
This is an electronic reprint of the original article.
This reprint may differ from the original in pagination and typographic detail.

Willberg, Elias; Tenkanen, Henrikki; Poom, Age; Salonen, Maria; Toivonen, Tuuli
Comparing spatial data sources for cycling studies – a review

Published in:
Transport in Human Scale Cities

DOI:
[10.4337/9781800370517.00025](https://doi.org/10.4337/9781800370517.00025)

Published: 12/08/2021

Document Version
Publisher's PDF, also known as Version of record

Published under the following license:
CC BY-NC-ND

Please cite the original version:
Willberg, E., Tenkanen, H., Poom, A., Salonen, M., & Toivonen, T. (2021). Comparing spatial data sources for cycling studies – a review. In *Transport in Human Scale Cities* (pp. 169-187). (NECTAR Series on Transportation and Communications Networks Research). Edward Elgar.
<https://doi.org/10.4337/9781800370517.00025>

This material is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the repository collections is not permitted, except that material may be duplicated by you for your research use or educational purposes in electronic or print form. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone who is not an authorised user.

14. Comparing spatial data sources for cycling studies: a review

Elias Willberg, Henrikki Tenkanen, Age Poom, Maria Salonen and Tuuli Toivonen

COLLABORATORS

Eran Ben-Elia, Ilil Beyer Bartana, Catarina Cadima, Tiziana Campisi, Gamze Dane, Elise Desjardins, Malene Freudendal-Pedersen, Zahra Hamidi, Ilari Heiska, Oddrun Helen Hagen, Olaf Jonkeren, Joris Klengen, Eeva Kostianen, Jukka M. Krisp, Kevin Krizek, Emilio Larrodé, Alain L'Hostis, Cyrille Medard de Chardon, Anna Nikolaeva, Samuel Nello-Deakin, Jesse Pappers, Petr Pokorny, John Pritchard, Ray Pritchard, Tanu Priya Uteng, Adam Radzinski, Mikko Raninen, Janika Raun, Gustavo Romanillos Arroyo, Federico Rupi, Erki Saluveer, Hans Skov-Petersen, Guilherme Tampieri, Maciej Tarkowski, Ainokaisa Tarnanen, João Filipe Teixeira, Ignacio Tiznado-Aitken, Kalle Vaismaa, Dea Van-Lierop, Elisabetta Venezia, Luca Zampanini, John Östh.

14.1 INTRODUCTION

Cycling is considered one of the core travel modes for developing human scale cities. With zero emissions and pollution, cycling has also been shown to increase physical activity levels and foster social connectedness (Garrard et al., 2012; te Brömmelstroet et al., 2017; Kroesen and De Vos, 2020). Compared to car travel, cycling saves space for other urban functions and acts as an important feeder mode for public transport (Geurs et al., 2016). Consequently, urban planning practice and mobility research now widely promote cycling as a sustainable transport solution (Martens, 2007; Pucher and Buehler, 2017). Wider acceptance and technological development have diversified the modes of cycling. Innovations such as electronically assisted bikes (e-bikes) and bike sharing systems (BSS) are rapidly being adopted, replacing the use of conventional bikes and, to a lesser extent, car trips (Bigazzi and Wong, 2020; Ma et al., 2020). Smart cycling technologies are getting more attention at different levels of governance, as they alter how this mode of travel is experienced,

understood and governed (Nikolaeva et al., 2019). Urban planning practice and mobility research share a growing need for understanding better the spatial, temporal and social aspects of cycling to respond to the rapidly changing scene in cities and attract more people onto bikes. To foster human-scale transportation, cities need to shift emphasis from the focus on understanding motorized transport to understanding cycling and active travel.

Despite the increasing need, lack of cycling data and access to it is still a common challenge. Absence of applicable data has been recognized as a limiting factor for transport and urban planning in many studies, also in mature cycling countries such as the Netherlands (Iacono et al., 2010; Aultman-Hall et al., 2012; Nello-Deakin and Harms, 2019; Aldred et al., 2019). While data collection processes with motorized transport are typically well established by transportation agencies, cycling-related data has rarely been included in these processes (Aultman-Hall et al., 2012). Aultman-Hall et al. (2012) note that the lack of established processes has led to this data being collected on a project-by-project basis at isolated locations, precluding system-wide extrapolation. Part of the challenge lies in the traditional ways of collecting data using travel diaries, surveys, interviews, static bike counters and various statistics. While these sources are useful, there are limitations in their availability and comprehensiveness. Some common challenges include laborious data collection, small sample sizes, insufficient spatio-temporal coverage, route and travel chain detection, and reliance on self-reporting (Bird et al., 2013; Handy et al., 2014).

The rapid development of sensor technologies has enabled mobility data collection from a variety of sources that uncover mobility patterns and daily movements of people (Kitchin, 2014; Romanillos et al., 2016; Pritchard, 2018). Cycling data sources originating from sensors and location-based services (Huang et al., 2018) include a variety of examples such as BSS (O'Brien et al., 2014; Fishman, 2015; Klingen and Van Ommeren, 2020), eye-tracking (Rupi and Krizek, 2019), GPS positioning through various devices or mobile applications (Romanillos et al., 2018; Rupi and Schweizer, 2018; Dane et al., 2020), social media (Das et al., 2019), street view imagery (Goel et al., 2018), video recording (Pokorny and Pitera, 2019) and virtual reality simulators (Leeuw and Kruijf, 2015). Furthermore, tracking the use of bikes is increasingly integrated into the Internet of Things (IoT) fabric of smart cities through smart applications and devices such as smart wearables (Nikolaeva et al., 2019), which further add to the increasing data flux.

