Evaluation report

Bicycle Data (bicycle-data.de)
University of Oldenburg

31/03/2022

Author: Griet Vanwynsberghe, Julie Tieberghien and Hans Vermeersch
Project coordinated by Province of Overijssel

This project is supported by the Interreg North Sea Region Programme (Priority 4, Promoting green transport and mobility) of the European Regional Development Fund of the European Union.

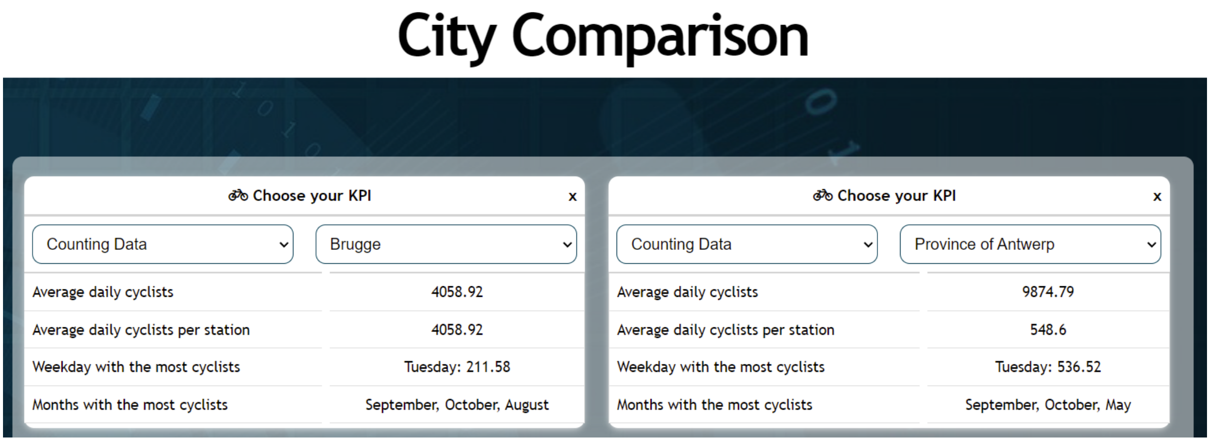
Description

What?

This report describes the Bicycle Data website, which is an extensive cycling data base. Within the scope of a student project at the University of Oldenburg (Business Informatics – Very Large Business Applications VLBA), under supervision of Prof. Dr. Ing. Jorge Marx Gómez and the research associates Johannes Schering and Harish Moturu, the website www.bicycle-data.de was developed. On this interactive website, cycling data gathered within the scope of the BITS project as well as other cycling data can be found. Users can compile and download the processed raw data and they can display analyses and visualizations. The data was enriched by geo coordinates and weather data. Data structures of similar types of data were harmonized (e.g. counting, near accidents) to make the database more valuable. Five types of cycling data are available: (1) bicycle counting, (2) bike sensors, (3) near accidents, (4) bicycle parking, and (5) bicycle app.

The Bicycle Data website was launched in spring 2021.

