WCTR 2013 RIO

SELECTED PROCEEDINGS

Leveranging data for the development of transport sustainability indicators

CAITLIN D. COTTRILL (CORRESPONDING AUTHOR), LECTURER SCHOOL OF GEOSCIENCES, UNIVERSITY OF ABERDEEN, SCOTLAND CONTACT: C.COTTRILL@ABDN.AC.UK

SYBIL DERRIBLE, ASSISTANT PROFESSOR CIVIL AND MATERIALS ENGINEERING, UNIVERSITY OF ILLINOIS AT CHICAGO, USA CONTACT: DERRIBLE@UIC.EDU

This is an abridged version of the paper presented at the conference. The full version is being submitted elsewhere. Details on the full paper can be obtained from the author.

ISBN: 978-85-285-0232-9

13th World Conference on Transport Research

www.wctr2013rio.com

15-18 JULY 2013

Rio de Janeiro, Brazil

Review Response and revisions, "Leveraging Data for the Development of Transport Sustainability Indicators" (C. Cottrill and S. Derrible)

Thank you for reviewing our paper, "Leveraging data for the development of transport sustainability indicators." We found the comments and suggestions quite helpful, and have responded in the following ways:

Reviewer 1:

- Privacy needs to be considered more - there is a section, but not integrated into the rest of the paper.
    - We have added statements to the sections on New Technologies (p. 6) to better tie this concern throughout the paper. In particular, we have addressed the privacy concerns related to use of GPS loggers (p. 7) and cellular and smartphones (p. 9).
- Quality of data needs to be brought more to the fore - particularly in cities where the GPS signal can be lost.
    - We have added a section on "Data quality and management" (p. 15) that addresses the need to develop accurate metadata and data management processes to ensure that data are handled properly. In addition, a sentence has been added to P. 7 specifically addressing the question of GPS.
- How to analyse the data once scaled up to a full survey of (say) 100 people - its easy to analyse for 1 person when that person is the author, but how to analyse for 100.
    - This has been addressed in the section on Assessment (p. 9) and in "Data quality and management" (p. 15).

Reviewer 2:

- … the authors list the uses of indicators on page three. These are not really returned to later when assessing potential for the future (page 9). Since project evaluation and other listed applications are critical in transport policy, perhaps these should be discussed in greater detail in the context of new data or simply omitted. The list on page three features many bullet points, only some of which are relevant in the current draft.
    - We have added text under the "Assessment" heading (pp. 9 and 10) to better clarify how we are addressing the indicator uses listed on p. 3.
    - In addition, on p. 10 an example has been added addressing how ongoing collection of data may allow for more accurate data to be collected on the effects of disruption. We hope that this better address the potential for project evaluation.
- I also hope to see a bit more about data cleaning and processing. Perhaps a table of descriptive statistics can show how the data fits within conventional transport models (or if new models are needed). In addition, some example of the anomalies noted on page 12 in bullet points would be nice.
    - We have added a section on "Data quality and management" (p. 15) to address this issue. In the interest of space, we have not included the suggested table, but we feel that the broader discussion is a valuable addition to the overall paper.

LEVERAGING DATA FOR THE DEVELOPMENT OF TRANSPORT SUSTAINABILITY INDICATORS

Caitlin D. Cottrill (Corresponding Author), Lecturer
School of Geosciences, University of Aberdeen, Scotland
Contact: c.cottrill@abdn.ac.uk

Sybil Derrible, Assistant Professor
Civil and Materials Engineering, University of Illinois at Chicago, USA
Contact: derrible@uic.edu

ABSTRACT

While myriads of people are putting much effort into making transportation more sustainable, measuring the results (through models, evaluation, or scenarios) is not trivial. In fact, not only is the selection of indicators challenging, but efforts made to design useful indicators are often hampered by the presence of data that are erroneous or incomplete. Nevertheless, in this era of Big Data, the significant penetration of new technologies such as smartphones and smart infrastructure could hold the key to developing more relevant and comprehensive indicators. In this paper, we review commonly used indicators and discuss their limitations with respect to the data upon which they are built. We then describe several new technologies that hold promise for the collection of more pertinent and accurate data sets upon with indicators may be built. Finally, we discuss their potential for the future and illustrate a hypothetical scenario by reviewing a one-day GPS traces of one of the authors, followed with a short discussion of the privacy implications of these methods. While the first and obvious application of new technologies will be to improve much needed accuracy, successfully combining different sources together could hold much potential from model calibration to real time operations.

