Chapter 22
The Benefits of Open Data in Urban Traffic Network

Miroslav Vujic, Luka Dedic, Martina Tomicic Furjan, and Igor Pihir

22.1 Introduction

In the last decade, the demand for big data collected by public services and used by third parties has significantly increased. Big data is data on a scale or of a complexity that makes it challenging to use [1]. Key characteristics of "big data" are high volume (1 GB–1 mil GB), great variety (data sets that are sometimes inaccurate and unreliable), and time variability [2]. The transport sector has always collected and analyzed large quantities of data, such as data from timetables, traffic news, and air schedules [3]. However, using that data as the third party for providing transport services of added value such as routing vehicles or intermodal rout guiding within public transport services was often expensive and unprofitable. To expand the usage and re-usage of big data, the first requirement is that this kind of data is open for use. Opening nonpersonal data in the public sector are seen to improve public services, increase the accountability of government, and generate economic benefit through the reuse of public data to develop new products [4, 5]. Among mentioned, the Public Sector Information Directive (PSI) [6] is one of the reasons why country members of the EU are introducing "open by default" principles in their databases. "Open by default means that, when creating a new governmental data set, publishing it under an open license becomes the default process" [7].

M. Vujic (✉)
Faculty of Transport and Traffic Sciences, University of Zagreb, ZUK Borongaj, Zagreb, Croatia
e-mail: miroslav.vujic@fpz.hr

L. Dedic
PROMETIS d.o.o, Zagreb, Croatia

M. T. Furjan · I. Pihir
Faculty of Organization and Informatics, University of Zagreb, Zagreb, Croatia

L. Knapčiková et al. (eds.), *5th EAI International Conference on Management of Manufacturing Systems*, EAI/Springer Innovations in Communication and Computing, https://doi.org/10.1007/978-3-030-67241-6_22

In this paper, we analyzed the amount and type of open data provided by the government services that can be used for the development of applications for optimizing urban traffic network. It is conceived as an overview of so far the progress of unlocking traffic data in Croatia, with emphasis on potential benefits on urban traffic network based on experience from the world.

22.2 Related Work

The authors [8] have presented research results on the open data project conducted in Ames, Iowa. The main goal of the project was to establish the open traffic data service, which will allow vendors and agencies to provide near real-time, proactive alerts to commercial drivers regarding traffic conditions along their routes. Authors stated that with enabled access to relevant traffic data and value-added data feeds, multiple benefits for external and internal users are achieved.

The qualitative research approach consisted of in-depth interviews with six market players and 27 governmental data owners in Flanders, a region in Belgium, which was presented by the authors [7]. The main goal of the research was the assessment of the needs for the reuse of data among commercial multimodal route planners. Besides confirming the moderate needs for reuse of data, the research resulted in identifying three possible constraints in the process of data reusing. It is stated that not every data set will or can be reused as there is a cost for adoption, all data need to be high in quality, and metadata (data about data) needs to be obtained for the data to be reused.

According to authors [9, 10], analysis of public transport data can help to understand transport users' journey patterns across the transport network, in terms of where users travel, what mode they choose, how frequently they travel, and how reliable their journeys are. Moreover, results presented by authors [11], show that data on public transport users can inform transport agencies and operators how people of different demographic groups use a service and with that provide a better insight in customer's needs.

Paper presented by authors [12] included a demo version of Linked Open Data used for intermodal route planning. Throughout an affordable publishing method called Linked Connections, data on transit stops, and time of departure are merged, providing the execution of the route planning algorithm. In this paper route planning algorithm is a basic Connection Scan Algorithm—CSA [13] that solves the earliest arrival problem by organizing connections (transit stops) as one single array, which then scans to compute journeys to all stops of the network.

According to [14], a part of the UK government's strategy to improve transparency and encourage economic growth is "opening up" data in transport by making it more widely available and linking it with data from other sectors. Moreover, the UK government has established a Transport Systems Catapult, overseen by the Technology Strategy Board (TSB), which has specific objectives to encourage the analysis of big data.

The same principle had Austroads, the peak organization of Australasian road transport and traffic agencies. In the report presented by the authors [15], it is stated that Governments can assist industry, researchers, and the public to develop innovative solutions to transport problems by providing open access to transport data. Two goals of the research were to identify key issues in the usage of open data by connected automated vehicles (CAV) and to provide a recommendation for next steps in the usage of that kind of data.

From related work, it can be said that usage and re-usage of open data can provide many benefits. By connecting open data, intermodal trips can be generated. Moreover, analyzed data on time and distance on the passenger's trip can help in providing better public service. Opening data by default can create added value through third parties that are developing applications for traffic optimizations and thus increase its usage.

Stated benefits influence traffic congestion reduction and overall optimization of the urban and suburban traffic network. However, to ensure the use of open data, the decision must come from the government and be legally constituted through the decision legislation.

