Using GPS Technology for Demand Data Collection

SUTP.
Sustainable Urban Transport Project

Sustainable Urban Transport Project

SUTP supports decision-makers worldwide to plan and to implement innovative and sustainable mobility solutions. SUTP offers a comprehensive knowledge platform, capacity development, hands-on advice and networking opportunities. Within the past 16 years, more than 5 000 decision-makers, planners and students have benefited from our training offers. We've produced a rich library of Sourcebook Modules, Technical Documents, Case Studies, Factsheets, Policy Briefs and Reading Lists. All documents are accessible through our webpage, along with a comprehensive photo collection and a video channel.

Be invited to use and distribute them!

http://www.sutp.org

About the authors

Jakob Baum is a transport policy advisor with a background in data-enabled mobility consultancy. He has been part of *GIZ's Sustainable Urban Transport Project* from 2015 onwards. In this role, he has been supporting various projects in employing GPS data collection techniques for demand data generation. Before joining GIZ, he worked as an independent consultant in corporate mobility management and municipal bicycle planning. With his specialization in data analysis, he also supported the implementation of car sharing schemes from both, the private and municipal site. Jakob holds a master's degree in transport system planning from the *Technical University of Berlin*. He also studied psychology and urban studies at the *University of California* and the *Free University of Berlin*.

Enrico Howe is a mobility researcher and consultant. He has worked for different Berlin-based research institutions and environmental consultancies within the fields of sustainable mobility and renewable energy supply. In 2013, he started working for the *Innovation Centre for Mobility and Societal Change* and focused his work on user acceptance among different sustainable mobility solutions. Among other topics, he is focusing on smartphone based mobility surveys, electric mobility, shared mobility and international mobility solutions. He is co-responsible for the product development of the smartphone tracking app modalyzer and is applying the app for research and consultation purposes e.g. in Ukraine, Colombia, Germany and Canada. GIZ uses the approach to support the development of *Sustainable Urban Mobility Plans* in Ukraine and Colombia.

Using GPS Technology for Demand Data Collection

Sustainable Urban Transport Project Technical Document #17

Travel demand data is a necessary basis for urban mobility planning, but especially in developing and emerging economies data availability is often weak or non-existing. The Global Positioning System (GPS) technology offers a cheap alternative for data collection to traditional diary or survey methods. This document elaborates on advantages and disadvantages of the approach. Also different aspects of the post-processing of GPS data in order to determine trips, mode choice and trip purposes are discussed.

Disclaimer

Findings, interpretations and conclusions expressed in this document are based on information gathered by GIZ and its consultants, partners and contributors.

GIZ does not, however, guarantee the accuracy or completeness of information in this document, and cannot be held responsible for any errors, omissions or losses which emerge from its use.

TABLE OF CONTENTS

Data Necessities for Transport System Planning

The goal of every transport planner is to meet the citizen's demand for mobility. In order to provide a reliable, safe and efficient transport system, they have a set of measures to choose from – ranging from different infrastructure measures that enhance the system, but also travel demand management measures such as congestion pricing and many more.

Facing the decision between such an array of possibilities, the planner needs to evaluate and appraise the options and their impact on the system, its users and the society as a whole. This is done through varying appraisal methods, including cost-benefit-analyses that are based on models. To run these models, three types of data are needed:

1. **Existing network and prices**
   Especially in developing and emerging economies data availability reveals a serious challenge: while network information is often freely available through OpenStreetMap (given precautions in quality control), travel demand data and behavioural models are very rare.

2. **Behavioural models**
   Behavioural models determine how a person travels given certain circumstances. Most modelling software on the market has a behavioural model implemented that is applied in projects located all over the world. Obviously, this lacks to account for cultural and local characteristics of travel decisions. For example, the travel time cost varies largely between different countries, age and sex groups (Litman 2009, chap. 5.2.7).

3. **Travel demand**
   This paper will guide through some aspects of GPS tracking as a travel demand data collection methodology. The most common travel demand format is an Origin-Destination-matrix (OD-matrix), describing the amount of people traveling between different places with a given mode of transport

Classical Approach to Data Collection

With the rise of transportation science in the 1950s, transport models have been based on data obtained from household surveys and travel diaries (Axhausen 1995). In these paper- or telephone-based collection formats, the participants reveal information on each conducted trip at the given day. This method produces a dataset with relatively high information density by using detailed questionnaires. Representative surveys are comparatively easy to conduct and expensive at the same time.

