OPTIMAP: A Dataset for Open Public Transport Infrastructure and Mobility Accessibility Profiles

Maximilian T. Fischer*, Daniel Fürst, Yannick Metz, Manuel Schmidt, Julius Rauscher, and Daniel A. Keim

Version: 2025-02-01

1 Introduction

This dataset provides a comprehensive assessment of public transport connectivity across Germany by analyzing both walking distances to the nearest public transport stops as well as the quality of public transport connections for daily usage scenarios with housing-level-granularity on a country-wide scale. The data was generated through a novel approach ([MAKF24]) that integrates multiple open data sources, simulation models, and visual analytics techniques, enabling researchers, policymakers, and urban planners to identify gaps and opportunities for transit network improvements.

Why does it matter? — Efficient and accessible public transportation is a critical component of sustainable urban development. However, many transit networks struggle to adequately serve diverse populations due to infrastructural, financial, and urban planning limitations. Traditional transit planning often relies on aggregated statistics, expert opinions, or limited surveys, making it difficult to assess transport accessibility at an individual household level. This dataset provides a data-driven and reproducible methodology for unbiased country-wide comparisons.

2 Key Facts and Download

Acronym:	OPTIMAP
Title:	OPTIMAP: A Dataset for Open Public Transport Infrastructure and Mobility Accessibility Profiles
Authors:	M. T. Fischer, D. Fürst, Y. Metz, M. Schmidt, J. Rauscher, D. A. Keim
Download:	https://mobility.dbvis.de/data-results/OPTIMAP_v2025-02-01.parquet
License:	Datenlizenz Deutschland - Namensnennung - Version 2.0 (dl-de-by/2.0)

Please cite the dataset as:

Maximilian T. Fischer, Daniel Fürst, Yannick Metz, Manuel Schmidt, Julius Rauscher, and Daniel A. Keim. *OPTIMAP: A Dataset for Open Public Transport Infrastructure and Mobility Accessibility Profiles*. Zenodo, 2025. DOI: 10.5281/zenodo.14772646

*Corresponding author: max.fischer@uni-konstanz.de

3 Dataset Description

The dataset in the PARQUET format includes detailed accessibility measures for public transport at a fine-grained, housing-level resolution. It consists of four columns:

- `lat`, `lng` (`float32`): GPS coordinates (`EPSG:4326`) of each house in Germany, expensively compiled from the house coordinates (HK-DE) data provided by the 16 federal states under the EU INSPIRE regulations.
- `MinDistanceWalking` (`int32`): An approximate walking distance (in meters) to the nearest public transport stop from each registered building in Germany.
- `scores_OVERALL` (`float32`): A simulated, demographic- and scenario-weighted measure of public transport quality for daily usage, considering travel times, frequency, and coverage across various daily scenarios (e.g., commuting, shopping, medical visits). The results are represented in an *artificial* time unit to allow comparative analysis across locations.

4 Methodology

The dataset was generated using a combination of open geospatial data and advanced transport simulation techniques:

- **Data Sources:** Public transit information from the German national access point (DELFI NeTEx), housing geolocation data from various state authorities, and routing information from OpenStreetMap.
- **Walking Distance Calculation:** The shortest path to the nearest transit stop was computed using the Dijkstra algorithm on a graph network of publicly available pathways sourced from OSM, considering the ten aerial-nearest public transport stops.
- **Public Transport Quality Estimation:** The dataset incorporates a scenario-based simulation model, analyzing weight-averaged travel times and connection frequency to typical daily POIs such as the individually nearest train stations, kindergartens, schools, institutions of higher education, fitness, cinemas, places of worship, supermarkets, shopping malls, restaurants, doctors, parks, and culture institutions. It includes walking distances to the start and from the destination public transport stops as well as the averaged travel and waiting times on the shortest route calculated via a modified Dijkstra algorithm. The results are aggregated using a demographically- and scenario-weighted metric to ensure comparability. The value is in the unit of time, although it should *not* be interpreted directly as real minutes.
- **Visualization and Validation:** A WebGL-based interactive tool and static precomputed maps were developed to allow users to interactively explore transport accessibility metrics dynamically, available at mobility.dbvis.de.

5 Potential Applications

The dataset enables multiple use cases across research, policy, and urban planning:

- **Public Accessibility Studies:** Provides insights into transport equity by evaluating mobility gaps affecting different demographic groups and regional areas, and comparing county and state efforts in improving public transport quality.
- **Urban Planning and Transport Policy:** Supports data-driven decision-making for optimizing transit networks, adjusting service schedules, or identifying underserved areas.
- **Smart City Development:** Assists in integrating mobility analytics into broader smart city initiatives for efficient resource allocation and sustainability planning.
- **Academic Research:** Facilitates studies in transportation engineering, urban geography, and mobility behavior analysis.