22.3 Open Data and Big Data in the Republic of Croatia

Open data in the Republic of Croatia is summarized on the central state portal called the Open Data Portal of the Republic of Croatia [16]. The organization for the collection of statistical data in the Republic of Croatia is the Croatian Bureau of Statistics [17]. For many years, the Croatian Bureau of Statistics collects transport statistics that include data used for the strategic traffic development of the country, such as monitoring the border and border traffic of people and goods in the road, rail, water, and maritime transport. The collected statistics are used by competent services to develop a strategy for expanding transport infrastructure. Additionally, to enhance strategic development in terms of investment in new infrastructure, data on the amount and type of traffic can help policymakers to maintain existing infrastructure.

Open data in urban environments is used to expand intermodality as the future of passenger transport. Passenger transport can set several requirements on the system that need to be met to bring the passenger within the given limits from the starting point to the destination. Providing alternatives such as shorter travel time, route selection, or mode of transport makes the transport system more attractive to use, which results in better system utilization. For such alternatives to be possible, it is necessary to consolidate a large set of data on the transport of passengers and their needs, but also the capacity and utilization of transport infrastructure and the superstructure by which the mentioned transport processes are performed. Data aggregation is possible through the development of platforms that enable the connection of open data and the use of big data technologies. The need for such technologies has grown exponentially in recent years, which has led decision-makers to encourage openness and availability of data through legal frameworks. Increasing the quality of legal frameworks such as open by default policies or increasing the degree of openness of data increases both the usability and usefulness of open data.

22.3.1 Legal and Legislative Aspect

Although earlier and valid strategic documents such as the Croatian Strategy for the Development of Public Administration for the period from 2015 to 2020 [18], e-Croatia 2020 Strategy [19], Open Government Partnership Action Plan [20], Anti-Corruption Strategy 2015–2020 [21] included measures and activities dedicated to creating conditions for opening data and opening individual data sets, until 2018 the Republic of Croatia was the only EU Member-State that did not adopt full access to open data through a unified policy at the national level. The lack of such a document has resulted in insufficiently focused development of open data, insufficiently developed an awareness of public authorities about the need and ways of opening data, with the possibility of consequently lagging regarding other European countries.

Recognizing this problem, but also the importance of data available to public sector bodies and the potential of open data, the Government of the Republic of Croatia adopted the Open Data Policy in July 2018, a document whose purpose is to build and permanently develop an environment conducive to creating new social and economic value using public sector data [22]. The basic principles of open data are contained in the Open Data Charter [23] adopted by the G8 in June 2013, which was later acceded to by many other countries, as well as the European Union, and was basically taken over by the Croatian Open Data Policy. These principles are:

- Open by default—Given that the use of data requires that public sector data are available and easily accessible, the policy stipulates that open data is generally made public and in accordance with set standards, taking into account the fact that in exceptional cases for justified reasons, certain data cannot be published. Also, policy measures will encourage the publication of all state data on the Open Data Portal.
- Timeliness and completeness—Open data policy measures will encourage the publication of as many data sets as possible and ensure that published data are timely, comprehensive, and accurate. Also, for users to have the best possible insight into the content of the data, data sets should be described in metadata, written in clear, and simple language. Finally, the policy will ensure the up-to-datedness of published data sets and enable feedback on the content of published data sets through the Open Data Portal and other appropriate mechanisms. To increase the value of open data, policy measures will support the publication of data without restrictions or fees for the reuse of information, for noncommercial or commercial purposes. Also, special attention will be paid to ensuring the

publication of data in machine-readable formats, as well as the publication of data through the application program interface (API), where appropriate, to ensure easy access to updated data.

- Accessibility and usability—Policy measures will seek to make data available to the widest range of users in the widest range of purposes in appropriate (open) formats to ensure that files can be easily retrieved, downloaded, indexed, and searched using all common applications. Where free access cannot be offered, benefits will be promoted and free access to data will be encouraged.
- Comparability and interoperability—Policy measures will ensure the sharing of experiences and exchange of open data at the national level, encourage cooperation of public sector bodies with civil society organizations and citizens, and enable the interested public to provide feedback on the most important data they want to be made public.

In the Republic of Croatia, the reuse of information and the opening of data for use for commercial and noncommercial purposes has been an obligation of public authorities since March 2013, when the Right to Access Information Act (OG 25/13) entered into force. The Act was amended in 2015 by transposing the amendments to the 2013 PSI Directive [6], which significantly expanded the obligation to publish open data as well as provide reuse information on request.

In accordance with the requirements of the PSI Directive, the Act prescribes the publication of data sets in machine-readable and open form, in accordance with the open standard, establishes the obligation to publish lists of databases and data sets, provides for the establishment of the Open Data Portal, costs for users, established possibilities of restricting use, as well as the basis for prescribing conditions and costs and granting exclusive rights, a regulated procedure for exercising reuse on request and protection of rights and supervision over the implementation of the Act by an independent body [22].

