Mobisuite: A User-Friendly Tool to Exploit E-Ticketing Data and Support Public Transport Planning

Alexander Fazari, Maurizio Arnone(✉), Cristiana Botta, Brunella Caroleo, and Stefano Pensa

Links Foundation, Via Pier Carlo Boggio, 61, 10138 Torino, Italy
{maurizio.arnone,cristiana.botta,brunella.caroleo, stefano.pensa}@linksfoundation.com
https://linksfoundation.com/

Abstract. Smart cards for public transportation have become a significant source of information for analyzing mobility trends and assisting public authorities and transportation organizations in making decisions. This article describes a visualization tool developed by LINKS Foundation that exhibits spatial and temporal characteristics, with the goal of demonstrating the potential of analysing and visualising smart card validation and transit data towards better decision making. The Piedmont region's transit data was used as a case study. The developed public transport visualization tool (called Mobisuite) represents the transit system supply, the mobility demand, and the ticketing system through three primary visualization modules. The tool employs a virtual check out algorithm, which estimates the users' travel diary and aids in the calculation of various transit indicators, e.g. passenger kilometers, maximum load, number of on-boarding passengers, average load. Mobisuite can be a helpful tool for displaying the public transport system for the whole transport network at a glance, and it can be used to simplify communication between transit operators and public agencies regarding the public transport planning, operation and fares.

Keywords: Smart-card ticket · Decision support systems · Fare collection · E-ticketing data · Transit planning · Public transportation · Transportation demand · Intelligent transportation systems

1 Introduction

In the last decade, Public transportation authorities are increasingly relying on smart card electronic fare collection systems [1]. In urban public transport, smart cards collect data of millions of users boarding vehicles over the network every day.

The large amount of daily collected data from smart card validations has created an opportunity to feature data engineering in order to better understand the complexities of urban mobility, which is a challenge for transportation researchers and stakeholders. The term "big data" emerged to describe the

A. L. Martins et al. (Eds.): INTSYS 2021, LNICST 426, pp. 149–161, 2022.
https://doi.org/10.1007/978-3-030-97603-3_11

massive amount of data generated, which exposes a significant new age transformation and reflects a significant change in how people live, act, and think [2]. Although the big data is varying in terms of source and type, it remains data. Thus, data processing and analysis are required including data acquisition, data harmonization, data storage and management, data processing (including data mining and smart analysis) and data presentation [3].

The big data on human mobility has been used to study human behaviors [4], allowing the discovery of new complex mobility patterns in the form of insightful indicators, that can be useful to manage and improve mobility in cities and in urban areas. In essence, these big data sources may help respond to the needs of planners and decision makers. In fact, planning is a decision process that needs to mine rich information to serve plan development, decision making and plan implementation evaluation [3].

Surveys have usually been used to determine demand and supply in the transportation industry. However, since surveying entails primarily in-person interviews and a heavy workload, it is an expensive and restricted form of data collection [5], as more datasets related to the use of transportation systems and human mobility have been created and made available, new opportunities beyond survey towards data-driven research has been emerged, especially data visualization tools. Data visualization converts these datasets into appropriate digital representations, which can lead to improvements in transportation system design, administration, and operations [6]. Bridging the demand-supply gap in transit service is critical for public transportation management, as planning steps can be taken to increase supply in high-demand areas or boost inadequate transit service deployment in low-demand areas [7].

Demand in transport systems can be very sensitive to sudden disruptive events. For example, during COVID-19 pandemic, public transport was one of the most disruptive sectors with early estimates suggesting that the drop in ridership during lockdown periods has been as much as 80%–90% in major cities in China, Iran, Italy and the United States, and as much as 70% for some operators in the United Kingdom [8]. Different measures have been taken into consideration by service providers, which include the reduction of service frequencies, the change of timetables and vehicle schedules, and the reduction of the total duration of the daily operations. All of these measures translate into the dimensioning of service capacity. Several transport operators have reduced their service frequencies to less than one-third of the pre pandemic time. Nonetheless, such decisions are rarely made based on a system-wide analysis [9]. In consequence, public transportation providers and local authorities have an urge for tools to help them make decisions on how to use their existing services more efficiently.

