A Spatial Algorithm for Dataset Integration Applied to Cycling Data

Radoslav Eliáš, Juraj Lazúr, Jiří Hynek, Tomáš Hruška

***Abstract*—One of the main problems of modern Smart Cities is the constant increase in traffic volume and travel times. City governments are trying to address these problems by promoting alternative modes of transport, including cycling. The infrastructure of modern Smart Cities is planned on the basis of analyses obtained from various data sources. In the case of cycling, suitable data comes from e.g. automated counters, manual counts or training applications. The first step in the effective use of data is typically to assign the data to the corresponding physical infrastructure elements in real space. Thus, the geographic component of datasets plays an important role in linking them. However, different data sources describe the same infrastructure elements differently, which makes it impossible to link the datasets in a straightforward way. The city of Brno had to deal with such a problem when trying to use the available data sources to improve the quality of cycling infrastructure. The purpose of this paper was to propose a method for transforming the different input datasets describing cycling transport into a common dataset that will serve as a basic data source for the analyses. The result is an integration algorithm that links the data describing the same infrastructure elements. The output of the algorithm is a model mapping the individual datasets into one common mapping network. This mapping can then be applied to both existing and future data. The implemented solution was tested in cooperation with the Brno City Council, where the integration of 5 different datasets was tested. The use of the integrated data was subsequently tested within two implemented dashboards.**

***Index Terms*—Data transformation, Data handling, Geospatial data, Data models, Cycling transport, Open Street Map**

I. INTRODUCTION

The rise of the Big Data concept also has a significant impact on cycling transport [1]. The simplification of obtaining quantitative as well as qualitative data on cycling has brought new opportunities to better analyse and plan for this integral mode of transport in modern Smart Cities. Traditional methods of data collection in the form of manual counts have gradually been joined by automated counters and data from health or training applications. The use of raw cycling data from a variety of sources finds its most common application in two complementary areas.

The first of these areas focuses directly on the cyclist and their individual health. GPS data from individual rides combined with physiological indicators can be used to analyse neurological and biological indicators [7]. The combination of these data helps, for example, in improving fitness or detecting various diseases [6]. A second area of application of cycling data is closely related to infrastructure. In modern Smart Cities, the planning and development of infrastructure is closely intertwined not only in the field of cycling with the need for good quality input data. Individual datasets coming from different sources are thus used e.g. for planning new cycle paths [3], analysing traffic accidents [5] or evaluating the success of implemented changes [2]. For all analyses and applications, however, it is necessary to logically link the individual datasets [1] at the beginning.

In the case of integrating multiple datasets, the most commonly used aspect, not only in the area of cycling transport, is the geographic component of the data [5]. For applications focusing on cyclists or their health, time can also be used in dataset integration. However, in the area of infrastructure planning and management, geography-based integration is essential. Individual datasets typically contain the location, and in some cases the specific infrastructure elements to which the record relates. However, the different representation of the same infrastructure elements across datasets complicates the proven integration methods used in other geographic data domains.

In the field of geographic data, there are proven methods, such as spatial join [4] shown on Figure 1, that integrate different datasets into one common space. However, in the case of transport data, which includes cycling, the use of these methods is not always possible. The problem lies mainly in the ambiguous representation of the physical infrastructure. Thus, one element, e.g. a street, may be defined differently in different datasets. In order to link these datasets, it is thus necessary to find all representations of the same infrastructure element in different datasets. While in the case of simple points it is possible to use, for example, the *k*-nearest neighbours method, in the case of polylines, i.e. streets, the problem is more complex.

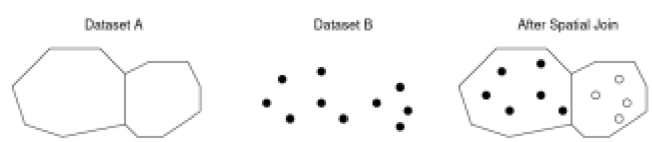

Fig. 1, One of the most common methods of geographic data integration is Spatial Join. Its essence is the integration of datasets based on the position of features in space. As can be seen in the figure, after integrating dataset A containing polygons with dataset B containing points, it is possible to obtain which points belong to which polygon using Spatial Join.

