ENVIRONMENTAL RESEARCH
LETTERS

ACCEPTED MANUSCRIPT • OPEN ACCESS

Data gaps in transport behavior are bottleneck for tracking progress towards healthy sustainable transport in European cities

To cite this article before publication: Guillaume Chevance *et al* 2024 *Environ. Res. Lett.* in press https://doi.org/10.1088/1748-9326/ad42b3

Data gaps in transport behavior are bottleneck for tracking progress towards healthy sustainable transport in European cities

Guillaume Chevance[1], Mark Nieuwenhuijsen[1,2,3], Kaue Braga[4], Kelly Clifton[5], Suzanne Hoadley[6], Lynn Kaack[7], Silke K. Kaiser[7], Marcelo Lampkowski[8], Iuliana Lupu[9], Miklós Radics[10,11], Daniel Velázquez-Cortés[1,2], Sarah Williams[1], James Woodcock[12], Cathryn Tonne[1,2,3]

[1]ISGlobal, Barcelona, Spain
[2]Universitat Pompeu Fabra, Barcelona, Spain
[3]CIBER Epidemiología y Salud Pública, Madrid, Spain
[4]C40 Cities
[5]University of British Columbia, Vancouver, Canada
[6]POLIS, Brussels, Belgium
[7]Hertie School, Berlin, Germany
[8]ICLEI European Secretariat, Freiburg, Germany
[9]DG Mobility and Transport, European Commission
[10]University of Cantabria, Santander, Spain
[11]European Commission, Joint Research Centre, Seville, Spain
[12]University of Cambridge, Cambridge, UK

Correspondence: Cathryn Tonne, ISGlobal, C/ del Rosselló, 132, Barcelona, Spain; cathryn.tonne@isglobal.org, +34 932 27 18 06

Introduction

More than 70% of the European population live in urban areas.[1] Cities are recognized as engines of innovation and wealth creation. They can also concentrate environmental hazards such as air and noise pollution and non-optimal temperature through urban heat islands, largely a consequence of urban and transport planning which prioritizes privately-owned motorized vehicles.[2] Transport emissions account for around 25% of the EU's total greenhouse gas emissions, with increasing trends in recent years.[3] Road transport constitutes the largest share of transport emissions, emitting 76% of all EU transport-related greenhouse gases. Urban mobility accounts for 40% of all road transport CO_2 emissions and up to 70% of other pollutants from transport. Decreasing car dependency to the profit of active and public transportation in European cities holds a dual objective: to meet greenhouse gas emissions reduction targets and deliver health co-benefits through increased physical activity and reduced air and noise pollution.[2]

A major barrier to monitoring progress towards healthy sustainable transport in urban areas is the absence of harmonized passenger transport data across European cities. Today, most studies in the transport domain are case studies of individual cities with limited generalizable findings at the EU level. The current fragmented data landscape impedes coordinated efforts across cities and countries. Harmonized data are essential for monitoring progress, informing planning and design of transport infrastructure, identifying key drivers of change, and incentivizing city-level action through benchmarking. Harmonized data across European cities could help in monitoring and identifying what works in different contexts and ultimately assist in scaling up coordinated climate and health solutions at the EU level.

To identify approaches to overcome this barrier, we organized a workshop bringing together a panel of diverse experts working on issues related to healthy, sustainable urban transportation. The workshop was organized as part of the UBDPolicy and CATALYSE projects which estimate the health burden of urban and transport planning practices in European cities and the health benefits of climate action in Europe respectively. This Perspective aims to (*i*) provide an overview of initiatives to monitor active and public transport and map currently available databases; (*ii*) identify promising approaches to data generation in the short- and medium-term; and (*iii*) identify potential applications of new data streams.

What data are needed?

We first identified features of our ideal data. The most valuable data would meet the following criteria: collected using comparable methods; trip-based; derived from (at least) annual time series allowing

for comparisons over time; spatially representative (i.e., between and within cities); and informative about who is traveling, how, where, and when. Harmonized survey/self-reported data should at least provide information on the number of trips made on a given day, travel mode, trip duration and distance, and ideally the trips' purpose. Comprehensive analysis of travel behavior at the EU level requires detailed information on spatially explicit places of residence, trip origins and destinations, parking circumstances, alongside individual-level demographic and socioeconomic factors. From a technical standpoint, to maximize scientific and societal impact, these data should be publicly available with few barriers to access (i.e., legal, and technical) and accessible in a single repository. Data should be in English, well structured (i.e., according to the "FAIR" principles, Findable, Accessible, Interoperable, Reusable) and machine-readable. The same data would be available for all cities in Europe and collected using standardized methodology at least annually to allow monitoring over time.