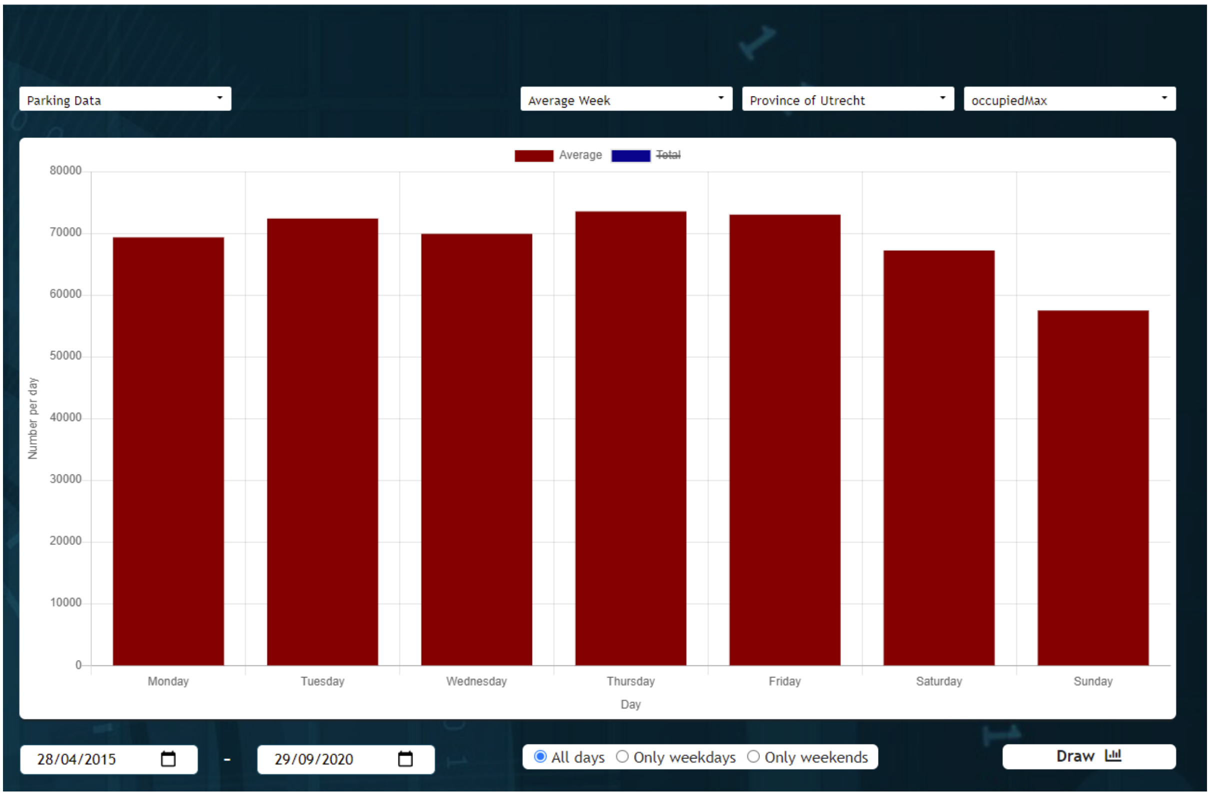
The website contains different tools. A first tool is the city comparison. With this tool the data of different regions and cities can be compared on all the KPI’s integrated in the website. Also different types of data can be compared, for instance the bike app with the sniffer bike. In the figure below, an example of the city comparison tool is shown. Here the example of counting data from Bruges and the Province of Antwerp is shown. The user can compare the average daily cyclists, the average daily cyclists per station, the weekday with most cyclists and the months with most cyclists.



In a second tool, the city analysis, users can analyse data of different cities and regions in more detail. A KPI and a city or region has to be chosen, next to time and date, at least if you want to have a look at the numbers of a certain year or month. The numbers in the city analysis tool provide more in-depth information and are more detailed compared to the city comparison tool.

Thirdly, in the graphical representation all the data previously mentioned can be visualised in bar diagrams. Trends in the data over days or years can be visualized easily. In this tool, different cities can be compared in terms of cycling amount or air quality as well. In the figure below, an example of parking data in the province of Utrecht is given.

Graphical Representation



Fourthly, in the tab Open Bicycle's Data specific data can be selected and can be downloaded. The user can choose between different KPI's, cities or regions, dates, locations and format of the data. Compared to other open data portals, the opportunity to choose between these categories is a clear benefit. It makes this data portal far more dynamic than other data portals. This data portal also has a generate link button. This makes it possible to share the data sets of the data base with other open data portals. This open data portal is also the connection and the interface to the CDH. Also on the website more information on the BITS project (and on specific pilots of the BITS project) and the Smarthelm project can be found.

Added value

The Bicycle data website has several added values. First and most importantly, it is the first platform that makes this much European cycling data openly available in a standardized and harmonized format. Never before a similar platform with so many data was built. Moreover, the raw data is freely accessible for everyone. The fact that KPI's were integrated only increases the value of this tool. These KPI's were developed in order to make the differentiated cycling data across cities and regions more measurable and comparable. Using the KPI's, it becomes possible to compare counting data collected in, for instance, the Province of Antwerp, the city of Bruges, Withernsea (UK) and the city of Oldenburg. The KPI's were determined following long internal discussions with BITS partners and even external stakeholders. This enhances another added value: besides BITS partners, other stakeholders, mainly German were engaged and were critical counterparts in the development of the platform.

Secondly, the platform not only contains the raw data, it also consists of enriched data sets. Other data, such as weather data and geo coordinates were afterwards added to data sets, which makes the data sets even more valuable and qualitative. To give a specific example: the weather conditions were added to the counting data collected in Withernsea (UK). These weather conditions can add additional hypotheses when interpreting the data. To illustrate: windy and rainy conditions lead to shorter trips and durations. In addition, the COVID-19 pandemic has increased the durations and distances what can be a sign for more recreational bicycle use. This was also confirmed by the counting values.

