

# Creation of a Mobility Data Curation Dashboard

## Bachelor's Thesis of Beatriz Contin Dorigan

### Mentoring:

M.Sc. Santiago Álvarez-Ossorio Martínez

M.Sc. Victoria Dahmen

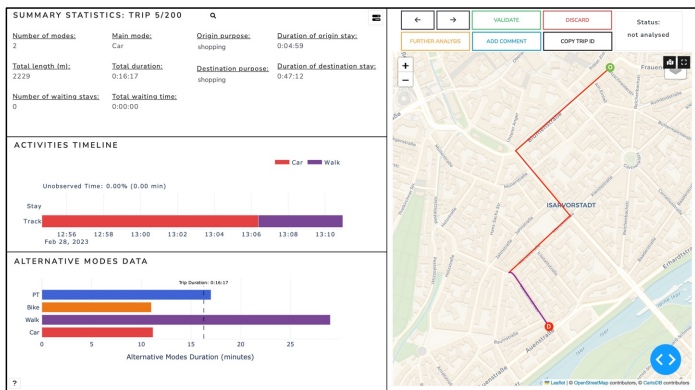


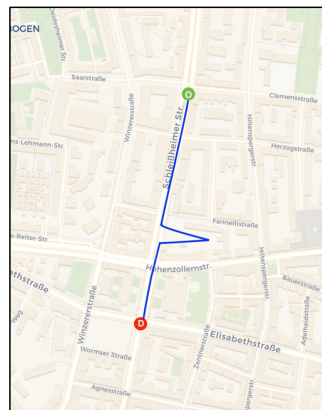
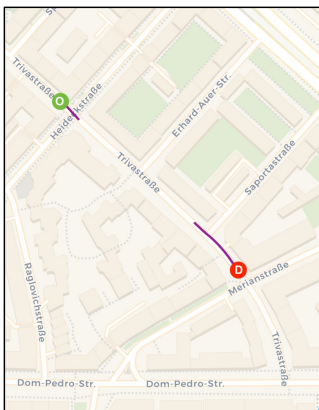
Figure 1: Final dashboard for geospatial data curation

### Methodology

- Establish the most commonly occurring errors in travel diaries
- Select features that allow the errors to be identified
- Implement the features in a dashboard layout using the Python framework Dash
- Optimise the friendliness and interactivity of the interface to improve user experience
- Test the dashboard using data from the *Mobilität.Leben* project regarding its effectiveness to facilitate the data curation

### Common errors in travel diaries

- Gaps within the trip
- Sudden route deviations
- Inconsistencies in mode and/or location
- Abnormal track lengths and sequences
- Durations that deviate from the expected



Figures 3-4: Examples of trips with GPS errors: trip with a gap (fig. 3) and trip with a sudden route deviation (fig. 4)

### Context

In 2022, the German government implemented two significantly cheaper public transport fares, the *9-Euro-Ticket* and the *Deutschlandticket*. To investigate how these new policies affected the mobility behaviour of the population, the *Mobilität.Leben* project was initiated. The location of the study participants was constantly tracked by their smartphones and travel diaries that describe their commutes were generated. Due to the large quantities of data collected and the susceptibility of imprecise recordings, ensuring the quality of the travel diaries is a significant challenge. Manual validations are time-consuming and automated ones are not entirely accurate.

### Motivation

This bachelor's thesis aims to develop a dashboard to facilitate the curation of geospatial data. The tool should display individual trips and their most important attributes, allowing the user to easily identify errors and remove data that does not meet sufficient quality standards.

trip_id	status	comment	further_analysis_tag
98df203c-ef43-4b02-ppt9-aed47bb7b87a_2022-12-17_2	validated		
0f36d7d3-1fws-4d2e-8d22-473686c5729c_2023-03-23_2	validated		
1e24927a-ef3d-432b-82d1-1c91552f7a65_2022-09-03_1	discarded		
3061572f-6016-opl7-8734-348a34b43a8f_2022-09-22_8	further analysis		reason2
07cf4774-0a61-45cf-88d5-babd6a5fc8dc_2023-02-28_3	discarded		
6552eedf-be9c-u8gi-8e12-e56d65050f1e_2023-05-19_9	not analysed	sample comment	
8068f92e-2sfj-45e5-9bb9-8ea954855b47_2023-05-16_4	not analysed		

Figure 2: Output file generated after validation with the dashboard

### Results

Errors that compromise the quality of the dataset were well identified by using the dashboard and the process to validate and discard trips is efficient. The tool is easy to use and runs without technical constraints. Still, the quality of the analysis is dependent on the user's effort to find errors within the data.

### Conclusions

The objective of this thesis was met as the dashboard allows travel diaries to be analysed in a structured way, facilitating the curation of geospatial data. Furthermore, data validated with the dashboard can be compared to processed data, verifying the reliability of the algorithms used for autonomous validation.

Nevertheless, the manual aspect of the process implies that it is still susceptible to human error, as anomalies within the data might go unnoticed. To overcome this limitation, it is advisable for clear validation criteria to be established and error-detecting features to be included in the future.

The dashboard reduces the time investment required for curating geospatial data and improves its overall quality. In practice, developing validation tools may encourage further research in the field of mobility.