The Future Mobility Survey: Balancing accuracy and participant burden in a smartphone survey

Francisco Pereira[1], Caitlin D. Cottrill[2], Fang Zhao[3], Chris Zegras[4], Moshe Ben-Akiva[5]

[1,2,3]Future Urban Mobility IRG, Singapore-MIT Alliance for Research and Technology, Singapore

[4]Department of Urban Studies and Planning, Massachusetts Institute of Technology, Cambridge, MA

[5]Intelligent Transportation Systems Lab, Massachusetts Institute of Technology, Cambridge, MA

Email for correspondence: camara@smart.mit.edu

1. Objective

Technological advances, such as GPS devices and smartphones, have had numerous impacts in the realm of travel surveys. These advances bring both challenges and opportunities as we work to balance the potential to collect nearly unlimited amounts of data with the need to not overburden transport survey participants. This paper addresses the steps currently taken to manage this balance in the context of the Future Mobility Survey (FMS), a smartphone based travel survey being developed in the Singaporean context.

2. Methodology

The FMS survey (Cottrill et al, forthcoming) uses a combination of a downloadable smartphone app, available for Android and iOS, and an online prompted recall survey to collect both demographic and travel data from participants (Figure 1). Data collected from the smartphone application are uploaded to a central server, mapped, automatically cleaned and analysed and made accessible to the participant from the project website, where he or she is asked to provide detailed travel information via a prompted-recall survey.

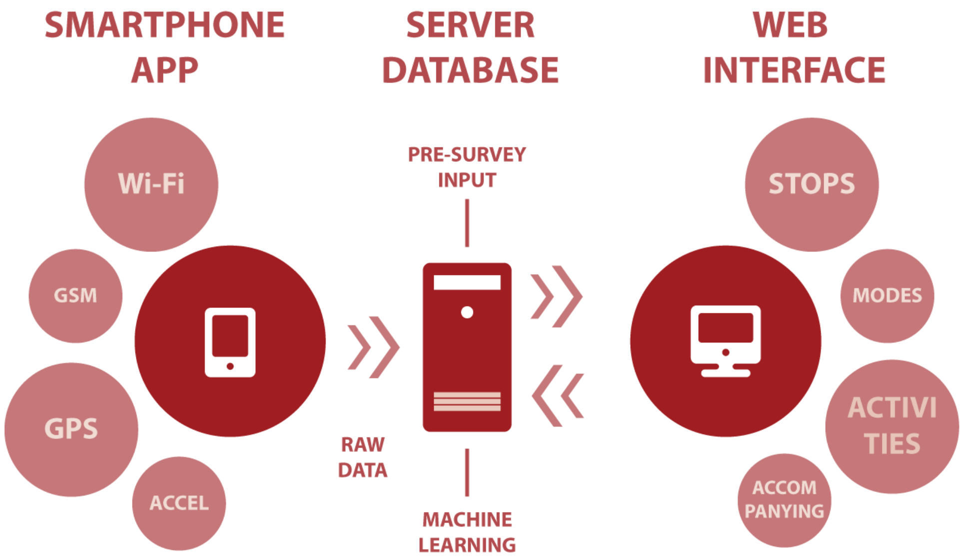

Figure 1: Survey architecture

In developing this system, we have worked to balance participant burden and data accuracy. Achieving this balance has entailed a three-part strategy: first, we have carefully developed the app for ease of use, non-intrusiveness and low power consumption; next, we implemented data analysis techniques that aim to prepare the web interface in such a way that the user needs minimal interaction; finally, we designed the web interface to become as intuitive and visually pleasant as possible in order to engage the user in activity validation. Developing, implementing and testing this strategy has provided considerable advancement in the use of ICT to support and enhance activity and travel data collection.