First submitted 14. 2. 2025.
This work was supported by project Smart information technology for a resilient society, FIT-S-23-8209, funded by Brno University of Technology.
Radoslav Eliáš is with the Brno University of Technology
Juraj Lazúr is a doctoral student at Brno University of Technology, Brno, 612 00 Czechia, (e-mail: ilazur@fit.vutbr.cz)
Jiří Hynek and Tomáš Hruška are with the Brno University of Technology (e-mails: hynek@fit.vutbr.cz and hruska@fit.vutbr.cz respectively).

This problem is even more acute in the context of periodic processing of input data, where, unlike single-impact analysis, the integration process has to be repeated periodically, which increases the computational complexity of the whole process.

In order to transform the individual datasets into one, it is necessary to initially determine the logical relationships between the individual records. Therefore, the proposed solution is based on an algorithm which initially constructs a basic infrastructure map from the Open Street Map. The algorithm then searches for a representation of each infrastructure element from the base map in each datasource. The algorithm generates for each dataset a mapping of the individual dataset records to elements of the common underlying map. Then, individual records from different datasets can be grouped according to the infrastructure elements of the common model.

The verification of the proposed concept consisted in the implementation of the algorithm and testing over datasets describing the city of Brno, the second largest city in the Czech Republic. The resulting mapping model works on average with a 78% success rate in identifying different images of the same infrastructure elements. The testing shows that the proposed solution is able to transform a wide range of different data sources into a single dataset. The transformation into a common dataset facilitates and extends the capabilities of cycling data analysis, which has been verified in the implementation of two dashboards using the generated mapping model.

II. Cycling Data Collecting and Integration

The simplification and reduction in the cost of data collection methods has led to a significant increase of potentially useful data availability. Traditional methods of collecting transport data, such as manual census [10], have thus been replaced by automated counters and traffic cameras. In the cycling domain, besides the typical transport methods, potentially useful data can also be collected directly from cyclists or through crowdfunding campaigns.

Each cycling data collection method captures a different aspect. Manual censuses of cyclists [9], typically conducted only a few times a year, enable to obtain data with the highest granularity. Depending on the setting of the census, the data may include only the total number of cyclists who passed through a given section in a given time period, or it may include more detailed data such as bicycle type. The disadvantage of manual counts of cyclists is the long time intervals between counts, which can for example be addressed by automated counters. These devices, typically placed at busy locations, use a variety of methods to record the number of cyclists [13]. Their disadvantage compared to manual counts is precisely the absence of any details about cyclists. At the same time, both methods are static-oriented, as the data collected is location-specific.

In contrast, the data collected from various health and sport applications generally contain the entire route, i.e. the cyclist's movement over time. The disadvantage of these applications is the higher degree of unreliability of the data, mainly due to the imperfection of the different sensors. In addition, cycling data can also be obtained through various surveys, but the quality of the data obtained in this way depends on a well-chosen sample of users [10]. In practice, most applications use at least two different methods of collecting cycling data, which typically requires the integration of the different datasets. However, applications approach this integration differently.

Different data collection methods generate datasets describing different aspects of cycling. Some research papers use only one data source, such as the Strava mobile app, which researchers try to use as much as possible [12]. A different approach is to use the data for validation of predicted behaviour. In [11], the data from automated counters was used to validate the accuracy of simulated system behavior. However, the vast majority of studies use at least two different data sources, which need to be linked before use. For example, in the work of [8], sensor data is used together with a questionnaire, while position records from GPS are used to determine the speed of the bicycle. The integration of all 3 data sources is then based on timestamps. In the case of cycling infrastructure management and planning, one of the most powerful methods for integrating different data sources is to use the geographic component of the data [14].

Geographic datasets can be integrated in different ways. The traditional way is to use the ontology method, which consists in defining common elements and their representations in different datasets. In the case of geographic data, an ontology can be defined for different data dimensions such as geometry, topology or symbolic representation [17]. Then, for each dataset an application ontology is created, which is linked to the common ontology by abstract rules [16]. An example would be an abstract definition of a street in the common ontology, which is linked using abstract rules to the application ontology of a particular dataset, which represents the street as a set of two points. However, ontology-based approaches are inadequate for handling geospatial data with multiple representations [15].

