Big Data-driven Multimodal Traffic Management: Trends and Challenges

Ivana Semanjski*, **, Sidharta Gautama*, **,
∗ Department of Industrial Systems Engineering and Product Design, Ghent University,
Gent, Belgium
∗∗ Industrial Systems Engineering (ISyE), Flanders Make
Ghent, Belgium
e-mail:ivana.semanjski@ugent.be; sidharta.gautama@ugent.be

***Abstract*—Availability of big data on moving objects is strongly affecting the way we view and manage mobility. Mobility in urban areas is becoming more and more complex, posing the challenge of efficient multimodal traffic management. So far, existing solutions were mainly car oriented, failing to capture the full complexity of the mobility within the city. In this paper, we present the potential of big-data driven solution for multimodal mobility management that capitalizes on the existing data availability and is realized through the "guarded" data architecture.**

Keywords-big data; traffic management; platform architecture, autonomous vehicles; digital twin.

I. INTRODUCTION

Availability of big data on moving objects is strongly affecting the way we view mobility. More and more initiatives strive to enhance the traditional data gathering processes for transportation planning and mobility studies by integrating big data. As these data are more detailed than the traditionally collected ones with higher spatial and temporal resolution, new possibilities are also emerging. One of such domain is data driven mobility and traffic management. In this paper, we tackle the potential of the big driven traffic management, particularly focusing on urban areas. The main challenge here is the above mentioned data integration which is, in all spheres of mobility, still at quite early stage coupled with the multimodality, as the existing mobility management solutions are still mainly car oriented. To tackle this, we focus on identifying current trends and existing challenges on a way to fully data driven traffic management. Here, we identify six main challenges ranging from data, across mobility to the business related challenges and five major future trends in this domain. In our research, we focus only on land transport modes that are integrated into the urban traffic management ecosystem and propose a "guarded" data platform architecture that could facilitate the data-driven mobility management in urban areas.

The paper is organized as follows: in the Section 2, we give detailed state of the art literature review on big data and current status of the big data integration into mobility management domain. Following this, in Section 3 we tackle the traffic management and the existing challenges and trends. In Section 4, we present the "guarded" big data platform architecture as a potential solution for the existing challenges in the traffic management domain. In the final section, we give conclusion remarks and present the future work on the topic.

II. STATE OF THE ART

A. Big data definitions

Among first, and probably best known definitions of big data, is the 3Vs definition [1]. 3Vs defines big data as the increase of data volume (data scale becomes increasingly big), variety between the data (data comes as structured, semi-structured and unstructured data) and velocity of data generation (data collection-processing chain needs to be promptly and timely conducted to maximally utilize the potential value of big data). As a big data based analytics developed over time, one of the key elements distinguishing among, simply, a large dataset and the big data turned to be ability to extract intelligence and useful insight from the data itself. Following this idea, a 4Vs [2][3] definition was developed. In the 4Vs definition, original velocity is divided into velocity and value, with the intention to acknowledge the extraction of value from data as one of main big data characteristics. Hence, the 4Vs stand for volume (great volume), variety (various types of data), velocity (swift data generation), and value (huge information value with distributed with a very low density among data).

B. Big data in mobility studies

Traditionally, data on mobility are collected via travel surveys, interviews and diaries. These cyclic data collection processes are repeated every one, two, but more often, five or even ten years formatting time series of the noted travel behaviors. Based on the data collected in such a manner transportation planners and decision makers have a systematic overview of the mobility patterns and can follow the main trends observed in the travel behavior. However, numerous studies [4][5] have shown that data collected in this manner deviated systematically from the actual travel behavior. Today, more and more studies explore the potential of big data when it comes to replacing the traditional data collection methods for mobility studies [6]–[9]. The most frequently explored big datasets can be categorized as:

- Pure Global Navigation Satellite Systems (GNSS) data
- Mobile network data
- Mobile sensed data.

Pure GNSS data are collected via GNSS devices. The GNSS devices record the timestamp and positioning data of the moving object. Mainly these devices were installed in

 ISBN: 978-1-61208-681-1

vehicles [10]. However, with the technology development, portable handheld GNSS devices were also used. The use of GNSS data for mobility studies mainly reflects in complementing the traditionally collected data on mobility behavior with more detailed activity detection [6][11]. However, existing studies highlight several challenges: discipline required to carry the portable device [12] and lack of the ability to have a full multimodal overview when the device is installed only in the vehicles [13][14].