2.1. Smartphone application design

The key role of the smartphone devices in our project is to act as data loggers. Overall, the FMS platform is designed to allow other types of devices to upload data to the server, such as dedicated GPS loggers or older phones. Thus, our app is deliberately “silent” in the sense that nothing at all is expected from the user besides making sure it is running. However, since the Apple App Store requires availability of some minimal visible functionality, we included simple stats such as total amount of data collected and a map-based visualization of recent traces. For coherency, we assumed the same design for the Android version. In any case, the interface is designed to minimally influence travel behaviour and avoid distracting the user with any interaction.

In Figure 2, we provide a few snapshots of the current interface.

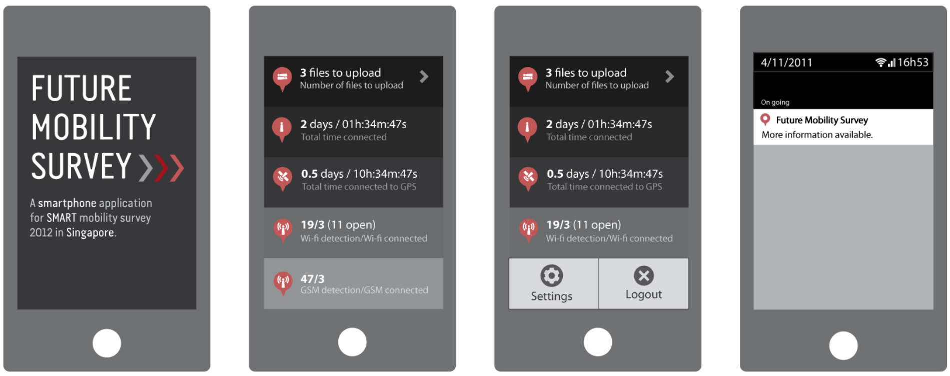

Figure 2. Smartphone app interface

A crucial challenge of this kind of technology is battery consumption, mainly due to the use of the GPS sensor. Effectively, this can become the major source of burden for participants while recording their trips and activities. On the other hand, since it is only available outdoors, the typical amount of GPS data collected during the day for a normal user is relatively small compared to the 24 hours of a day. As a result, a logger with fixed frequency GPS sampling will systematically fail to “get the first fix” (GTFF), which itself demands high battery consumption. Another limitation of fixed frequency is that the GTFF process needs some time to converge to a position estimate, a period that depends on many factors, such as number of visible satellites and their geometry, almanac correctness, ambient noise or sensor quality. So, if not enough time is given for each GTFF attempt, the logger will keep failing and the data will be poor while still draining the battery.

We apply two techniques to minimize this problem, namely, using the accelerometer for detecting “still”, or no-movement, stages (thus, GPS will not be needed) and what we call “phased sampling”. The concept is the following: we collect intensive GPS data (1Hz) for a continuous period of time and then deliberately sleep for a period. In this way, the number of GTFF processes will be much smaller than fixed sampling, while providing adequate time for location estimation. After a period of fine-tuning, current configurations used are 3 minutes for collecting, 2 minutes for sleeping.

2.2. Data analysis

The data analysis component serves to transform the logged raw data into understandable information for the user. It is particularly focused on inferring stops, modes and activities. Figure 3 shows the general process followed.