Keywords: sustainable transportation, indicators, new technologies, smartphones

INTRODUCTION

While one of the biggest transportation planning challenges of the 20th century was to provide efficient systems to accommodate the automobile revolution, the 21st century has seen a shift towards a focus on the sustainability criterion. Consequently, transportation planners worldwide have undertaken the task of increasing the sustainability of transportation systems and networks; nevertheless, whether at the design stage or when assessing system performance, *measuring* sustainability is not straight-forward (Zegras 2005). Indeed, while efficiency can arguably be measured by volume and flow, sustainability measurements are not limited to greenhouse-gases or vehicle-kilometers traveled. As a result, and despite the lack of agreement on a standard definition, a cornucopia of indicators have been developed and applied that try to quantify every facet of sustainable transportation.

One problem with these indicators, however, is that they tend to be heavily reliant on data, which is often erroneous or incomplete. There is hope for improvement, however, as the 21st century is also seeing the advent of pervasive computing and new technologies, often grouped under the term "Big Data", which can not only make current indicators more accurate but also potentially enable the development of new and more comprehensive indicators. There are seemingly endless opportunities to leverage the proliferation of data that may be collected via a dispersed set of sensors (for example, those available in cellular and smartphones), which simultaneously can be used to calibrate models and measure the impact of new policy and infrastructure undertakings. More generally, information technology and computer science have much to offer for a sustainable future (Millett, *et al.* 2012).

Recent technological advances have increased the availability and decreased the price of person-based location loggers (either stand-alone or based on a cellular or smart phone platform), which will greatly expand the ability to collect more detailed data on trips on all modes, and also provide the potential to collect data on a broader spectrum of travelers (for example, those who are non-drivers). This is but one example of the increasing availability and affordability of distributed sensors that may be used to gather data for the support of sustainability indicators. In general, the increasing availability of small, inexpensive sensors and data storage facilities, as well as decreasing costs of communicating and sending these data stores to interested parties, has greatly widened the scope for indicator development and use.

The impetus for this work is to open the discussion for more targeted research that concentrates on using new technologies as a source of data for sustainability indicators, both as primary resources, and as supplements to more traditional sources such as household travel surveys. In this article, we first recall why and where indicators are used. Secondly, we conduct a review of existing, commonly used indicators and the data upon which they are built. Third, we introduce new technologies and their potential in terms of data collection, and we go through several specific examples of their applications. Finally, we offer one example where the activities and traces of one of the authors are examined, followed by a short discussion of the privacy and management implications of such technologies and the data sets they collect.

Our analysis provides an initial platform that reviews the state of the practice and highlights some of the possibilities from new technologies for determining sustainability indicators. In addition, by focusing on small, inexpensive data collection technologies, these indicators provide useful suggestions for a wide variety of locations and contexts. Considering the proliferation of mobile phones, and the constant improvement of their sensors, tapping into this seemingly infinite database seems to be a must for the future of transport and land-use planning.

CURRENT PRACTICE

The need for indicators

Indicators are measures that provide specific information about the property or state of a system. Whether as simple measures (i.e. number of kilowatt-hours consumed), combinations of measures (i.e. floor-to-area ratio) or complex measures (i.e. gross domestic product), indicators are in constant use worldwide for a variety of purposes. While often regarded solely as a means to assess system performance, indicators can be used for many more applications, including:

- General comparison (ex: GDP per capita)
- Project design evaluation (ex: benefit-cost ratio)
- On-going project verification (ex: percentage of project completed to date)
- Project completion (ex: number of months a project was delayed)
- Real-time evaluation (ex: electricity demand by district)
- Model calibration, inputs and outputs
- Scenario development (ex: employment rate)
- System performance (ex: percentage of times a service was on-time)

Naturally, individual indicators are rarely capable of serving this entire spectrum. In fact, many indicators are the product of other indicators compiled or collected in different applications, unraveling a complex chain of interactions, which can make identifying helpful and accurate indicators a complicated task.

