deliveries would be made through. This means that private transportation companies will be able to make more precise projections about how long delays will be during a specific event, time, day or weather. However, this type of solution or algorithm could also be implemented as an API on other apps to help making delivery predictions. In this way, customers can always stay informed and plan accordingly. Ultimately, this idea of using open data will help private transportation companies to become better on planning their delivery routes and become more efficient in their processes, and consequently achieve higher customer satisfaction in their business.

From the sections on data analytics within both traffic and public transportation, business operations can evidently be improved by combining open data with predictive analytics. In order to offer reliable and timely services in the urban transport industry, companies will be in need of predicting potential disruptions that external sources such as traffic and weather can have on the flow of transportations. It has been shown that weather and traffic conditions can have significant impacts on delays caused by traffic congestion and reduced speeds and that these transportation disturbances can be quite accurately forecasted through the use of predictive analytics and machine learning using open data as input. These delay predictions could greatly facilitate the planning processes in companies where the number of vehicles required, timetables and alternative routing when possible can reduce the experienced disturbances as much as possible and hence provide an improved experience for the users of these services. Moreover, analyses on how public transportation is functioning could be of value for private transit companies such as Uber to gain insights in where to estimate demands in certain areas affected by disruptions in public transit, as well as to set their prices in accordance with the demand. Thus, open transportation data in combination with data analytics may bring big potential gains for both public as well as private companies operating in the public transit industry.

While the availability of open data has made it possible for several new businesses reliant on open data to emerge, the cases brought up in this study have shown that a lot of the value that can be withdrawn from it comes from the data that these businesses generate themselves. This data can be used e.g. to target advertising towards its user as a way of gaining revenues or use it for business improving purposes. Moreover, although open data is usually only available from public agencies, the companies that use it to create new business opportunities can create further value from it that other companies can benefit from. As seen from the case of Moovit the company has been involved in several collaborations with other companies to help them develop their businesses with help from the platform that Moovit has built up to a large extent from open data. This brings new business opportunities and a way of gaining revenues from the application in more ways than solely from its traditional users. Furthermore it allows new innovations to arise which benefits society as a whole. Similarly, Citymapper is offering access to their APIs and SDKs in order to let other businesses incorporate some of the features developed by the company into their own applications. Thus, like Moovit, the company has shown that there are additional ways apart from the main purpose of the application to find business opportunities. Waze has been exchanging data with event organizers in order to improve their own application while simultaneously helping the opposite parts improve their businesses. While this exchange of data differs from true open data in the sense that open data has to be publicly available for free, it highlights some of the possible benefits that sharing data among companies or organizations can create value for both parties without either one losing their competitiveness.

The findings of this study has shown that the value of the data generated by open data-driven data not only exists for other private companies, but for public organizations as well. As societies are becoming more and more digital, data can bring valuable contributions to the development of cities and their infrastructure as well as information and insights that can aid the planning for different publicly managed organizations. As companies like Citymapper, Waze and Sygic generate enormous amounts of data through their applications, a lot of information of how people are moving, the efficiency of urban mobility and accidents occurring can be collected. By giving access to public institutions to the crowdsourced data these companies generate, information regarding behavior and accidents can be used to make better city planning decisions and to reduce the amount of people being harmed from traffic and the spread of diseases than would perhaps have been possible through data gathered only through public funds. Thereby the government, which is most often the distributor of open data, and companies using it can thrive from each other and create a win-win situation where both involved parties can find increased success. Some of the problems highlighted with crowdsourced data are its reliability and validity. There may be a lesser degree of transparency and control of the data collected by private institutions than public ones and as data generated by individual application users can be difficult to evaluate in terms of reliability some actions may need to be taken in order for crowdsourced data to have a place in public planning. Hence, there may be an incentive of taking a collaborative approach between public organizations and private companies to ensure that the data gathered is reliable and that it can be compared with publicly collected data in terms of validity and quality. However, as for example the accident data collected by Waze is often reported quicker than data collected by Trafikverket, it may be of high relevance for society if a sufficient method for validating it is developed.

6. Future opportunities & Conclusion

A future increased implementation of open data in existing businesses may be seen as the awareness of it increases. With the large availability of open APIs, more companies may find increasing value in adopting the use of open data, not least as it is offered at no cost. One possible future opportunity of open data with regards to the open data portals that this study is centered around is the integration of timetables or traffic conditions through external websites. For instance, while booking a doctor’s appointment or event tickets through websites like Ticketmaster, the user could be provided with information about public transportation and how to get to the location of the event as efficiently as possible, or with information on traffic conditions alerted through e-mail during the time of the event. This could improve both the scheduling for the events as the responsible people to a higher degree will ensure that people arrive on time, as well as improving the level of satisfaction for the customer. Another possible application is for real estate websites to include public transport timetables that lets the user see the time it takes to travel to work or school which might be a parameter that is considered when choosing a place to live.