In the case of geographic data, the ontology approach has to deal with multiple object representations across different datasets. As an example, an intersection may be represented by a point in one dataset, while in another dataset, it is represented as a set of points as is shown on Figure 2. From an ontology perspective, these are the same category of objects, but from a geographic representation and data usage perspective, they are two completely different concepts [19]. One solution is to define separate ontologies for each dataset, where these ontologies are constructed from scratch or different international standards such as INSPIRE[1] are used. The different ontologies are then linked by semantic relations. However, some of the relationships between data, such as the more substantial representation of objects, cannot be addressed using semantic relations. As a solution, multiple representation databases are used, which directly model the relationships between different object representations in real space [18]. By using ontologies, it is possible to integrate

[1] https://knowledge-base.inspire.ec.europa.eu/index_en

different datasets into one. A completely different way is to integrate data using the concept of spatial join.

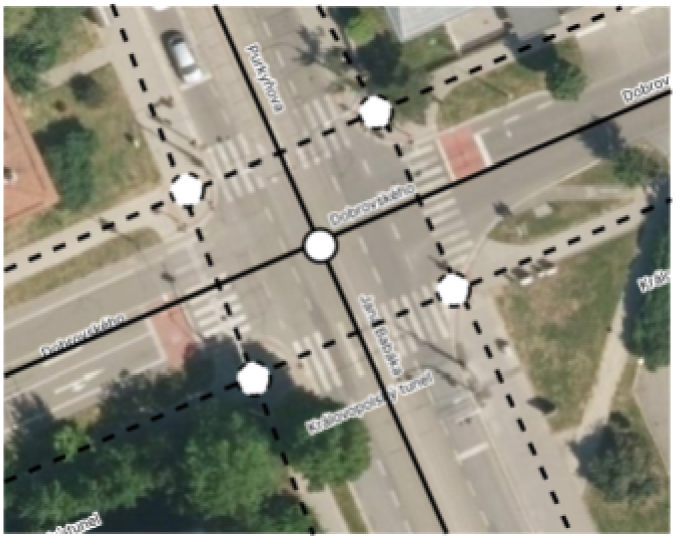

Fig. 2, The heterogeneous application ontology makes it difficult to integrate different datasets. While in the case of the first dataset represented by solid lines, the crossroad is defined by a single point, in the case of the second dataset represented by dashed lines, the crossroad is a set of 4 points. Both datasets work with the same concept, but implement it quite differently.

In many cases, not only in the transport sector, it is necessary to integrate point data with infrastructure elements or administrative areas [20]. A similar problem is the integration of several transport networks into one. These and similar tasks can be solved using techniques implementing the spatial join concept. It allows us to find all pairs of objects in the multidimensional plane that satisfy a given relation [21]. One of the most frequently solved problems using the spatial join methodology is finding overlapping objects of two distinct datasets. The advantage of the spatial join concept is precisely the ability to take into account multiple dimensions, such as elevation combined with area [22]. This feature of the spatial join concept allows the integration of different geographical datasets. It was also the problem of integrating datasets working with different transport networks that had to be solved first in the context of the Brno City Council's request to use the available cycling data.

III. Problem Definition

In order to make better use of cycling data in infrastructure analysis and planning, it is necessary to integrate the available datasets. However, there are specific cases where neither ontology nor spatial join methods can be used on their own. Besides the integration itself, its frequent repetition is also problematic, which increases the computational requirements. Thus, successful integration of atypical geographic datasets requires solving two fundamental problems.

The first problem lies in the integration of the datasets themselves. Since each dataset is generated by a different methodology, the individual records have a different structure. In addition, each dataset operates with its own, or none at all, underlying infrastructure map, which again makes the integration process difficult. At the same time, the different datasets contain different types of geographic objects, which, in addition to the ambiguous structure and underlying network, requires the integration of point and line data.