In recent years, many researchers approached the issue of visualizing information coming from public transport data in order to support public transport providers' decisions. For example, Prommaharaj et al. (2020) [10] developed a visualization tool for public transport data, but this tool was related only to the supply side (i.e. GTFS data) and not on the demand side (i.e. user validations).

Other research approached the problem of data fusion of APC (Automated Passenger Counting), Smart Cards and GTFS to visualize public transit use (Giraud et al., 2016) [11], but the resulting web interface was mainly aimed at supporting transit planners to see the movements of the users within the network. Also other studies developed open architecture platforms for transit data demonstration, analysis and visualization. For example, Kurkcu et al. (2017) [12] applied a bus trajectory data to develop a web-based tool to acquire, store, process and visualize bus trajectory data. Palomo et al. (2016) [13] proposed an online visualization tool, Trips Explorer (TR-EX), for analyzing reliability of transportation schedules. The tool allows users to compare planned timetables against real service, to analyze speed profile at route segments level, and to assess delay, wait time and reliability at station level. The AFC (Automated Fare Collection) data have been also explored by Anwar et al. (2016) [14] to develop a web-based application to monitor and visualize the performance of bus fleets.

Mobisuite fits into this line of research, seeking to combine - in addition to the analysis and visualisation of supply and demand - an analysis of ticketing fares related to the use of public transport, so that planners and transport operators can have a complete and integrated perspective of their services.

In this work Mobisuite is presented as a tool aimed to support transport planning to analyse and visualise transit data towards better decision making for the enhancement of the services of transport operators. Mobisuite is not a simulation or optimization tool. It analyzes and compares current supply and demand data. Moreover, it evaluates the impact of new integrated fare models on users and transport operator systems. Mobisuite reflects various transit system characteristics through three different interfaces: (1) users' demand: visualizes the current service usage (also enabling for temporal comparisons in terms of historical data), provides how much, when, where, and by whom the public transport (PT) services are used; (2) public transport supply: illustrates the structure of the PT service network (e.g. how many rides are provided, when and where); (3) ticketing: shows how different sorts of smart card titles are utilized (in terms of how much, when, where, and by whom). Thanks to Mobisuite, the PT planners and the transport operators will be able to have a full perspective of the system of services and their utilization. The insights they get from Mobisuite will help them fulfill the demands of transportation customers, allowing them to modify the services they offer and so provide a high-quality and tailored service.

The article is structured as follows: the data definition, architecture specifications and analysis indicators of Mobisuite are described in Sect. 2; the functioning of the tool and the results are discussed in Sect. 3; finally, Sect. 4 provides conclusions and future research directions.

2 Methodology

2.1 Business Understanding

BIP [15] - Biglietto Integrato Piemonte is the integrated electronic ticketing system of the Piedmont Region, located in the northwest of Italy, that allows, thanks to a contactless smart card, access to any public transport in any area of the regional territory. BIPEX [16] is the protocol for data exchange between transport companies, organizations and governments. BIPEX was developed using European and international open standards (TransModel, NeTEx, and SIRI) with the aim of enabling electronic ticketing systems to communicate with each other as widely as possible and harmonize data collection. BIP and BIPEX were created by 5T [17] (i.e. Telematic Technologies for Transport and Traffic in Turin) on behalf of the Piedmont Region and it is in use at PT companies (road and rail) operating in the Piedmont BIP electronic ticketing system. The contactless smart card based on BIP can be used to load both pay per use credit and travel passes to cover all possible trips and types of PT tickets available in the Piedmont Region, regardless of route, mode of transportation, or the PT company involved, as well as to use bike-sharing.

The Piedmont Region is used as a case study to show the development of a tool to help transportation planning by analyzing and visualizing transit data in order to make better decisions for the expansion of transportation operators' services.

2.2 Data Understanding

The data was provided by 5T which contains the metadata based on NeTEx standard. NeTEx [18] is a CEN "European Committee for Standardization" Technical Standard for exchanging Public Transport schedules and related data. It is divided into three parts, each covering a functional subset of the CEN Transmodel [19] for Public Transport Information:

1. Part 1 describes the Public Transport Network topology (CEN/TS 16614-1:2014);
2. Part 2 describes Scheduled Timetables (CEN/TS 16614-2:2014);
3. Part 3 covers Fare information (CEN/TS 16614-3:2015);

The metadata is based on a complete and flexible XML schema, the XML is further flattened and relevant fields are selected to create the input data for Mobisuite tool. Along with the Public Transport Network topology, Scheduled Timetables and Fare information, the passenger validation data and sales data were also provided for the months of January and February 2018, which Mobisuite used in the development phase.