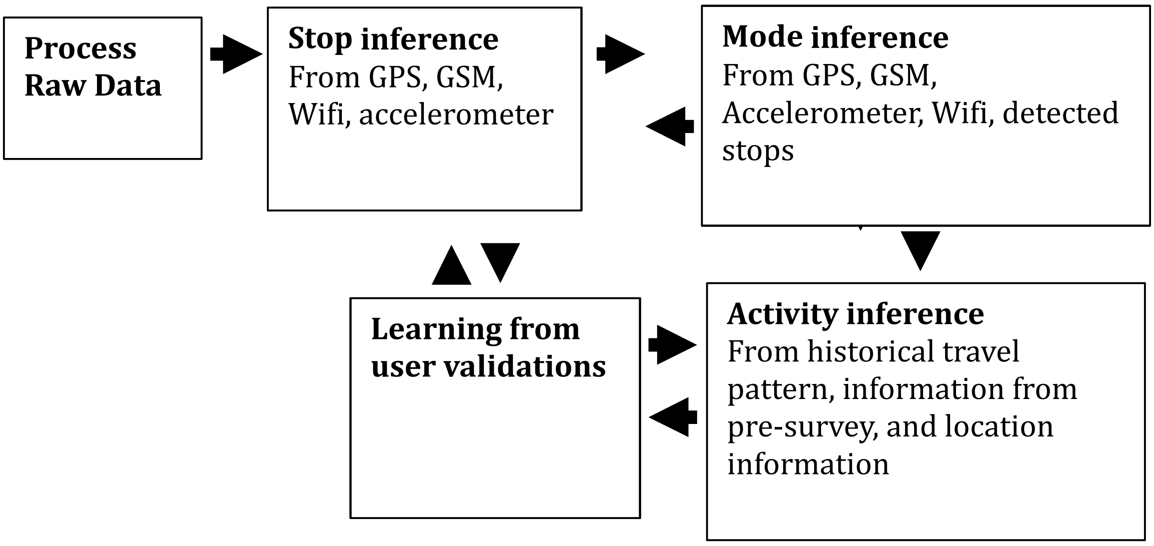

Figure 3. Data analysis process

The algorithms used in each step vary. The "process raw data" step consists of a series of scripts for cleaning, composing and temporally aligning the incoming data for use in the subsequent analysis steps. The "stop inference" applies a rule-based algorithm in two phases: first, it matches spatial/temporal windows (Hariharan and Toyama, 2004) to the data to obtain candidate stops; then, it uses wifi and GSM data to merge stops, particularly using accelerometer information to detect "still" periods (where, although the GSM is "jumping", the user has remained in the same place). This step also uses past validation information to match a user's recurrent places (e.g., home, work) with GSM signatures and adds/removes stops based on mode detection results (e.g. there must be a stop for change mode/transfer between any two different modes).

The "mode inference" step applies a machine-learning algorithm, support-vector machine or SVM (Cortes et al, 1995), to accelerometer and GPS data to identify the mode out of the set of car, bus, subway, walk, bicycle or motorbike. Finally, the "activity inference" matches the historical data, namely the previous validations, to current stops to identify recurrent locations. Current development of this module also considers contextual information such as the Points of Interest or the mode interchange areas.

The goal of the "learning from user validations" step, under development, is to systematically update these algorithms over time, i.e. perform online learning.

2.3. Web interface design

The third component of our system is the web interface. In this case, the major challenge lies on the side of the user interaction design. First of all, the notion of an activity diary is somewhat strange to the layman, thus the first difficulty for any regular user is simply to understand what to do. We note that, to properly fill the activity diary (i.e. to *validate* their data), the user needs to verify the following information:

- Stop locations

- Stop durations (start and end time)
- Activities performed in detected locations
- Travel modes taken between stops
- Costs/options associated with travel modes such as bus number, parking cost or accompanying persons

The first challenge is thus to communicate what each concept means, explain each individual task and options, and prompt the user with visual and textual cues to recall the actual activities and travel. Relying on lengthy tutorials to guide the user is known to have limited impact so we created a home page (Figure 4) with the essential recruitment information and links for a quick reference (*FAQ*) and detailed information (*Participate!*). We also provided leaflets, a video and a help line. Finally, we also added a personalized remote desktop help service that will be described below.