Several criteria have been proposed to assess the resourcefulness of an indicator (Keirstead and Leach 2008; Munda 2005) or to account for limits in decision-making power by various authorities (Derrible et al. 2010). Focusing on urban planning, Lautso, *et al.* (2002) identify the four key characteristics of relevance, representativeness, policy sensitiveness, and predictability as criteria for inclusion of indicators into an integrated sustainability model. While these criteria do not necessarily fit the panel of applications listed above, they constitute a starting block to select relevant indicators and reflect the complexity of factors that must be taken into account when developing useful indicators.

Indicators

Indicators in general have a mixed history, with Cobb and Rixford (Cobb and Rixford 1998) stating that, “Although humans have been using indicators since the dawn of history, the self-conscious use of indicators to judge social conditions is of much more recent origin.” The use of indicators in the planning and implementation of transport projects was pioneered in the 1950s, and initially, with the advent of the freeway era, was almost entirely concentrated on mobility-based variables (e.g., traffic flow and speed) as a means to reduce congestion. Today, we recognize that these indicators captured only part of the problem, and their solutions often made the transport system worse. The integration of land- and energy-use indicators, in addition to environmental indicators, was therefore recommended as a way in which the various interactions between transport systems and sustainability could be made more explicit (Newman and Kenworthy 1989), and more effectively integrated into transport modeling and development exercises. Additional integration of human health indicators, in particular those related to air and water quality, has further strengthened the use of indicators as methods by which to better link transport and sustainability (Litman 2007).

A vast literature exists on the topic, whether in the form of journal articles and proceedings (Borken 2003; Fabish and Haas 2011; Gudmundsson 2003; Samberg, *et al.* 2011), institutional and government reports and white papers (Bongardt et al. 2011; CIA-Asia and EMBARQ 2007; European Commission 2011; UNCDR 2010), and so on (Litman 2012). Note that (Bongardt et al. 2011) offers a particularly exhaustive list of current indicators.

Table 1 shows a non-exhaustive list of existing indicators. Most existing indicators seem to fall within one of the following six categories: accessibility, safety, environmental impacts, infrastructure, monetary, and institutional. These categories can in turn be classified within the three pillars of sustainability, society, environment, and economy, where we also need to add the role of governments due to their critical role in transportation planning and operations. Note that we also find some overlap between these categories and recognize that safety indicators can be both social and environmental indicators, and infrastructure indicators belong to both environmental and economic pillars. Although readers may disagree with part of this classification, no indicators seem to be able to represent a complete sustainability indicator, and those that have been described in the literature are simply not measurable (Zegras 2005).

Table – 1 Non-exhaustive list of existing indicators

Social	Accessibility	Distance to education and employment
		Household budget allocated to transportation
		Access to public transportation
		Costs per km driven
		Universal design
		Cultural interchange
	Safety	Deaths per kilometer of road
		Injuries per kilometer of road
		Traveler assaults
Environmental		Hazardous waste
		Health impacts
	Env. Impacts	Mode split
		Greenhouse-gas per capita
		Energy consumption per kilometer driven or per ton-kilometers
		Cleanliness of local vehicle stock (energy efficiency)
		Air quality
		Noise pollution

		Fuel quality
	Infrastructure	Share of land used for transportation
		Length of roads
		Length of transit lines by mode
Economic		Availability of bike paths
		Interconnectivity of transport infrastructure
		Quality of roads
	Monetary	Gross Domestic Product per capita
		Capital expenditure
		Traffic flow and speed
		Minutes lost in congestion
		Parking prices
		Transport efficiency for freight
		Cost per kilometer (fixed and variable)
	Institutional	Participatory transport planning
		Road safety laws
Inst.		Air quality laws
		National/local transit strategies
		Inspection and maintenance

Current Issues in Data Collection

Present indicators have been built upon data sets that are somewhat limited based on the ability of governmental agencies to design and fund systems, networks, and programs to collect information that may be incorporated into their development. Existing indicators reflect such limitations, as they are often built upon data such as traffic counts, crash reports, census data, and air quality information that are systematically collected through state, local or national government agencies.

Even the dominant "travel survey" offers only limited information; indeed travel surveys contain a very small sample size of the population (typically from 1 to 5%), are infrequent, rarely differentiate seasons and weather conditions, and offer a poor level of accuracy (for example, interviewed individuals often forget trips or do not recall the exact start and end times of their activities), to name a few.