2.3 Data Preparation

Extract and Harmonized Module. Mobisuite is a tool which makes use of BIP data to build various fundamental infrastructure modules shown in Fig. 1; the first module was created to extract and harmonize BIP raw data.

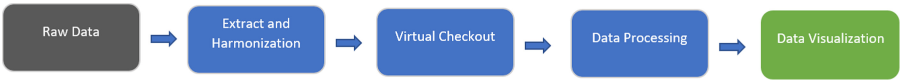

Fig. 1. Flow of the elaborations and analysis process of BIP data, from raw data to visualization within the Mobisuite tool.

Extract and harmonized module transforms the raw data collected with electronic smart cards into 8 unstructured data categories listed in Table 1 in a dictionary serialization using Python programming language, where each object has a key and a corresponding value represented as a pair (key: value). Python libraries such as Pandas, Numpy, and Geopy were utilized to elaborate and analyze the data.

2.4 Data Modeling

Virtual Check-Out Module: All the previous data extracted from the extract and harmonization module acts as an input to the virtual_check-out module which was created based on the trip-chaining model for estimation of the check out of the passengers[20]. The majority of passengers are travel pass holders (who just need to check-in), and there aren't enough pay-per-use credit transactions to allow for model training. The virtual check-out module was used to create the "travel diary," which is a list of all the trips carried out by a user on a given day, organized chronologically and with all essential information (recorded or calculated), such as: date, service, line, vehicle, operator, boarding bus stop (recognized by GPS system at check-in), boarding time (check-in time), alighting bus stop (estimated for travel pass users), alighting time (calculated for travel pass users based on the estimated bus stop alighting location and travel duration between boarding and alighting), kilometers traveled (the distance between the boarding and alighting bus stops) and time on board (the difference between the boarding and alighting times).

Virtual check-out trip chaining models are based on few key assumptions [20]:

1. A passenger's most likely alighting stop is the boarding stop of his next trip (or one close by);
2. Passengers do not use private cars between two consecutive public transportation trip legs;
3. Passengers do not move over a long distance;

Table 1. BIP data harmonization

Extract and Harmonized Module Output	Descriptions
Trip	Contains nested data described by a key of the Trip_id and data for Route_id, Line_id as values: (Trip_id: Trip_ id,Route_id, Line_id)
Calendar	Contains the dates where the trip is active, formatted in a dictionary with the Trip_id as key and a list of days with this format DD/MM/YYYY as the values: (Trip_id: [Date])
Stop code	Contains the stop number as key and the corresponding Stop_id as value: (Stop_number: Stop_id)
Trip_stops	Contains nested data identified by Trip_id as key, and the list of the stops_id in this trip as values. Then each stop_id has nested data describing the stop: (Trip_id:[Stop_id:[Stop_id, Trip_id, Date, Direction, Route_id, Line_id, Arrival_time, Stop_sequence]
Arch	Is a data structure, with a a couple of consequent Stops_id as key: (departure and arrival stops) and nested data related to the distance between the two stops, as a value: (Stop_departure, Stop_arrival, Arch_Km, Arch_id)
Programmed and definitive trip	Stores the trip_id in the definitive version (executed trip) and its corresponding programmed trip. It is used as a validator in the virtual check-out algorithm
Validation	A relational database containing the validation information for each smart card, including the card's Id, the time of validation, the validation procedure, the stop_id of the validation, the line_id, the trip_id and the route_id. The validation procedure differs between travel pass holders (who just need to check-in) and pay-per-use passengers (required to both check-in and check-out)
Sales data	Includes smart card title sales and passenger renewals, which were utilized in ticketing analyses

Virtual check-out algorithm follows several steps:

1. For each BIP smart card ID, a list of the validations and boarding stops sorted by time of check-in is created. Each check-in is associated with the corresponding service so that it is possible to identify the sequence of services used and boarding stops for each trip chain ("travel diary");
2. Services are matched with their route;
3. The BIP smart card ID corresponding to "single trips", whose destination cannot be estimated, are identified and eliminated from the starting database;
4. for each BIP smart card ID corresponding to a normal trip and for each consecutive couple of boarding stop and previous route used, the entry of the boarding stop of the next route and the previous route is the estimated alighting stop of user transferring from the route in day based on the fulfillment of the algorithm assumptions;
5. the alighting stops estimated represent the virtual check-outs and are inserted in the travel diary associated to each BIP smart card ID;

Data Processing Module: The last module is data processing, which does additional analysis and computation of various transport indicators before feeding it to the last mile visualization tools. The analyses are based on the output of the Virtual check-out trip chaining model and sales data.