However, like traditional cycling data, novel data sources have biases and usability challenges that relate to availability, cost and insufficient spatio-temporal coverage of data, and mode and route detection (Goel et al., 2018; Lee and Sener, 2020; Sun and Mobasher, 2017). There are also other challenges limiting novel data sources' potential, such as privacy

issues, demographic representativeness and generalizability, and requirements for computational processing capacity and know-how (Oksanen et al., 2015; Romanillos et al., 2016; Médard de Chardon, 2019). Despite a rapidly growing body of literature exploring the possibilities, a comprehensive view on how different novel cycling data can best benefit planning and research is still missing. Fostering such understanding can help researchers and planners to better evaluate and choose the most fitting data for different needs, which in turn can lead to saving resources and efforts in data collection and acquisition, and ultimately to more reliable and representative outcomes that support the growth of cycling in cities.

The present chapter provides a comparison of contemporary data sources available for cycling research and planning, building upon the expertise of a broad group of cycling experts. We review common data sources in the literature in relation to their capability to answer some of the most common questions on cycling asked by urban and transportation planners, including:

1. Where are people cycling (spatial patterns)?
2. When are people cycling (temporal patterns)?
3. Who are the cyclists (user demographics)?
4. What are their cycling types (work, leisure, and so on)?
5. Why do people choose to cycle (the motivations for and barriers to cycling)?

With each data source, we review separately different aspects such as availability; spatial, temporal and user information; and typical biases. First, we identified different empirical data sources from the recent cycling-related scientific literature. Next, we implemented an expert survey with a short questionnaire among researchers and practitioners with experience in using empirical cycling data sources. The survey was distributed among the Network on European Communications and Transport Activities Research (NECTAR) researchers, and their own networks of experts. The questionnaire allowed respondents to assess the strengths, weaknesses and usability of data sources, to give justifications and to point out relevant literature. Additionally, we asked respondents to assess different data sources on a five-point Likert scale (from very poor to very well) on how well the data sources fit for ‘where’, ‘when’, ‘who’, ‘what’ and ‘why’ types of questions. A set of example questions like those presented above was provided to direct evaluations. The responses allowed us to map the applicability of each data source. The questionnaire was completed by 58 respondents who provided 71 distinct assessments of cycling data sources. To test an open and co-creative writing process, every respondent was further invited to review the manuscript in a collaborative platform Authorea.

Before continuing, it is necessary to highlight some limitations of our approach. First, there are various cycling-related questions in planning and research that do not fall into our categories. Yet, the questions reviewed here represent the basic information typically needed by planners and researchers, being a prerequisite for more applied questions. Second, scale is a challenge. For example, a data source may excel at capturing spatial patterns at street level, but not be applicable for monitoring broader areal cycling patterns, which both nevertheless fall under the jointly evaluated ‘where’ category. Third, some of the reviewed data sources suffer from the small numbers of expert evaluations, which adds uncertainty to their evaluation.

14.2 DESCRIBING DATA SOURCES FOR CYCLING

Data sources mostly used in cycling analyses include three categories: (1) bike counters and manual observations; (2) BSS, Global Positioning System (GPS) tracking, sports applications; and (3) surveys, interviews and public participation geographic information system (PPGIS). These sources are described briefly below, and in more detail in Table 14.1 where their characteristics are divided into components of availability, typical attributes, spatial, temporal and user information, and finally providing literature examples for each data source. The list of the reviewed sources (Table 14.1) is by no means exhaustive and there are various other data sources that can benefit cycling research. Yet, the reviewed data sources represent commonly collected and available data, and were assessed by at least three different experts in our survey. Within the categories, the data sources are ordered by the number of expert evaluations.

14.2.1 Bike Counters

Counters are a simple way to collect data on cycling levels and to monitor bike traffic longitudinally. Automatic bike counters use a variety of detection techniques ranging from infrared and radar waves, pneumatic tubes, inductive loops to video or thermal imaging, magnetometers and piezoelectric detectors (Alta, 2016). Counters are statically attached to a specific location and commonly located in the hotspots of bike traffic. Typically, counter data includes the coordinates of the counter, and the data is collected by a fixed interval (for example, minute, quarter-hour, hour, day).

