D. Demmler, D. Krupka, H. Federrath. (Hrsg.): INFORMATIK 2022,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2022 783

Potentials of Bicycle Infrastructure Data Lakes to Support Cycling Quality Assessment

Johannes Schering [1], Jorge Marx Gómez[2] , Lena Büsselmann[3], Federico Alfaro[4], Jan Stüven[5]

Abstract: A data-driven quality assessment of bicycle infrastructure is necessary in times of crisis to support the decision-making process in cycling promotion. The INFRASense project was initiated to support the scoring of bike paths by providing new crowdsourcing data that is combined with other relevant data sources (traffic amount, accidents, citizen reportings etc.). The storage and processing of heterogeneous bike infrastructure data may be a challenge. With its flexibility a Data Lake could be an alternative to the traditional Data Warehouse. In the first step the paper gives an overview about data-driven initiatives in the use-case of bike infrastructure quality assessment and the recently started research project INFRASense. We will provide an overview about data sources that may potentially be included into the data driven quality assessment. Big Bicycle Data is available in many different structures and formats (CSV, XML, SHP etc.). In the second step the concepts of Data Lake and Data Warehouse are introduced. The benefits and weaknesses of these two solutions are shown followed by a discussion about which one of these is the best concept for storage, processing, and analysis of heterogeneous bicycle infrastructure data. In the last step we are providing an outlook how an efficient bicycle infrastructure data management system could be implemented.

Keywords: Bicycle Data, Data Lake, Data Warehouse, Bicycle Infrastructure, Quality Assessment

1 Introduction

In the past few years, a growing interest in cycling promotion can be perceived. The Corona pandemic has accelerated this trend. People more and more preferred individual means of transport such as bicycles or cars in order to maintain separation from other people. Especially the private car became even more popular with a high “feel good” factor [Ei21]. The city of Vienna that is traditionally famous for public transport

[1] University of Oldenburg, Department of Business Informatics VLBA, Ammerländer Heerstr. 114-118, Oldenburg, 26129, johannes.schering@uni-oldenburg.de, https://uol.de/vlba/personen/mitarbeiterinnen/johannes-schering
[2] University of Oldenburg, Department of Business Informatics VLBA, Ammerländer Heerstr. 114-118, Oldenburg, 26129, jorge.marx.gomez@uni-oldenburg.de
[3] University of Oldenburg, Department of Business Informatics VLBA, Ammerländer Heerstr. 114-118, Oldenburg, 26129, lena.buesselmann@uni-oldenburg.de
[4] University of Oldenburg, Department of Business Informatics VLBA, Ammerländer Heerstr. 114-118, Oldenburg, 26129, federico.alfaro@uni-oldenburg.de
[5] University of Oldenburg, Department of Business Informatics VLBA, Ammerländer Heerstr. 114-118, Oldenburg, 26129, jan.stueven@uni-oldenburg.de, https://uol.de/jorge-marx-gomez

doi:10.18420/inf2022_66

perceived a decrease of traffic participants in bus, tram etc. by 10 percent in the first year of the Corona pandemic [Sc22]. Due to multiple societal factors, it would be beneficial however, if more people were to change their mobility behaviour in favour of using the bike. For example, the use of active methods of transportation could provide health benefits, reduce operating costs of vehicles and reduce congestion [Li22]. The recent developments in energy and fuel costs have increased the magnitude of these benefits. To motivate more people to use the bike as a form of transport, the underlying biking infrastructure needs to be improved. It is proved by scientific research that better cycling infrastructure can encourage more cycling [HuOH14].

In the case of German cycling infrastructure, it is often the responsibility of municipalities to invest in and maintain cycling infrastructure. However, according to the German Association of Towns and Municipalities, German cities will experience a deficit of 19.4 billion € until 2024 as a consequence of the Corona crisis. At the same time, municipalities can expect increasing costs due to the consequences of the Ukraine war which can include care for refugees or energy costs [DSt22]. Therefore, municipalities need to focus on cost efficient expenditures with regards to cycling infrastructure. In order to facilitate this, municipalities require information regarding the existing cycling infrastructure. At this moment this information is not available or of poor quality. The available information is mainly limited to surveys, traffic countings, accident analysis and traffic observations [MT10]. In addition, the perspective of the cyclists is not considered in the planning process. Experts in bicycle infrastructure planning need standardized assessment criteria which are based on reliable data that reflects real problems in bicycle use such as waiting times and the perspective of the cyclists. A quality assessment based on crowdsourcing data could support the planning process of the bike infrastructure and the efficient investment in the network. As we will see in the following, some data sources that may support the bike infrastructure quality assessment are already available. From the data processing perspective, the question is how best to store and process heterogeneous data sets from cycling infrastructure domain. It will be discussed in this contribution whether a Data Lake may be an alternative to the traditional Data Warehouse.