Mobile network data are data collected by telecom operators as a byproduct of their daily activities. Most often, for mobility studies, these data include two elements. The first element corresponds to traces triggered by the serving network, so called network signalization data. The second one corresponds to traces triggered by the user interaction activity (e.g., call, SMS or data transfer activity), so called Call Detail Record (CDR). The geographical precision and time resolution of the CDR and network signalization data is somewhat lower than this is the case with the GNSS data. The reason for this lies in the fact that the network notes the base station (antenna) location, which covers the area where the user is located, and not the actual mobile phone device location. For this reason, the mobile network data seem to be a better candidate for longitudinal analysis of human mobility patterns [15]–[20].

Mobile sensed data are data collected from mobile phone sensors. This includes the above mentioned GNSS data as one of the possible sensors integrated into the mobile phone. However, the main characteristic of the mobile phone sensed data is the fusion of data sensed from multiple different sensors as accelerometer, microphone, gyroscope and others. One of the examples are positioning data that can be sourced from the GNSS sensor, Wi-Fi network location or mobile networks' base station location readings, or as the combination of any of the positioning sensors integrated into the mobile phone itself.

C. Big data in traffic management

When it comes to the use of big data in real-life applications, several topics seem to be of particular interest. The first one is the privacy related to the use of personal data and, when it comes to mobility, traces of the individuals' movements through the space [3][21]. The second one is the business related intelligences and value that one can gain from the big data oriented analytics [22]. The third one comes from the technical challenges that come with the big data integration as difficulties in regard to data capturing, data storage, data analysis and data visualization [23][24].

Concerning the integration of big data into the traffic management, literature shows that this area is still at quite early stage and that majority of papers in this domain are focused on car traffic and aspects as traffic flow prediction [25]. Here, methods mainly use so called shallow traffic prediction models and are still unsatisfying for many real-world applications. Several papers look at the deep learning methods, but still keep their focus on motorized transportation only [25]- [27]. Summary of these efforts can be found in comprehensive literature reviews that highlight advances and complexity in this domain [28][29].

In our paper, we focus on overcoming the limits of the existing solutions by considering the "guarded" data architecture that would be able to tackle the privacy issues through the guarded and data access control oriented solution. Furthermore, we focus on urban mobility from the end-users oriented point of view, considering not only car oriented traffic, but full spectrum of multimodal mobility that one can find in the cities of today.

III. TRAFFIC MANAGEMENT

Traffic management comprehends organization, arrangement, guidance and control of stationary and moving traffic [30]. The sub terms exist per different branches of transportation. Hence, the air traffic management is an aviation term encompassing all systems that assist aircraft to depart from an aerodrome, transit airspace, and land at a destination aerodrome. The sea traffic management defines a set of systems and procedures to guide and monitor sea traffic in a manner similar to air traffic management [31]. However, the mobility in urban areas has become so complex that the initial scope of the traffic management in urban areas, mainly focused on car transportation, has expanded to include a full spectrum of diverse transportation modes and numerus interactions between them. Hence, the traffic management in urban areas includes management of motorized vehicles, public transport, pedestrians, bicyclists and other flows and aims to provide safe, orderly and efficient movement of persons and goods, as well as efficient interaction between different transportation modes. For this reasons, the traffic management in urban areas is one of the most complex and challenging tasks when it comes to the traffic management.

A. Big data integration challenges

As mentioned above, current solutions for urban transportation management are mainly focused on motorized transportation, namely the cars. Systems monitor, record and/or guide car traffic flows by relying on sensing equipment installed on roads. Take-up of such system requires significant resources for city authorities. This investment is two-fold: (i) purchasing of the system that is vendor locked in and (ii) adjusting the, already existing equipment on roads, to the requirements of such vendor lock-in system. This often includes replacement of some costly elements as traffic lights and/or Variable Message Signs (VMS). It is a long-term commitment between the city authorities and the system provider that leaves the city with limited flexibility. On another hand, as cities strive to ensure higher quality of life in their area and well balanced transportation mode use that can ensure sustainability of the mobility system, such solutions only partially satisfies city needs as it mainly covers only one transportation mode. Hence, the challenges (C) towards a big data driven traffic management can be described as follows:

$$C = (MM, Open, B2B, P, DI, G)_{TM} \qquad (1)$$

where:

MM – stands for the multimodality as a challenge of having a full spectrum of different transportation modes present in the cities under one umbrella.