14.2.2 Manual Observations

Manual observations refer to a variety of data that is collected by an observant from single or multiple locations at a single occasion or at regular intervals. This type of data can include simple cyclist counts, subcounts (by age, gender,

Table 14.1 Typical characteristics of the reviewed cycling data sources

Data type	Accessing the data	Typical attributes	Spatial accuracy and coverage	Temporal information	User information	Example studies
Bike counter data	Through data portals of public sector agencies or through measuring campaigns	Counter ID, bike count, time	Spatial coverage is limited to the counter locations. Spatial accuracy is good when the exact locations of counters are known. No track records.	Automatic counters provide year-round continuous data with good temporal resolution.	Aggregated cyclist counts by time intervals, no individual user data.	Nordback et al., 2011; Rupi et al., 2019
Manual observation data	Through observation campaigns	Observation ID, lat./long. coordinate, time, observation count / description	Spatial coverage is limited to places the observers are located or visit. However, the locations/routes can be selected to suit the study purpose.	Temporal coverage is limited to observation periods, repeated observations possible.	Detailed individualized data in active ride-along observations; user counts and basic description of subcounts in passive observations.	Fraboni et al., 2018; Gamble et al., 2017; Van Duppen and Spierings, 2013
Bike sharing origin-destination trip data	Through the system operator or from open data portals	Trip ID, depart/return station, depart/return time, user ID	Typically OD records without route information. Spatial accuracy is high as it is linked to the station locations. Coverage bound to the station network.	Time stamps available for hires. Data collected throughout the system operation period, often data available from continuous years.	Typically anonymized individual-level data, sometimes enriched with basic demographic variables (e.g. age, gender) or subscription type.	Klingen and Van Ommeren, 2020; Médard de Chardon and Caruso, 2015; O'Brien et al., 2014

Data type	Accessing the data	Typical attributes	Spatial accuracy and coverage	Temporal information	User information	Example studies
GPS-tracking data	Through GPS-tracking campaigns or survey apps	User ID, track ID, lat./long. coordinates timestamp	High and accurate spatial coverage, track records show exact routes; potential data gaps or location errors due to app late start, irregular frequency of data collection, signal loss or multipath signals in urban canyons.	High temporal resolution depending on the data collection settings, typically from 1 second to 1 minute.	Detailed individualized data may be collected separately in the data collection campaign or survey app.	Dane et al., 2020; Romanillos et al., 2018; Rupi and Schweizer, 2018; Rötzer et al., 2019
Sports application data	Data products accessed by purchasing or through collaborations with responsible companies or public-private sector hybrid collaborations	Segment ID, time, cyclist count, cyclist count by trip types	High spatial accuracy with street segment-level data. Spatial coverage is typically good, but depends on the number of users of the given sports application in the study area.	High temporal resolution and coverage (raw data can contain minute-by-minute cycling counts of street segments).	Either aggregated user counts or limited anonymized user information (e.g., age, gender) depending on the data product.	Hochmair et al., 2019; Oksanen et al., 2015; Sun and Mobasheri, 2017

Data type	Accessing the data	Typical attributes	Spatial accuracy and coverage	Temporal information	User information	Example studies
Survey data	Through survey campaigns, open data portals or public sector agencies	Respondent ID, answers, comments	Typically limited spatial coverage (e.g. main origin and destination location, sometimes route characteristics). Population-level travel surveys include spatial data on a coarse level (e.g. postal code or census tract).	Limited temporal coverage. With recurrent surveys, longitudinal analyses possible.	Anonymous and detailed individual-level data.	Nello-Deakin and Harms, 2019; Nielsen and Skov-Petersen, 2018; Silva et al., 2019
Interview data	Through interview surveys	Respondent attributes, answers to questions	Generally limited spatial coverage and accuracy. May still provide detailed individual descriptions on specific areas or routes.	Generally limited temporal coverage and accuracy. Longitudinal data possible if interviews are repeated.	Detailed individualized data.	Freudendal-Pedersen, 2015; Nello-Deakin and Nikolaeva, 2020; Médard de Chardon, 2019
PPGIS data	Through PPGIS-survey campaigns, open data portals or public sector agencies	User ID, marker ID, answers, comments	Based on markings/drawings on a background map. Spatial accuracy and coverage typically varies by user. Spatial data quality is uncertain.	Limited temporal coverage unless the PPGIS survey is designed to capture variation in time with specific questions.	Anonymous individual-level data, data collection incorporated to survey questions.	Pritchard et al., 2019; Snizek et al., 2013

Note: OD = origin-destination. ID = identification.

BSS user/regular, and so on) or bike type counts (for example, regular, sport, cargo bike), recognizing whether people cycle in groups or alone, or evaluating cycling behaviour, for example at an intersection to recognize potential hazard risk. Manual observations can also take the form of ride-along observations where the observant is cycling with the study participant.

14.2.3 Bike Sharing Trip Data

BSS have become commonplace in many cities. Typically, these systems collect information on bike rental trips with each trip as an individual record. These records typically reveal the origin and the destination station, depart and return time, and the user and bike identifications (IDs). Depending on the data provider and the system registration settings, trip-level datasets may be linked to basic user variables such as age, gender, postal code and subscription type. Trip datasets from docked systems do not contain exact routes between the stations unless the GPS data is separately collected. Application-based dockless BSS, in turn, track users' movements through positioning, but these datasets are rarely publicly available.

14.2.4 GPS Tracking

GPS tracking is increasingly integrated into various cycling data sources. Here we refer to GPS tracking that is recorded with a separate GPS device or a specific location tracking application on a portable device (for example, mobile phone or tablet) even though GPS tracking can also be integrated to other sources of data reviewed in this chapter, such as to BSS and sports application data. GPS tracking records spatially and temporally precise location track points or line sequences. GPS devices can be handheld and carried by the cyclist or be attached to the bike frame. Tracking data is linked to a particular user and contains time-stamped geographical coordinates for every track point or line sequence, measured typically with a short interval ranging from seconds to minutes.