22.3.2 Open Data

The Croatian Bureau of Statistics [17] is an organization that conducts statistics at the state level. Data in the field of transport contain data on rail transport, road infrastructure, road transport of passengers and goods, border traffic, urban and suburban transport, pipeline transport, maritime and coastal transport, traffic at airports and seaports, drivers, registered vehicles, and road accidents. Although statistics are kept extensively at the state level and data are open and available for the period of last 10 years, the level of data openness (at least transport data) is low. All available and open data are published under the Creative Commons Attribution [24] license, and the openness of the data is level 2, which would correspond to structured data in a closed format (.xls, .xlsx). For easier comparison, the levels of data openness according to the degree of openness are shown in Fig. 22.1.

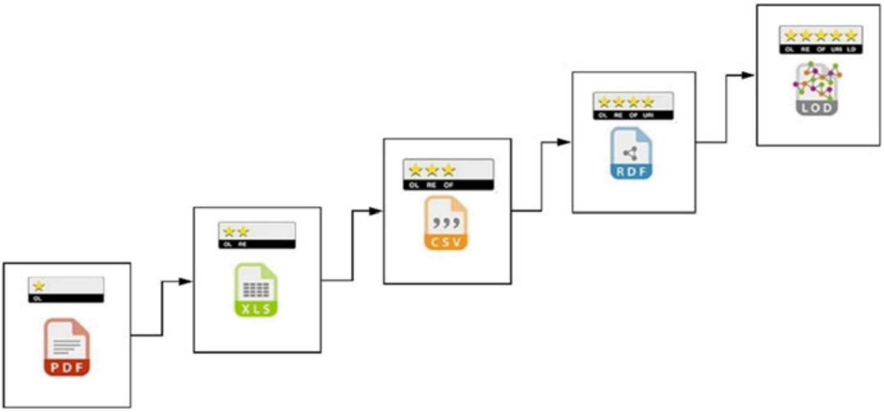

Fig. 22.1 The 5-Star Open data Model [25]

Data available on the Open Data Portal are not linked to other data, nor are detailed with metadata (hour, day, location of data). Open data regarding the traffic on urban roads are collected at the state level and are divided into the following groups

- Transport equipment in road urban and suburban transport
- Kilometers traveled, and passengers carried in road urban and suburban transport
- Vehicle fleet and lines in road urban and suburban transport
- Employees in road urban and suburban transport

The layout of the open data is shown in the original form in Table 22.1

Due to low openness, low detail of data (data available only in summarized form), and their up-to-datedness (data published on an annual or quarterly basis), the usability of data for the improvement of transport processes of passenger or goods at a level lower than the strategic almost nonexciting.

For the development of the systems that use real-time data, it is necessary to use sources that are connected and have updated, usable, and accessible metadata. In the field of transport in the Republic of Croatia, there is an insignificant number of projects aimed at expanding the usability of open data in traffic and transport.

To boost the supply and use of open government data in Croatia and beyond, the EU launched the H2020 project “Twinning Open Data Operational” (TODO). With the support of key organizations in the Croatian open data ecosystem and esteemed national and international experts, TODO will enhance the research capacity and research excellence in open data research of University of Zagreb (UNIZG) and its staff through the partnership with two leading universities in the open data domain, University of the Aegean (UAEGEAN) and Delft University of Technology (TUDELFT).

Through activities directed at training, knowledge exchange, collaboration, outreach, and long-term sustainability, TODO will develop and implement an inter-

Table 22.1 The open data presentation from Croatian Bureau of Statistics

Croatian Bureau of Statistics		2010	2011	2012	2013	2014	2015	2016	2017	2018
Trams										
Lead cars	No.	325	317	313	306	306	303	303	303	295
Pass. places	No.	51,653	50,667	50,339	49,615	49,615	49,306	49,306	49,306	48,479
Trailers	No.	82	77	70	63	63	58	58	58	51
Pass. places	No.	9253	8683	7900	7117	7117	6547	6547	6547	5759
Buses										
	No.	1240	1196	1196	1202	1180	1191	1247	1262	1264
Pass. places	No.	119,269	114,669	114,716	113,104	108,530	108,418	112,973	113,347	112,495
Lines										
Trams	No.	21	21	21	21	21	21	21	21	21
Length	Km	232	232	232	232	232	232	232	232	228
Bus	No.	505	518	479	488	484	481	504	507	514
Length	Km	9282	9525	8281	8307	8129	7956	8718	8881	9822

disciplinary multi-domain open data research approach to increase the maturity of the concept and impact of the open data ecosystem in Croatia and beyond. Through research, TODO will explore the gaps between supply and demand for open data and build an understanding of an open data ecosystem in Croatia. Participation in existing international networks, together with the newly built scientific excellence and innovation capacity and increased mobility of UNIZG staff, will stimulate success in attracting research and education funding and the establishment of a sustainable academic open data research ecosystem in the UNIZG [26].

22.3.3 Big Data

Big data technologies represent great opportunities because they help develop new creative products and services, such as mobile apps or business intelligence products. Such technologies can boost job growth and development, but also improve the quality of life of Europeans. Big data technologies are used to collect, process, and analyze large amounts of data. Data are diverse, structured, and unstructured, generated, and arrives at high speed and different intervals (often in real-time), making them very complex to analyze. The collected data can come from various devices or systems. Some of them are smartphones, smart bracelets or watches, smart refrigerators, ATMs, GPS systems, etc.