Open – stand for the challenge of integrating different open datasets that might come from the city authorities or different crowdsourcing campaigns and hence, can be structured with clear data quality standards and responsibilities or semi-structured / unstructured with the "best effort" data quality responsibilities.

B2B – stands for business related challenges, as such system would require business-to-business cooperation among potential market competitors. Hence, clear IPR (Intellectual Property Rights), licensing, data access and processing conditions need to be integrated.

P– stands for data privacy challenge as big data often include handling of moving object (e.g., private mobile phones or cars data), camera feeds etc.

DI –stands for data integration challenge. This challenge is mainly related to the lack of uniform data standards across different transportation modes and/or geographic regions.

G – stands for data governance and traffic management governance among different regions, as big data driven traffic management relies also on traffic flows that approach cities from regions outside of the city authority governance (e.g., regional roads etc.). Hence, synchronization of the traffic management strategies is needed among different areas.

B. Trends

The existing trends in the big data driven traffic management include:

$$T = (T, S, MMI, AV, BS)_{TM} \tag{2}$$

where:

T – stands for transition of the traffic management strategies across different transportation modes. As it was the case in history that, for example, the aviation has taken over some lessons learned and best practices from maritime traffic navigation, telecommunications have taken over the data flow management from air navigation, it is becoming inevitable to translate and test different traffic management strategies across multimodal data. One of the potential polygons for this might be developing digital twins of urban areas.

S – stands for different infrastructure access strategies and levels of services that can be developed through such integrated overview of urban mobility. For example, city authorities can implement strategy to reward sustainable mobility behavior by granting the access to the city events, public transport tickets or high speed lines for private cars.

MMI – stand for integration of multimodal multisource data on mobility.

AV- as it is expected that autonomous driving will heavily rely on data, big data driven traffic management plays an important role in the transition towards fused traffic flows (combination of autonomous and traditional vehicles) and purely autonomous vehicles traffic flows.

BS- stands for braking the silos and allowing the cities to have a flexible solution without vendor lock-ins.

IV. BIG DATA DRIVEN TRAFFIC MANAGEMENT PLATFORM

It is not a rare case that cities have started collecting data on multimodal mobility by themselves and providing them as an open data [32]–[34]. However, there is only a limited level of adoption of such data sources in the traffic management market. For this reason, it this paper, we tackle the potential of integrating big data into a transportation management framework that could provide cities with the flexible and multimodal overview.

Figure 1 shows an example of the "guarded" big data platform architecture for transport and mobility management. The architecture includes several layers, namely:

- Raw data layer – an entry point to any raw (low lever, unprocessed) data on mobility that are being integrated into the architecture
- Data quality check layer – guards the input to the architecture by checking the quality of the raw data (e.g., is data structure and data collection frequency of adequate level). Such layer has the possibility to let the data further into the architecture structure (with the notion of the initial data quality) or to generate the report on the data quality so that the detected issues can be corrected.
- Standardized and normalized data layer – preforms data cleaning and standardization so that, regardless of the initial data source, at this layer all the data on the same topic (e.g., speed data) have the same format and can be processed further in a same manner.
- Data quality check layer (can be seen as a sub layer of the Standardized and normalized data layer) – layer where all the data transformation achieved in the previous layer are noted. For example, to get the speed data of the same format maybe one needed to reduce the temporal resolution of the initial dataset, hence has lower the quality of the initial data set.
- Processed insights layer – layer where insights of the higher level are created. For example, speed data were used to produce the travel time map.
- Access control layer – guards the access to the processed data and insights.
- User insights layer – layer where different users can have an access to the data. These user groups can be city operators, authorities, citizens' initiatives or private citizens. They can all have an access to the subset of data (or all) depending on their access rights (guarded by previous layer). One example of this can be police officer who has authorization to view the camera feed data, whereas city operator or private citizen does not has this authorization due to the privacy legislation. Another example can be the

presence of the license protected data, hence one can purchase the license and access the data or not.

Next to the bottom (data quality check) and top (access control) guarding layers, the architecture is guarded horizontally on both ends. This is because it is foreseen that data can enter the architecture at different levels. For example, one can provide the platform with the already normalized and standardized data, but the platform horizontally checks the input quality of this data and collects or provides quality feedback to the data provider. The same goes for each level of the architecture. On the other end, based on the licensing conditions, one can collect the data from the platform at each level. For example, the developers might be granted the access to the standardized data on mobility that they can use to produce different services. This is done through the horizontal access control layer.