Thirdly, the Bicycle Data platform has the potential to publish all their data sets on other portals, such as for example the Cycle Data Hub. In this way, they contribute a lot (enriched) data sets to the Cycle Data Hub. The possibility to share the Bicycle Data data sets increases the availability of cycling data in general.

Fourthly, the data on the platform can be downloaded in different machine readable formats. Next to csv, data is also available in Json. In other words, a user of the platform always has the choice to pick a format that fits with his needs.

Next, the Bicycle Data website is the result of a master's student project within the University of Oldenburg. This means that it is the result of a larger and longer cooperation between 11 master students and their supervisors. Their final goal was to create an interactive website with raw and visualized data. The project was a win-win for both students and supervisors: the supervisors received support in developing the tool and the students were able to cooperate to an actual, authentic research question in real life applications and learned how to work in a larger collaboration project.

Also, since the platform is developed by a university team, there were a lot of opportunities to valorise the efforts in developing the platform and the outputs of the platform in scientific publications and conference proposals. It is the achievement of the VLBA team that a long list of publications is available nowadays. At the end of this report an overview of these publications can be found.

Finally and not least importantly, one of the merits of this project is the increased awareness for cycling data that has been created. Similarly to the CDH, the past years a larger data reflex within the domain of cycling data has been developed. For the first time all of these diverse cycling data sets are available at one certain spot what improves the accessibility and the awareness for cycling data. There is still a long way to go, however, compared to the context when the BITS project started in 2019, many changes and progresses have been made.

Lessons learned

Below, the main lessons learned during the process of developing the Bicycle Data website are summarized.

Firstly, the developers of the platform concluded that it is important to take into account any changes in the delivery of the data. Therefore, they recommend to **think about the type of the data to be collected on beforehand**. The team encountered many practical problems due to changed data structures during the process of the data collection and data analysis, which leaded to (huge) data problems and delays.

A second lesson learned concerns the request for **standardization** in cycling data. The same concepts are often measured differently. To give an example, near accident data collected in Zwolle and in Antwerp can often not easily be compared, due to two different causes. On the one hand the data structure differs (different attributes are included) and is not easily comparable. On the other hand concepts can differently be defined. In this example, when can we define a near accident as a near accident? Based on what criteria? Zwolle can use other criteria (different types of road users, different time intervals, different measurement periods) to define a near accident compared to Antwerp and thus the collected data from both cities cannot be compared. Therefore, more standardization is necessary. There is no common understanding to measure certain concepts. Every city or provider has another approach, which leads to the lack of comparability.

Thirdly, the project managers repeat the call to **make all cycling data available**. The more shared data on this platform, the more valuable it becomes. More data will attract more users, will bring more insights and comparisons and this will generate more new delivered data sets. And moreover, the more available data and comparisons, the more useful it becomes for data driven policy and policy recommendations.

Another lesson learned concerns the **user friendliness**. The platform was firstly developed within a student project. Afterwards, the developers presented the platform to different audiences, in the scope of the BITS project, on international conferences etc. They learned from the feedback of users and listeners on how the platform could be improved not at least concerning its user friendliness. The feedback made clear that users were not really familiar with the tool as they were not always able to find the data they needed. As a result, the full potential of the platform was not always seen by the users. This process of feedback shows the importance of learning by doing and how the user friendliness can be improved in a very efficient way.

Another major lesson learned concerns the **management of privacy relevant data** in terms of GDPR. A lot of data is collected and presented on the platform and some data is privacy sensitive (especially the tracking data). The VLBA team experienced the different perspectives and approaches concerning GDPR in the different NSR countries. The need to have a profound debate concerning GDPR is stressed. Especially in the field of anonymization of tracking data there are different philosophies and approaches how to deal with that. It is not clear what the users may accept and what not. Finding a good balance between GDPR and making tracking data available is very important.

A sixth lesson learned deals with the **need of expertise in the domain of cycling data**. The VLBA team has a lot of experience with the integration, processing and analysis of large data sets, but not necessarily in the domain of cycling data. Sometimes during the development of the platform, they missed an expert who could support them in terms of content. For example, the sniffer bikes collected data on air quality. As a non-expert in air quality data, it was hard to make interpretations about how bad or moderate air quality had been. Or another example: what are realistic outliers in these kind of data collections and which are not realistic anymore? Support of a domain expert is desirable in order to make more and better interpretations of the data.

A following lesson learned concerns the **dependency of and cooperation with partners** within such a huge project. The developers of the platform were dependent for the data sets of the BITS partners and their data collections. As long as they didn't deliver the data to the team in Oldenburg, the team couldn't continue developing and expanding their platform. This caused some unexpected delays. Likewise, given that the platform was developed in cooperation with a team of university students, timing was even more important here. Participation by students is always bounded to the calendar of an academic year. More sensitivity on this is needed in further projects.