2 Data Driven Quality Assessment

2.1 State of the Art

The existing approaches to score bike infrastructures are described in the following. The overview helps to understand the research gap and the relevance of a recently started research project that is further described below. Many municipalities are providing platforms for citizen communication. Inhabitants and visitors have the opportunity to get in touch with the local administration. As an example, the city of Osnabrück provides the

EreignisMeldeSystem Osnabrück (EMSOs, *Event Reporting System Osnabrueck*)[6] portal. Users can report incidents and choose between different categories (roads, garbage disposal, green areas etc.), mark the location on the map and write a text to describe the event or shortcoming. In terms of cycling, the *Allgemeiner Deutscher Fahrrad-Club* (ADFC, *German Bicycle Club*) provides platforms to enable cyclists to send information about critical bike infrastructure and (perceived) shortcomings. At the *ADFC Mängelmelder* ((*ADFC Deficiency indicator*)[7] for the city of Bremen cyclists may choose a location, select different (sub-)categories (traffic lights, construction sites, road damages, illegal parking etc.) and write a description. The limitation here is that the reports only reflect the subjective perceptions of individual persons. The feedback is not suitable for deriving general and comparable information about cycling quality.

The *Happy Bike Index* provides a first data driven and objective quality assessment for inner city intersections based on data from the municipal administrations.[8] In the current version the assessment considers (perceived) cycling safety regarding automobile traffic. The used indicators are the allowed maximum speed level and the daily traffic volume. At the moment, the beta version of the index does not consider construction conditions of the bike infrastructure like width, surface quality, or state. These data sets should be considered as soon these are available. Still missing are real problems on the bike trips that are detected by cycling people (e.g. waiting times at intersections).

CROW Fietsberaad which is a cycling expert committee from the Netherlands, has made a suggestion of how bike paths can be classified into quality levels on a range from A (=very good) to E (=very bad). The central aspect is the bike path width. As a result of different studies this aspect was identified as key for comfort and traffic safety. It becomes more and more relevant with the uptake of pedelec sells – because there are more and more bicycles with different speed levels on the road. If there is enough space, there is room for overtaking or evasion movements. As another aspect, the traffic amount is considered. If motorized bikes are allowed on the bike paths this has also an influence on the quality assessment. The amount of cargo and racing bikes should also be considered [VB21]. Similar to the Happy Bike Index, data of real bike trips and cyclist feedback is not considered in the methodology.

The creation of a holistic quality assessment of the bicycle infrastructure may be a challenge as we have different types of bike paths and alignment of these. The cycling network can be differentiated into advisory bike lanes and bike lanes on the road, one- and two-direction bike lanes, mixed pedestrian and bike path, bicycle boulevards that prioritize cyclists and road segments where cyclists are riding on the road together with motorized traffic without any protection [Ko13]. The *Forschungsgesellschaft für Straßen- und Verkehrswesen* (FGSV, *Research Association for Road and Traffic Systems* in Germany) that brings together various experts from the bicycle traffic planning