The second issue is the applicability of the solution in the longer term. While most works implement the integration of cycling data once at the beginning of the analysis, new records are added every day in the case of data describing the city of Brno. This implies the need to potentially repeat the integration on a regular basis, resulting in increased demands on computational capacity, which will steadily increase as the volume of data grows.

The nature of these two problems precluded the direct use of ontology methods as well as spatial join at the beginning of the solution design. The proposed solution shaped by these problems therefore seeks to combine both proven and widely used methods of geographic data integration in order to integrate the available cycling datasets.

IV. Proposed Solution

The essence of the proposed solution was to create a tool for regular automated integration of datasets describing bicycle transport in Brno, which will enable better knowledge extraction from these data. The result of the work is an integration algorithm combining ontology and spatial join methods. Then, the practical output is a set of scripts that implement the individual steps of the algorithm. The algorithm itself can be divided into several steps, as shown in Figure 3.

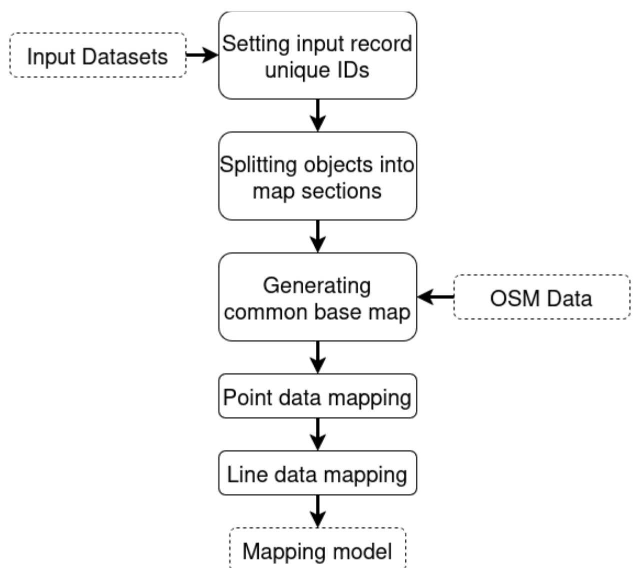

Fig. 3, The proposed integration algorithm first divides the input dataset records themselves into map sections. Subsequently, a common base map is generated. The data integration itself consists in creating a mapping of the individual records to the underlying map, and the output of the algorithm is this mapping.

In order to reduce the number of comparisons when integrating datasets, the state space—the city of Brno—is divided into square sections of equal size. This creates sections into which individual points are then sorted based on their coordinates. In the case of lines, where it is possible that a line extends into more than one section, the bottom left point of the line is decisive. In this way, all records from the input datasets are sorted into sections. The purpose of introducing sections is to optimize and speed up the integration itself, since by assigning a record to a section, the record itself will only be matched against a subset of the elements of the common base infrastructure map.

The division into sections is followed by the generation of a common base map. As the problem definition shows, the purpose of using the data is to support the planning and management of cycling infrastructure. Therefore, the proposed solution integrates the input datasets into a common underlying network representing the infrastructure. Since the individual datasets work with different underlying networks, the proposed solution transforms the datasets into a new underlying network based on OpenStreetMap data. This creates a source of truth, which is used by the proposed solution as the basis for the data integration itself. Due to the different types of geographic objects in the input datasets, the proposed solution splits the integration into two steps.

In the first step, point data, typically represented by automatic counters, are integrated. First, each element of the common underlying map as well as the integrated dataset is tagged with a unique ID. Then, for each record, the absolute distance to all elements that are in the section to which the record has been assigned is calculated. Based on the selection of the smallest distance, a mapping between the IDs of the individual elements and records of the dataset is gradually created. The output is thus an accurate mapping of which records belong to which infrastructure elements. Only the newly added records are then mapped during updates. However, this approach requires some consistency in the tagging of records from the input datasets.

In the second step, line data, where the most typical example is the recording of the cyclist's ride progress from a sports application, are integrated. Similar to point data integration, line data integration works with unique IDs. The unique ID of the infrastructure elements in the common underlying map is shared obtained from the previous step. The essence of the proposed algorithm is to assign a record, a street, to one of the candidates from the common base map. The selection of the most correct infrastructure element that the examined record describes consists of several sub-steps.