14.2.5 Sports Applications

Activity, sports and health monitoring applications such as Endomondo, Strava or Sports Tracker are increasingly used as a source of cycling data. Depending on the purpose and the sophistication of the application, they allow users to track their movements, heart rate, altitude, travelled speed and distance, and set up performance goals and share achievements with fellow users (Piwek et al., 2016). As the data is almost exclusively collected by private organizations, derived data products have variation in relation to availability, data type, form,

pricing and variable characteristics. Due to privacy concerns, most datasets are aggregated. For example, the Strava Metro data product consists of street geometries with minute-by-minute aggregated user counts, but without individual user information.

14.2.6 Surveys

Survey data on cycling can vary from short and targeted questionnaires focused on a single issue to large country-level mobility surveys focused on general travel patterns of the population including cycling. Survey data are primarily collected in a disaggregated form; however, when shared through intermediate sources, survey data are often aggregated to protect respondents' privacy. Surveys are increasingly implemented with online forms, which has made it easier to distribute them and collect data from wider groups. National travel surveys are typically collected yearly or every 2–3 years during a few days of extensive data campaigns.

14.2.7 Interviews

Interviews provide qualitative data in the form of shared experiences and stories, typically from a small sample of people. Interviews can take place individually between the interviewer and the interviewee, or in a group environment where multiple interviewees come together. Structure of interviews varies from non-structured and free-flowing discussions to semi-structured and structured interviews with standardized questions presented to each interviewee. In the cycling context, interviews can shed light, for example, on the reasons and incentives for as well as the barriers to cycling.

14.2.8 PPGIS

Public participation GIS (PPGIS) mapping refers to geospatial techniques that allow the collection of both qualitative and quantitative data through active involvement of individuals (Brown and Kytä, 2014). Using a typical PPGIS application, participants can map places (points), routes (lines), areas (polygons) or combinations of these, together with responding to survey questions. In the cycling context, PPGIS can be used to collect information on, for example, origins and destinations of trips, route choices or areas of interest, and to link this information to individual place-related experiences and preferences attached to the mapped entities.

14.2.9 Other Sources

Other sources of cycling data mentioned in the expert survey include cycling statistics, BSS station count data, Google Street View images, eye-tracking and video recording of cyclists, travel diaries, social media posts extracted from platforms such as Twitter and virtual reality (cycling simulators). While many of these may be potential for studying cycling, we do not review them here due to the small number of responses.

14.3 ASSESSING DATA SOURCES FOR CYCLING

Our survey respondents assessed the cycling data sources' applicability for different types of analytical questions. The results, shown in Figure 14.1 with radar charts, present a clear distinction between more quantitative sources (BSS trip data, bike counters, GPS tracking, manual observations and sports applications) and qualitative sources (surveys, interviews and PPGIS). The radar charts are ordered by the number of assessments, starting from the top left. All data sources in the former category exhibit better applicability to spatio-temporal 'where' and 'when' types of questions, while the latter category of data sources has the opposite patterns with strong points in 'who', 'what' and 'why' types of questions. The results point to a trade-off with using sources that are able to measure precise spatio-temporal locations or even cycling routes, but often lack individual and semantic information; and correspondingly using sources that are richer and more nuanced in user profile and information on motivation, but may lack spatio-temporal coverage and accuracy.

We further asked the survey respondents to report strengths, weaknesses and biases with different cycling data sources based on their experiences. Table 14.2 summarizes the results of these responses. The results capture similar differences of analytical capability between the data sources as Figure 14.1, but bring forward challenges of data collection and availability. The availability of data or the ease of data collection is listed as a strength for bike counter, BSS, PPGIS and survey data. This implies that these sources are either commonly available openly or for research purposes, or can be collected with a relatively small effort. As with GPS tracking, interview, manual observation and sports application data, data access or collection effort is considered as a weakness requiring considerable human or financial resources and therefore limiting their use.

Notable aspects in relation to reported biases are the user-related biases inherent in many data sources. Assessments show that GPS tracking, bike sharing trips and sports application datasets often suffer from participation bias, as most of the data is recorded by active users who may represent only



Figure 14.1 Expert assessments on the fitness of cycling data sources for answering different types of cycling-related questions

Table 14.2 Expert-reported strengths and limitations of cycling data sources

Data	Main strengths	Known biases/limitations
Bike counter data (n = 11)	<ul style="list-style-type: none"> – Unbiased data – Reliability typically good – Temporally comparable data – Increasingly available data 	<ul style="list-style-type: none"> – Lack of spatial coverage – Lack of user information – Accuracy of counters varies depending on the counting method – Poor reliability with some counters when snowing – Data gaps typical – Maintenance can be costly – No semantics (e.g., trip purpose)
Manual observations (n = 4)	<ul style="list-style-type: none"> – Particularly useful for conflict observation or hazard identification – Can be carried out anywhere – Variety of different kind of observations can be captured at once 	<ul style="list-style-type: none"> – Limited spatio-temporal reach – Only samples – Subjective measurer bias inherent – Designing a successful observation campaign, which captures targeted issue at the right time and place is challenging – Requires lot of human resources
Bike sharing OD trip data (n = 14)	<ul style="list-style-type: none"> – Large sample sizes, records may contain even millions of trips – Potentially high spatial coverage (with extensive systems) – High temporal resolution and coverage – Individualized, yet anonymous data – Increasing availability from many cities – Shows activity (i.e., single trips can be distinguished) 	<ul style="list-style-type: none"> – Lack of socio-demographic data – No route information – User demographics often biased to certain groups – Super-users and/or residents near stations may generate majority of data – No semantics (e.g., trip purpose)
GPS tracking (n = 10)	<ul style="list-style-type: none"> – High spatial and temporal coverage and detailness – Reveals cycling routes and route choices – Differentiation of cyclists' types (e.g., commuters, leisure) possible with enough data – Speed information included – Can be collected through mobile phone apps 	<ul style="list-style-type: none"> – Difficult to collect large data samples – Requires dedicated data collection campaigns that need resources – Participant bias (e.g., by age or gender or enthusiasts only) – Privacy issues – No semantics (e.g., trip purpose) – App malfunctions may impair results