On the other hand, its reliability is often very weak as it relies on stated preferences: The high manual workload for the participants may result in low information quality on trips as well as their distances and durations. Also un-intentional misreporting, especially regarding multi-modal and short trips by foot, is a challenge.

The costs implied with a representative, traditional data collection are often close to prohibitive for cities in developing and emerging economies. In order to overcome the difficulties on data reliability of traditional data collection, transport researchers and practitioners started to add supplemental technology into their methodological designs.

Global Positioning System (GPS) – Based Data Collection

For the first time, GPS-technology was used in 1997, in a context of a mobility survey in order to automate data collection and to enhance its robustness (Battelle Transport Division 1997 as cited by Schönau 2016, 34). This first study already showed a main advantage of GPS tracking over self-reported travel diaries: the share of short trips in this survey was higher than in previous, traditional surveys. Many short trips have simply not been included in household surveys or traffic diaries – probably because they are not perceived as 'real' trips by the participants in the first place or because making the effort of filling out a diary for a short trip is not perceived as necessary. An overview of advantages and disadvantages is given in Table 1.

Travel Diary or Survey

- *low reliability*

- *no information on chosen travel route*

- *low scalability*

- *high collection cost*

+ *information on chosen mode*

+ *information on trip purposes*

GPS-based

+ *high reliability on trip times and durations*

+ *travel route information*

+ *easily scalable*

+ *low collection cost*

+ *inter-modality detected with higher accuracy*

- *complex post-processing needed:*

- *no direct information on used mode*

- *no direct information on trip purpose*

Table 1: Travel Diary vs. GPS-based Data Collection (own work, based on Schönau 2016)

Initially, GPS-based data collection took mostly place with special in-vehicle or handheld devices, guaranteeing a high data accuracy. Their handling is easy as they have to be simply carried by the participants without further need for interaction. Energy consumption is, in congruence with the small array of functionality, limited, even though the capture frequency is high and the generated tracks are usually very accurate. Downsides are the burden to carry an additional device for the participants (which may lead to unaccounted trips if (un)intentionally left at home) as well as the implied costs of acquisition.

Smartphones: the omnipresent Data Collection Device

With the dissemination of smartphones, their potency as personal tracking objects has become tremendous. Without need to carry an additional device and hence no further costs, barriers for their utilization are very low. Therefore, the strongest argument for the use of smartphones is the ease of acquisition of participants and scalability through simple software distribution.

However, today's smartphone batteries are not designed to facilitate constant tracking applications. On account for this, existing tracking software often decreases the frequency of GPS tracking points and the amount of requested satellites. Therewith the track quality decreases substantially compared to specialized handheld devices. Also, low battery status or anxiety for it can lead to interruption of tracking.

Another challenge for GPS tracking is the loss of signal. GPS devices need constant view to multiple satellites. Tracking underground or in an urban environment with

many skyscrapers is therefore difficult. Smartphones have the advantage to be able to use auxiliary sensors such as Wifi, mobile network or Bluetooth signals to locate their user.

A general challenge is the misrepresentation of population subgroups in the sample. This bias may be reinforced by using smartphones as the tracking device, as varying dissemination rates and disproportionate ownership of smartphones in the population groups are to be expected. Especially in many developing countries, a large share of the population doesn't own a smartphone. Also motivation to participate varies between different socio-demographic groups. This has to be accounted for in the investigation design, e.g. through accompanying traditional interviews, targeted recruiting of underrepresented groups, supply of smartphones and finally weighting in the data analysis. Using data from the network provider (Call Detail Records) instead may also be an option – however, that comes with its own challenges.