[6] https://geo.osnabrueck.de/emsos/
[7] https://adfc-bremen.mängelmelder.de/
[8] https://fixmyberlin.de/zustand

domain in Germany is responsible to define new guidelines and standards for planning. In 2021, the expert team published a new guideline regarding the consistent and standardized assessment of bike paths to create a solution for the challenge described above. The *Hinweise zur einheitlichen Bewertung von Radverkehrsanlagen* (H EBRA, *Suggestions for the Consistent Assessment of Bike Paths*) gives an overview about the most important criteria and the weighting of these from the planning perspective. There are three dimensions as assessment aspects: 1) alignment of the bike path, 2) state of the bike path and 3) attractiveness. The alignment is related to a) motorized traffic and b) pedestrian traffic. The roads are classified to congestion classes (traffic amount). The pedestrian traffic is considered when roads have a lot of shops, a lot of pedestrians in need of protection (children, handicapped people etc.), important connections for bicycle traffic, steep declivity, house and property entrances at narrow side paths, bus and tram stops, high number of pedestrians in relation to side path width etc. The state of the bike path is divided into shortcomings at a) specific locations (potholes, bollards, root breakup etc.), b) over the whole length of the path that causes time losses (surface types with high vibration level, missing markings etc.), c) junctions (missing bicycle traffic light signaling, long waiting times at red lights etc.) and d) overtakings (bike path width, illegal parking cars etc.). The attractiveness is divided into travel experience for cyclists (fresh air, green environment, low sound emissions etc.) and the perceptibility of the travel experience for cyclists (only fulfilled when whole quality level is good, pedestrian and/or bicycle traffic amount is not to high etc.) [FGSV21].

The H EBRA guideline includes an overview about demanded data sources. Geographic Information Systems (GIS) could give insights to length, inclines and alignment of bike paths. Construction criteria such as junction type, bike path width, surface state and type need to be taken into account. The traffic amount can be measured by day. Maximum speed levels for motorized vehicles may be considered. Bicycle counters deliver daily cyclist numbers at local spots of the city. When the side path is used by cyclists and pedestrians the amount of people that are walking and related conflict situations are important. The different types of shortcomings and the travel experience for cyclists that are described above should also be part of the data sources. Especially when it comes to problems on real bike trips it remains unclear what kind of data may be used [FGSV21]. This kind of “real cyclists” data is also not considered in the Dutch approach of CROW Fietsberaad described above.

2.2 INFRASense

To find new solutions to the challenges described above, a new bicycle data driven research project INFRASense[9] was initiated. The main goal is the development of a sensor-based quality assessment of the bike infrastructure which is based on crowdsourcing data and consists of a software tool including a mobile app for the analysis and the definition of quality levels (A = very good to F = very bad). The

[9] www.infrasense.de

software tool *Bicycle Lane Quality Evaluation* (BIQE) will be a new basis for the decision making in the bike infrastructure planning process. The involvement of real cyclists is a very important aspect to create an assessment tool that reflects the user perspective. At least 1.000 citizens will participate in the data collection using a specialized sensor box that measures GPS, acceleration (which allows the analysis of vibrations on the bicycle) and environmental data directly on the bike. Further information about the bike sensors can be found in Schering et al. [Sc21]. Another 750 citizens will use the user interface of the software tool to provide feedback about the quality assessment to the project. With the user interface citizens can send information about individual bike rides and quality assessment The feedback will be used to make the quality assessment more accurate. Machine Learning methods for the automatized classification of the bike paths will be applied. As a result of the project we will have a system that consists of a software solution, bike sensors and analysis methods that provide crowdsourcing data-based information about cycling quality. The provided Key Performance Indicators (waiting and loss times, surface qualities, traffic flow, weather influence, comparison of bad bike paths quality with the number of accidents etc.) and visualizations will support the automatic quality assessment.

The development requires a comprehensive data management system in the background. To process all the external and internal data sets an architecture for data integration, -analysis and -provision is needed. The following working steps include the construction of a database, the implementation of interfaces to municipal GIS or open data portals, automated sensor export, etc. The data management should enable the data analysis of several criteria relevant to cycling quality.

3 Data Sources and Approaches for Data Processing

3.1 Data sources

Many different data sources are needed to enable the quality assessment of bicycle infrastructure. On the one hand, the bike sensor is very important for obtaining user-specific information about the quality of a bike trip. The data can be used, for example, to answer questions about where cyclists have to stop particulary often or at which intersection waiting times are very long. On the other hand, external data sources contain data on the bike path network (material, path type, width), traffic planning (bike counting, traffic flow, occupancy rate etc.), shape files of municipal GIS (bus stops, neighbourhood borders etc.) and accident data from the police. In the following, Tab. 1 provides an overview of the available data sources that may be used in the quality assessment.