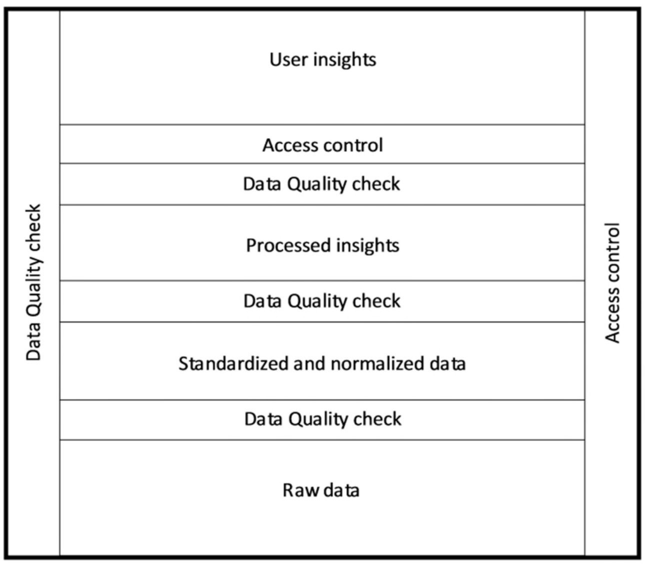

Figure 1. “Guarded” big data platform architecture

Figure 2 gives an example of the processing chain thought the “guarded” architecture. In this example, a public transportation company can integrate their data into the mobility platform that has the “guarded” architecture. These data can be integrated as one of the accepted standards for the architecture (e.g., European NetEx standard). Data can then be used by a local SME (Small Medium Enterprise) that develops a routing app. For commercial purposes, the use of the public transport data might be restricted by the license conditions. Hence, the SME gets the granted the access to the date if the license conditions are meet. The same dataset, might also be processed thought the architecture into the insights related to an public event organized by the city. Hence, the traffic management officer in charge of the event organization might have access to these data under different conditions and mighty view them in a different form. For example, these data can be an input to the alerting system that indicates if the crowdedness in affected area is above the certain threshold. In this context, the city event manager can request additional public transportation vehicle in the area or sent request to the police to regulate the smooth passing of the public transportation vehicles.

V. Conclusion And Future Work

With the current developments in the mobility and big data domains, it seems inevitable to move towards a data driven traffic management framework. Such a framework could allow cities and other mobility related authorities to capitalize on already existing data sources to improve mobility conditions in their areas. Furthermore, it is a reasonable step towards integration of strongly data driven, autonomous vehicles into the existing traffic flows, either as part of the mixed traffic flows or as purely AV flows. Although this is something that is not expected to happen in the near future, industry developments strongly point in this direction. Hence it is necessary to put cities in a position to think forward on how to integrate such developments and still efficiently manage overall multimodal mobility in their areas. In this paper, we have identified the main trends and challenges in this domain and suggested a “guarded” data platform solution that could meet these challenges. Furthermore, we recognized digital twins of urban areas as a forward thinking solution that could be used by urban authorities as a big data driven option for multimodal mobility management and a testing polygon for implementation of different strategies in real-life urban environment.