2.5 Evaluation

The Virtual check-out trip chaining models has been tested and validated using the travelling data collected in the province of Cuneo, located within the Piedmont Region, at the end of 2017 where the algorithm was able to estimate the alighting bus stop of 94% of the valid travel pass holder validations[20]:. The model's reliability was assessed in particular by using the same technique to a data sample of the pay-per-use credit (where both check-in and check-out are necessary) and comparing estimated alighting bus stops with those recorded automatically. More than 70% of the estimated alighting bus stops in the case study matched the one reported or was at a maximum distance of 500 m from it.

As a result of the accuracy and reliability achieved by the algorithm, the trip-chaining model was chosen to estimate the pass holder passengers' travel diary for Mobisuite.

Mobisuite has revealed that, for various reasons, the data collected do not allow to trace 100% of the trip legs. In the case study it was possible to estimate 95% and 94% of the total travel pass holder validations, equal to 97880 and 98830 validations respectively, in January and February 2018. With the use of the virtual checkout algorithm it was possible to reconstruct from the correct validations 94% and 93% of the trip legs (sequence of check-in and check-out). In particular, 84% and 83% respectively in January and February 2018 did not require any elaboration whereas 10% in both months showed errors that were corrected with algorithms developed ad hoc (e.g. single trips and trips where the distance between consecutive stops is higher than 1 km), it was necessary to use another estimation method (i.e. OD matrix expansion using the matrix obtained through the check-out algorithm as base matrix). Only 6% and 7% from the correct validations could not be reconstructed, respectively in January and February 2018.

2.6 Deployment

The visualization interface was created using Microsoft PowerBI, which is designed with three main modules that depict several variables related to user demand, public transit supply, and ticketing.

3 Results and Discussion

Mobisuite visualizes transit system indicators from the viewpoints of public transportation lines, routes, journeys and stops. Figure 2 depicts Mobisuite, which has four main display tabs: home, users demand, public transport supply and ticketing, based on data retrieved and prepared for each module. The interface has been developed in Italian to facilitate the interpretation by the end user; the interested reader will find the English translation of the main terms in the glossary at the end of this article.

Microsoft PowerBI was used as a visualization tool, where it was designed with four main displays: home display, user's demand module display, supply display and finally the ticketing display. The home displays in Fig. 2 provides an overview of all the lines and routes, including a set of indicators such as bus-km, passenger-km, maximum load, number of on-boarding passengers, average load, and commercial km, the results of which may be shown based on the transport agency, year, month, line, and route.

Average load can be filtered monthly and daily with hour reference, which is an essential indication for determining the efficacy of lines, routes, and trips. Moreover, it determines the temporal and spatial characteristics of transportation demand, which leads to improved transportation service planning [21]. Understanding the average load, can be used to reduce the gap between supply and demand by adapting the fleets to the actual demand, resulting in a reduction in overall CO2 emissions pro capite [22].

In Fig. 3, the user's demand module display shows several indications for the selected line and route, as well as the number of trips carried out by each user categorised by smart card title. Other information, such as the distribution of km range vs. number of trips traveled, and the number of people going between zones, are also provided. The insights from Mobisuite user's demand module will assist in meeting the needs of transportation users, allowing the PT planner and transport operator to adjust the services they provide and so provide a higher-quality service that is suited to their customers' needs.

