

Big Data Collection using Smartphone Based Surveys and Open APIs: the Future Mobility Sensing for London

Availability of rich and high quality data is a key driver in every sector. Transportation is no exception, where the way data is created, collected, analysed and shared is immensely evolving. Technological progress has allowed us to shift away from historically manual and spatially fixed data collection to dynamic methods. In the field of travel behaviour, traditional face-to-face paper based surveys have developed to exploit the advancements in information and communication technologies (ICT). The rapid penetration of location-aware smartphones has made researchers realise the potential of these devices to collect individual level travel data. As such, numerous smartphone based travel survey tools have emerged in recent years that track the trips of individuals and offer precise spatial and temporal information. These new collection methods also have significant benefits regarding the survey experience of the user. Respondent burden is decreased by the tools automatically recording exact times and locations without user input, and by machine learning aiding long-term data collection. Further, these computerised systems also have huge untapped potential, as they can be used to harness information from any source.

One of these sources is the vast amount of open data and APIs (Application Programming Interfaces) available for developers. The quality and quantity of such open datasets are constantly improving, expanding and becoming more dynamic through live feeds

Given the wealth of open transport APIs, the first aim of this study is to investigate and propose a methodology to link open APIs to smartphone based travel survey tools to improve the quantity and the quality of the collected data and to further reduce response burden. By connecting the diary with openly available transport operators' APIs, rich data can be harnessed on top of the locational data, such as travel costs, network conditions, availability and characteristics of alternative transport modes. Greater London is taken as a case study where the Future Mobility Sensing tool is connected with Transport for London's (TfL) open data and APIs as well as Google open APIs. The complex way public transport fares are calculated makes it very difficult to get accurate cost estimates from respondents. However, for any modelling exercise travel costs are an essential variable. Against this background, the focal point of this research is to connect the public transport operator's (TfL's) cost API's to create instantaneous, automatic and accurate travel cost values for public transport trips taking into account factors such as zoning, transfers and peak/off-peak pricing.

Besides decreasing respondent burden, survey experience can also be enhanced by making the questionnaires more interactive. Providing users with in-survey feedback can increase the respondents' willingness to participate (2). These methods can increase response and completion rates, as it gives respondents motivation and a reward. Thus, the second aim of this research is to utilize the above-mentioned travel costs, as well as other information collected, and describe an approach for presenting respondents with summary records of their travel.