In essence, this situation results in data sets that are collected and updated sporadically, making them outdated almost as soon as they are presented in a workable form. As Pei, *et al.* (Pei et al. 2010) note, "Often, the comprehensiveness is limited by the availability of data or the high costs of data collection." Moreover, data sets are often not released in their entirety or are made scarcely available to the public, which can bias results from studies that use the data directly. Such release-lags and limits on availability are becoming especially problematic at the moment, where fast computing allows for high levels of disaggregation.

These issues do not even include the vast amount of data that is not currently collected and yet retains a critical role. Arguably one of the biggest gaps in current data collection techniques is the lack of information on interpersonal relationships between individuals. Many trips involve the co-location of two individuals whose decision to meet at a specific point was not determined by their respective utilities or shortest paths. This is especially problematic in this age of social networks that can have strong impacts on personal mobility (Sharmeen, *et al.* 2010; Silvis, *et al.* 2006). Moreover, although an old topic (Lynch 1960), the study of the impact of urban form on travel behaviors is another area that suffers from the lack of adequate data.

An additional concern has been inequalities in respect to the types and coverage of data that have been collected. Although a more novel technique, the use of GPS loggers offers a good example. When initially incorporated into household travel surveys, GPS loggers functioned as in-vehicle devices and thus were primarily used to collect data on motorized vehicle trips (Wolf and Lee 2008). While these devices provided an excellent supplement to traditional paper-based activity diaries and showed marked influence in improving trip reporting (Wolf, *et al.* 2001; Zmud and Wolf 2003), they offered little supporting data for walking, bicycle and transit trips. In addition, vehicle-based GPS loggers created an inherently biased sample, as their use was limited to drivers.

Nowadays, with the emergence of person-based sensors, such as those included in cellular and smartphones, it is possible for data to be collected in a way that is cost-efficient and represents a more inclusive approach to indicator development. Additional information on these and other emerging technologies will be discussed in the next section.

NEW TECHNOLOGIES

The availability of quantifiable data sets that may be used for purposes of developing, testing and validating sustainability indicators is providing scope for increased quality of models and forecasts into the overlap between transport and the environment. Widespread deployment and use of a variety of static sensors that collect environmental data (including basic information on temperature, humidity, and light, as well as more detailed data on CO2 emissions and particulate matter, water quality, traffic patterns, and other data) in conjunction with the growth in person-based sensors on platforms such as smartphones and tablet computers, has created a wide range of data sets that may be mined and analyzed to create useful indicators for purposes of modeling and crafting a variety of sustainability indicators. While focusing on human-based data, rather than infrastructure or vehicle-based information, brings with it a new set of issues (in particular, concerns regarding privacy – discussed below), it also opens up a wealth of more geographically refined data, that is more illustrative of the specific environment in which data are generated. Here, we will describe and discuss some extant and emerging technologies and the data sets that they make available.

GPS loggers

As noted above, GPS loggers are increasingly being used for location-related data collection, particularly in household travel surveys. While it is observed that there are some limitations in these data collection efforts (namely, in scope and data quality depending upon the geography of the area under study), it is evident that such efforts have had great benefits in the ability to develop precise, measurable indicators related to travel and location. One of the greatest benefits of GPS data availability with respect to indicators has been the ability to participate in more "fine-grained" transportation analysis. The traditional four-step model is based on the fairly coarse grain of origin and destination zones, which may under- or over-emphasize certain links in the network. GPS data allow trips to be assigned to specific network links, and anchored at precise origin and destination locations. The availability of this data has led to a greater emphasis on activity- and agent-based models, which also allow for more integrated consideration of non-transport factors (such as air quality and exposure). Here, GPS data may be used both as an indicator in itself to be incorporated into models (for example, transit travel time), as well as a factor in developing indicators from model outputs (for example, CO2 exposure along a specific network link). One concern evident in the use of highly detailed GPS data, however, is the potential for identification of individual travelers in the network (for example, recent work by (de Montjoye et al. 2013) indicates that as few as four spatio-temporal points can be used to identify up to 95% of individual travelers), which highlights the need for caution as data are collected, used and disseminated. In addition, GPS traces often need to be processed and "cleaned", especially in densely built areas. Although this is a non-negligible hurdle, this problem does not seem to pose any significant threat to the usefulness of GPS traces.