Data	Main strengths	Known biases/limitations
Sports application data (n = 4)	<ul style="list-style-type: none"> – Large masses of data – High spatial and temporal coverage and preciseness – Street segment-level cycling count analyses possible – Some data products can differentiate between different activities (e.g., commuting, leisure) 	<ul style="list-style-type: none"> – Data products typically aggregated, no individual data – Data not openly available as collected by private actors – Typically heavily biased to athletic cyclists and male users – Active user bias – No semantics (e.g., trip purpose)
Survey data (n = 10)	<ul style="list-style-type: none"> – Can be targeted to specific groups – Provides understanding on perceptions, motivations and barriers, i.e., unmeasurable issues – The data readily available if official survey – Easy to collect online – Comparability with other survey data 	<ul style="list-style-type: none"> – Response bias (e.g., by age or gender or enthusiasts only) – Sample sizes often limited – Answers might not reflect true behaviour – Trip chains and short trips often not reported – Poor temporal and spatial coverage and resolution
Interview data (n = 7)	<ul style="list-style-type: none"> – Rich contextual information on behaviours, experiences, choices and perceptions – Uncovers complexity (e.g., interplay of actors, realities of transport services) – Unbiased sample – Emphasis on human experience 	<ul style="list-style-type: none"> – Small sample sizes – Subjective bias – Expensive and time-consuming to collect and process – Difficult to obtain exact spatio-temporal data
PPGIS data (n = 3)	<ul style="list-style-type: none"> – Can distinguish spatially exact information together with rich user information – Easy-to-collect online – Can be targeted to specific groups/ areas 	<ul style="list-style-type: none"> – Typically small sample sizes – Uncertainty of data quality – Subjective bias – Uneven spatial coverage and resolution: most users map only a few locations/routes, disproportionate data from active users – Poor temporal coverage and resolution

certain population groups. To some extent, these biases are produced by uneven cycling demographics common in many places, but they may also be attributed to the uneven use of mobility tracking technologies among different demographic groups. Some respondents refer to sporty and male users being over-represented in sports application data, while among BSS users over-representation may be related to groups for whom the system is accessible. There are still ways to mitigate these biases. For example, with Strava data, Fischer et al. (2020) show that using only the trips labelled as commuting may better represent ridership patterns of population. User-related biases are differently present in interview, manual observation, PPGIS and survey data,

which may suffer from small sample sizes and subjective biases, which impair the generalizability of results obtained from these sources.

14.4 DISCUSSION

Emerging data sources have facilitated the increasing body of cycling research and broadened the research perspectives. However, our results highlight the need for careful consideration of data availability, representativeness and fitness for different analytical needs. Difficulties in data collection, lack of suitable data and poor accessibility to it were identified as the main limitations of several reviewed data sources, causing analytical complications in understanding cycling. Some of the critical data needs include representative information from large populations on detailed cycling routes and trip chaining behaviour, as well as the representation of people in the survey data who are less active as cyclists or do not cycle at all. In order to avoid the biases of single data sources, it is often crucial to combine multiple data sources, both traditional and novel (Tenkanen and Toivonen, 2020).

Many transport and planning agencies share cycling data that is often limited to counter or survey data. Increased data sharing such as making BSS trip datasets accessible might provide a good strategy for cities as it can generate novel research from them and therefore benefit the public policy (Willberg et al., 2021). Similarly, publishing anonymized or aggregated research data sets by researchers would benefit the understanding of cycling through new research opportunities. A few open data initiatives such as the European Cycling Challenge (2017) and Bike Data Project (2020) have also motivated citizens to share their cycling data, but coordinated efforts by cities to collect comparable, socially representative and longitudinal data are mostly lacking.

However, cycling data with necessary spatio-temporal detail and coverage is mostly collected by private companies. In many cases, individuals share their data with the companies via free sport or navigation applications, but getting access to this data for research and transport planning purposes is possible only at high cost or not at all. Developing transparent, privacy-respectful and sound data products as well as ways to access these could bring new opportunities to fill some current cycling data gaps.

Similarly to other types of personal data, collecting individual-level data on cyclists' behaviour requires balancing between privacy protection and the benefits that detailed data offers for understanding cycling patterns. For example, linking BSS records with user-specific health and socio-demographic data would be useful for revealing specific social, behavioural and health associations (Romanillos et al., 2016). In addition, detailed semantic information collected on top of quantitative trip data enables obtaining a more complete picture on the situational, socio-spatial and economic aspects of cycling.