Handheld Tracking Devices	*Smartphones*
+ easy to handle for participants	*+/- knowledge about smartphones required*
+ long battery lifetime	*- short battery lifetime*
+ high tracking quality	*- lower tracking quality*
- need to carry additional device	*+/- personal disposability of smartphones varies largely between countries (smartphone penetration rate)*
- higher cost of provision	*+ lower cost of provision*
+ information on trip purposes	*- complex post-processing needed:*
- tracking capability in challenging environments (high building density, underground)	*+ WiFi and other signals can be used in challenging environments*
	- bias in sample / ownership

Table 2: Handheld Tracking Devices vs. Smartphones (own work, based on Schönau 2016)

Why and How did you move? The Need to Post-Process GPS-data

A GPS track has information about location, elevation and time attached to each data point. From this, speed, heading and acceleration can easily be derived. Obviously, a high frequency in data point collection leads to higher resolution of location, speed and acceleration data.

segments of a multimodal trip, the threshold can be decreased: More segments are detected, analysed and merged again afterwards. (Shen and Stopher 2014, 324)

Main challenges arise from signals loss and inaccuracy. However, existing algorithms are able to segregate up to 98% trips and stops correctly (H. Gong et al. 2012).

```
[...]
<trkpt lat="46.57638889" lon="8.89302778"><ele>2374</ele><time>2017-01-14T10:13:20Z</time></trkpt>
<trkpt lat="46.57652778" lon="8.89322222"><ele>2375</ele><time>2017-01-14T10:13:48Z</time></trkpt>
<trkpt lat="46.57661111" lon="8.89344444"><ele>2376</ele><time>2017-01-14T10:14:08Z</time></trkpt>
[...]
```
Box: Example of an array of three GPS-points stored in a GPX-file

On a trip level, the track gives valuable and highly accurate insights into route choice, trip distance and duration. However, it doesn't reveal used modes and trip purposes directly. Therefore, post-processing algorithms have been developed deducting this information from the tracks.

Identifying Trips

A raw GPS track consists of an array of points. In order to divide those points into different trips and stops, a set of criteria is applied as Gong et al. show in their literature review (2014): most researchers define that one trip has ended and a stop has been reached when a certain "dwell" time has been reached. The threshold is usually set between 120 and 300 seconds. Furthermore, the change of location, heading and the density of track points are common criteria. In order to account for different

Identifying Mode Choice

After trips have been extracted, the used mode has to be identified for each trip. In an urban context, foot, bike, car and different forms of public transport are the most commonly distinguished modes.