RFID

Radio Frequency Identification (RFID) technologies also have great potential as data sources for indicators, as they are already in common use for identification and tracking purposes. Particularly with emerging interest in incorporating sustainability into supply-chain management and planning, RFID tags can provide detailed data on the transport and movement of goods across space, which greatly supplement current data sources. RFID records are already commonly used for the development of performance indicators for logistics service providers, including equipment utilization rates, empty-to-loaded backhaul mile index, and others (Krauth et al. 2005); however, the potential to use RFID data to develop additional indicators for use in a variety of fields is growing. RFID systems tend to be relatively inexpensive and, unlike GPS, do not need a direct line-of-sight to function effectively. These benefits, in addition to their use flexibility, indicate that data from RFID sensors may be incorporated into the development of sustainability indicators for such features as tracking and modeling efficient use of parking facilities, freight logistics (for example, efficient and effective tracking and routing of perishable goods), and as cost-effective probes for collecting data on traffic flow, congestion, and incidents. Such data can be collected from a widely dispersed network of sensors depending upon coordination of agencies and organizations across space, and has the potential to greatly enhance the variety of data-rich indicators that can be developed.

Real-time infrastructure data collection

Although mostly at the conceptual phase, real-time infrastructure seems to hold great promise as a source of data (Millett et al. 2012). The best known applications are the smart grid and smart metering (Grob 2010), which will enable collection of daily information on electricity consumption, in particular for electric vehicles, and more accurate estimation of greenhouse-gas emissions. We will also be able to locate when and where a vehicle is unplugged and plugged in again, which will allow us to better estimate travel demand, travel times, and energy consumption patterns, to name a few. Real-time infrastructure also applies to public transportation, where one can easily imagine better indicators to react instantly to service disruptions, adjust service to real time demand, and so on (Sandblad et al. 2010). The coupling of real-time infrastructure data with other sources could unleash vast possibilities to generate and calculate multiple indicators in real-time and at a highly disaggregate level, which offers clear benefits compared to the current practice of using mainly annual averages.

Cellular and smartphones

The growth in mobile phone subscription rates has vastly outperformed that of traditional land line phones in recent years, with Bohlin, *et al.* (2010) estimating that as of the fourth quarter of 2009 there were roughly 4.6 billion mobile phone subscribers worldwide, accounting for nearly 67% of the world's population. According to the International Telecommunication Union (ITU), such growth has been particularly prevalent in developing countries, which have seen a nearly 180% rise in number of mobile phone subscriptions from 2006 to 2011 (developed countries have seen a roughly 30% growth over the same time period) (ITU 2012). Over the same time period, fixed land line subscriptions have dropped by roughly 13% for developed countries and 4% for developing countries (ITU 2012). Penetration rates of smartphones (defined as, "a mobile telephone with computer features that may enable it to interact with computerized systems, send e-mails, and access the web." (Dictionary.com 2012)) are also growing, with Gartner estimating that worldwide sales increased by 44.7% from the first quarter of 2011 to the first quarter of 2012 (Gartner 2012). The increased use of these devices will likely have significant impacts on the availability of data that may be used for the development of sustainability indicators. In particular, sensors on the phone (which may include a gyroscope, accelerometer, digital compass, proximity/IR sensor, light sensor, camera, and microphone) provide the potential to gather data on both the activities and habits of the user, as well as on the environment in which the user is located or travelling. Such possibilities, in conjunction with the widespread use of mobile phones, will greatly enhance the development of sustainability indicators for modelling purposes.

The sensors included in smartphones, and to a lesser extent standard cellular phones, allow a great deal of contextual data to be gathered at the microscopic level, as opposed to the more current standard of the meso- or macroscopic scale. In passive use, i.e., through no active involvement of the user beyond carrying the phone while it is turned on (also known as opportunistic sensing), the user's location over time may be gathered via location-detecting sensors including GPS, GSM, Wi-Fi and accelerometers. Once having this data, it is then possible to determine, with a fair degree of accuracy, the following:

- Travel mode and path
- Location of stops and activities
- Locations of "favourite places"