At the beginning, geographic coordinates are rounded to 5 decimal places. This rounding represents a negligible loss of precision in the context of this work, but greatly simplifies the removal of parallel objects that actually represent the same object. An example is a pair of streets that are parallel in terms of input data, but in reality represent a single object. By rounding the geographic coordinates, the similarity score is increased and the objects can be related to each other. The second sub-step is to constrain the angle that the compared objects take.

An important assumption of this sub-step is that different representations of the same object have very similar orientations in space. Since most of the line objects in practice are represented by a set of lines, it is not possible to simply compare the angle between two lines. Therefore, the proposed solution works with a bounding box enclosing each line. The angle itself is then computed based on the diagonals of the bounding boxes of the compared objects, as the example in Figure 4 shows. If the angle is greater than 45° then the comparison is stopped. However, the actual determination of the similarity to the candidate infrastructure elements is only determined in the last sub–step.

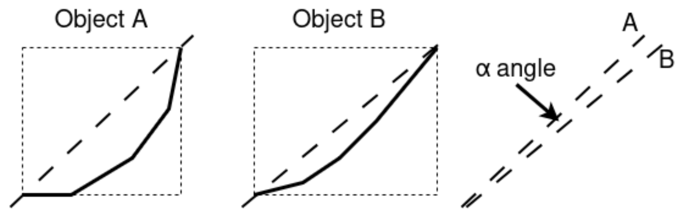

Fig. 4, Determine the similarity of objects A and B based on the angle α formed by the lines A and B, which are the diagonals of the bounding boxes of each object. A smaller angle in this case means a higher probability that the objects represent a single real infrastructure element.

While the first two substeps mainly aim at candidate reduction, the actual similarity score is only determined in the third sub–step. Thus, the similarity of the examined record to the candidate infrastructure features is determined based on the overlap of the individual bounding boxes. Thus, the examined record is generally assigned to the infrastructure element with which its bounding box overlaps the most, as illustrated in Figure 5. The result, as in the case of point data, is a set of ID pairs that determines the relationships between the common underlying map and the integrated dataset. In the case of a mapping update, only new objects cannot be mapped, since individual records within datasets are not identified by a persistent id. Thus, the output of the entire algorithm is a common underlying map that contains for each dataset the mapping of individual records to infrastructure elements.

The integration of geographic data using ontology or spatial join methods is well researched and widespread. However, there are cases that require a specific approach combining these methods. One such case is the integration of data describing the bicycle transport of Brno city.

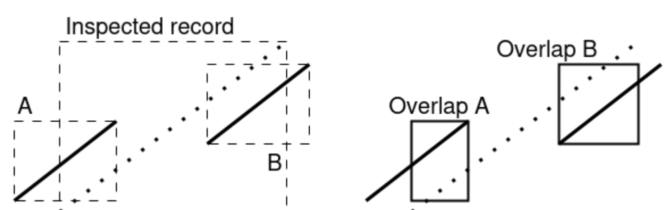

Fig. 5, The final assignment of a record from the input dataset to an element of the underlying map depends on the degree of overlap of the bounding boxes. The left part of the figure shows the bounding boxes of the investigated record as well as the candidate elements from the underlying map A and B. The right part then shows the overlap, in which case the examined element would be assigned to element B.

The proposed solution exploits the capabilities of geographic data integration methods, which are combined by the proposed algorithm in order to integrate different structured and ontologically heterogeneous datasets.

V. Results and Evaluation

Since the solution itself was designed based on the requirements of the Brno City Council, testing of the proposed algorithm and its implementation were carried out in cooperation with the planning department. Within the testing, both the accuracy of the generated mapping and the usability of the integrated data were investigated by means of the implementation of an analytical dashboard.

OpenStreeMap data were used for modelling the common base map, while more than 65 000 streets were processed within the Brno agglomeration. Integration testing was performed over 4 publicly available and 1 proprietary dataset. The datasets available included multi-year data from 19 bicycle counters, anonymised data from the Strava app, multi-year data from the BKOM survey, data from the ŘSD road survey and from the Bike to Work campaign.