In recent years, the amount of data created, stored, and processed has grown exponentially. In the EU, the analysis of big data in 2018 was done by 12% of

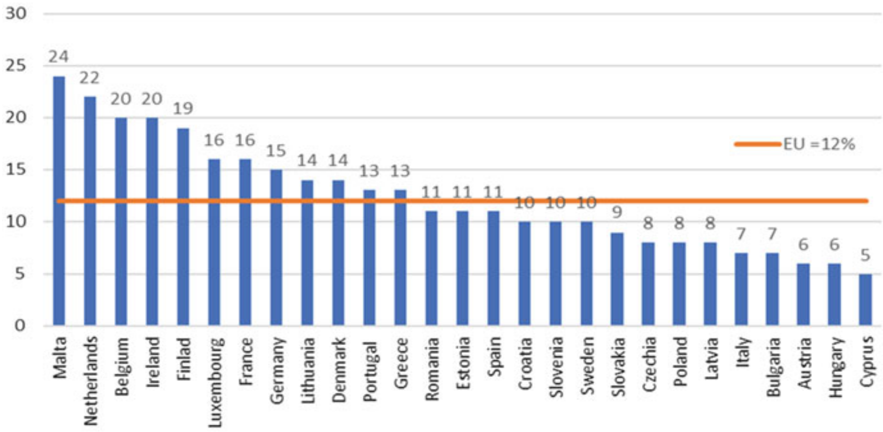

Chart 22.1 Use of big data analysis by enterprises in the EU Member-States, 2018 (% of enterprises) [28]

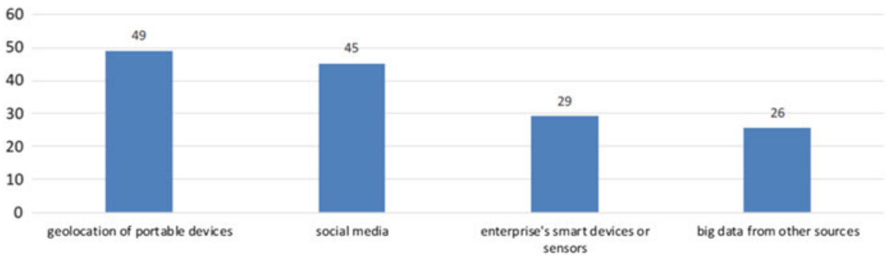

Chart 22.2 Use of big data analysis in the EU by data source, 2018 (% of enterprises analyzing big data) [28]

companies with at least ten employees. Big data analyzes were mostly performed by large companies (33%) or medium-sized companies (19%), with 8% of the analyses performed by employees in companies, while 5% of companies hired an external company as an associate in big data collection [27, 28]. Chart 22.1 shows the ranking of the Republic of Croatia in the use of Big Data in correlation to other EU member-states and the EU average.

According to Eurostat statistics [28] in Republic of Croatia, nearly 10% of companies used the big data analysis, which is below the average of the EU (12%). Chart 22.2 shows the percentage of the big data collected sorted by the data source.

Based on data provided in Chart 22.2. It can be concluded that almost half of the companies (49%) that apply big data analytics analyze geolocation data collected via mobile devices. This is followed by an analysis of data collected through social networks (45%), and less than one-third of companies analyzed data collected with their own smart devices or sensors (29%). Data from other sources are analyzed by 26% of companies [28].

Although the use of big data technology in the country is below the EU average, the Republic of Croatia often and actively participates in projects implemented among EU members. Some of them were launched to maintain ecosystems such as AFTERLIFE, a project that develops technologies that filter and convert wastewater from the food and beverage industry into bioplastics and food additives, while others focus on medical developments such as EUTHYROID, a project that has enabled the establishment of an infrastructure for data collection and analysis to monitor iodine deficiency across Europe [29].

To our knowledge, there are currently no European projects in the Republic of Croatia (except TODO) that could or will significantly affect the development of big data technologies and applications based on such technologies and are related to transport and traffic processes. As already mentioned, the use of big data is considered the future of the development of traffic processes. Therefore, it is necessary to enable the greatest possible usability and transparency (accuracy and up-to-datedness of data) that can contribute to the development of transportation systems.

22.4 The Benefits of Using Open Data and Big Data in Traffic

The European Union considers open data to be an important resource for the development of a digital European society and market, contributing to the creation of competitive advantages for innovation, job development, and the improvement of services for citizens and the economy. According to EU estimates expressed in the document Creating Value Through Open Data [30], the Open data market in 2020 is at least 75.7 billion euros and will potentially open the possibility of new employment for 25,000 people. For the Republic of Croatia, the EU estimates the creation of around 1000 new jobs by the end of 2020, and the market value is estimated at 186 million € [22].

The advantages of using big data technologies have been recognized at the EU level, which has launched several projects related to the use of open data and big data technologies in transport. Although examples of the use of big data technologies in the field of transport in the world are limited and most approaches still use traditional data sources such as inductive loops, travel surveys, etc. according to [31] it is possible to single out applications of big data technologies in transport presented in three main categories:

- Urban planning
- Transportation operation
- Traffic safety.