Mode distinction can be done via different approaches such as machine learning, probabilistic methods or criteria-based algorithms or a mix of the above. Criteria-based (or rule-based) methods often work with speed patterns among other criteria such as transit network data. Some modes have distinctive characteristics in their speed and acceleration profile. A common speed pattern of a car can easily be differentiated from a pedestrian or a bicycle (see Figure 1). The probability that an algorithm detects the right mode is very high in this case.
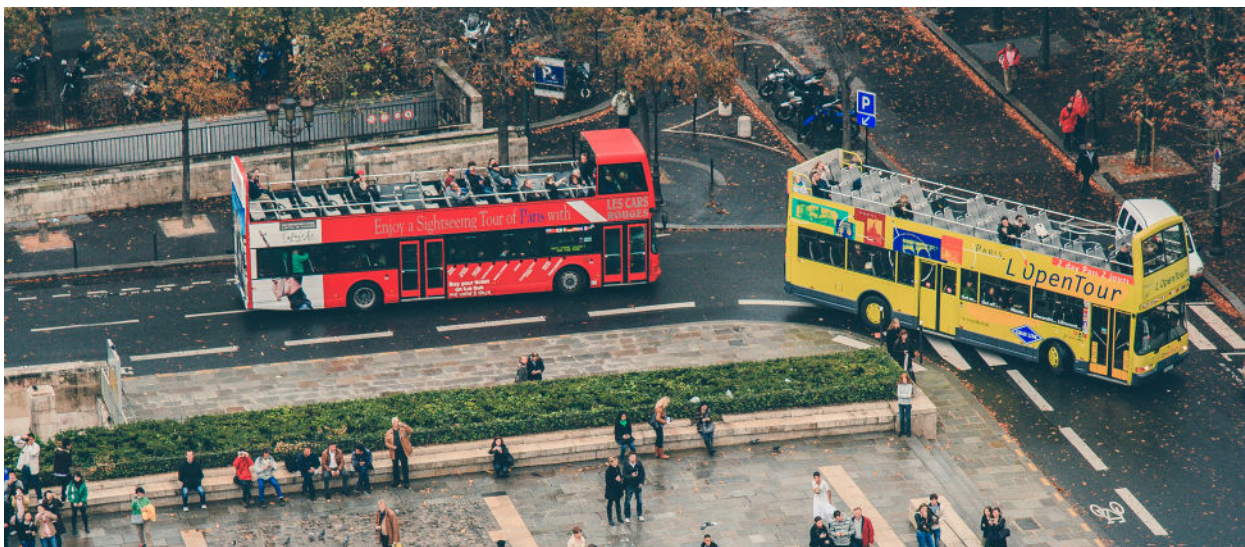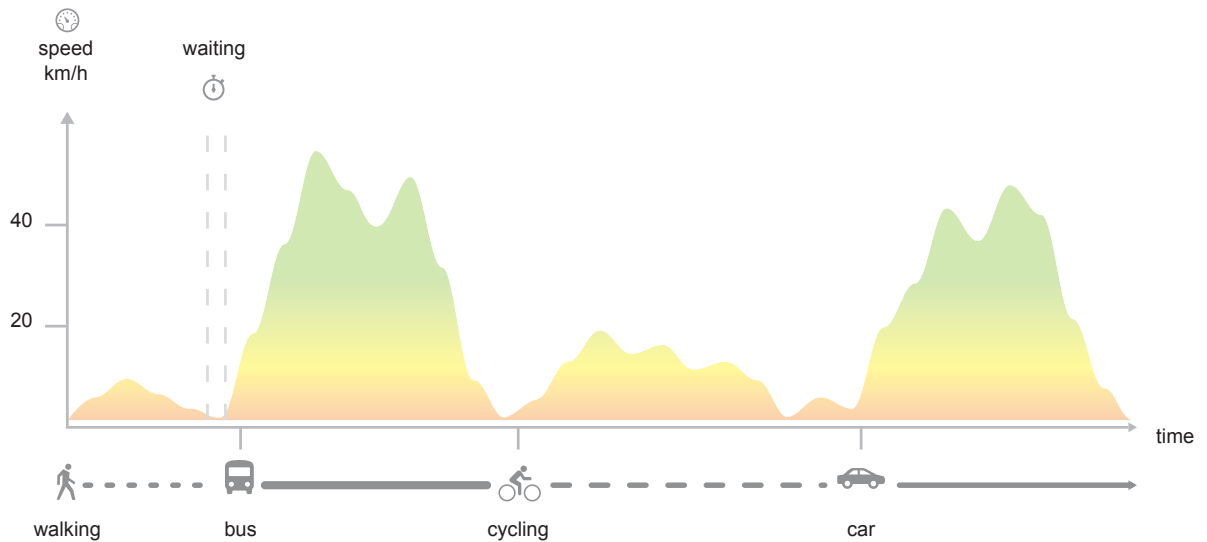
Figure 1: Speed Patterns of Different Modes

Detecting public transport use in delimitation to car usage is already more complex since speeds of the two modes are similar. Regular stops are a potential way to identify public transit – but how to differentiate between cars and buses being stuck in traffic?

In order to tackle this issue, many algorithms rely on spatial data as a second source of information. By matching the GPS track with road networks as well as public transport routes and schedules, the algorithms can e.g. detect if a person followed a bus route and stopped in proximity to public transport stations, which makes it very likely that transit was being used. Open source solutions like the Open Street Map public transportation

database reveal important data such as routes, stops and public transportation mode. This data set varies globally in accuracy and information density but can be used as a solid input source for public transportation recognition. If necessary, local planners can edit the Open Street Map data base.

More sophisticated algorithms are able to detect the mode of more than 90% of trips correctly (Shen and Stopher 2014, 325ff.). However, the described limitations in mode detection show that active user integration for tasks like track validation or transportation mode editing can prove beneficial to the aim for adequate data sets.