What data are currently available, and what are the key gaps?

Currently, no dataset approximating our ideal could be identified. Overall, there are few coordinated data collection initiatives across Europe. The table below summarizes types of data sources discussed in the workshop and their strengths and limitations.

Data source	Main outcome(s)	Strengths and limitations
Surveys	Mode share (self-reported)	Can be population representative; relatively easy to deploy; offer a complete view of any transport-related behaviors and potential determinants (e.g., travel-related attitudes and motivations). Surveys often conducted at single point in time, therefore, do not allow for continuous monitoring; need to be harmonized if the survey items differ; not always made openly available by the cities. Self-reported data (e.g. trip duration) may be subject to self-perception bias compared to more objective data.
Mobile phones	Number and duration of trips, sometimes mode of transport and routes taken	Data harmonized; outcomes objectively measured; longitudinal assessment at the individual level possible. Requires technical skills to handle data size and type; privacy and availability can pose issues. High cost of acquiring data from private sector for research purposes.
Street view data	Mode share	Harmonized data available in many cities across Europe; however, not at individual or trip-level (e.g., duration, purpose); not regularly updated and may have limited spatial coverage.

Pedestrian/ Bicycle counters	Count data	Time series data allowing longitudinal analysis; data can be harmonized across cities. Does not capture characteristics of the traveler (e.g. socio-demographics) or the trip (e.g. duration, purpose); focused mainly on bike (e-scooters or other type of mobility cannot be differentiated).

There is a clear need for regulatory or other (e.g., voluntary) mechanisms to standardize reporting of what data are collected, frequency of collection; and to make data easily available in the public domain. Surveys are the main source of currently available data. For example, DG MOVE coordinated an extensive EU-wide survey in 2021, including information on transport-related behaviors.[4] The survey included 112,000 interviews across all EU Member States. However, there are no plans to routinely carry out the survey, impeding monitoring of indicators over time. Many individual cities conduct surveys covering similar aspects of transport behaviors to develop transport models or to monitor changes in modal share or travel patterns. However, survey instruments and methodologies are not yet well coordinated at the European level, making comparisons over time and between cities challenging. Not all cities make data openly accessible and there is no single data repository, although initiatives are ongoing (e.g. an attempt to create a common European mobility data space[5]). Many public entities mask the geographic locations of travel due to concerns about privacy, which limits the possibility to link active travel with characteristics of the built, social, and natural environments.

ICLEI, a global network of more than 2,500 local and regional governments committed to sustainable urban development, provides harmonized transport data collected from its members via the CDP-ICLEI Track platform.[6] Surveys are performed at the city level, i.e., respondents are city representatives, not individuals. Transport data collected via this platform include freight and passenger travel, their respective modal shares, greenhouse gas emissions from the transport sector, the existence of low/zero-emission zones or restricted zones for logistics, and the number of electric vehicle charging points. In 2021, 205 European cities completed the survey and reported information relevant to tracking transport-related behaviors in cities. CDP-ICLEI recommends annual reporting, but it is not compulsory. Local governments have the flexibility to decide the frequency of reporting their climate actions and greenhouse gas emissions data. The data reported by cities via the CDP-ICLEI Track platform is accessible through the CDP Open Data Portal.[6]

Academic researchers are also contributing to the generation of open access, multi-city travel behavior datasets. A recent study[7] pooled self-reported mode share data from 797 cities, about a

third of which were European cities. Data were obtained via the European Platform on Mobility Management (EPOMM) and the final dataset is openly available online. Another study[8] pooled surveys from 19 European cities. While these efforts by academic researchers are valuable, more sustainable initiatives that support routine updating of datasets over time are needed.

Beyond surveys, several European cities (e.g., Paris, Vienna) are starting to experiment with data derived from smartphones using cellular and/or Global Navigation Satellite System (GNSS) data. These initiatives are intended to analyze travel demand, with a particular interest in active modes such as cycling and walking. Early experience has identified pros and cons for each type of data: cellular data alone cannot differentiate between travel modes, but this may be possible when combined with other datasets, such as tram network graph and passenger count data to distinguish tram passengers from cyclists. GNSS data have high spatial resolution and are more effective at mode discrimination than cellular data, although it performed better at identifying pedestrians than cyclists in Paris.[9] The main drawbacks of GNSS data are too few users and under-representativeness in less-privileged areas. Since GNSS and cellular data are procured from private suppliers, data are typically not in the public domain, unless specifically negotiated in the contract. As public authorities increasingly move towards third party data sources and "data as a service" models, the issue of open data is becoming more pressing. Some midway solutions are emerging such as private companies sharing aggregated or raw data with selected individuals under specific conditions.[9]