However, due to the sensitivity of these joint datasets, collecting and linking various individual-level data requires well-justified reasoning, obtaining consent from the individuals, setting strict access rules to raw data and ensuring full compliance with research ethics. Thus, it is not possible to share rich individual-level data with the research community, and shared data products need to be either fully anonymized and limited in scope or aggregated.

Increased standardization of approaches could systematize cycling data collection and sharing, and facilitate comparative studies. Despite some efforts (e.g., Alta, 2016; Toole Design, 2019), the cycling data standards are not similarly developed as with motorized transport. Critical aspects for standardization include finding meaningful aggregation levels to retain privacy without compromising sensitiveness, explicitness of the spatial and temporal attributes, and inclusion of categorized demographic information when possible. Cycling data standards should also focus on data findability and accessibility. One interesting recent example is the 'Talking bike' initiative by the Dutch national government, which intends to create General Data Protection Regulation (GDPR) compliant and coordinated cycling data market for governments and municipalities to purchase standardized cycling data from selected private partners (NDW, 2020). In general, docking station-based BSS trip data is a good example of a solid and privacy-respectful cycling data type. In addition to being longitudinal and having a decent sample size, BSS data are inherent in spatial anonymity as docked stations seldom represent the true trip origin or destination. However, BSS data tends to be available mainly in cities in which the system is managed by a public sector institution. Public interest may still open new horizons in collecting contextual information on cycling such as equipping shared bikes with environmental sensors.

While any current cycling data source alone cannot provide comprehensive knowledge on cycling, novel datasets have increasingly helped to fill some gaps, especially in the spatio-temporal corners of cycling research. Yet, traditional data sources are still indispensable, especially in shedding light on the underlying reasons and motivations for cycling. By reflecting the dynamics of current cycling practices, novel and traditional datasets together empower transport and land use planners in their efforts to increase cycling and to transform urban transport systems to more sustainable, healthier and essentially more human scale systems. There is still a long way to go in bridging the gap in cycling data availability to facilitate the use of the most fitting data according to the need. We call upon everyone collecting and studying cycling to contribute to responsible data sharing and open science practices whenever possible.

REFERENCES

- Aldred, R., Watson, T., Lovelace, R., and Woodcock, J. (2019). Barriers to investing in cycling: stakeholder views from England. *Transportation Research Part A: Policy and Practice*, 128, 149–159.
- Alta (2016). *Innovation in Bicycle and Pedestrian Counts – A Review of Emerging Technology*. Retrieved from <https://altaplanning.com/wp-content/uploads/Innovative-Ped-and-Bike-Counts-White-Paper-Alta.pdf>.
- Aultman-Hall, L., Dowds, J., and Lee, B.H.Y. (2012). Innovative data collection for pedestrians, bicycles, and other non-motor vehicle modes. *TR News*, 280, 8–9.
- Bigazzi, A., and Wong, K. (2020). Electric bicycle mode substitution for driving, public transit, conventional cycling, and walking. *Transportation Research Part D: Transport and Environment*, 85, 102412. <https://doi.org/10.1016/j.trd.2020.102412>.
- Bike Data Project (2020). Retrieved from <http://bikedataproject.com/#donate>.
- Bird, E.L., Baker, G., Mutrie, N., Ogilvie, D., Sahlqvist, S., and Powell, J. (2013). Behavior change techniques used to promote walking and cycling: a systematic review. *Health Psychology*, 32(8), 829–838.
- Brown, G., and Kytta, M. (2014). Key issues and research priorities for public participation GIS (PPGIS): a synthesis based on empirical research. *Applied Geography*, 46, 122–136.
- Dane, G., Feng, T., Luub, F., and Arentze, T. (2020). Route choice decisions of E-bike users: analysis of GPS tracking data in the Netherlands. In: P. Kyriakidis, D. Hadjimitsis, D. Skarlatos and A. Mansourian (eds), *Geospatial Technologies for Local and Regional Development*. AGILE 2019. Lecture Notes in Geoinformation and Cartography. Cham: Springer. https://doi.org/10.1007/978-3-030-14745-7_7.
- Das, S., Dutta, A., Medina, G., Minjares-Kyle, L., and Elgart, Z. (2019). Extracting patterns from Twitter to promote biking. *LATSS Research*, 43(1), 51–59.
- European Cycling Challenge (2017). Retrieved from <https://civitas.eu/event/european-cycling-challenge>.
- Fischer, J., Nelson, T., and Winters, M. (2020). Comparing spatial associations of commuting versus recreational ridership captured by the Strava fitness app. *Transport Findings*. <https://doi.org/10.32866/001c.16710>.
- Fishman, E. (2015). Bikeshare: a review of recent literature. *Transport Reviews: A Transnational Transdisciplinary Journal*, 1647, 1–22.
- Fraboni, F., Marin Puchades, V., De Angelis, M., Pietrantonio, L., and Prati, G. (2018). Red-light running behavior of cyclists in Italy: an observational study. *Accident Analysis and Prevention*, 120, 219–232. <https://doi.org/10.1016/j.aap.2018.08.013>.
- Freudental-Pedersen, M. (2015). Whose commons are mobilities spaces? The case of Copenhagen's cyclists. *ACME: An International Journal for Critical Geographies*, 14(2), 598–621.
- Gamble, J., Snizek, B., and Nielsen, T.S. (2017). From people to cycling indicators: documenting and understanding the urban context of cyclists' experiences in Quito, Ecuador. *Journal of Transport Geography*, 60, 167–177. <https://doi.org/10.1016/j.jtrangeo.2017.03.004>.
- Garrard, J., Rissel, C., and Bauman, A. (2012). Health benefits of cycling. In: J. Pucher and R. Buehler (eds), *City Cycling* (pp. 31–56). Cambridge, MA, USA and London, UK: MIT Press.