Categories are analyzed separately, but their symbiosis, i.e., the use of different data clusters can create a system that uses big data technologies in order to speed up the transport of passengers and goods, while considering the cost and safety of the transportation process.

22.4.1 Urban Planning

The use of big data technologies for urban planning is possible through the analysis of travel demand and the prediction of movement patterns using location data from mobile phones or call details records.

The authors [32] in their research showed that the estimated origin–destination flows derived from the aggregation of the trips from millions of individual mobile phone users in the Boston Metropolitan area correlate well with the US Census estimations. Moreover, compared to traditional census survey data, presented estimations allow capturing weekday and weekend patterns as well as seasonal variations. These features could make methods for origin–destination flow estimation based on opportunistically collected mobile phone location data a critical component for transportation management and emergency response.

A similar study was conducted by the authors [33] using mobile phone call detail records (CDR) and limited traffic counts methodology to develop OD matrices. CDR, which consists of time-stamped tower locations with caller IDs, are analyzed first and trips occurring within certain time windows are used to generate tower-to-tower transient OD matrices for different time periods. An optimization-based approach, in conjunction with a microscopic traffic simulation platform, is used to determine the scaling factors that result in best matches with the observed traffic counts. The methodology is demonstrated using CDR from 2.87 million users of Dhaka, Bangladesh, over a month, and traffic counts from 13 key locations over 3 days of that month.

Based on the methodologies used from the examples shown, it is possible to plan a transport network, predict the time spent in the network, or design a route selection model for cyclists. Transport network planning using the mentioned methodologies is explained on the example of research [34] where the authors presented the possibility of designing a transport network with deriving frequent patterns of movements from anonymized mobile phone location data and merging them to generate candidate route designs. Additional routines for optimal route selection and service frequency settings are then employed to select a network configuration made up of routes that maximize systemwide traveler utility. Using data from half a million mobile phone users in Abidjan from the telco operator Orange provide resource-neutral system improvement of 27% in terms of end-user journey times is demonstrated.

Prediction of time spent in the network is shown in the study [35] where the authors show a methodology based on data collected from taxicabs in New York City. Taxicabs equipped with global positioning system (GPS) devices provide the locations of origins and destinations, travel times, fares, and other information on taxi trips. The new model infers the possible paths for each trip and then estimates the link travel times by minimizing the error between the expected path travel times and the observed path travel times. The model was evaluated using a test network from Midtown Manhattan. Results indicate that the proposed method can efficiently estimate hourly average link travel times.

The design of the cyclist route selection model is presented in the research [36]. Based on a detailed GPS-data analysis for the bike-sharing system, mobility patterns of the usage were identified. Depending on different factors like weather conditions, time of the day, and holidays/weekends, a demand model was created to obtain an optimal distribution of bikes within the operating area. At the end of this paper, an application of an operator-based relocation strategy is given. With partial relocation of the fleet, it is ensured that the demand for bikes is optimally satisfied regarding time and space.

From the presented examples, it is possible to conclude that the use of open data and big data can greatly contribute to the optimization of traffic processes. Predicting network time for a large fleet of vehicles, allocating superstructural resources based on estimated use by location or time, and designing transportation networks that minimize travel times, travel predictions, and OD matrices reduce operating costs and potentially increase the utilization of existing infrastructure and superstructure.

22.4.2 Transportation Operation

Transport-related services based on the use of big data technologies are focused on providing decision support services in traffic processes or are support for Advanced Traveler Information Systems (ATIS). The most well-known researchers in this area are focused on travel time prediction, traffic incident and anomaly detection, anticipatory vehicle routing, and predicting bus bunching in the network using smart card data [31].

An example of travel time prediction based on big data technologies is presented in a study [37], where the authors developed a particle filter approach for real-time short- to medium-term travel time prediction using real-time and historical data. A 95-mile freeway stretch from Richmond to Virginia Beach along I-64 and I-264 is used to test the proposed algorithm. The confidence boundaries of the predicted travel times demonstrate that the proposed approach provides good accuracy in predicting travel time reliability. Lastly, fast computation time and online processing ensure the method can be used in real-time applications.

Recognizing incidents and detecting traffic anomalies can greatly help in transportation operations. The authors [38] presented research related to incident detection using an analysis of published data on social media. A comprehensive approach has been developed to extract and analyze real-time traffic-related twitter data for incident management purposes. The developed approach was implemented at the District of Columbia Department of Transportation for incident management. Data validation has been conducted against the real-world incident database. The preliminary results of the analysis have shown that social media data are promising for early incident detection and can be used as a supplemental source for incident data collection.