Innovative, big data approaches such as utilizing street view data (e.g., Google Street View) hold promise to estimate transport mode share in cities. Mode share can be estimated using manual annotation[10] or more sophisticated machine learning techniques to automatically extract key features.[11] A study in Britain found good predictive power when comparing Google Street View with census commute mode share.[10] Limitations include the frequency of data updates and how comprehensive the coverage is (e.g., cycle and walking specific paths). Other sources of image data include fixed cameras. These are used to estimate flow mode shares at specific points and in network-based analyses flows across the city.[12] Such data are often publicly available but not stored (i.e., retrospective analyses are not possible, unlike with Google Street View). Video recordings also allow to estimate cyclist speed.[13] Satellite and aerial images are increasingly of sufficient quality to estimate modes and are more widely available. The association between features of the built environment and travel modes is well established.[14] Given the increasing availability of rich datasets on the built environment, these can potentially be used to predict mode share, e.g., a study

showed how an increase in cycling infrastructure would lead to a rise in cycling, but equally, it could be used to estimate cycling share in cities for which data are not available.[14]

Bicycle counters represent a harmonized source of open data at the European level that can be used in machine learning techniques to estimate bicycle traffic volume in areas without counters.[15] For example, a study conducted during the COVID-19 pandemic pooled daily bicycle counts from 736 bicycle counters in 106 European cities.[16] APIs can be developed to automatically extract this data from open data repositories, generating a longitudinal indicator of cycling across different European countries. However, a limitation of this approach is that it only identifies cycling (or any other mode used on cycling path such as e-scooter), without information on other modes needed to calculate mode share.

Other promising data sources include public transport flow from smart card data and automated passenger counts frequently used to study public transportation usage.[17] Provided that these data are made open access in several European cities (see for example in China)[18], this could constitute a relevant indicator to track changes in public transportation utilization in Europe. Further, user-generated data that are voluntarily made open access or shared with research teams, notably from apps and wearables, could yield large datasets relevant for active transportation[19]; however, so far, these methods come with significant barriers, including population representativeness, sustainable funding of crowdsourcing tools, privacy concerns, and the requirement for technical skills to handle data.[20] Finally, data from shared micro-mobility services (e.g., bike sharing, electric scooters) represent a valuable data source to track mobility behavior within cities.[21]

Who is best placed to develop data resources?

Different actors can play a role in filling data gaps. These include national and local governments, city networks, the private sector and researchers. Currently, at the EU level there is no regulated and harmonized data collection covering motorized and non-motorized (e.g. walking, cycling) road transport modes. In the absence of legally binding reporting, city networks, such as ICLEI, can play a role in harmonizing and coordinating data collection on a voluntary basis to support benchmarking and sharing best practices. City networks or civil society organizations can play a role in consolidating data, such as cycle counts, into easily accessible data repositories. The private sector is uniquely placed to generate datasets based on cellular and GNSS networks. Making these data available in an appropriate way, sensitive to privacy concerns would allow cities and researchers to use these data for practical problem solving and knowledge generation. Values and priorities to guide the responsible use of data and privacy protection of mobility data have been developed by different

actors.[22] Public institutions can also play a role in driving legislation for private operators willing to share data, for example the public sector initiative *Green Data for Health* which aims to create a common environmental data space for health in France.[23]

Conclusion

Lack of harmonized, easily accessible data on travel behavior is a major bottleneck to monitoring progress towards healthy sustainable transport. Priorities for addressing these gaps include (*i*) ensuring travel surveys conducted at the city level are harmonized (i.e., data should at least provide information on the number of trips performed for a given day, trips' mode, trips' length and distance, and eventually trips' purpose) and regularly repeated to support longitudinal monitoring; (*ii*) promoting open access for user generated data from mobile apps (*iii*) expansion of street view data to include active travel infrastructure and (*iv*), and expansion of bicycle counter networks and the sustainable development of open access data repositories for these data.

Acknowledgements. This work has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreements No 101057131, Climate Action To Advance HeaLthY Societies in Europe (CATALYSE) and No 101094639, Urban Burden of Disease Estimation for Policy Making (UBD Policy).

We also acknowledge support from the Spanish Ministry of Science and Innovation and State Research Agency through the “Centro de Excelencia Severo Ochoa 2019-2023” Program (CEX2018-000806-S), and support from the Generalitat de Catalunya through the CERCA Program. The first author has also been awarded with the grant RYC2021-033537-I, supported by MCIN/AEI/10.13039/501100011033 and by the European Union "NextGenerationEU"/PRTR”; and from the Ministry of Research and Universities of the Government of Catalonia (2021 SGR 01563).