Provider	Contents	Data Format	Interface
CoSynth GmbH (internal)	Bike sensor data: Accelerometer, GPS etc.	CSV	API (regular)
Project internal data collection	Camera data, images	MP4	Reading out (per experiment)
Planungsbüro VIA (internal)	Bike path width	Excel	Will be collected once and transmitted by email
City administration (Department of traffic guidance)	Locations of intersections	PDF	e-mail (once)
City administration (Department of traffic guidance)	Traffic volume, occupancy rate	CSV	API (daily)
City administration (Department of Mobility)	Temporary and permanent Bike countings	PDF, CSV	Webportal (Download, e.g. per month), email
City administration (Department of Mobility)	List of bike paths (obligatory use), bridges, snow-plowing service	PDF	e-mail (once)
City administration (Department of Maintenance)	Road register	Excel	e-mail (once)
City administration (Department of communication)	Reportings about shortcomings	CSV	*not defined yet
City administration (Department of Geoinformation and Statistics)	Geoinfornation: Cycling network, bike parking facilities, bus stops, topography, districts, road classification …	SHP/SHX/ SBN/CPG/ DBF/ PRJ	e-mail (once)
Police Inspection	Historical bicycle accident data	Excel	e-mail (once per year)
Visual Crossing	Weather data	CSV	API (daily)

Tab. 1: External data sources

3.2 Data Warehouses

As we learnt from the overview above, the project has to deal with heterogeneous data sources. These consist of CSV, XML, Word, PDF, SHP or GeoJSON files. The different data formats have the effect that data management must be able to handle structured, semi-structured and unstructured data sets. The heterogeneity could be a challenge in the data integration process for traditional data management and analysis systems. There are different concepts and technologies for data storage. In the past, files, data bases, data warehouses, cloud systems and data lakes have been proposed [ZKBG21].
The Data Warehouse can be defined as an integrated and historical storage system [SD21]. It can be used for data analysis purposes and was developed to support the decision-making processes in businesses [Kr13]. Relevant data sets are extracted from different source systems, are transformed into the schema of the target Data Warehouse and are saved in it (in batches, in real time etc.) [DS16]. The Data Warehouse has four central characteristics. According to a definition of Bill Inmon (1996) who is known as the father of Data Warehousing it is subject oriented (e.g., biking), integrated (transforming data from different sources into a consistent format), non-volatile (data in the warehouse should not change anymore) and time-variant (changes over time may be observed) [In96]. All data sets with relation to the same topic will be interconnected and standardized as part of the Extract, Transform, Load (ETL) process. The data sources can be provided by different operational systems and may have variations in file layouts, code representation and header names [Po10]. As part of the ETL process the data sets are transformed and transferred at defined time intervals. The historical data will be saved and will not be changed anymore. The storage of historical data is a key characteristic of the Data Warehouse [Po10].

3.3 Data Lake

A potential alternative for the storage and integration of big and heterogeneous data could be a Data Lake, which functions as a raw data repository to store all data sets in any format [Ch20]. The Data Lake concept was introduced in 2010 by James Dixon [Ma17, MT16]. Since then, the concept has become well established in the industry. In the research domain, however, acceptance of the Data Lake is still in the very beginning [Zo21]. The implementation of Data Lake technologies for processing and storage of cycling data was not considered in research. Depending on the data sets of the Data Lake and the use case scenario there are different requirements regarding a Data Lake architecture [Gi21]. As bicycle data is still a quite young research domain it is not clear how suitable Data Lake approaches from the literature could be applied in the context of the bike infrastructure analysis and quality assessment.

A Data Lake can be defined as a central location for data storage. Data can be stored and applied in the Data Lake in any format [SD21, TR21]. In addition, data sets can be added to the Data Lake without changing its data structure [MT16] to enable an analysis without any pre-defined use cases. The Data Lake is a collection of different data sets [TN20]. Data import may be realized by any method [Ma17]. An important difference compared to the Data Warehouse is, that the ETL process will not be applied directly. In contrast the data sets will be extracted from their origin source and will be stored in the Data Lake. In the next step data can be transformed in its structure. This process may be called Extract, Load, Transform (ELT) [Ka18]. Information can be extracted from the Data Lake as part of data processing and analysis [Ma17]. The Data Lake may deal with structured, semi-structured and unstructured data. It is often applied in the case of a fast-growing amount of big unstructured data sets, which makes it different to the Data Warehouse [TN20].