Advanced vehicle guidance systems use real-time traffic information to route traffic and to avoid congestion. Unfortunately, these systems can only react upon the presence of traffic jams and not to prevent the creation of unnecessary congestion. Authors [39], in their research, presented the anticipatory vehicle routing approach that allows directing vehicle routing by accounting for traffic forecast information. It is based on delegate multiagent systems, i.e., an environment-centric coordination mechanism that is, in part, inspired by ant behavior. Antlike agents explore the environment on behalf of vehicles and detect a congestion forecast, allowing vehicles to reroute. The approach is evaluated in a simulation of a real-world traffic environment, and results indicate a considerable performance gain compared with the most advanced strategy under test, i.e., a traffic–message–channel-based routing strategy.

The public transport timetable from a theoretical point of view often does not correspond to one in reality. The main reason for the frequent delays of tram or bus lines is increased congestion and the inability to maintain the required travel speed, especially in road segments where public transport corridors are not separated from private transport. Such situations often lead to the grouping of public transport vehicles, which creates vehicle redundancy and lowers PT efficiency. To address the problem of vehicle redundancy, authors [40] in their research presented a predictive framework to capture the stop-level headway irregularity based on transit smart card data. Historical headway, passenger demands, and travel time are utilized to model the headway fluctuation at the following stops. An empirical experiment with two bus routes in Beijing is conducted to demonstrate the effectiveness of the proposed approach. The predictive method can successfully identify more than 95% of bus bunching occurrences in comparison with other well-established prediction algorithms. Moreover, the detection accuracy does not significantly deteriorate as the prediction lead time increases. Instead of regularizing the headways at all costs by adopting certain correction actions, the proposed framework can provide timely and accurate information for potential bus bunching prevention and inform passengers when the next bus will arrive. This feature could greatly increase transit ridership and reduce operating costs for transit authorities.

In addition to the mentioned research, the use of open data and big data technologies is possible for developing a dynamic congestion charging system that can be used to determine the price classes of congestion charging depending on the traffic load of the toll zone. Also, the same technologies can be used to develop a demand-responsive parking pricing system that could inform drivers about the availability of parking spaces near the destination and thus reduce the generation of traffic caused by searching for available parking spaces.

22.4.3 Traffic Safety

Traffic safety is one of the important segments, which is considered when designing all traffic processes. Technological evolution and advances in vehicle sensors have also increased the level of safety provided to the driver. Research related to the use of big data technology in the field of traffic safety is addressed to exploring the critical situations arising from the design of infrastructure, understanding behaviors of drivers, and developing models for traffic sign detection or crash prediction based on real-time data mining.

Exploring the critical situations arising from the design of the infrastructure can be used for assessing road safety, behavioral studies, and traffic flow model validation. Researchers [41] in their paper presented a practical framework for the implementation of an automated, high-resolution, video-based traffic-analysis system. The system collects large amounts of microscopic traffic flow data from ordinary traffic using CCTV and consumer-grade video cameras and provides the tools for conducting basic traffic flow analysis as well as more advanced, proactive safety and behavior studies. In addition to providing a rich set of behavioral data about time-to-collision and gap times at nearly 40 roundabout weaving zones, some data validation is performed using the standard measure of tracking accuracy with results in the 85–95% range.

Driving styles can be broadly characterized as calm or volatile, with significant implications for traffic safety, energy consumption, and emissions [42]. Understanding the behaviors of drivers can contribute to increasing traffic safety, decreasing energy consumption, and exhaust emissions. Authors [42], in their research, investigated how to quantify the extent of calm or volatile driving and explore the correlations between these two types of drivers. A fundamental understanding of instantaneous driving behavior is developed by categorizing vehicular jerk reversals (acceleration followed by deceleration), jerk enhancements (increasing accelerations or decelerations), and jerk mitigations (decreasing accelerations or decelerations). Volatility in driving decisions, captured by jerky movements, is quantified using data collected in Atlanta, GA, during 2011. The database contains 51,370 trips and their associated second-by-second speed data, totaling 36 million seconds. Rigorous statistical models explore correlates of volatility that include socioeconomic variables, travel context variables, and vehicle types. The study contributes by proposing a framework that is based on defining instantaneous driving decisions in a quantifiable way using big data generated by in-vehicle GPS devices and behavioral surveys.

Driving behavior models can be developed using real-time data mining technologies. Authors [43] in their research presented an improved framework for real-time crash prediction models. The model has been constructed using high-resolution detector data collected from Shibuya 3 and Shinjuku 4 Expressways under the jurisdiction of Tokyo Metropolitan Expressway Company Limited, Japan. It has been specifically built for the basic freeway segments, and it predicts the chance of formation of a hazardous traffic condition within the next 4–9 min for a 250-m-long road section. The performance evaluation results reflect that at an average threshold value, the model can successfully classify 66% of the future crashes with a false alarm rate of less than 20%. The data used to evaluate the model are provided by Tokyo Metropolitan Expressway, and it contain detector data (data of speed, vehicle count, occupancy, and number of heavy vehicles per lane for every 8 ms round the clock for 24 ha day, 365 days a year) and corresponding crash data.