The implementation of the proposed algorithm consists of a set of Python scripts using Jupyter Notebook technology. The NumPy and GeoPandas libraries are also used in the implementation. The source code as well as the documentation for the whole tool can be found in the public repository[2]. The analytical dashboard was then implemented using ArcGIS, as this system is widely used within the Brno City Council.

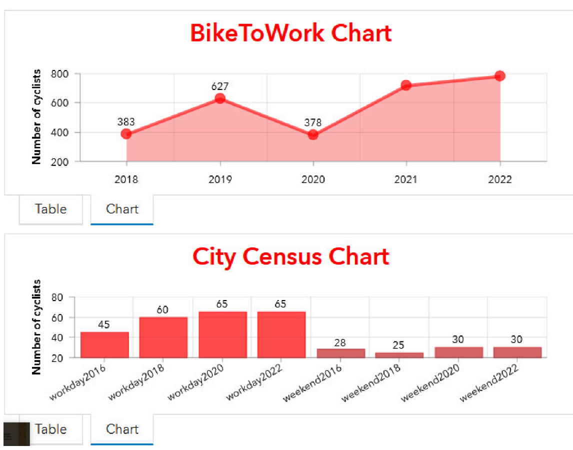

Fig. 6, Two charts from the dashboard displaying numbers of cyclists detected on Královopolská street in the May of 2022.

Verifying the accuracy of the generated mapping required manual annotation of the data, as this is the only way to verify the accuracy of the automatically generated mapping. However, manual annotation of 65,000 streets would be very difficult. Therefore, a small representative sample of 21 elements was selected. The individual elements were selected with an emphasis on as much variability as possible. Thus, streets from the centre of Brno with multiple parallel streets as well as country roads from the suburbs were included. Each selected element had a different length, angle and trajectory. The annotation itself consisted in visualizing the selected features in all tested datasets to make the selection of matching records as accurate as possible.

[2] https://github.com/Radluy/Cycling-traffic-intensity-Brno

The results of the comparison of the expected and actual mappings agreed on average 78% after testing. While for example the Bike to Work dataset the agreement averaged around 76%, for the BKOM census the success rate was 81%. Particularly, the streets whose trajectory was very close to the horizontal or vertical direction proved to be problematic, which meant a significant reduction in the content of their bounding box. The second problematic group of streets where the generated mapping did not match the expected mapping included streets with parallel pavement. In these cases, the algorithm tied the mapping to the sidewalk instead of the street.

Different views of the integrated data were modelled as a part of the analytical dashboard implementation using ArcGIS. While basic visualizations included, for example, the total number of bicyclists crossing a selected street at a specific time shown in Figure 6, more advanced analyses focused on, for example, comparing bicycle traffic volumes between weekday and weekend, or heatmap of cycling intensity as shown on Figure 7 and Figure 8.

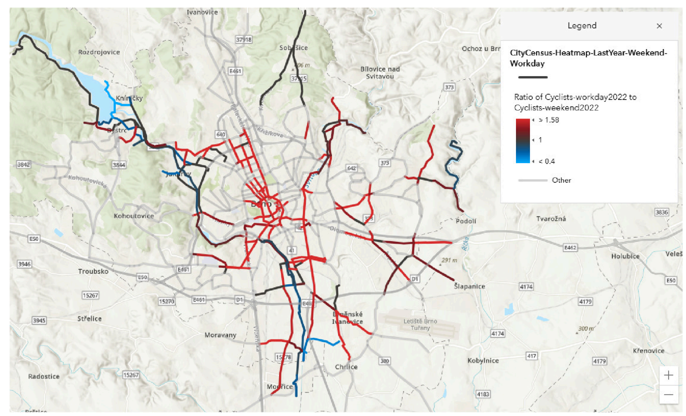

Fig. 7, Comparison of cycling numbers during one workday and one weekend day in 2022 from the City Census datasets. The red indicator represents that workday numbers have the majority while blue is the other way around. The rest is defined as a spectrum between these two colors shown in the legend.