There is undoubtedly potential value to be gained in multiple parts of society from open data both from the point of view of private and public companies and institutions, and both for improving and creating new businesses. However, there are some challenges with open data that seem relevant to address. Firstly, a lot of open data is available but far from all open data is well structured and available through easily accessed portals and APIs. Furthermore, open data is arguably not used to its full potential, perhaps because not all data has any relevance for other actors than the one providing it, or because the awareness and knowledge about how to use it is not as high as it could be. There are also big debates on whether companies could potentially lose competitiveness if they publish their data openly which might be why the very majority of open data today consists of data available from governmental institutions. In order to maximize the potential of open data, the awareness may need to increase as well as the availability of open data from private actors in addition to public ones. Moreover, the quality and structure of the available open data may need to be improved in order for more actors to be willing to use it but as the level of dissemination continues to be what seems to be quite low, governmental institutions may find it difficult to motivate spending resources on this issue. Hence, the future potential of open data will likely be dependent on whether people or companies will find it increasingly valuable or not.

Concludingly, data is available today in extremely large quantities, presenting businesses with a plethora of business prospects. All of this data, which is often referred to as big data, is meaningless if it cannot be exploited. One of the biggest drawbacks of open data is that it is often available in vast amounts and is unorganized. This is a significant cost to businesses in terms of profitability, since it takes a significant amount of time and effort to locate and organize essential data so that it can be used for the right purpose. Another challenge with open data is the low level of awareness which leads to a lot of its potential not being put to use in the business world. Since open data is offered to everyone, it is also hard for companies to only rely on using it to become competitive on the market. Another issue is the privacy of releasing data openly for the community. The risk of wrongly leaking information regarding individuals could have fatal consequences for many stakeholders. That is why there needs to be processes that search through this type of data before making it publicly available. However, using open data correctly and efficiently can be a very strong tool which companies can use to stay innovative and improve their operational processes.

References

Alam, O., Kush, A., Emami, A., Pouladzadeh, P. (2021). Predicting irregularities in arrival times for transit buses with recurrent neural networks using GPS coordinates and weather data. *Journal of Ambient Intelligence and Humanized Computing, 12*, 7813–7826. https://doi.org/10.1007/s12652-020-02507-9.

Appel, M. (2013). *Så funkar appen Waze som varnar för bilköer.* PC för Alla. https://pcforalla.idg.se/2.1054/1.743573/black-friday-router-natverk

Arabghalizi, T. Labrinidis, A. (2019). How full will my next bus be? - A framework to predict Bus Crowding Levels. School of Computing and Information - Department of Computer Science.

Attaran, M., & Attaran, S. (2019). Opportunities and challenges of implementing predictive analytics for competitive advantage. *Shah, M. & Yeoh, W (Eds.), Applying Business Intelligence Initiatives in Healthcare and Organizational Settings,* 64-90 IGI Global. doi: 10.4018/978-1-5225-5718-0.

Balali, F., Nouri, J., Nasiri, A., & Zhao, T. (2020). *Data Intensive Industrial Asset Management. Cham, Springer: IoT-based Algorithms and Implementation.* doi: 10.1007/978-3-030-35930-0.

Balbin, P., Barker, J., Leung, C., Tran, M., Wall, R., Cuzzocrea, A. (2020). Predictive analytics on open big data for supporting smart transportation services. *Procedia Computer Science, 176,* 3009-3018. https://doi.org/10.1016/j.procs.2020.09.202

Bus-News. (2021). Moovit Unveils New Features: Real-Time Crowding and Accessibility. Science & Technology. https://bus-news.com/moovit-unveils-new-features-real-time-crowding-and-accessibility/

Charalabidis, Y. & Zuiderwijk, A. (2012). Benefits, Adoption barriers and Myths of Open Data and Open Government. *European research on electronic citizen participation and engagement in public policy making*. doi: 10.1080/10580530.2012.716740.