22.5 The Benefits of Using Open Data and Big Data in Traffic

Until June 2018, the Republic of Croatia was the only EU Member-State that did not adopt full access to open data through a unified policy at the national level. Open data in the Republic of Croatia are summarized on the central state portal called the Open Data Portal of the Republic of Croatia. Data in the field of transport contain data on rail transport, road infrastructure, road transport of passengers and goods, border traffic, urban and suburban transport, pipeline transport, maritime and coastal transport, traffic at airports and seaports, drivers, registered vehicles, and road accidents. Statistics are kept extensively at the state level, and data are open and available for the period of the last 10 years, but the level of data openness is low. All available and open data are published under the Creative Commons Attribution license, and the openness of the data is at level 2. Due to low openness, low detail of data (data available only in summarized form) and their up-to-datedness (data published on an annual or quarterly basis), the usability of data for the improvement of transport processes of passenger or goods at a level lower than the strategic is almost nonexciting. When developing systems that use real-time data, it is necessary to use sources that are connected and have updated, usable, and accessible metadata.

In the Republic of Croatia, nearly 10% of companies used the big data analysis, which is below the average of the EU (12%). For the Republic of Croatia, the EU estimates the creation of around 1000 new jobs by the end of 2020 related to open data and big data technologies, and the market value of 186 million €. The advantages of using big data technologies have been recognized at the EU level, which has launched several projects related to the use of open data and big data technologies in transport. On the examples of the use of big data technologies in the field of transport in the world, urban planning, transportation operation, and traffic safety can be singled out as three main categories.

Based on the methodologies used from the examples shown, it is possible to plan a transport network, predict the time spent in the network, or design a route selection model for a specific mode of transport. Transport-related services based on the use of big data technologies are focused on providing decision support services in traffic processes or are support for Advanced Traveler Information Systems (ATIS). The benefits in this field can be found through travel time prediction, traffic incident and anomaly detection, anticipatory vehicle routing, and detection of PT vehicle bunching.

Research related to the use of big data technology in the field of traffic safety is addressed to exploring the critical situations arising from the design of infrastructure, understanding behaviors of drivers, and developing models for traffic sign detection or crash prediction based on real-time data mining.

As presented in this paper, it can be concluded that the Republic of Croatia still lags in the usage of open data and big data technologies. One of the main reasons is the low level of openness of available data collected on the state level. To encourage the expansion of usage of open data and big data technologies, decision-makers need to provide more open data that is connected, usable, free of charge, and describe in detail with metadata.

Acknowledgments This research is realized within the EU's Horizon 2020 research and innovation program under Grant Agreement Number 857592—Twinning Open Data Operational (TODO) and supported by the University of Zagreb Program Funds Support for scientific and artistic research (2020) through the project: "Innovative models and control strategies for intelligent mobility."

References

1. Big and Open Data in Transport, The parliamentary office of science and technology London (2014)
2. M. Schroeck, M.R. Shockley, J. Smart, D. Romero-Morales, P. Tufano, *Analytics: the real-world use of big data* (IBM Institute for Business Value, 2012)
3. L. Henry, Analytics and big data—Rail Public Transportation is a Leader, in *APTA 2013*, (Rail Conference, Philadelphia, 2013)
4. UK Government Page, https://www.gov.uk/government/publications/open-data-white-paper-unleashing-the-potential. Accessed 13 Aug 2020
5. UK Government Page, https://www.gov.uk/government/publications/department-for-education-open-data-strategy-june-2012. Accessed 13 Aug 2020
6. Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the reuse of public sector information
7. P. Colpaert, M. Van Compernolle, N. Walravens, R. Verborgh, P. Mechant, E. Mannens, L. De Marez, P. Ballon, R. Van de Walle, Open transport data as an enabler for multimodal route planning, in *Proceedings of the 11th ITS European Congress*, (2016)
8. D.O.T. Iowa, *Traffic operations open data service* (Iowa State University, Institute for Transportation, Iowa, 2018)
9. M. Batty, Big data, smart cities and city planning. Dialog Human Geogr **3**(3), 274–270 (2013)
10. D. Uniman, Service reliability measurement framework using smart card data: application to the London underground. M.Sc. Thesis, Massachusetts Institute of Technology, USA (2009)
11. M. Ortega-Tong, Classification of London's public transport users using smart card data. M.Sc. Thesis, Massachusetts Institute of Technology, USA (2013)
12. P. Colpaert, *Publishing transport data for maximized reuse. Web book* (Ghent University, Ghent, 2017)
13. J. Dibbelt, T. Pajor, B. Strasser, D. Wagner, Intriguingly simple and fast transit routing, in *Experimental algorithms*, (Springer, New York, 2013), pp. 43–54
14. Five-Year Delivery Plan to March 2018. Transport systems catapult, UK (2013)
15. A. Somers, *Connected and automated vehicles (CAV) open data recommendations, research report AP-581-18* (Austroads Ltd, Sydney, 2016)
16. Croatian Open Data Portal Homepage, https://data.gov.hr/. Accessed 13 Aug 2020
17. Croatian Bureau of Statistics Homepage, https://www.dzs.hr/default_e.htm. Accessed 13 Aug 2020
18. L. Lepri, Croatian strategy for the development of public administration for the period from 2015 to 2020. INFuture2015: e-Institutions—Openness, accessibility, and preservation (2015), DOI: https://doi.org/10.17234/INFUTURE.2015.5.