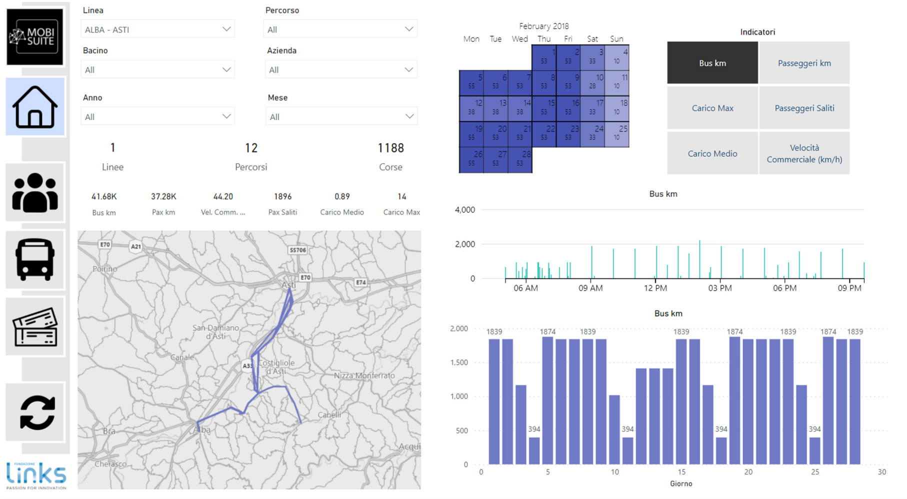

Fig. 2. Mobisuite interface - Screenshot of the home page of the tool.

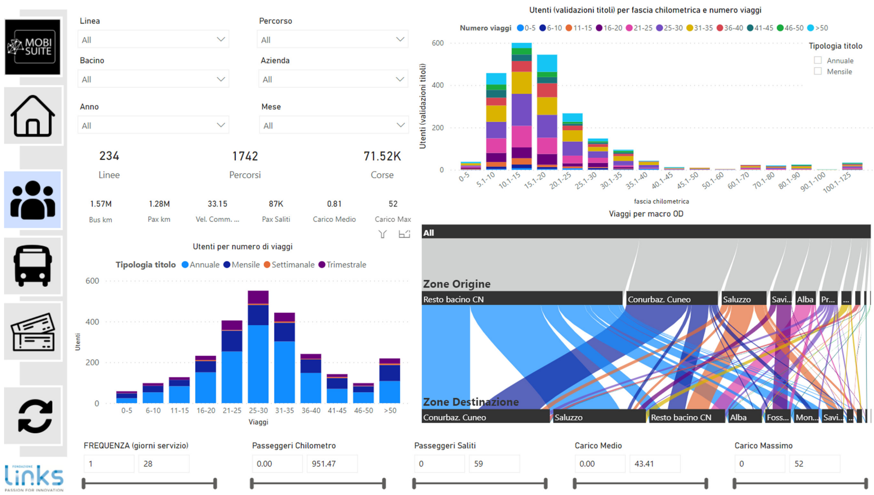

Fig. 3. Mobisuite interface - Screenshot of the user demand analysis page of the tool

The supply display Fig. 4, contains all accessible service statistics, such as frequency of service and number of stops, as well as bus km, passenger km, maximum load, number of on-boarding passengers, average load, and commercial km. Through the supply display, the PT planner and transport operator can get a comprehensive view of the whole PT system features.

Fig. 4. Mobisuite interface - Screenshot of the supply analysis page of the tool

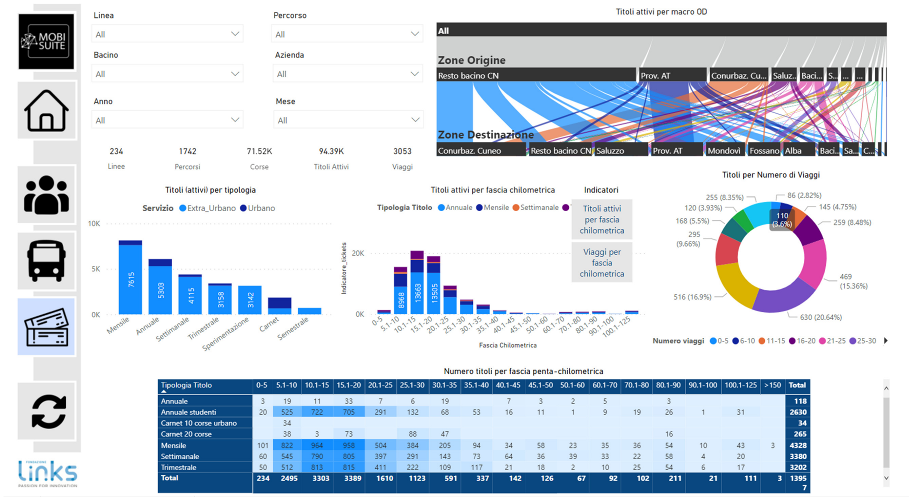

Fig. 5. Mobisuite interface - Screenshot of the ticketing analysis page of the tool

Finally, the ticketing display Fig. 5 shows statistics related to the active smart card titles in urban and extra-urban services, such as the active smart card titles for trip ranges traveled, and the active smart card titles for the macro Origin/Destinations pairs (OD).