Tracking: Mon Jan 7, 2018

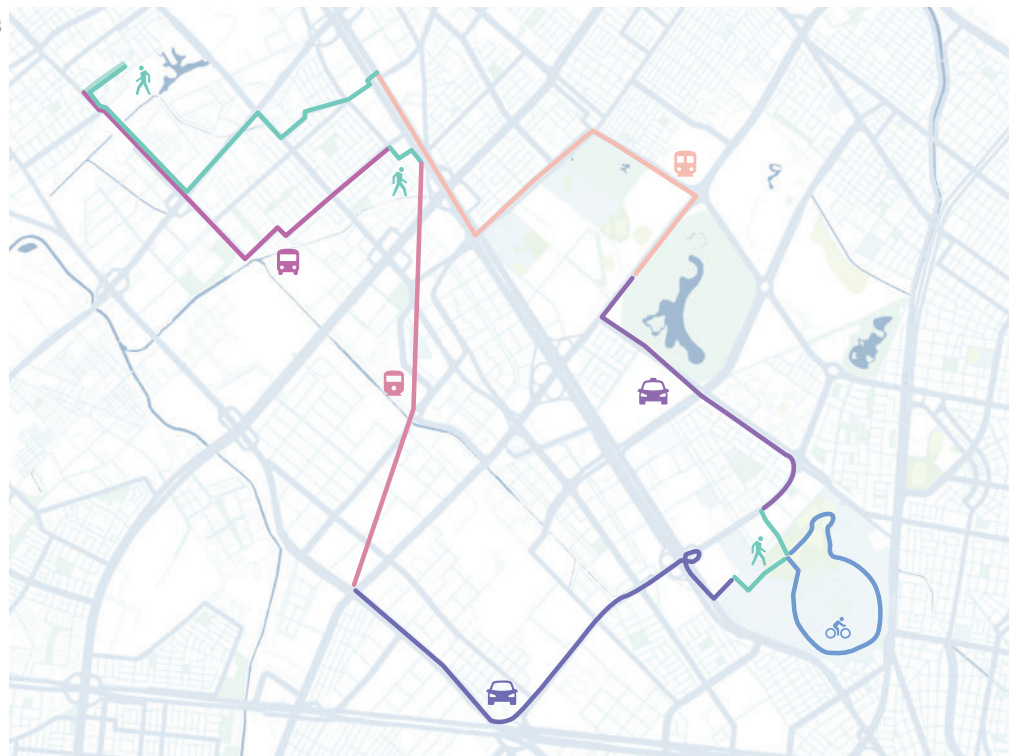|		
Walking	3.6 km	1h 30 m
Cycling	1.5 km	24 m
Car	2.3 km	18 m
Taxi	1.8 km	14 m
Bus	2.1 km	20 m
Metro	5.2 km	8 m
Train	4.3 km	12 m

Figure 2: A GPS-Track of a Person using different Modes

Identifying Trip Purpose

Next to the transport mode choice, the identification of trip purposes - the reason why trips were made - is a major issue in the creation of a reliable data basis for transport system planning. Common categories are home, job and leisure trips.

In order to detect a trip's purpose, the GPS data is combined with spatial data that includes land-uses (e.g. residential, industrial) and points of interests (e.g. restaurants, shops). If a trip starts at a residential area and ends at a school, chances are very high that it was a home-school or home-work trip. By adding a portion of personal data, e.g. the home and work address of a person, the predictions can be made even more precise.

However, the accuracy of automated detection of trip purposes is not yet as high as the precision of mode choice detection and therefore much less applied (Schönau 2016, 48). With very high data quality, just above 70% of accuracy has been reached, with many studies being in the range of 40-60% accuracy (Bohte and Maat (2009) and McGowen and McNally (2007), cited by Shen and Stopher (2014, 328f.)).