19. e-Croatia 2020 Strategy, *Ministry of public administration* (Croatia, Zagreb, 2017)
20. Open Government Partnership Page, https://www.opengovpartnership.org/process/action-plan-cycle/. Accessed 13 Aug 2020
21. Open Government Partnership Page, https://www.opengovpartnership.org/members/croatia/commitments/HR0038/. Accessed 13 Aug 2020
22. A. Musa, D. Bevanić, D. Herak, L. Jandrijević, M. Kovačić, Z. Luša, N. Vrček, *Priručnik o ponovnoj uporabi informacija za tijela javne vlasti: Otvoreni podaci za sve [Handbook on re-use of information for public authorities: Open data for all]* (Information Commissioner, Republic of Croatia, Zagreb, 2018)
23. Open Data Charter Page, https://opendatacharter.net/principles/. Accessed 13 Aug 2020
24. Open Knowledge Foundation Page, http://opendefinition.org/licenses/. Accessed 13 Aug 2020
25. Five Star Open Data Homepage, http://5stardata.info/en/. Accessed 13 Aug 2020
26. European Commission Page, https://cordis.europa.eu/project/id/857592. Accessed 13 Aug 2020
27. Stranica Europske Komisije [European Commission page], https://ec.europa.eu/croatia/basic/everything_you_need_to_know_about_big_data_technology_hr. Accessed 13 Aug 2020
28. Eurostat Newsrelease, https://ec.europa.eu/eurostat/documents/2995521/9447642/9-13122018-BP-EN.pdf/731844ac-86ad-4095-b188-e03f9f713235. Accessed 13 Aug 2020
29. European Commission Page, https://ec.europa.eu/research/infocentre/article_en.cfm?&artid=49217&caller=other. Accessed 13 Aug 2020
30. W. Carrara, E.S. Chan, S. Ficsher, E. van Streenbergen, Creating value through open data. Eur. Union (2015). https://doi.org/10.2759/328101
31. S. Amini, I. Gerostathopoulos, C. Prehofer, Big data analytics architecture for real-time traffic control, in *5th IEEE International conference on models and technologies for intelligent transportation systems (MT-ITS)*, (2017). https://doi.org/10.1109/MTITS.2017.8005605
32. F. Calabrese, G. Di Lorenzo, L. Liu, C. Ratti, Estimating Origin—Destination flows using opportunistically collected mobile phone location data from one million users in Boston Metropolitan Area. IEEE Pervasive Comput. **99**, 36–44 (2011)
33. M.S. Iqbal, C.F. Choudhury, P. Wang, M.C. González, Development of origin-destination matrices using mobile phone call data. Transp. Res. Part C Emerg. Technol. **40**, 63–74 (2014)
34. F. Pinelli, R. Nair, F. Calabrese, M. Berlingerio, G. Di Lorenzo, M.L. Sbodio, Data-driven transit network design from mobile phone trajectories. IEEE Trans. Intell. Transp. Syst. **17**(6), 1724–1733 (2016)
35. X. Zhan, S. Hasan, S.V. Ukkusuri, C. Kamga, Urban link travel time estimation using large-scale taxi data with partial information. Transp. Res. Part C Emerg. Technol. **33**, 37–49 (2013)
36. S. Reiss, K. Bogenberger, GPS-data analysis of Munich's free-floating bike sharing system and application of an operator-based relocation strategy, in *Intelligent Transportation Systems (ITSC), IEEE 18th International Conference*, (2015), pp. 584–589
37. H. Chen, H.A. Rakha, Real-time travel time prediction using particle filtering with a non-explicit state-transition model. Transp. Res. Part C Emerg. Technol. **43**, 112–126 (2014)
38. K. Fu, R. Nune, J.X. Tao, Social media data analysis for traffic incident detection and management, in *Transportation Research Board 94th Annual Meeting no. 15–4022*, (2015)
39. R. Claes, T. Holvoet, D. Weyns, A decentralized approach for anticipatory vehicle routing using delegate multiagent systems. IEEE Trans. Intell. Transp. Syst. **12**(2), 364–373 (2011)
40. H. Yu, D. Chen, Z. Wu, X. Ma, Y. Wang, Headway-based bus bunching prediction using transit smart card data. Transp. Res. Part C Emerg. Technol. **72**, 45–59 (2016)
41. P. St-Aubin, N. Saunier, L. Miranda-Moreno, Large-scale automated proactive road safety analysis using video data. Transp. Res. Part C Emerg. Technol. **58**, 363–379 (2015)
42. X. Wang, A.J. Khattak, J. Liu, G. Masghati-Amoli, S. Son, What is the level of volatility in instantaneous driving decisions? Transp. Res. Part C Emerg. Technol. **58**, 413–427 (2015)
43. M. Hossain, Y. Muromachi, A Bayesian network based framework for real-time crash prediction on the basic freeway segments of urban expressways. Accid. Anal. Prev. **45**, 373–381 (2012)