The ticketing display not only categorizes different types of travel pass holders, but also shows how a subset of passengers used them in terms of number of trips, kilometers traveled, and cost, particularly across distinct urban and extra-urban zones, with the goal of evaluating the impact of new integrated fare models on the user and on the PT system.

4 Conclusion

Mobisuite is a tool that allows to compare the supply of public transport services (i.e. routes, number of rides, daily frequency) with the related demand (i.e. passengers per route, passenger*km) and to assess how different types of tickets or transport credit are used (i.e. number of trips, kilometers traveled, cost) by different subsets of users (i.e. students and employees) across urban and extra-urban zones.

Thanks to Mobisuite, the public transport planners and the transport operators will be able to have a full perspective of the system of services and their utilization.

The insights that can be gained from interpreting the Mobisuite calculations help planners to better meet the needs of public transport customers, as the ability to clearly analyze the relationship between supply and demand, allows to redesign services in order to offer quality, tailor-made services. Reducing the gaps between supply and demand allow to limit the CO2 emissions pro capite and the operational costs in the face of limited travel demand.

Mobisuite user-friendly interface and easy-to-read elaborations (charts, maps, tables) ensure easy use, even by less experienced users. Moreover, the data and elaboration visualization helps to improve communication between transit operators and local authorities regarding public transportation planning and operation.

The tool is still under improvement with the integration of a module that focuses on analyzing specific stops and trips. Another area that might be improved in the future is the use of static data paired with real-time data to create additional visual analytics modules that can conduct a range of real-time strategies and actual system performance.

Glossary

Linea. Name of the Bus Line
Percorso. Name of the Line Path. Each line could have different path according, for instance, to time of the day in which the service is operated
Bacino. Basin, the geographical area in which services are operated
Azienda. Company that provide the services
Anno. Year
Mese. Month
Linee. Lines, number of lines that match the filtered search
Percorsi. Paths, number of line paths that match the filtered search
Corse. Trips, number of trips that match the filtered search
Pax km, Passeggeri km. Passengers - km, total distance in kilometres traveled by all passengers on board on a trip or a set of trips
Vel. Comm.Velocita Commerciale (km/h). Commercial Speed
Pax Saliti, Passeggeri Saliti. Boarded Passengers, total number of passengers boarded on a trip or a set of trips

Carico Medio. Average Load, average number of passengers aboard a trip or a set of trips
Carico Max. Maximum load, maximum number of passengers on board on a trip or a set of trips
Indicatori. Indicators, indexes
Utenti per numero di viaggi. Users per numbers of trips
Utenti (validazioni titoli) per fascia chilometrica e numero viaggi. Users (validated tickets) per distance range and number of trips
Viaggi per macro OD. Number of trips per aggregated OD
Frequenza (giorni servizio). Frequency, number of days per month in which the service is operated
Numero fermate. Stops, number of stops per trip
Titoli Attivi. Active Tickets, tickets validated at least once in the reference period
Viaggi. Trips
Titoli (attivi) per tipologia. Active Tickets, per typology of subscription
Titoli attivi per fascia chilometrica. Active Tickets, per distance range
Viaggi per fascia chilometrica. Trips, per distance range
Titoli per Numero di Viaggi. Tickets/Subscriptions per numbers of trips
Annuale. Annual Subscription
Annuale Studenti. Annual Subscription for students
Carnet 10 corse urban. 10 Urban Ticket Carnet
Carnet 20 corse. 20 Ticket Carnet
Mensile. Monthly Subscription
Settimanale. Weekly Subscription
Trimestrale. Quarterly Subscription
Totale. Total