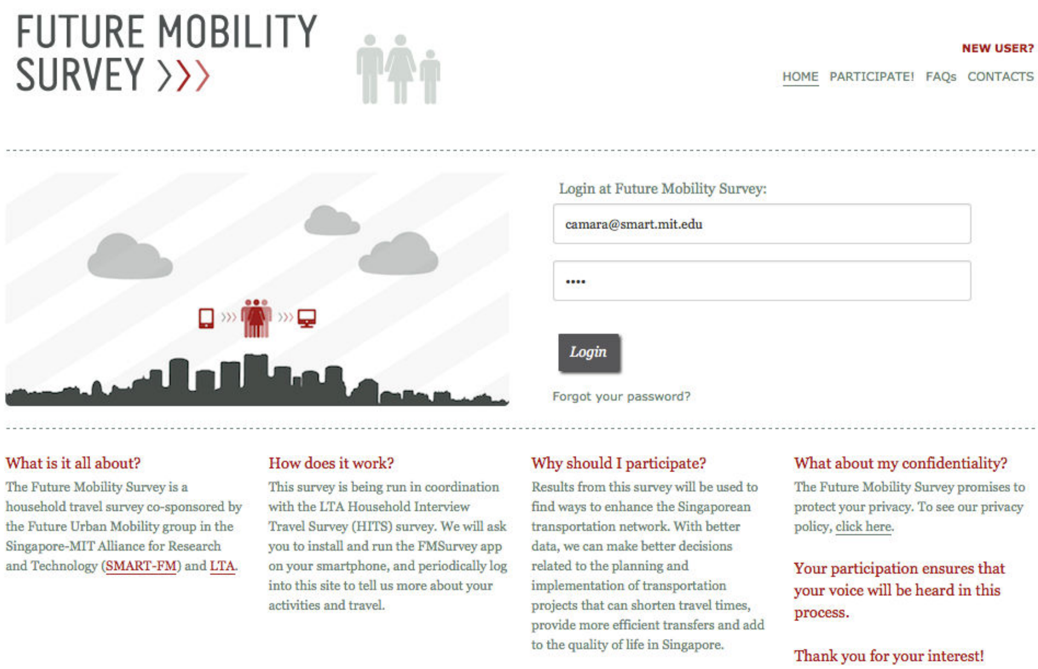

Figure 4. Opening page of FMS

Even with the above information and other types of material (a leaflet and video), the task is still understandably challenging for the vast majority of users given the amount of information that is required from them to confirm/change. We designed several user interface proposals that considered aspects such as grouping (all stops together; individually; grouped by tours), sequencing (stops, activities, mode; stops, modes, activities, etc.), validation mode (on the map, using textual interface), icons, colors and font sizes. We then conducted usability tests for the different web interface versions. We summarize a few of the key findings:

- Understanding the continuity of time is essential, particularly that each activity/trip starts when the previous trip/activity ends. Previous simplified interfaces that only showed times for either activities or trips were deemed incomplete/confusing;
- Unless the traces have high quality, their visualization on the map can become confusing. GSM or low quality GPS data can become quite "jumpy". After trying other intermediate solutions (e.g. represent the trace precision with different line types; only show high precision), the least disturbing solution was found to be adding a "show traces" option;
- The ability to understand the map varies widely among users; this implies that the text descriptions need to be as self-explanatory as possible;

- The task of deleting stops/trips is much easier than adding them. As a result, we fine-tuned the stop detection algorithm to generate more false positives than false negatives. This generates more stops than what we would intuitively desire but the interface burden is lower than otherwise;
- Users are sensitive to font size as well as to amount of information on the screen, especially in the first uses of the interface;
- Elaborate interface features, such as right-click, mouse-hover feedback or drag/drop are inappropriate for a vast amount of the public so we restricted ourselves to more traditional solutions.

In Figure 5, we show the current activity diary interface as the user sees it after selecting the validation day.