VI. Discussion

The result of this work is an algorithm that allows the integration of different datasets describing bicycle transport in Brno. The proposed solution has been validated by implementing the algorithm and testing it on a selected set of manually annotated data. The usability of the integrated data was then tested by implementing an interactive dashboard. The resulting algorithm can generate reusable mappings of different geographic datasets to a common underlying network with an accuracy of 78%. Due to its generality, the proposed algorithm can also be used in other cases of geographic dataset integration. The solution itself simplifies the process of regular integration of new data, which can contribute to simplify the accessibility and increase the quality of planning and management of cycling infrastructure. On the other hand, the proposed algorithm is not yet able to deal successfully with some anomalies such as very close parallel streets.

Further development of the proposed solution should focus on two areas. The first area is to increase the reliability of the algorithm itself, for example by introducing a bounding box in the form of an oval, instead of a rectangle. Such a solution could help to increase the overall accuracy, especially in specific cases with significantly smaller bounding box sizes. The second area of further development should focus on defining a custom method for indexing records of integrated datasets. Custom indexing would allow better detection of new objects, thus reducing the proportion of generated mappings that need to be updated.

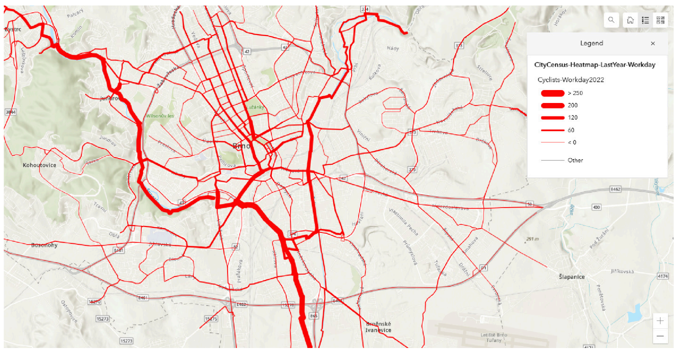

Fig. 8, Heatmap visualizing the number of cyclists from the City Census dataset in one workday of 2022.

VII. Conclusion

The aim of this work was to find a way to integrate different datasets describing cycling data of Brno city. The proposed solution consisted in the design of an algorithm that combines two of the most commonly used methods for the integration of geographic data. The output of the algorithm is a mapping that allows the repeated integration of different datasets into a single underlying traffic network. The design of the solution has been validated by implementing the algorithm and testing with annotated data. Subsequently, the usability of the integrated data was tested by implementing an interactive dashboard in close cooperation with the Brno City Council. The resulting solution demonstrates the way in which a combination of ontology and spatial join methods can be used to solve problems that are not solvable using only one of these methods. The proposed algorithm enables automated integration of different datasets, thus simplifying the process of planning and management of cycling infrastructure, which can contribute to increasing the satisfaction and safety of cyclists.

References

[1] G. Romanillos, M. Zaltz Austwick, D. Ettema, and J. De Kruijf, “Big Data and Cycling,” Transport Reviews, vol. 36, no. 1, pp. 114–133, Sep. 2015, doi: 10.1080/01441647.2015.1084067.

[2] A. Hull and C. O’Holleran, “Bicycle infrastructure: can good design encourage cycling?,” Urban, Planning and Transport Research, vol. 2, no. 1, pp. 369–406, Jan. 2014, doi: 10.1080/21650020.2014.955210.

[3] J. Dill, “Bicycling for Transportation and Health: The Role of Infrastructure,” Journal of Public Health Policy, vol. 30, no. S1, pp. S95–S110, Jan. 2009, doi: 10.1057/jphp.2008.56.

[4] “Spatial Join of Big Data,” Encyclopedia of GIS, pp. 2032–2032, 2017, doi: 10.1007/978-3-319-17885-1_101270.

[5] M. Dozza, “Crash risk: How cycling flow can help explain crash data,” Accident Analysis & Prevention, vol. 105, pp. 21–29, Aug. 2017, doi: 10.1016/j.aap.2016.04.033.

[6] H. Charvátová, A. Procházka, S. Vaseghi, O. Vyšata, and M. Vališ, “GPS-based analysis of physical activities using positioning and heart rate cycling data,” Signal, Image and Video Processing, vol. 11, no. 2, pp. 251–258, Jun. 2016, doi: 10.1007/s11760-016-0928-z.