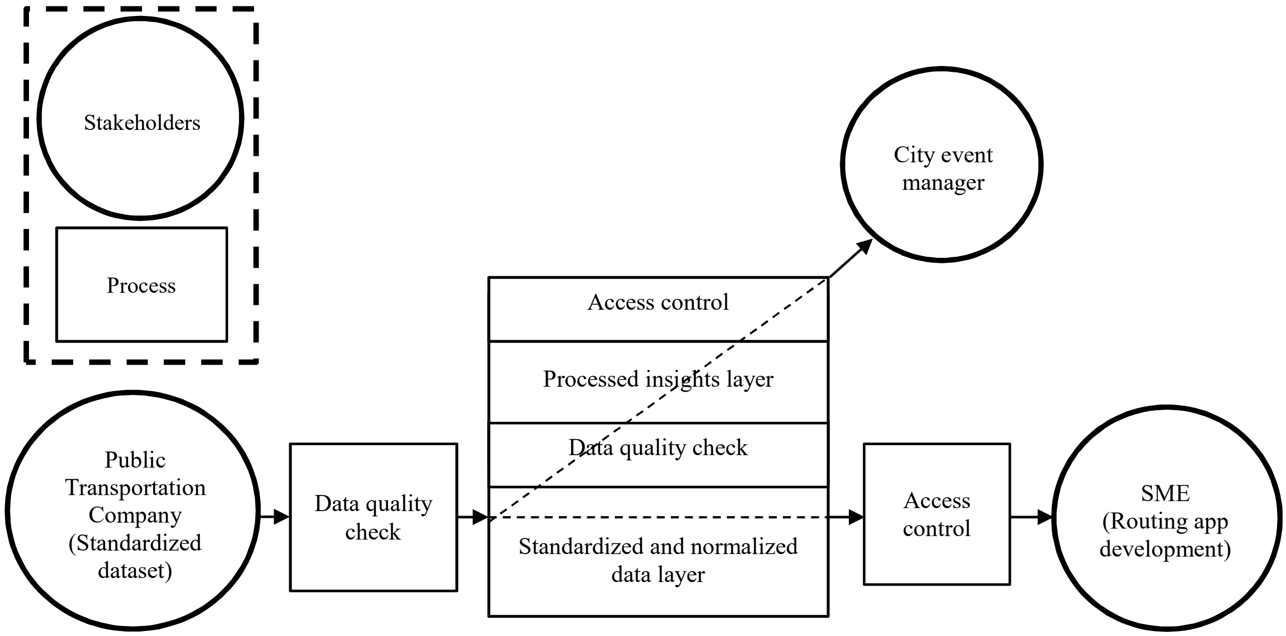

Figure 2. Process example

ACKNOWLEDGMENT

This research is funded by the EU Urban Innovation Actions (UIA) under the TMaaS (Traffic Management as a Service) project.

REFERENCES

[1] D. Laney, “3-d data management: controlling data volume, velocity and variety,” 2001.

[2] B. J. Gantz and D. Reinsel, “Extracting Value from Chaos State of the Universe : An Executive Summary,” IDC iView, no. June, pp. 1–12, 2011.

[3] M. Chen, S. Mao, and Y. Liu, “Big data: A survey,” Mob. Networks Appl., vol. 19, no. 2, pp. 171–209, 2014.

[4] P. R. Stopher and S. P. Greaves, “Household travel surveys: Where are we going?,” Transp. Res. Part A Policy Pract., vol. 41, no. 5, pp. 367–381, Jun. 2007.

[5] D. Ettema, H. Timmermans, and L. van Veghel, “Effects of Data Collection Methods in Travel and Activity Research,” Eur. Inst. Retail. Serv. Stud. Eindhoven, Netherlands, vol. 2000, no. i, 1996.

[6] J. Wolf, M. Loechl, J. Myers, and C. Arce, “Trip Rate Analysis in {GPS}-Enhanced Personal Travel Surveys,” Transp. Surv. Qual. Innov., vol. 2000, no. August 2001, pp. 483–498, 2003.

[7] S. Amin et al., “Mobile Century Using GPS Mobile Phones as Traffic Sensors : A Field Experiment,” 15th World Congr. Intell. Transp. Syst., pp. 8–11, 2008.

[8] H. Gong, C. Chen, E. Bialostozky, and C. T. Lawson, “A GPS/GIS method for travel mode detection in New York City,” Comput. Environ. Urban Syst., vol. 36, no. 2, pp. 131–139, Mar. 2012.

[9] I. Semanjski, S. Gautama, R. Ahas, and F. Witlox, “Spatial context mining approach for transport mode recognition from mobile sensed big data,” Comput. Environ. Urban Syst., vol. 66, 2017.

[10] S. Turner, W. Eisele, R. Benz, and D. Holdener, Travel Time Data Collection Handbook. Arlington: Texas Transportation Institute, 1998.

[11] T. Feng and H. J. P. Timmermans, “Detecting activity type from gps traces using spatial and temporal information,” Eur. J. Transp. Infrastruct. Res., vol. 15, no. 4, pp. 662–674, 2011.

[12] I. Semanjski and S. Gautama, “Sensing Human Activity for Smart Cities ’ Mobility Management”, Mobility Management, Rijeka, Croatia, InTech, 2016. .

[13] T. Feng and H. J. P. Timmermans, “Transportation mode recognition using GPS and accelerometer data,” Transp. Res. Part C Emerg. Technol., vol. 37, pp. 118–130, Dec. 2013.