Extrapolating from this data, and incorporating land-use and roadway and transit networks, it is possible to ascertain more detailed information about the user, including travel preferences, residential and work locations, activity habits (such as social, religious or recreational activities), and patterns of behaviour. With basic input from the user, it is also possible to collect data on such activities as social interactions (such as the number of persons with whom she is travelling and events to which he is going) and contextual decisions (such as altering a planned route due to a traffic disturbance). Finally, in some cases it is possible for the user to participate to an even greater degree. In "participatory sensing" the user may make use of smartphone applications (or "apps") to directly provide data of interest to involved parties. Burke, *et al.* (Burke et al. 2006) state, "Participatory sensing will task deployed mobile devices to form interactive, participatory sensor networks that enable public and professional users to gather, analyse and share local knowledge." Examples of this type of application include the following:

- Biketastic: An application which asks bike commuters to log their bike route via their phone's GPS and provide geo-tagged annotations (such as camera images and text-based notes) about the route conditions and experience. These data are then linked to automatic sensor data collected by the phone to infer the path's roughness and the surrounding traffic density. (http://biketastic.com/)
- A variety of governments and organizations have developed applications or services by which citizens may directly report the presence of such hazards as potholes, graffiti and downed trees by taking a picture with their camera-phones and sending it directly to the agency or organization. Pictures taken in such a manner are geotagged, thus making it possible for the agency to directly respond to the hazard location.
- Environment oriented apps such as California's "Creek Watch," which asks users to take and send pictures of water body and provide information on its water level, flow rate, and amount of debris present to state officials (Woody 2010); or "Noise Meter," which uses the built-in microphone on a device to allow the user to collect information on the ambient noise levels in the user's immediate environment. Collected data are then uploaded, mapped, and made available to anyone with interest.
- Information services such as Farmer's Friend, a Uganda-based text message and phone service which provides information and advice on matters pertaining to farming and agriculture (Economist 2009), have the ability to track query topics to determine current local climate conditions based on search terms and categories of interest.

Apps and monitors such as these provide a layer of data that is not otherwise available for modelling purposes due to financial or resource limitations. In addition, data gathered via their use may be further joined with other existing data sets to add value to both. Here, however, the issues of privacy and data management again raise concern, as such detailed data sets may impinge on the privacy of users. Further discussion of this issue is included below.

POTENTIAL FOR THE FUTURE

Assessment

The potential for using new technologies as sources of data collection for sustainable transportation indicators is vast, though not trivial. In particular, while Big Data seems to carry unlimited possibilities, processing millions rows of raw data is not without issues, as touched upon briefly in the last section. In spite of this, and recalling the uses for indicators listed above, new technologies show particular benefits for system performance, general comparison and model calibration.

To date, a number of research endeavors have already used mobile phone data to study mobility patterns (Gonzalez, *et al.* 2008; Di Lorenzo and Calabrese 2011; Simini, *et al.* 2012; Steenbruggen, *et al.* 2011); the gargantuan volume of data, however, makes it difficult to scrupulously identify the patterns involved, especially since the data sets used lack socio-economic data. Nevertheless, these data sets will become more common and include crucial additional data, which will provide not only socio-economic data, but also details of communication between individuals.

Once again, the first and obvious opportunity relates to data accuracy. Although higher accuracy does not necessarily imply better indicators, benefits are self-evident. To take a simple example, the traditional vehicle kilometer traveled (VKT) measurement is usually calculated by multiplying average number of trips per day with average trip length. The average number of trips per day is usually known from travel surveys that sometimes omit trips or contain erroneous information. Average trip length is calculated by estimating trip length using transport assignment software and not the actual route taken. Although it may be satisfactory for a general city-to-city comparison, it is not sufficient to make a more detailed zone-by-zone comparison. Getting better data will therefore immediately improve the current practice and enable much more disaggregate analyses.

Moreover, more data will enable a more accurate study of each individual indicator listed in table 1 from a statistical standpoint, i.e. beyond simply looking at averages. For instance, what statistical distribution (e.g. normal, power-law) do these indicators follow within or across cities, and how do their parameters differ? By then comparing these parameters (e.g., variance, scaling factor), targeted policies can be implemented to not only affect the *average* of a quantity, but its entire distribution, which is particularly relevant for the VKT example described in the previous paragraph. Statistically speaking, we may also see improvement by simply avoiding the pitfalls of Simpson's paradox.