The application of the Data Lake brings new challenges for storage, processing and search of data [TR21]. An important challenge relates to the metadata of the data sources. There may not be a complete data catalog for the data sources in the Data Lake. In addition, the data possibly does not contain metadata relevant to its use or utilizes different formats to describe its metadata [Na18]. Finding data may be difficult because no global schema exists [TR21]. Without the appropriate structure and organization of the data, a Data Lake can turn into a so-called Data Swamp, which limits the use of the data [KW18, TR21, TN20]. Further problems arise in the implementation of Data Lakes as there is no universal data lake architecture that covers all necessary aspects [Gi19].

3.4 Differentiation between Data Lake and Data Warehouse

The concepts of the Data Lake and the Data Warehouse are often compared to each other as a system to store data for data analysis [TR21]. However, the data processing approaches are very different. As described above, the Data Warehouse relies on the ETL, the Data Lake on the ELT process. The Data Lake imports the raw data as it is and the structure is only transformed when there is demand for that. That makes the concept more flexible and agile [KW18]. The Data Warehouse is characterized by less flexibility. The schema of the Data Warehouse is developed and determined before the data is loaded. The Data Warehouse is efficient for the storage and the processing of structured data from operative systems [SD21]. Semi- and unstructured data may be a challenge because of its heterogeneity and complexity [Gi19, ML16].

The Data Lake may contribute to find a solution to these challenges which arise from the growing amount of available (big) data. There is a broad consensus in the literature that the architectures currently available to enable the integration of this data is primarily the Data Lake [CCWQ20]. As we already learned from the data overview in Tab. 1 above, in the specific use case of bike infrastructure there a quite a lot of data sets available. Bike infrastructure already falls to the category of Big (Bicycle) Data simply because of the volume involved [Ro16]. Therefore, it can be assumed that the Data Lake approach may be a suitable solution for the realization of a holistic system for the quality assessment of the bike infrastructure.

The approach that is discussed in this contribution allows further data processing regarding the quality of the bicycle infrastructures. A lot of measures that may contribute to a better bike path network are imaginable. Based on the detected waiting times at intersections at frequently used roads the traffic signalling may be adjusted. Surface types can be changed (e.g. to asphalt) or the state of the surface can be renewed at locations with a high number of bumps on the road. Evasion movement hotspots can be an indicator that space is not sufficient (e.g. because of a mixed space scenario with pedestrians). As a consequence, the bike path or side walk may be broadened. All these measures may support cyclists to avoid to have time losses and to make cycling more attractive. Traffic safety is another important factor whether people are using the bike or not. If the data processing reveals roads with many truck-bicycle accidents the bike path alignment could be changed (e.g. to a protected bike lane). In roads with many car accidents the allowed maximum vehicle speed level may be reduced from 50 to 30 km/h. If we identify locations where many people are complaining about parking violators, the number of parking spaces could be reduced. That will contribute to the availability of space, less conflicts with pedestrians and hopefully to a more liveable city.

4 Conclusion and Outlook

A data-driven quality assessment of bicycle infrastructure is necessary in times of crisis to support the decision-making process in cycling promotion. Existing approaches are limited to the subjective perceptions of cyclists or only take into account traffic safety or construction criteria data. The H EBRA guideline gives an orientation of how a holistic quality assessment could be realized. The INFRASense project was initiated to support this process by providing new crowdsourcing data that is combined with other relevant data sources. Since we have to deal with a huge amount of data to create a holistic quality assessment of road segments, the question arises what could be the most suitable solution for data storage and processing. Data Warehouses have a long tradition in this field as they are able to restructure and process huge amounts of data. For the development of the data driven quality assessment, a Data Lake could be an alternative. More flexibility is facilitated because it is able to handle heterogeneous.