References

1. Pelletier, M.P., Trépanier, M., Morency, C.: Smart card data use in public transit: a literature review. Transp. Res. Part C Emerg. Technol. **19**(4), 557–568 (2011)
2. Mayer-Schonberger, V., Cukier, K.: Big Data: A Revolution that Will Transform How We Live, Work and Think. Houghton Mifflin Harcourt Publishing Company, Boston (2013)
3. Yu, W., Mao, M., Wang, B., Liu, X.: Implementation evaluation of beijing urban master plan based on subway transit smart card data. In: 22nd International Conference on Geoinformatics (2014)
4. Chen, C., Ma, J., Susilo, Y., Liu, Y., Wang, M.: The promises of big data and small data for travel behavior (aka human mobility) analysis. Transp. Res. Part C Emerg. Tech. **68**, 285–299 (2016)
5. Ferreira, N., Poco, J., Vo, H.T., Freire, J., Silva, C.T.: Visual exploration of big spatio-temporal urban data: a study of New York city taxi trips. IEEE Trans. Visual Comput. Graph. **19**(12), 2149–2158 (2013)
6. Chen, W., Guo, F., Wang, F.Y.: A survey of traffic data visualization. IEEE Trans. Intell. Transp. Syst. **16**(6), 2970–2984 (2015)

7. Kaeoruean, K., Phithakkitnukoon, S., Demissie, M.G., Kattan, L., Ratti, C.: Analysis of demand-supply gaps in public transit systems based on census and GTFS data: a case study of Calgary, Canada. Publ. Transp. **12**(3), 483–516 (2020)
8. Batsas, M.: Public transport authorities and COVID-19, impact and response to a pandemic. International Association of Public Transport, Australia/New Zealand. https://australia-newzealand.uitp.org/sites/default/files/V1_COVID-19%20impacts_AJ_v03.pdf. Accessed 17 May 17 2020
9. Gkiotsalitis, K., Cats, O.: Public transport planning adaption under the COVID-19 pandemic crisis: literature review of research needs and directions. Transp. Rev. **41**(3), 374–392 (2021)
10. Prommaharaj, P., Phithakkitnukoon, S., Demissie, M.G., Kattan, L., Ratti, C.: Visualizing public transit system operation with GTFS data: a case study of Calgary, Canada. Heliyon **6**(4), e03729 (2020)
11. Giraud, A., Trépanier, M., Morency, C., Légaré, F.: Data fusion of APC, smart card and GTFS to visualize public transit use. Centre interuniversitaire de recherche sur les réseaux d'entreprise Ldots, Technical report, CIRRELT (2016)
12. Kurkcu, A., Miranda, F., Ozbay, K., Silva, C.T.: Data visualization tool for monitoring transit operation and performance. In: 2017 5th IEEE International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS), pp. 598–603. IEEE (2017)
13. Palomo, C., Guo, Z., Silva, C.T., Freire, J.: Visually exploring transportation schedules. IEEE Trans. Visual Comput. Graph. **22**(1), 170–179 (2015)
14. Anwar, A., Odoni, A., Toh, N.: Busviz: big data for bus fleets. Transp. Res. Rec. **2544**(1), 102–109 (2016)
15. BIP: BIP Piedmont. http://www.5t.torino.it/progetti/bip-piemonte/. Accessed 19 May 2021
16. BIPEX: BIPEX Piedmont. http://www.5t.torino.it/progetti/bipex/. Accessed 26 Oct 2021
17. 5T: 5T. http://www.5t.torino.it/. Accessed 22 May 2021
18. NeTEx: Netex. http://www.netex-cen.eu/. Accessed 26 Oct 2021
19. Transmodel: Transmodel. http://www.transmodel-cen.eu/. Accessed 26 Oct 2021
20. Arnone, M., Delmastro, T., Giacosa, G., Paoletti, M., Villata, P.: The potential of e-ticketing for public transport planning: the piedmont region case study. Transp. Res. Proc. **18**, 3–10 (2016)
21. Gettleman, J., Schultz, K.: Modi orders 3-week total lockdown for all 1.3 billion Indians. New York Times **24** (2020)
22. Fong, W.K., Matsumoto, H., Lun, Y.F.: Application of system dynamics model as decision making tool in urban planning process toward stabilizing carbon dioxide emissions from cities. Build. Environ. **44**(7), 1528–1537 (2009)