Another potential that might be overlooked at the moment is that of route choice. Current models estimate route choice through shortest path algorithms and equilibrium constraints, without paying attention to urban form. Getting more precise demand figures for each road in a city will not only help to better calibrate transport assignment software (perhaps through machine-learning) but give us a better sense of the relationship between travel behavior and urban form. Because current technologies tend to be used consistently, it will also be possible to collect data across time (as opposed to a periodic survey), which will allow for impacts of implemented projects to be more accurately evaluated. In the case of route choice, for example, impacts of disruptions such as road work or inclement weather, or changes to the public transport system can be evaluated based on ongoing panel data from users, allowing for better indicators to be developed regarding behavior change.

Moreover, one of the biggest untapped potentials for the future is the use of indicators for real-time operations. In the past two decades, *Intelligent Transport Systems* (ITS) have been used successfully to inform the traveler (e.g., information board on a highway, time before next bus at a stop, etc.). In the future, however, sensors will be able to estimate demand in real time, which will therefore enable authorities to take actions immediately, for instance by modifying traffic signal cycles or dispatching more buses. Suitable indicators will evidently need to be developed and compiled for this purpose.

Arguably, one of the biggest benefits is simply the ability to combine the data from new technologies. For example, data on a person's daily travel patterns may be joined with environmental data on air quality and noise to estimate exposure levels to environmental hazards daily and over time. This data, in turn, may be used as inputs into scenario modelling, thus allowing the transport community to better evaluate present actions that may be taken to address a spectrum of potential future concerns. Particularly as these technologies have evolved so quickly, the development of comparative indicators will be enhanced by the incorporation of a wide variety of location data and other sensor input from these affordable, highly dispersed technologies.

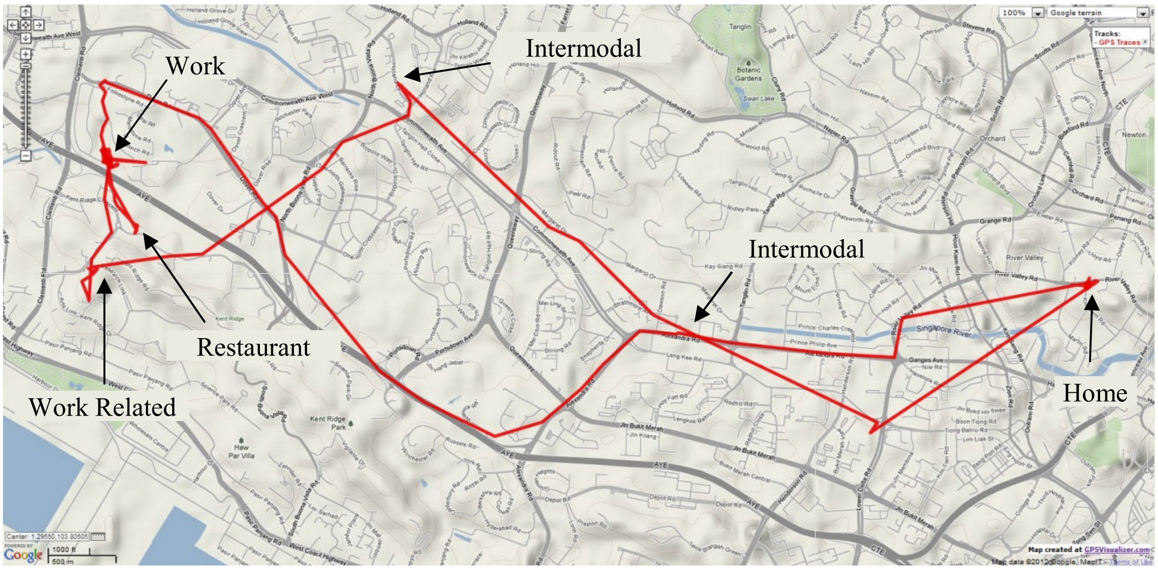
Example

As a means to illustrate some of the opportunities described above, we chose to present a specific example using data that was gathered from one of the authors. Here, we have focused in large part on the use of smartphones for use in data collection for the development of sustainability indicators. Such a focus is predicated, largely, on three factors:

- Current and emerging smartphone sensor availability
    - Currently available sensors include GPS, GSM, Wi-Fi, accelerometers, compasses, ambient light sensors, microphones, cameras and gyroscopes, among others.
    - Sensors and services in development include indoor navigation, air quality sensors, particulate matter sensing, thermometers, gas sensors and others.
- Cell phone ubiquity and growth of smartphone penetration rates (as discussed in the new technologies section).
- Growing availability of apps and services that can relate, link and transmit data between agents and actors.