[14] J. Wolf, S. Bricka, T. Ashby, and C. Gorugantua, “Advances in the Application of GPS to Household Travel Surveys,” Househ. Travel Surv., 2004.

[15] R. Ahas et al., “Feasibility Study on the Use of Mobile Positioning Data for Tourism Statistics - Consolidated Report,” no. 30501, p. 46, 2014.

[16] F. Calabrese, M. Diao, G. Di Lorenzo, J. Ferreira, and C. Ratti, “Understanding individual mobility patterns from urban sensing data: A mobile phone trace example,” Transp. Res. Part C Emerg. Technol., vol. 26, pp. 301–313, Jan. 2013.

[17] D. E. Seidl, P. Jankowski, and M.-H. Tsou, "Privacy and spatial pattern preservation in masked GPS trajectory data," Int. J. Geogr. Inf. Sci., vol. 30, no. 4, pp. 785–800, 2016.

[18] A. Vij and K. Shankari, "When is big data big enough? Implications of using GPS-based surveys for travel demand analysis," Transp. Res. Part C Emerg. Technol., vol. 56, pp. 446–462, Jul. 2015.

[19] O. Järv, R. Ahas, and F. Witlox, "Understanding monthly variability in human activity spaces: A twelve-month study using mobile phone call detail records," Transp. Res. Part C Emerg. Technol., vol. 38, pp. 122–135, Jan. 2014.

[20] A. Mishra, "Fundamentals of cellular network planning and optimisation-," Commun. Eng., p. 277, 2004.

[21] S. Sagiroglu and D. Sinanc, "Big Data : A Review," 2013 International Conference on Collaboration Technologies and Systems (CTS), IEEE, pp. 42–47, 2013.

[22] H. Chen, R. H. L. Chiang, and V. C. Storey, "Business Intelligence and Analytics: From Big Data to Big Impact," MIS Q., vol. 36, no. 4, pp. 1165–1188, 2012.

[23] C. L. Philip Chen and C. Y. Zhang, "Data-intensive applications, challenges, techniques and technologies: A survey on Big Data," Inf. Sci. (Ny)., vol. 275, pp. 314–347, 2014.

[24] A. Labrinidis, Y. Papakonstantinou, J. M. Patel, and R. Ramakrishnan, "Exploring the inherent technical challenges in realizing the potential of Big Data," Commun. ACM, vol. 57, no. 7, pp. 86–94, 2014.

[25] Y. Lv, Y. Duan, W. Kang, Z. Li, and F. Wang, "Traffic Flow Prediction With Big Data : A Deep Learning Approach," IEEE Trans. Intell. Transp. Syst., vol. PP, no. 99, pp. 1–9, 2014.

[26] L. Li, X. Su, Y. Wang, Y. Lin, Z. Li, and Y. Li, "Robust causal dependence mining in big data network and its application to traffic flow predictions," Transp. Res. Part C Emerg. Technol., vol. 58, pp. 292–307, Sep. 2015.

[27] A. Abadi, T. Rajabioun, and P. A. Ioannou, "Networks With Limited Traffic Data," Ieee Trans. Intell. Transp. Syst., vol. 16, no. 2, pp. 653–662, 2015.

[28] Brian L . Smith ! and Michael J . Demetsky, "Traffic flow forecasting : comparison of modeling approaches," Transportation (Amst)., vol. 123, no. 4, pp. 261–266, 1997.

[29] G. de Jong, A. Daly, M. Pieters, S. Miller, R. Plasmeijer, and F. Hofman, "Uncertainty in traffic forecasts: Literature review and new results for The Netherlands," Transportation (Amst)., vol. 34, no. 4, pp. 375–395, 2007.

[30] R. T. Underwood, Traffic management: an introduction. North Melbourne, Victoria, Australia: Hargreen Publishing Company, 1990.

[31] European Commission, "MONALISA project," 2018. [Online]. Available: http://www.sjofartsverket.se/en/MonaLisa/. [retrieved: September, 2018].

[32] Open Knowledge International, "Open Up Public Transport Data." [Online]. Available: http://opendatahandbook.org/solutions/en/Public-Transport-Data/. [retrieved: August, 2018].

[33] US City Open Data, "Datasets / Transit," US City Open Data Census, 2017. [Online]. Available: http://us-city.census.okfn.org/dataset/transit. [retrieved: September, 2018].

[34] European Union, "European Open Data in European cities," no.6, pp. 2016–2018., 2016.