- Geurs, K.T., La Paix, L., and Van Weperen, S. (2016). A multi-modal network approach to model public transport accessibility impacts of bicycle-train integration policies. *European Transport Research Review*, 8(4), 1–15.
- Goel, R., Garcia, L.M.T., Goodman, A., Johnson, R., Aldred, R., et al. (2018). Estimating city-level travel patterns using street imagery: a case study of using Google Street View in Britain. *PLOS ONE*, 13(5). <https://doi.org/10.1371/journal.pone.0196521>.
- Handy, S., van Wee, B., and Kroesen, M. (2014). Promoting cycling for transport: research needs and challenges. *Transport Reviews*, 34(1), 4–24.
- Hochmair, H. H., Bardin, E., and Ahmouda, A. (2019). Estimating bicycle trip volume for Miami-Dade county from Strava tracking data. *Journal of Transport Geography*, 75, 58–69. <https://doi.org/10.1016/j.jtrangeo.2019.01.013>.
- Huang, H., Gartner, G., Krisp, J.M., Raubal, M., and Van de Weghe, N. (2018). Location based services: ongoing evolution and research agenda. *Journal of Location Based Services*, 12(2), 63–93.
- Iacono, M., Krizek, K.J., and El-Geneidy, A. (2010). Measuring non-motorized accessibility: issues, alternatives, and execution. *Journal of Transport Geography*, 18(1), 133–140.
- Kitchin, R. (2014). Big Data, new epistemologies and paradigm shifts. *Big Data and Society*, 1(1), 205395171452848. <https://doi.org/10.1177/2053951714528481>.
- Klingen, J., and van Ommeren, J. (2020). Urban air pollution and time losses: Evidence from cyclists in London. *Regional Science and Urban Economics*, 81, 103504. <https://doi.org/10.1016/j.regsciurbeco.2019.103504>.
- Kroesen, M., and De Vos, J. (2020). Does active travel make people healthier, or are healthy people more inclined to travel actively? *Journal of Transport and Health*, 16, 100844. <https://doi.org/10.1016/j.jth.2020.100844>.
- Lee, K., and Sener, I.N. (2020). Emerging data for pedestrian and bicycle monitoring: sources and applications. *Transportation Research Interdisciplinary Perspectives*, 4, 100095. <https://doi.org/10.1016/j.trip.2020.100095>.
- Leeuw, G. de, and Kruijf, W. de. (2015). The act of (future) cycling: testing urban designs and conducting research in virtual reality. *Proceedings of the European Transport Conference*. Frankfurt.
- Ma, X., Yuan, Y., Van Oort, N., and Hoogendoorn, S. (2020). Bike-sharing systems' impact on modal shift: a case study in Delft, the Netherlands. *Journal of Cleaner Production*, 259, 120846. <https://doi.org/10.1016/j.jclepro.2020.120846>.
- Martens, K. (2007). Promoting bike-and-ride: the Dutch experience. *Transportation Research Part A: Policy and Practice*, 41(4), 326–338.
- Médard de Chardon, C. (2019). The contradictions of bike-share benefits, purposes and outcomes. *Transportation Research Part A: Policy and Practice*, 121, 401–419.
- Médard de Chardon, C., and Caruso, G. (2015). Estimating bike-share trips using station level data. *Transportation Research Part B: Methodological*, 78, 260–279.
- Nello-Deakin, S., and Harms, L. (2019). Assessing the relationship between neighbourhood characteristics and cycling: findings from Amsterdam. *Transportation Research Procedia*, 41, 17–36.
- Nello-Deakin, S., and Nikolaeva, A. (2020). The human infrastructure of a cycling city: Amsterdam through the eyes of international newcomers. *Urban Geography*, 1–23. <https://doi.org/10.1080/02723638.2019.1709757>.
- NDW (2020). National Database Road Traffic Data (NDW) working on making bicycle data accessible. Retrieved from <https://northsearegion.eu/bits/bits-best>