A final lesson learned concerns **the delivery and the integration of the data**. The Bicycle Data team had to send out several requests to receive the data sources from the partners. And next to this, the integration of the data sets had to be done manually and takes a lot of time. Moreover, this is a potential source of error. Thoroughly preparing this process in advance in future projects is a lesson learned and is an advice to future developers.

Conclusion

To conclude, a progressive, unique data platform was created by the VLBA team at the university of Oldenburg. The tool contains many valuable, often enriched data sets and thanks to e.g. the KPI's, comparisons of cycling data are more straightforward. The development process also learned that standardization of cycling data is needed to be able to compare the data more easily. Also lack of data and inconsistency concerning GDPR are issues to be aware of. In short, taking into account the BITS objectives, the pilot definitely reaches its goals. This tool may contribute indirectly to more cyclists and less CO2 emission. The overview of the data and the insights learned from the data offers valuable opportunities to be used in future policy recommendations and decisions. Next to the Cycle Data Hub, this tool encourages and contributes to the overall data reflex needed in the domain of cycling data.

Publication list

Höper, T., Schering, J., & Marx Gómez, J. (2021). Can smartphone generated cycling data contribute to the improvement of the bicycle infrastructure?. *INFORMATIK 2021. Gesellschaft für Informatik*, Bonn, pp. 127-135. DOI: 10.18420/informatik2021-009.

Schering. J., Janßen, C., Kessler, R., Dmitryev, V., Marx Gómez, J., Brouwer, W., Kamermans, A., & Verweij, L. (2021). Bike sensor applications in Germany and the Netherlands gather new data on bicycle use. In Halberstadt, J. (ed.). YEEES Conference 2021– A German-African Digital Conference: Resilience through Sustainable Entrepreneurship, Digital Solutions and Education. Booklet of Extended Abstracts. Vechta: pp. 76-78.

Schering, J., Janßen, C., Kessler, R., Dmitriyev, V., Marx Gómez, J., Stehno, C., Pelzner, K., Bankowsky, R., & Hentschel, R. (2021). ECOSense and its preliminary findings: Collection and analysis of bicycle sensor data. In A. Kamilaris, V. Wohlgemuth, K. Karatzas, & I. Athanasiadis (Ed.), *EnviroInfo 2020 Environmental Informatics – New perspectives in Environmental Information Systems: Transport, Sensors, Recycling.* Adjunct Proceedings of the 34th EnviroInfo conference. Düren: Shaker Verlag, pp. 145-153.

Schering , J., Janßen, C., Kessler, R., Dmitriyev, V., Stüven, J., Marx Gómez, J., van Dijk, E., Brouwer, W., Kamermans, A., Verweij, L., & Janssen, G. (2022). ECOSense and Sniffer Bike: European bike sensor applications and its potential to support the decision making process in cycling promotion. In J. Marx

Gómez, & M. R. Lorini (Ed.), *Digital transformation for sustainability. ICT-supported Environmental Socio-economic Development*. Cham: Springer Nature Switzerland. (in publication process)

Schering. J., & Marx Gómez, J. (2021). BITS: Open Bicycle Data and its potential for the improvement of the bicycle infrastructure. In A. F. Valderrama Pineda (Ed.), *Book of Abstracts of the 5th Annual Meeting of the Cycling Research Board Conference.* Copenhagen: pp. 78-79.

Schering, J., & Marx Gómez, J. (2022). A New Research Approach to Harmonize European Bicycle Counting Data from Germany, Belgium and the UK. In J. Halberstadt (Ed.), *YEEES Conference 2021 Proceedings*. Cham: Springer Nature Switzerland. (in publication process)

Schering, J., & Marx Gómez, J. (2022). The BITS Project – Making cycling data available and comparable on a European scale. *Transportation Research Procedia 60*, pp. 424-431. ISSN 2352-1465. https://doi.org/10.1016/j.trpro.2021.12.055.
(https://www.sciencedirect.com/science/article/pii/S235214652100956X)

Schering, J., Marx Gómez, J., Soetens, S., Verbeek, K., & Singh, A. (2021). How to measure safety risks for cyclists at intersections? In J. Halberstadt, J. Marx Gómez, J. Greyling, T.K. Mufeti, H. Faasch (Ed.), *Resilience, Entrepreneuship and ICT - Latest Research from Germany, South Africa, Mozambique and Namibia*. Cham: Springer Nature Switzerland, pp. 263-278.