[7] A. Procházka, O. Vyšata, O. Ťupa, M. Yadollahi, and M. Vališ, “Discrimination of axonal neuropathy using sensitivity and specificity statistical measures,” Neural Computing and Applications, vol. 25, no. 6, pp. 1349–1358, May 2014, doi: 10.1007/s00521-014-1622-0.

[8] S. Joo and C. Oh, “A novel method to monitor bicycling environments,” Transportation Research Part A: Policy and Practice, vol. 54, pp. 1–13, Aug. 2013, doi: 10.1016/j.tra.2013.07.001.

[9] P. Ryus et al., “Guidebook on Pedestrian and Bicycle Volume Data Collection,” Jan. 2014, doi: 10.17226/22223.

[10] K. Lee and I. N. Sener, “Emerging data for pedestrian and bicycle monitoring: Sources and applications,” Transportation Research Interdisciplinary Perspectives, vol. 4, p. 100095, Mar. 2020, doi: 10.1016/j.trip.2020.100095.

[11] G. Wallentin and M. Loidl, “Agent-based Bicycle Traffic Model for Salzburg City,” GI_Forum, vol. 1, pp. 558–566, 2015, doi: 10.1553/giscience2015s558.

[12] FRANCKE, Angela; LIßNER, Sven. Big data in bicycle traffic. *Technical Paper BMVI*, 2017.

[13] E. S. Willberg, H. Tenkanen, A. Poom, M. Salonen, and T. Toivonen, “Comparing spatial data sources for cycling studies – a review,” Jan. 2021, doi: 10.31235/osf.io/ruy3j.

[14] K. Janowicz, “The role of space and time for knowledge organization on the Semantic Web,” Semantic Web, vol. 1, no. 1,2, pp. 25–32, 2010, doi: 10.3233/sw-2010-0001.

[15] L. van den Brink, P. Janssen, W. Quak, and J. Stoter, “Towards a high level of semantic harmonisation in the geospatial domain,” Computers, Environment and Urban Systems, vol. 62, pp. 233–242, Mar. 2017, doi: 10.1016/j.compenvurbsys.2016.12.002.

[16] H. Uitermark, “Ontology Construction for Geographic Data Set Integration,” The Ontology and Modelling of Real Estate Transactions, pp. 149–161, Oct. 2017, doi: 10.4324/9781315237978-11.

[17] RODRÍGUEZ, F. Hernández; ARANDA, G. Bravo; NAVARRO, A. Martín. An ontology-based approach to spatial information modelling. In: *Advances in GIS Research II: Proceedings of the Seventh International Symposium on Spatial Data Handling. London, Taylor and Francis*. 1996. p. 13-26.

[18] C. B. Jones, D. B. Kidner, L. Q. Luo, G. Ll. Bundy, and J. M. Ware, “Database design for a multi-scale spatial information system,” International journal of geographical information systems, vol. 10, no. 8, pp. 901–920, Dec. 1996, doi: 10.1080/02693799608902116.

[19] W. Huang, K. Kazemzadeh, A. Mansourian, and L. Harrie, “Towards Knowledge-Based Geospatial Data Integration and Visualization: A Case of Visualizing Urban Bicycling Suitability,” IEEE Access, vol. 8, pp. 85473–85489, 2020, doi: 10.1109/access.2020.2992023.

[20] S. You, J. Zhang, and L. Gruenwald, “Large-scale spatial join query processing in Cloud,” 2015 31st IEEE International Conference on Data Engineering Workshops, pp. 34–41, Apr. 2015, doi: 10.1109/icdew.2015.7129541.

[21] E. H. Jacox and H. Samet, “Spatial join techniques,” ACM Transactions on Database Systems, vol. 32, no. 1, p. 7, Mar. 2007, doi: 10.1145/1206049.1206056.

[22] H. M. Veenhof, P. M. G. Apers, and M. A. W. Houtsma, “Optimization of spatial joins using filters,” Advances in Databases, pp. 136–154, 1995, doi: 10.1007/bfb0000545.