6 Conclusion

By offering high-resolution public transport accessibility data at housing-level granularity, this dataset contributes to a more transparent and objective understanding of urban mobility challenges. The integration of simulation models, demographic considerations, and scalable analytics provides a novel approach to evaluating and improving public transit systems. Researchers, city officials, and policymakers are encouraged to leverage this dataset to enhance transport infrastructure planning and accessibility.

This dataset contains both the approximate walking distances in meters and a weighted overall quality score in an artificial time unit for each individual house in Germany. More advanced versions are currently *not* publicly available. This base dataset is publicly available and adheres to open data licensing principles, enabling its reuse for scientific and policy-oriented studies.

Source Data Licenses

While not part of this dataset, the scientific simulation used to create the results leverages public transit information via the National Access Point (NAP) DELFI as NeTEx, provided via GTFS feeds of Germany (CC BY 4.0). Also, routing information used during the processing was based on Open Street Map contributors (CC BY 4.0).

Primarily, this dataset contains original and slightly processed housing locations (`lat`, `lng`) that were made available as part of the **EU INSPIRE** regulations, based on Directive (EU) 2019/1024 (*of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information (recast)*.

In Germany, the respective data is provided individually by the 16 federal states, with the following required attributions and license indications:

- BB: EU INSPIRE / © GeoBasis-DE/LGB, dl-de-by/2.0 (data modified)
- BE: EU INSPIRE / © Geoportal Berlin / Hauskoordinaten, dl-de-by/2.0 (data modified)
- BW: EU INSPIRE / © LGL, www.lgl-bw.de, dl-de-by/2.0 (data modified)
- BY: EU INSPIRE / © Bayerische Vermessungsverwaltung, CC BY 4.0 (data modified)
- HB: EU INSPIRE / © Landesamt GeoInformation Bremen, CC BY 4.0 (data modified)

- HE: EU INSPIRE / © HVBG, dl-de-by-zero/2.0 (data modified)
- HH: EU INSPIRE / © FHH (LGV), dl-de-by/2.0 (data modified)
- MV: EU INSPIRE / © LAiV M-V, CC BY 4.0 (data modified)
- NI: EU INSPIRE / © LGLN 2024, CC BY 4.0 (data modified)
- NW: EU INSPIRE / © Geobasis NRW, dl-de-by-zero/2.0 (data modified)
- RP: EU INSPIRE / © GeoBasis-DE / LVermGeoRP 2024, dl-de-by/2.0 (data modified)
- SH: EU INSPIRE / © GeoBasis-DE/LVermGeo SH, CC BY 4.0 (data modified)
- SL: EU INSPIRE / © GeoBasis DE/LVGL-SL (2024), dl-de-by/2.0 (data modified)
- SN: EU INSPIRE / © GeoSN, dl-de-by/2.0 (data modified)
- ST: EU INSPIRE / © GeoBasis-DE / LVermGeo LSA, dl-de-by/2.0 (data modified)
- TH: EU INSPIRE / © GDI-Th, dl-de-by/2.0 (data modified)

Original Research

The methodology and techniques are described in an original research article published in 2024. When referring to our approach, please cite the following publication:

Please cite the original research as:

Yannick Metz, Dennis Ackermann, Daniel A. Keim, and Maximilian T. Fischer. "Interactive Public Transport Infrastructure Analysis through Mobility Profiles: Making the Mobility Transition Transparent". In: *2024 IEEE Visualization in Data Science (VDS)*. VDS. IEEE, 2024, p. 9. DOI: 10.1109/VDS63897.2024.00006

BibTeX (Original Research):

```
@inproceedings{MobilityProfiles.VDS.2024,
 author = {Metz, Yannick and
 Ackermann, Dennis and
 Keim, Daniel A. and
 Fischer, Maximilian T.},
 booktitle = {2024 IEEE Visualization in Data Science (VDS)},
 doi = {10.1109/VDS63897.2024.00006},
 pages = {9},
 publisher = {IEEE},
 series = {VDS},
 title = {Interactive Public Transport Infrastructure Analysis
 through Mobility Profiles: Making the Mobility Transition
 Transparent},
 year = {2024}
}
```

References

[FFM+25] Maximilian T. Fischer, Daniel Fürst, Yannick Metz, Manuel Schmidt, Julius Rauscher, and Daniel A. Keim. *OPTIMAP: A Dataset for Open Public Transport Infrastructure and Mobility Accessibility Profiles*. Zenodo, 2025. DOI: 10.5281/zenodo.14772646.

[MAKF24] Yannick Metz, Dennis Ackermann, Daniel A. Keim, and Maximilian T. Fischer. "Interactive Public Transport Infrastructure Analysis through Mobility Profiles: Making the Mobility Transition Transparent". In: *2024 IEEE Visualization in Data Science (VDS)*. VDS. IEEE, 2024, p. 9. DOI: 10.1109/VDS63897.2024.00006.