Best Practice: Using GPS Tracking in the Ukraine and Colombia

The Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) supports local governments in shaping sustainable mobility systems worldwide. In Ukraine and Colombia, the development agency GIZ cooperated with modalyzer, a smartphone-based traffic data collection app developed by the Innovation Centre for Mobility and Societal Change (InnoZ). It has been implemented in four cities within the project "Integrated Urban Development in Ukraine" to get a better understanding of the mobility demand, identifying travel patterns and behaviour of the population. Additionally the approach has been applied to Colombia in the 550,000 inhabitant city of Ibagué. The generated data will finally be used for the development of integrated urban mobility concepts.
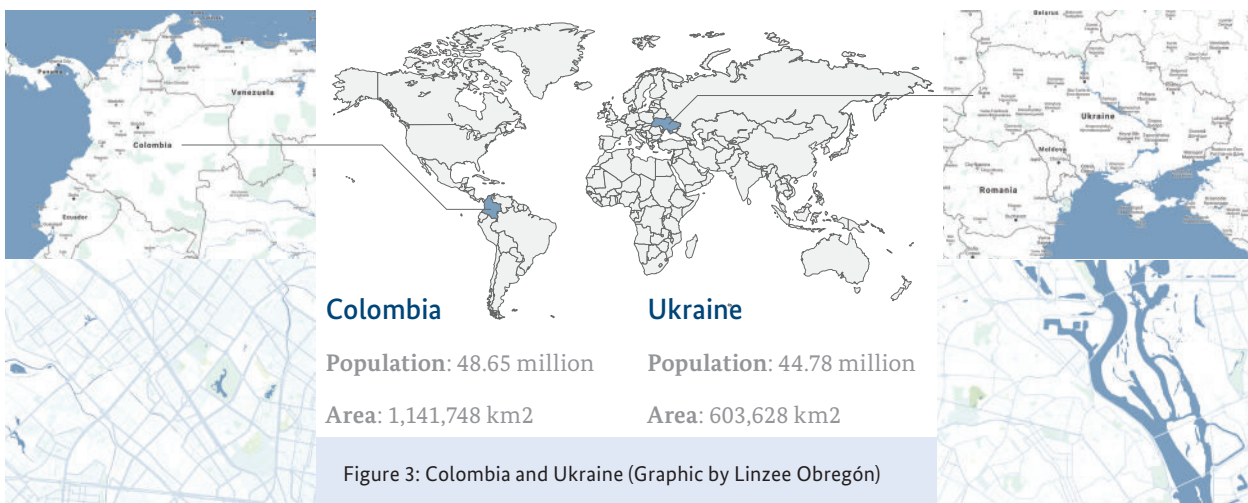
Modalyzer has been adapted to Ukrainian and Colombian needs and automatically identifies nine transportation modes and transit types. Further modes can be added manually – e.g. users can specify and edit trips by tagging them with the local transportation mode marshrutka, a local form of minibus service. The individual users of modalyzer can benefit from the app by monitoring their travel patterns through user friendly diagrams and tables, including information on their CO2 footprint and kilometres travelled by each mode.

In both countries, the tracking activity has been very high and the recruitment successful. Overall, more than 1,500 users tracked over 87,000 tracks and more than 210,000 km. This implies a significant user contribution to the planning in these cities.

Verifying GPS-tracking: Prompted Recall Surveys

As elaborated above, GPS post-processing algorithms are already capable of revealing information on trip mode and purpose to a large extent. In order to verify these findings or to gather further information, e.g. on reasons why a person choses a certain mode, so-called Prompted Recall Surveys (PRS) can be used. This increases the burden on the participant slightly, but ensures a higher prediction quality and may be used to enhance the algorithms, too.

PRS has also been used in the Ukrainian and the Colombian case: modalyzer prompts the result of the tracking to the user at the end of each day and the user ensures its quality by verifying the detected trips.

Colombia

Population: 48.65 million

Area: 1,141,748 km2

Ukraine

Population: 44.78 million

Area: 603,628 km2

Figure 3: Colombia and Ukraine (Graphic by Linzee Obregón)