Chowdhury, M. Apon, A. Dey, K. (2017). *Data Analytics for intelligent transportation systems*. Elseiver Inc. Cambridge, MA, United States. ISBN: 978-0-12-809715-1

Citymapper. (n.d.a). *Making Cities Usable.* Citymapper. https://citymapper.com/company

Citymapper. (n.d.b). *Powered by Citymapper.* Citymapper. https://citymapper.com/poweredby

Citymapper. (2019). *Citymapper PASS.* Medium. https://medium.com/citymapper/citymapper-pass-17c56da5dfa0

D'Addario, J. (2021). *Case study: The value of data sharing in the private sector.* The Open Data Institute.
https://theodi.org/article/case-study-the-value-of-data-sharing-in-the-private-sector/

data.europa.eu. (2019). *Find your way with Waze: a navigation app.* Data.europa.eu.
https://data.europa.eu/it/news/find-your-way-waze-navigation-app

data.europa.eu. (n.d.a). *Sygic.* data.europa.eu.
https://data.europa.eu/sites/default/files/use-cases/global_-_sygic.pdf

data.europa.eu. (n.d.b). *Vélib' Métropole.* data.europa.eu.
https://data.europa.eu/sites/default/files/use-cases/france_-_velib_metropole.pdf

Darsena, D. Gelli, G. Iudice, I. Verde, F. (2021). Enabling and Emerging Sensing Technologies for Crowd Management in Public Transportation Systems: A Review. IEEE Sensors Council.

Edelstein, S. (2019). *Wae offers cities a trove of data to help fight traffic.* Digitaltrends.
https://www.digitaltrends.com/cars/waze-for-cities-data-puts-traffic-data-on-google-cloud/

Eriksson, I. (2019). *Towards Integrating Crowdsourced and Official Traffic Data.* [Master's Thesis, Uppsala Universitet]. DiVA.
https://www.diva-portal.org/smash/record.jsf?pid=diva2%3A1336673&dswid=-6624

Fu, Q. & Easton, J. M. (2017). Understanding data quality: Ensuring data quality by design in the rail industry. *2017 IEEE International Conference on Big Data (Big Data)*, pp. 3792-3799.

Gaia. (2021). Website url: https://gaia.se/gaia-public-transport/

Gerstle, D. (2012). Understanding bus travel time variation using AVL data. Massachusetts Institute of Technology. Department of Civil and Environmental Engineering.

Gudivada, V., Apon, A. & Ding, J. (2017). Data Quality Considerations for Big Data and Machine Learning: Going Beyond Data Cleaning and Transformations. *International Journal on Advances in Software 10.1*, pp. 1-20.

Jansen-Young, C. (2021). Real-time traffic data - a trending topic. Intertraffic organization - rai-amsterdam.
https://www.intertraffic.com/news/big-data/real-time-traffic-data-developments-2021/

Juddoo, S., George, C., Duquenoy, P., & Windridge, D. (2018). Data Governance in the Health Industry: Investigating Data Quality Dimensions within a Big Data Context. *Applied System Innovation,* 1(4), 43. doi:10.3390/asi1040043

Lenkei, Z. (2018). *Crowdsourced traffic information in traffic management: Evaluation of traffic information from Waze.* [Master's Thesis, KTH]. DiVA. https://www.diva-portal.org/smash/record.jsf?pid=diva2%3A1264077&dswid=-925

Leung, C., Elias, J., Minuk, S., De Jesus, A., Cuzzocrea, A. (2020). An Innovative Fuzzy Logic-Based Machine Learning Algorithm for Supporting Predictive Analytics on Big Transportation Data. *2020 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE),* 1-8, doi: 10.1109/FUZZ48607.2020.9177823.

Li, L., Lin, H., Wan, J., Ma, Z., Wang, H. (2020). MF-TCPV: A Machine Learning and Fuzzy Comprehensive Evaluation-Based Framework for Traffic Congestion Prediction and Visualization. *IEEE Access, 8,* 227113-227125, doi: 10.1109/ACCESS.2020.3043582.

Li, Z., Cui, X., Zhang, Z. (2021). Short-term Traffic Flow Prediction Based on Recurrent Neural Network. *2021 International Conference on Computer Communication and Artificial Intelligence (CCAI),* 81-85, doi: 10.1109/CCAI50917.2021.9447466.