The next step is to decide which ETL or ELT approach is best suited for processing of bicycle infrastructure data. The external data sources need to be combined with the bike sensor data to enable more specific data analysis. Potential connections between the data sources could be established using the time stamps, geo coordinates and road names. For example, road lengths could be linked to sensor or accident data in order to calculate how many brakings, stops or accidents occur per meter on a specific road segment. These associated Key Performance Indicators support the bicycle infrastructure assessment process and will be discussed in further publications. Another requirement is that GIS data may be stored and processed as part of the data management.

One potential option for the implementation could be a Data Lake in combination with a

Data Warehouse. As described above, the Data Lake stores all the raw data sets regardless of their format and structure without changing them. The Data Lake has an in-build API functionality to enable the implementation of the ETL or ELT process. The API allows communication between raw data storage and ETL or ELT methodology. In addition to that a Data Warehouse could be implemented. In the Data Warehouse the data may be transformed into a specific format, taking into consideration the specific requirements of the domain experts in the field of traffic planning. More specific research questions can be answered as part of the data analysis. In any case, these results will be presented in further publications.

Acknowledgements

INFRASense is funded by the Bundesministerium für Digitales und Verkehr (BMDV, *German Federal Ministry of Digital and Transport*) as part of the mFUND program (project number 19F2186E) with a funding amount of around 1.2 Mio. Euro. As part of mFUND the BMDV supports research development projects in the field of data based and digital mobility innovations. Part of the project funding is the promotion of networking between the stakeholders in politics, business, administration and research as well as the publication of open data on the mCLOUD portal (which will be superseded by Mobilithek in 2022/23).

References

[BSS21] Bauer, M.; Sanchez, L.; Song, J.: IoT-Enabled Smart Cities: Evolution and Outlook. Sensors 21, Nr. 13, pp. 1-29, 2021.

[Ch20] Chiusano, S.; Cerquitelli, T.; Wrembel, R.; Quercia, D.: Breakthroughs on Cross-Cutting Data Management, Data Analytics, and Applied Data Science. Information systems frontiers 23, Nr. 1, pp. 1-7, 2020.

[DS16] Dhanda, P.; Sharma, N.: Extract Transform Load Data with ETL Tools. International Journal of Advanced Research in Computer Science 7, Nr. 3, pp. 152-160, 2016.

[DSt22] Deutscher Städte- und Gemeindebund: Kommunale Finanzen: Einführung, https://www.dstgb.de/themen/finanzen/, accessed 10/05/2021.

[Ei21] Eisenmann, C.; Kolarova, V.; Nobis, C.; Winkler, C.; Lenz, B.: DLR-Befragung: Wie verändert Corona unsere Mobilität? Verkehrsmittelnutzung, Einkaufs-, Arbeits- und Reiseverhalten, https://verkehrsforschung.dlr.de/de/news/dlr-befragung-wieveraendert-corona-unsere-mobilitaet, accessed 04/03/2021.

[FGSV21] Forschungsgesellschaft für Straßen- und Verkehrswesen (FGSV): Hinweise zur einheitlichen Bewertung von Radverkehrsanlagen (Ausgabe 2021). FGSV Verlag, Köln, 2021.

[Gi21] Giebler, C.; Gröger, C.; Hoos, E.; Eichler, R.; Schwarz, H.; Mitschang, B.: The Data Lake Architecture Framework: A Foundation for Building a Comprehensive Data Lake Architecture. In: Proceedings der 19. Fachtagung fuer Datenbanksysteme fuer

Business, Technologie und Web (BTW 2021), pp. 351-370, 2021.

[Gi19] Giebler, C.; Gröger, C.; Hoos, E.; Eichler, R.; Schwarz, H.; Mitschang, B.: Leveraging the Data Lake: Current State and Challenges. In: Ordonez, C.; Song, I.-Y.; Anderst-Kotsis, G.; Tjoa, A M.; Khalil, I. (ed.): Big Data Analytics and Knowledge Discovery. Springer Nature Switzerland AG, pp. 179–188, 2019.

[HuOH14] Hull, A.; O'Holleran, C.: Bicycle infrastructure: can good design encourage cycling? Urban, Planning and Transport Research. Routledge, 2(1), pp. 369–406, 2014. doi: 10.1080/21650020.2014.955210.

[In96] Inmon, W.H.: Building the Data Warehouse. John Wiley & Sons, Inc., New York, 1996.