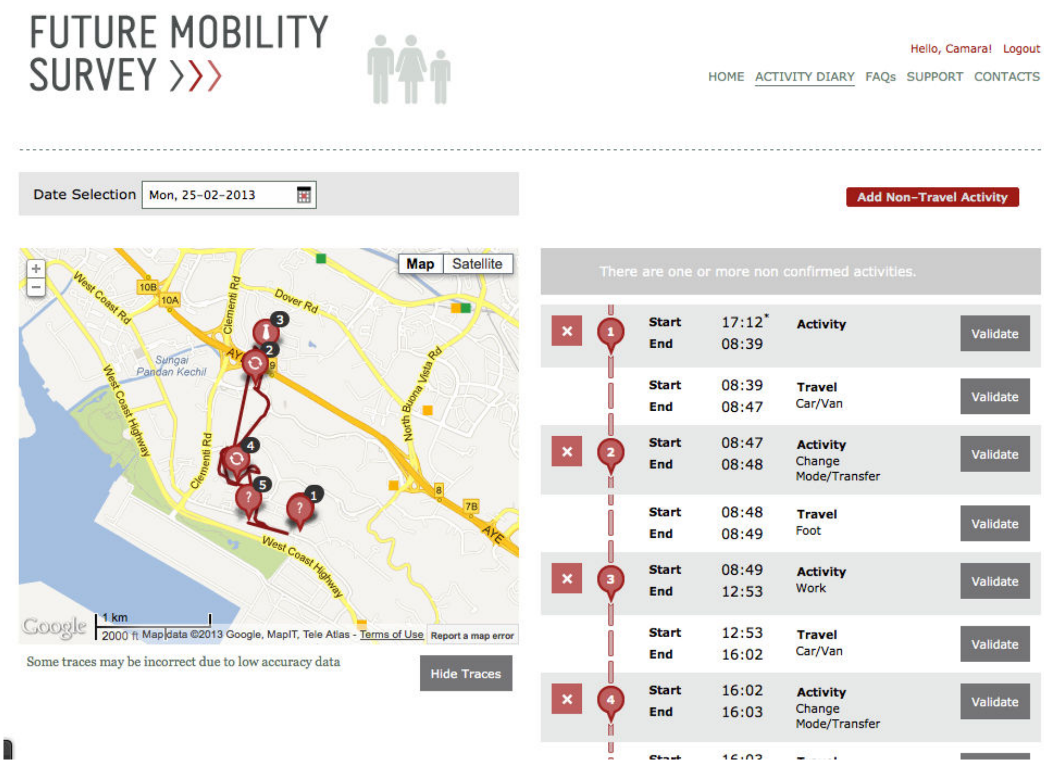

Figure 5. Activity diary main screen. The day started at 08:39am (after arriving to that location the day before at 17:12) with a car trip to stop 2, followed by a 1 minute walk to stop 3, etc.

The task is to click on all the "Validate" buttons in the intended order and edit or confirm each stop/activity and trip. Figures 6 and 7 show the validation of stop 1 and the subsequent trip, respectively.