Lubian, M. Morin, C. (2021). Integration och API:er på Västtrafik - Västra Götalandsregion. https://pulsen.se/download/18.370ae51160f495f829178ba/1537252837583/13+-+1045+-+INTEGRATION+-+Västtrafik+Integration+och+API+Pulsendagarna+20180913.pdf

Maddison, J & D'Addario, J. (2020). *Case study: The value of sharing data benchmarking and insights.* The Open Data Institute. https://theodi.org/article/case-study-the-value-of-sharing-data-for-benchmarking-and-insights/

McCarthy, R. McCarthy, M. Ceccucci,W. Halawi, L. (2019). *Applying predictive analytics - finding value in data.* Cham: Springer. doi: 10.1007/978-3-030-14038-0.

Moovit. (2021.a). *Arriva partners with Moovit for its first national Mobility as a Service launch in The Netherlands* [Press release]. https://moovit.com/press-releases/moovit-arriva-maas-partnership-netherlands/

Moovit. (2021.b). *Aberdeenshire Council partners with Moovit to launch the region's first enhanced demand-responsive transport service* [Press release]. https://moovit.com/press-releases/moovit-aberdeenshire-council-partnership/

Moovit. (2020). *In Their Own Words: Mooviter Community Volunteers on Their Great Work During COVID-19.* Moovit. https://moovit.com/blog/mooviter-community-during-covid-19/

Moovit. (n.d.a). *About Moovit.* Moovit. https://moovit.com/about-us/

Moovit. (n.d.b). *Join the Mooviter Community.* Moovit. https://editor.moovitapp.com/web/community?campaign=companywebsite&__hstc=171844735.62789c7e0da6566c868468fda8cb818a.1637421951374.1637421951374.1637421951374.1&__hssc=171844735.18.1637421951375&__hsfp=1861757312

Patterson, C. (2013). The six primary dimensions for data quality assessment: defining data quality dimensions. DAMA, United Kingdom. https://damauk.wildapricot.org/resources/Documents/DAMA%20UK%20DQ%20Dimensions%20White%20Paper2020.pdf.

Prekopová, M. (2019). *Avoid emission zones restricted for your vehicle type.* Sygic. https://www.sygic.com/blog/2019/avoid-emission-zones-that-are-restricted-for-your-vehicle-type

Q. Shi, M. Abdel-Aty. (2015). Big Data applications in real-time traffic operation and safety monitoring andimprovement on urban expressways,, Transp. Res.

Ramasamy, A. & Chowdhury, S. (2020). Big Data Quality Dimensions: A Systematic Literature Review. *Journal of Information Systems and Technology Management, 17,* 1-13. doi:10.4301/S1807-1775202017003.

Rudmark, D. (2013). Att öppna datakällor för tredjepartsutveckling - rekommendationer för kollektivbranschen. https://www.diva-portal.org/smash/get/diva2:1131851/FULLTEXT02

Runkler, T. A. (2020). Data Analytics. Wiesbaden: Springer Vieweg.

Sabo, S. (2018). Hur man kan använda Google Maps API:er för att beräkna och analysera restidskvoter. Ett kandidatarbete inom data vetenskap.

Sall, E. Zorn, L. Cooper, D. Sana, B. Coe, S. (2016). *Making open transportation data useful and accessible: Recommendations for good practices in open data standards management. UrbanLabs LLC*. Standing committee on Library and Information Science for Transportations.

Samal, S R., Gireesh Kumar, P., Cyril Santhosh, J., Santhakumar, M. (2020). Analysis of Traffic Congestion Impacts of Urban Road Network under Indian Condition. *IOP Conf. Series: Materials Science and Engineering, 1006,* doi: 10.1088/1757-899X/1006/1/012002.

Samtrafiken. (n.d). *Vi skapar världens smartaste kollektivtrafik.* Samtrafiken. https://samtrafiken.se/om-oss/

Santos, G. Nikolaev, N. (2021). Mobility as a Service and Public Transport: A Rapid Litterature Review and the Case of Moovit. School of Geograpgy and Planning, Cardiff University, UK.

Sawers, P. (2016). *Waze partners with marathons and other events to improve traffic and road closure data.* VentureBeat. https://venturebeat.com/2016/08/11/waze-launches-events-partner-program/

Sawers, P. (2019). *Waze integrates city data-sharing program with Google Cloud.* VentureBeat. https://venturebeat.com/2019/10/01/waze-integrates-city-data-sharing-program-with-google-cloud/

Scherrer, L., Tomko, M., Ranacher, P., Weibel, R. (2018). Travelers or locals? Identifying meaningful sub-populations from human movement data in the absence of ground truth. *EPJ Data Science,* 7. https://doi.org/10.1140/epjds/s13688-018-0147-7

Smith, G. (2015). *Marknadsplatser för öppen data - en fallstudie kring användare, användning och upplevt värde.* Viktoria Swedish ICT.