- practices/national-database-road-traffic-data-ndw-working-on-making-bicycle-data-accessible/.
- Nielsen, T.A.S., and Skov-Petersen, H. (2018). Bikeability – urban structures supporting cycling. Effects of local, urban and regional scale urban form factors on cycling from home and workplace locations in Denmark. *Journal of Transport Geography*, 69, 36–44.
- Nikolaeva, A., te Brömmelstroet, M., Raven, R., and Ranson, J. (2019). Smart cycling futures: Charting a new terrain and moving towards a research agenda. *Journal of Transport Geography*, 79, 102486. <https://doi.org/10.1016/j.jtrangeo.2019.102486>.
- Nordback, K., Piatkowski, D., Janson, B.N., Marshall, W.E., Krizek, K.J., and Main, D.S. (2011). Using inductive loops to count bicycles in mixed traffic. *Journal of Transportation of the Institute of Transportation Engineers*, 2(1), 39–56.
- O'Brien, O., Cheshire, J., and Batty, M. (2014). Mining bicycle sharing data for generating insights into sustainable transport systems. *Journal of Transport Geography*, 34, 262–273.
- Oksanen, J., Bergman, C., Sainio, J., and Westerholm, J. (2015). Methods for deriving and calibrating privacy-preserving heat maps from mobile sports tracking application data. *Journal of Transport Geography*, 48, 135–144.
- Piwek, L., Ellis, D.A., Andrews, S., and Joinson, A. (2016). The rise of consumer health wearables: promises and barriers. *PLOS Medicine*, 13(2), e1001953. <https://doi.org/10.1371/journal.pmed.1001953>.
- Pokorny, P., and Pitera, K. (2019). Observations of truck-bicycle encounters: a case study of conflicts and behaviour in Trondheim, Norway. *Transportation Research Part F: Traffic Psychology and Behaviour*, 60, 700–711.
- Pritchard, R. (2018). Revealed preference methods for studying bicycle route choice – a systematic review. *International Journal of Environmental Research and Public Health*, 15(3), 470.
- Pritchard, R., Frøyen, Y., and Snizek, B. (2019). Bicycle level of service for route choice – a GIS evaluation of four existing indicators with empirical data. *ISPRS International Journal of Geo-Information*, 8(5). <https://doi.org/10.3390/ijgi8050214>.
- Pucher, J., and Buehler, R. (2017). Cycling towards a more sustainable transport future. *Transport Reviews*, 37(6), 689–694.
- Romanillos, G., Moya-Gómez, B., Zaltz-Austwick, M., and Lamíquiz-Daudén, P.J. (2018). The pulse of the cycling city: visualising Madrid bike share system GPS routes and cycling flow. *Journal of Maps*, 14(1), 34–43.
- Romanillos, G., Zaltz Austwick, M., Ettema, D., and De Kruijff, J. (2016). Big Data and cycling. *Transport Reviews*, 36(1), 114–133.
- Rötzer, M., Kadriu, J., Duchêne, M., Essig, M., & Krisp, J. M. (2019). Sensor-based analysis of bike lane quality. In T. Toivonen, K. Geurs, and E. Willberg (eds), *Proceedings Towards Human Scale Cities – Open And Happy, 15th Biennial NECTAR Conference* (pp. 155–157). University of Helsinki. Retrieved from https://www2.helsinki.fi/sites/default/files/atoms/files/nectar2019_abstract_book.pdf.
- Rupi, F., and Krizek, K. (2019). Visual eye gaze while cycling: analyzing eye tracking at signalized intersections in urban conditions. *Sustainability*, 11(21), 6089. <https://doi.org/10.3390/sul1216089>.
- Rupi, F., Poliziani, C., and Schweizer, J. (2019). Data-driven bicycle network analysis based on traditional counting methods and GPS traces from smartphone. *ISPRS International Journal of Geo-Information*, 8(8), 322.
- Rupi, F., and Schweizer, J. (2018). Evaluating cyclist patterns using GPS data from smartphones. *IET Intelligent Transport Systems*, 12(4), 279–285.

- Silva, C., Teixeira, J., Proença, A., Bicalho, T., Cunha, I., and Aguiar, A. (2019). Revealing the cycling potential of starter cycling cities: Usefulness for planning practice. *Transport Policy*, 81, 138–147. <https://doi.org/10.1016/j.tranpol.2019.05.011>.
- Snizek, B., Sick Nielsen, T.A., and Skov-Petersen, H. (2013). Mapping bicyclists' experiences in Copenhagen. *Journal of Transport Geography*, 30, 227–233.
- Sun, Y., and Mobasheri, A. (2017). Utilizing crowdsourced data for studies of cycling and air pollution exposure: a case study using strava data. *International Journal of Environmental Research and Public Health*, 14(3). <https://doi.org/10.3390/ijerph14030274>.
- te Brömmelstroet, M., Nikolaeva, A., Glaser, M., Nicolaisen, M.S., and Chan, C. (2017). Travelling together alone and alone together: mobility and potential exposure to diversity. *Applied Mobilities*, 2(1), 1–15.
- Tenkanen, H., and Toivonen, T. (2020). Longitudinal spatial dataset on travel times and distances by different travel modes in Helsinki Region. *Scientific Data*, 7(1), 1–15. <https://doi.org/10.1038/s41597-020-0413-y>.
- Toole Design (2019). *Nonmotorized Data Collection and Monitoring: Program Guide and Implementation Plan*. Retrieved from https://www.michigan.gov/documents/mdot/Non_Motorized_Data_Collection_and_Monitoring_Report_681956_7.pdf.
- Van Duppen, J., and Spierings, B. (2013). Retracing trajectories: the embodied experience of cycling, urban sensescapes and the commute between 'neighbourhood' and 'city' in Utrecht, NL. *Journal of Transport Geography*, 30, 234–243.
- Willberg, E., Salonen, M., and Toivonen, T. (2021). What do trip data reveal about bike-sharing system users? *Journal of Transport Geography*, 91, 102971. <https://doi.org/10.1016/j.jtrangeo.2021.102971>.