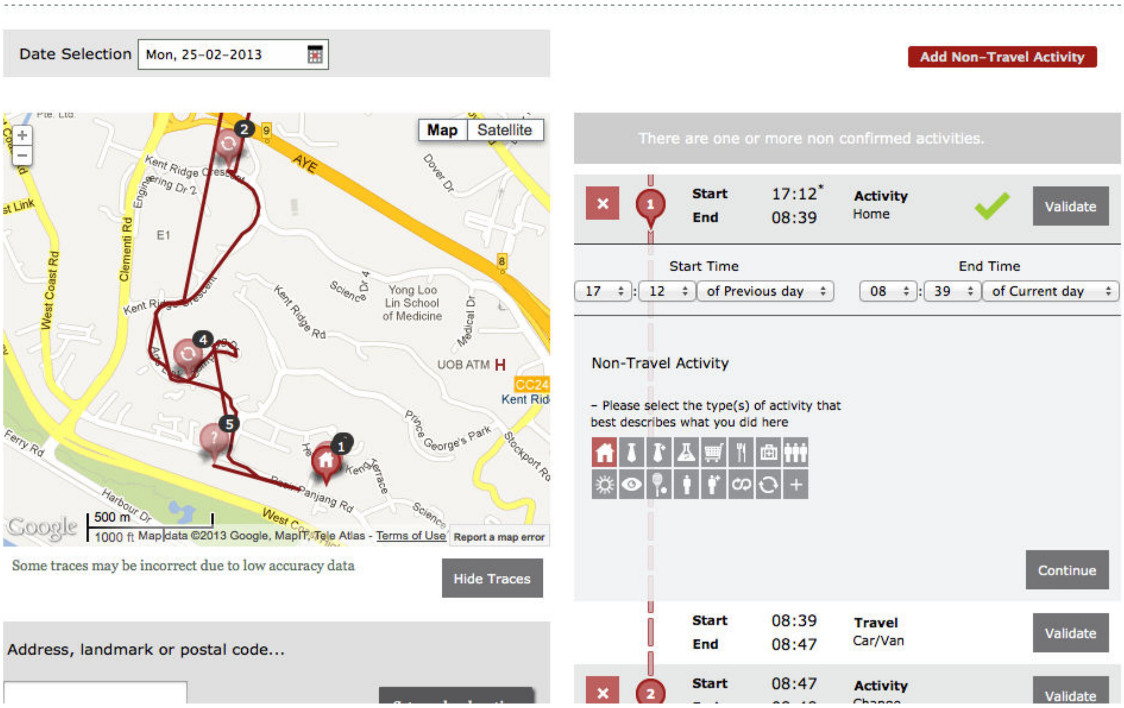

Figure 6. Validating stop 1

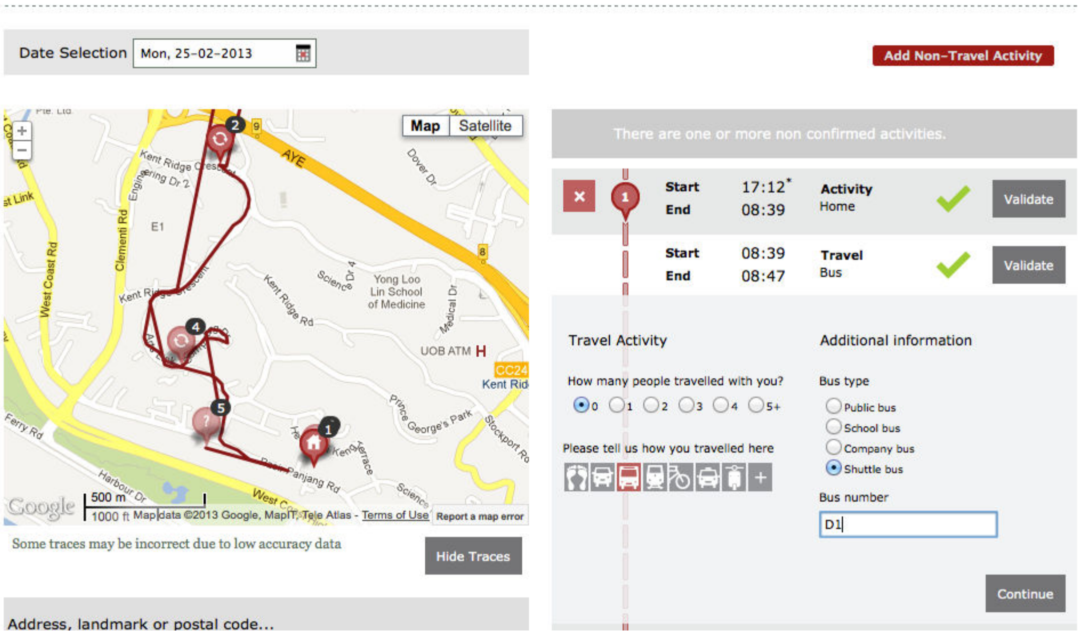

Figure 7. Validating a trip

Finally, we reiterate the fact that, for the majority of the users, the learning curve can still be steep. In response to this, we added a new feature that allows for remote desktop help, based on the Firefly® technology[1]. With this feature, the user can provide an FMS representative (the "helper") temporary remote access to her FMS page within the browser. The helper can then either use a chat window or call the user to guide her through the

---

[1] http://usefirefly.com/

interface. The helper can see the user's FMS page and highlight certain pre-defined areas. To initiate the process, the user only needs to click on the "need help?" button on the left-hand side of the window (Figure 8). Then, she can either talk to the helper immediately (if available) or schedule a tutorial session (Figure 9).