Sun, F., Pan, Y., White, J., Dubey, A. (2016). Real-Time and Predictive Analytics for Smart Public Transportation Decision Support System. *2016 IEEE International Conference on Smart Computing (SMARTCOMP),* 1-8, doi: 10.1109/SMARTCOMP.2016.7501714.

Sygic. (n.d.a). *Long Haul.* Sygic. https://www.sygic.com/enterprise/use-case/gps-for-truck-fleets

Sygic. (n.d.b). *Delivery.* Sygic. https://www.sygic.com/enterprise/use-case/delivery

Sygic. (n.d.c). *Smart Mobility Platform.* Sygic. https://www.sygic.com/enterprise/use-case/smart-city-mobility

Sygic. (n.d.d). *Driver Scoring.* Sygic. https://www.sygic.com/enterprise/driver-scoring-telematics

Sygic. (n.d.e). *Insurance Telematics.* Sygic. https://www.sygic.com/enterprise/insurance-telematics

Taulli, T. (2019). *Artificial intelligence: The basics - a non technical introduction*. Apress.

Tavmen, G. (2020). Data/infrastructure in the smart city: Understanding the infrastructural power of Citymapper app through technicity of data. *Big Data & Society,* 1-15, https://doi.org/10.1177/2053951720965618.

Taylor, M. (2021). *How to save Citymapper.* Wired. https://www.wired.co.uk/article/how-save-citymapper

Trafiklab. (2016). *Citymapper.* Trafiklab. https://www.trafiklab.se/sv/fallstudier/2016/11/citymapper/

Trafiklab. (n.d.a). *Om Trafiklab.* Trafiklab. https://www.trafiklab.se/sv/om-trafiklab/.

Trafiklab. (n.d.b). *Realtime data.* Trafiklab. https://www.trafiklab.se/sv/api/trafiklab-apis/gtfs-regional/realtime/

Trafiklab. (n.d.c). *Route-planning.* Trafiklab. https://www.trafiklab.se/docs/using-trafiklab-data/the-right-data-type-for-your-project/route-planning/

Trafiklab - developer. (n.d.c). Passagerarinformation API:er - Gaia Public Transport. Trafiklab. https://developer.trafiklab.se/tags/passagerarinformation

Trafikverket. (n.d.a). *Vår verksamhet.* Trafikverket. https://www.trafikverket.se/om-oss/var-verksamhet/

Trafikverket. (n.d.b). *Läs om våra data.* Trafikverket. https://www.trafikverket.se/tjanster/data-kartor-och-geodatatjanster/las-om-vara-data/

Trafikverket. (n.d.c). *Hämta eller beställ data.* Trafikverket. https://www.trafikverket.se/tjanster/data-kartor-och-geodatatjanster/hamta-var-oppna-data/

Trafikverket. (n.d.d). *API.* Trafikverket. https://api.trafikinfo.trafikverket.se/API

Van Bodegom, D. (2021). *How Predictive Analytics Is Reshaping Transportation Management.* Industry Europe. https://industryeurope.com/sectors/transportation/how-predictive-analytics-are-reshaping-transportation-management/

Vélib’ Métropole. (n.d.a). *About Vélib’.* Vélib’ Métropole. https://www.velib-metropole.fr/en/service

Vélib’ Métropole. (n.d.b). *Bienvenue sur la page Open Data du service Vélib' Métropole.* Vélib’ Métropole. https://www.velib-metropole.fr/donnees-open-data-gbfs-du-service-velib-metropole

Västtrafik AB. (2020). Annual report of 2020. https://www.vasttrafik.se/globalassets/media/dokument/styrelsehandlingar-och-protokoll/moten-2021/mote-26-februari/nr-07.6-arsredovisning-for-verksamhetsaret-2020.pdf

Waze. (n.d). *About Us.* Waze. https://www.waze.com/about

Ylijoki, O. & Porras, J. (2016). Perspectives of defining big data: A mapping study and Discussion. *Journal of innovation management, 4*(1), 69-91. doi: 10.24840/2183-0606_004.001_0006.

Zhang, D., Qu, X., Fang, Y. (2020). Traffic Congestion Analysis Based on Deep Neural Networks. *Proceedings of 2020 the 10th International Workshop on Computer Science and Engineering (WCSE 2020),* 52-56, doi: 10.18178/wcse.2020.06.009.