Conclusion

GPS tracking technology has reached usability for travel demand data collection. It is cheaper and, when used appropriately, more detailed than traditional survey methods. However, in order to reach high accuracy in detecting trips, modes and trip purposes, high-quality spatial data and public transportation data is needed. Particularly in cities of the Global South, data availability on informal transit is often low. The Ukrainian and Colombian examples have shown that implementing Prompted Recall Surveys in a data collection application for smartphones is a viable option to overcome the issue of data availability and to enhance data reliability. Limited smartphone ownership is a remaining challenge.
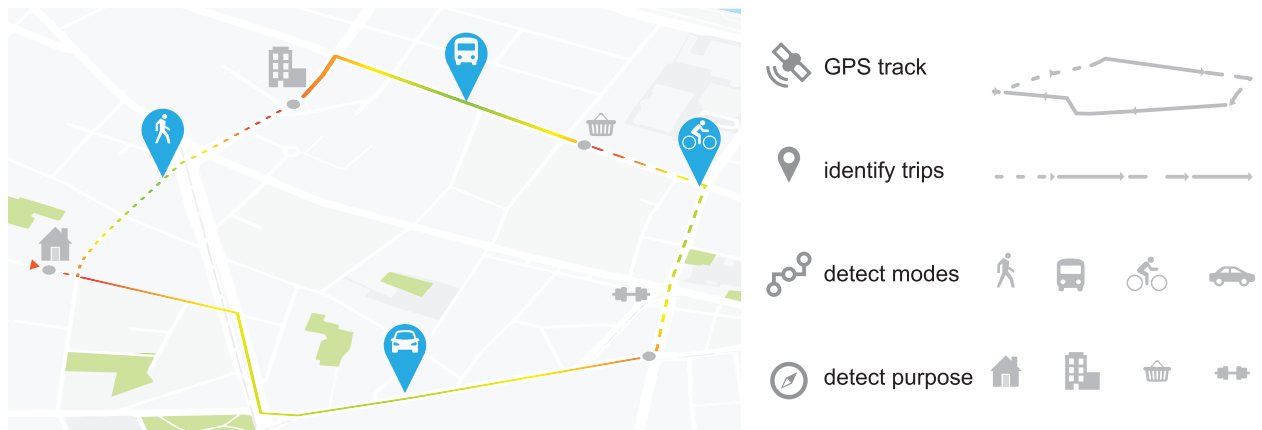

Figure 4: Overview on One Day of Tracking

How to apply all this?
Recommendations for transport planner

The evolution of GPS tracking has reached a stage that makes it suitable for large-scale data collection. But how can we as practitioners make use of the technology without re-inventing the wheel by programming another tracking application?

Option 1: Use existing Apps

Android smartphones make it very convenient for app developers to differentiate between walking, cycling and cars: Google implemented an algorithm in the operating system that is able to distinguish between these modes and made it easily available. Hence, there is a range of apps offering a tracking service relying solely on those categories. However, as they don't include public transport, they are worthless when it comes to professional data collection. Also, most apps you will find through the app stores are targeted on private use and do not include a feature to share the aggregated data with you.

Conclusion: not feasible

Option 2: Cooperate with Research Institutes

Academia has been involved with GPS tracking technology for a long time. Most findings that are presented here are results of research projects, which were most often conducted in cooperation with local authorities.

A main advantage in cooperating with research institutes is the support in methodical terms that the academic partner can provide. For example, Singapore's Land Transport Authority (LTA) has developed a smartphone-based transportation activity survey system (Yu Xiao et al. 2012), being now called "Future Mobility Survey" and applied it in Singapore (Cottrill et al. 2013). The Massachusetts Institute of Technology, KTH in Stockholm and the University of Queensland (ATLAS) have developed tools which implement some of the algorithms mentioned above as well.

In the end, your organization's staff should be able to analyze and make use of the collected data without needing to be data scientists. Therefore, it is important that the institutes hand over the data in an appropriate format and provide suitable tools for your staff. This

should also include software that supports aggregating, analyzing and visualizing the vast amount of data. This will enable your experts to use the data for their decisions.

In order to ensure a mutual beneficiary cooperation, these items should be agreed on in advance. This already requires expert knowledge – so make sure to have a capable person in your team.

Conclusion: much potential, but may need experts in municipal's rows

Examples:

MIT Singapore's Future Mobility Survey: https://happy-mobility.org/

MIT's Co2Go: http://senseable.mit.edu/co2go/

KTH's MEILI: https://www.kth.se/profile/yusak/page/trialling-smartphone-based-travel-data-c

Univ. of Queensland's ATLAS: http://www.civil.uq.edu.au/atlas

Option 3: Get the full package

In order to use the technology and make best use of the results, you can ask professional services to support you through the whole process. This includes the provision of an application, tools to sift through the data and advisory on how to gain insights.