References

1. Europe and its Urban Development, 1970 to 2020 [Internet]. European Investment Bank. [cited 2023 Jun 1]. Available from: https://www.eib.org/en/essays/the-story-of-your-city

2. Nieuwenhuijsen MJ. Urban and transport planning pathways to carbon neutral, liveable and healthy cities; A review of the current evidence. Environ Int. 2020 Jul 1;140:105661.

3. Transport and the Green Deal [Internet]. [cited 2023 Dec 1]. Available from: https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/european-green-deal/transport-and-green-deal_en

4. European Commission, Directorate-General for Mobility and Transport, Armoogum, J., Garcia, C., Gopal, Y. et al., *Study on new mobility patterns in European cities – Final report. Task A, EU wide passenger mobility survey*, Publications Office of the European Union, 2022, https://data.europa.eu/doi/10.2832/728583

5. PrepDSpace4Mobility [Internet]. PrepDSpace4Mobility. [cited 2023 Dec 1]. Available from: https://mobilitydataspace-csa.eu/

6. CDP Open Data Portal | CDP Open Data Portal [Internet]. [cited 2023 Dec 1]. Available from: https://data.cdp.net/

7. Prieto-Curiel R, Ospina JP. Large cities are less efficient for sustainable transport: The ABC of mobility [Internet]. arXiv; 2023 [cited 2023 Dec 1]. Available from: http://arxiv.org/abs/2212.13956

8. Berrill P, Nachtigall F, Javaid A, Milojevic-Dupont N, Wagner F & Creutzig F. Comparing urban form influences on travel distance, car ownership, and mode choice. Transp R part D. 2024;128:104087

9. EIB Technical Note on Data Sharing in Transport. Available from : https://www.eib.org/en/publications/eib-technical-note-on-data-sharing-in-transport

10. Goel R, Garcia LMT, Goodman A, Johnson R, Aldred R, Murugesan M, et al. Estimating city-level travel patterns using street imagery: A case study of using Google Street View in Britain. PloS One. 2018;13(5):e0196521.

11. Doiron D, Setton EM, Brook JR, Kestens Y, McCormack GR, Winters M, et al. Predicting walking-to-work using street-level imagery and deep learning in seven Canadian cities. Sci Rep. 2022 Nov 1;12(1):18380.

12. Zhang Z, Li M, Lin X, Wang Y. Network-wide traffic flow estimation with insufficient volume detection and crowdsourcing data. Transp Res Part C Emerg Technol. 2020 Dec 1;121:102870.

13. Kassim A, Tayyeb H, Al-Falahi M. Critical review of cyclist speed measuring techniques. J of Traffic and Transp Engineering. 2020 Dec 7;1:98110.

14. Yang Y, Wu X, Zhou P, Gou Z, Lu Y. Towards a cycling-friendly city: An updated review of the associations between built environment and cycling behaviors (2007–2017). J Transp Health. 2019 Sep 1;14:100613.

15. Kaiser S, Klein N, Kaack L. Predicting cycling traffic in cities: Is bikesharing data representative of the cycling volume? 2023 [cited 2023 Dec 1]; Available from: https://opus4.kobv.de/opus4-hsog/frontdoor/index/index/docId/4942

16. Kraus S, Koch N. Provisional COVID-19 infrastructure induces large, rapid increases in cycling. Proc Natl Acad Sci. 2021 Apr 13;118(15):e2024399118.

17. Welch TF, Widita A. Big data in public transportation: a review of sources and methods. Transp Rev. 2019 Nov 2;39(6):795–818.

18. Wei S, Xu J, Sun J, Yang X, Xin R, Shen D, et al. Open big data from ticketing website as a useful tool for characterizing spatial features of the Chinese high-speed rail system. J Spat Sci. 2018 Jul 3;63(2):265–77.

19. Nelson T, Ferster C, Laberee K, Fuller D, Winters M. Crowdsourced data for bicycling research and practice. Transp Rev. 2021 Jan 2;41(1):97–114.

20. Keseru I, Wuytens N, Macharis C. Citizen observatory for mobility: a conceptual framework. Transp Rev. 2019 Jul 4;39(4):485–510.

21. Wang R, Wu J, Qi G. Exploring regional sustainable commuting patterns based on dockless bike-sharing data and POI data. J Transp Geo. 2022 June 102: 103395.

22. Privacy Principles for Mobility Data. 2024 [cited 2024 March 13]; Available from: https://www.mobilitydataprivacyprinciples.org/#principles

23. Green Data for Health ; Legal tools. 2024 [cited 2024 March 13] ; Available from: https://gd4h.ecologie.gouv.fr/juridique