[KW18] Khine, P.; Wang, Z.: Data lake: a new ideology in big data era. ITM Web of Conferences 17, pp. 3025, 2018.

[Ko13] Köhler, U.: Einführung in die Verkehrsplanung. Grundlagen, Modellbildung, Verkehrsprognose, Verkehrsnetze. Fraunhofer IRB Verlag, Stuttgart, 2013.

[Kr13] Krishnan, K.: Data Warehousing in the Age of Big Data. Morgan Kaufmann, Boston, 2013.

[Li22] Litman, T.: Evaluating Active Transport Benefits and Costs. Guide to Valuing Walking and Cycling Improvements and Encouragement Programs. Victoria Transport Policy Institute, Victoria, 2022.

[ML16] Madera, C.; Laurent, A.: The Next Information Architecture Evolution: The Data Lake Wave. In: 8th International Conference on Management of Digital EcoSystems, MEDES 2016, pp. 174–180, 2016.

[Ma17] Mathis, C.: Data Lakes. Datenbank-Spektrum 17, Nr. 3, pp. 289-293, 2017.

[MT10] Mettenberger, T.; Thiemann-Linden, J.: Bewertung der Radverkehrsqualität. Methoden der Wirkungskontrolle und stadtweite Fahrradbilanzen. Forschung Radverkehr, O-2, Deutsches Institut für Urbanistik –Difu-, Berlin, 2010.

[MT16] Miloslavskaya, N.; Tolstoy, A.: Big Data, Fast Data and Data Lake Concept. In: Procedia Computer Science 88, pp. 300-305, 2016.

[Na18] Nargesian, F.; Zhu, E.; Miller, R.J.; Pu, K.Q.; Arocena, P.C.: Data Lake Management: Challenges and Opportunities. In: Proceedings of the VLDB Endowment, Nr. 12, pp. 1986-1989, 2018.

[Po10] Ponniah, P.: Data Warehousing Fundamentals for IT Professional. 2nd ed. John Wiley & Sons Inc., Hoboken, New Jersey, 2010.

[Ro16] Romanillos, G.; Zaltz A.M.; Ettema, D.; De Kruijf, J.: Big Data and Cycling. Transport Reviews 36, Nr. 1, pp. 114–133, 2016.

[SD21] Sawadogo, P.; Darmont, J.: On Data Lake Architectures and Metadata Management. Journal of Intelligent Information Systems 56, Nr. 1, pp. 97-120, 2021.

[Sc21] Schering, J.; Janßen, C.; Kessler, R.; Dmitriyev, V.; Marx Gómez, J.; Stehno, C.; Pelzner, K.; Bankowsky, R.; Hentschel, R.: ECOSense and its preliminary findings: Collection and analysis of bicycle sensor data. In: Kamilaris, A.; Wohlgemuth, V.;

Karatzas, K.; Athanasiadis, I. (ed.). EnviroInfo 2020 Environmental Informatics – New perspectives in Environmental Information Systems: Transport, Sensors, Recycling. Adjunct Proceedings of the 34th EnviroInfo conference. Shaker Verlag, Düren, pp. 145-153, 2021.

[Sc22] Scherreiks, W.: Weiter zur smarten Stadt – Zukunftsplan Wien. VELOPLAN 01/22, pp. 18-27, 2022.

[TR21] Taniar, D.; Rahayu, W.: Data Lake Architecture. In Barolli, L.; Natwichai, J.; Enokido, T. (ed): Advances in Internet, Data and Web Technologies. Springer Nature Switzerland, pp. 344-357, 2021.

[TN20] Thomas, K..; Nair, P.: Data Lake: A Centralized Repository. International Research Journal of Engineering and Technology (IRJET) 07, Nr. 5, pp. 2978-2981.

[VB21] Veroude, B.; van Boggelen, O.: Discussienotitie - Actualisatie aanbevelingen voor de breedte van fietspaden. CROW Fietsberaad, Utrecht, 2021.

[Zo21] Zouari, F.; Boukadi, K.; Kabachi, N.; Ghedira-Guegan, C.: Data Management in the Data Lake: A Systematic Mapping. ACM International Conference Proceeding Series, pp. 280-284, 2021.