Conclusion: recommended, if you don't have neither time nor experts in your organization

Examples:

InnoZ's Modalyzer: https://www.modalyzer.com

Option 4: Ask Google, your local mobile network operators or others

Google is collecting the movement data of Android phones all the time. By enabling the 'timeline' option in the Google Maps App, you can review where you spend your time and how you moved from A to B. Also mobile network operators (MNO), such as Orange, Verizon and others, collect cell phone movement data automatically. In an aggregated form, this data reveals demand and travel pattern in neat detail.

However, in practice it has proven difficult for many municipals to get access to this data. Google doesn't provide an interface. Some mobile network operators sell their movement data, but only in raw form. Small companies have focused on transforming it into OD-data.

Conclusion: while most comprehensive, access is very limited.

Example:

Positium (transforming network data into OD): https://www.positium.com/

References

- Axhausen, K. W. 1995. "*Draft Travel Diaries: An Annotated Catalogue 2nd Edition.*" ResearchGate. https://www.researchgate.net/publication/237558340_Draft_Travel_Diaries_An_Annotated_Catalogue_2nd_Edition.

- Battelle Transport Division. 1997. "*Lexington Area Travel Data Collection Test, Final Report.*" https://www.fhwa.dot.gov/ohim/lextrav.pdf.

- Bohte, Wendy, and Kees Maat. 2009. "*Deriving and Validating Trip Purposes and Travel Modes for Multi-Day GPS-Based Travel Surveys: A Large-Scale Application in the Netherlands.*" Transportation Research Part C: Emerging Technologies 17 (3): 285–97. doi:10.1016/j.trc.2008.11.004.

- Cottrill, Caitlin, Francisco Pereira, Fang Zhao, Iněs Dias, Hock Lim, Moshe Ben-Akiva, and P. Zegras. 2013. "*Future Mobility Survey.*" Transportation Research Record: Journal of the Transportation Research Board 2354 (October): 59–67. doi:10.3141/2354-07.

- Gong, Hongmian, Cynthia Chen, Evan Bialostozky, and Catherine T. Lawson. 2012. "*A GPS/GIS Method for Travel Mode Detection in New York City.*" Computers, Environment and Urban Systems, Special Issue: Geoinformatics 2010, 36 (2): 131–39. doi:10.1016/j.compenvurbsys.2011.05.003.

- Gong, Lei, Takayuki Morikawa, Toshiyuki Yamamoto, and Hitomi Sato. 2014. "*Deriving Personal Trip Data from GPS Data: A Literature Review on the Existing Methodologies.*" Procedia - Social and Behavioral Sciences, The 9th International Conference on Traffic and Transportation Studies (ICTTS 2014), 138 (July): 557–65. doi:10.1016/j.sbspro.2014.07.239.

- Litman, Todd. 2009. "*Transportation Cost and Benefit Analysis: Techniques, Estimates and Implications.*" Victoria Transport Policy Institute.

- McGowen, Patrick Tracy, and Michael G. McNally. 2007. "*Evaluating the Potential To Predict Activity Types from GPS and GIS Data.*" In . https://trid.trb.org/view.aspx?id=802616.

- Schönau, Manuela. 2016. "*GPS-Basierte Studien Zur Analyse Der Nachhaltigen Urbanen Individualmobilität.*" Dissertation, Universität Ulm. https://oparu.uni-ulm.de/xmlui/handle/123456789/4189.

- Shen, Li, and Peter R. Stopher. 2014. "*Review of GPS Travel Survey and GPS Data-Processing Methods.*" Transport Reviews 34 (3): 316–34. doi:10.1080/01441647.2014.903530.

- Yu Xiao, David Low, Thusitha Bandara, Parth Pathak, Hock Beng Lim, Devendra Goyal, Jorge Santos, et al. 2012. "*Transportation Activity Analysis Using Smartphones.*" In , 60–61. IEEE. doi:10.1109/CCNC.2012.6181051.