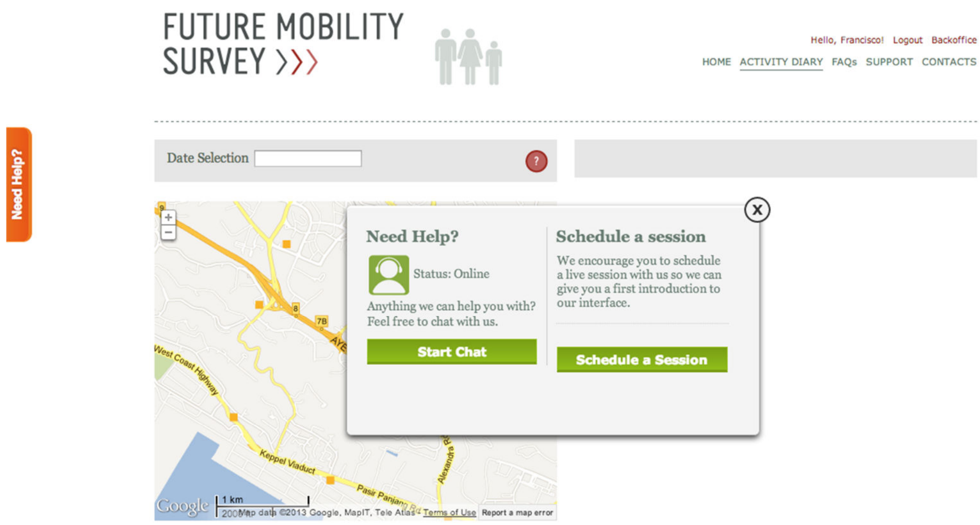

Figure 8. Accessing Firefly based help.

Figure 9. Scheduling a help session.

3. Conclusions

At the time of this writing, FMS is running in Singapore, with a targeted sample size of 1000 subjects. The findings and decisions reported here correspond to an on-going process currently being thoroughly tested. Upcoming publications shall bring a more conclusive and verified set of conclusions, particularly in comparison with the Household Interview Travel Survey (HITS 2012), a *traditional* type of survey simultaneously being run in Singapore.

Our iterative process of testing, improving and refining the system revealed that implementing a smartphone-based activity survey requires various trade-offs in order to be palatable to survey participants and useful for data collection purposes. In this paper, we report on the critical needs to develop a comprehensive survey platform that respects this balance.

The ubiquity of advanced technologies in the mobile environment reveals great potential for expanding data collection methods. Taking advantage of such potential, however, will require careful attention to competing needs. This paper presents a comprehensive approach for the development of a basic data collection platform that may be modified or expanded for use in a variety of contexts or situations.

References:

Cortes, Corinna; and Vapnik, Vladimir N.; "Support-Vector Networks", Machine Learning, 20, 1995.

Cottrill, C., Pereira, F. C., Zhao, F., Dias, I., Lim, H. B., Ben-Akiva, M. and Zegras, C. “Future Mobility Survey: Experience in Developing a Smart-Phone-Based Travel Survey in Singapore”. Transportation Research Record: Journal of the Transportation Research Board. Forthcoming.

Hariharan, R. and K. Toyama. Project Lachesis: Parsing and Modeling Location Histories. in Geographic Information Science: Third International Conference, GIScience 2004. 2004. Adelphi, MD, USA: Springer-Verlag GmbH.