These considerations make smartphones a compelling example of how technology may be used to leverage emerging data collections for the development of more targeted, accurate and effective sustainability indicators. In the hypothetical example below, we describe the potential indicators that could be developed given current and emerging sensor technologies.

Figure 1 shows the activity path of one author for one day. More precisely, the activity path consisted of travel to work, work related business and a meal, and two modes (bus and foot) were used during the course of the day. The path shown on the map, while not completely accurate (in some areas GPS connection was lost, resulting in a "straight line" trace), provides a good representation of the author's travel within the transport network, including information on specific times of activity stops and the mode used on individual legs of the travel. Such data, if collected widely and aggregated, may be used to develop more accurate indicators of, for example, VKT, distance travelled by mode, path activity and network linkages, etc. While not *new*, improvements to the accuracy of such data will likely be a boon to the efficiency and usefulness of those indicators that use them for development. In addition, repeated collection of such data will make it possible to accurately infer the likely mode and/or activity in which the subject has participated. In this case, the likelihood that the subject's home-to-work and work-to-home trips are repeated with a good amount of regularity (in particular with respect to travel mode and times), means that it will be fairly simple to determine these patterns over the history of collected data.



Activity	Start	End	Arrival mode	Comment
Home		10:11		leaving home in morning
Intermodal	10:12	10:16	walk	walk to bus stop
Intermodal	10:46	10:48	bus	take bus
Intermodal	11:02	11:03	bus	change bus
Work related	11:04	11:13	walk	walk to destination
Work	11:25	18:54	bus	bus to work
Restaurant	18:56	20:09	bus	bus to restaurant
Work	20:14	22:22	walk	walk to work
Intermodal	22:26	22:31	walk	walk to bus stop
Intermodal	22:56	23:10	bus	take bus
Intermodal	23:34	23:34	bus	change bus
Home	23:39		walk	walk home

Figure 1 – Activity-based travel survey with GPS traces using GPSVisualizer (Schneider 2012)

Regarding new indicators, if we posit that the traveler makes use of the smartphone in a "typical" manner (i.e., sending and receiving calls and text messages, surfing the web, using social networking apps, etc.), then additional data may also be gathered and transmitted. Say, for example, that on the way to the bus stop from work the traveler receives a call from a cell in the city center, and then places a call to a nearby cell. After hanging up, he opens an app on his phone that allows him to search for nearby restaurants, followed by sending a text to the same cell to which he made the phone call. After receiving and making these calls and texts, he walks to a bus stop that is more distant than the nearest stop where he can catch his intended bus, and waits while two buses pass before boarding the third. From here, he takes the bus to a local restaurant, where he spends just over one hour. Given this series of events, several assumptions can be made (such as a cancellation of plans from the first caller, resulting in plans made with the person called, with the longer walk and waiting time reflecting travel coordination with the second companion); however, we are most interested in the data that can be collected during this series of events. First is, as noted, the traveler's location data and the time of travel; however, the following information could also be gleaned given emerging and extant sensors (and appropriate agreements with the data collection agencies):

- Influence of social network on travel behavior and plan changes (may be indicated by timing of heading changes, waiting times, or as determined by activity pattern in conjunction with call/SMS history)
- Ambient light in areas where smartphones are being used
- Communications network strength and connectivity
- Counts of particulate matter and air pollution
- Patterns of internet search histories, including search terms and "hits" of interest
- Periods and places of "active" travel – or non-motorized transport such as walking, biking, skateboarding, etc.

Such data sets can be used to develop a number of indicators, including the following:

- Light availability: Measured by amount of ambient light in a particular area during times which would naturally be "dark". If joined with GPS or other mapping data showing light features and/or wildlife migratory patterns, this may be used to determine areas for investment in artificial light reduction measures where excess light is present (for example, in areas where streetlights are badly directed), or to provide information on areas where a transit passenger may feel safer waiting for a bus (generally where more light is available).
- Air pollution exposure levels of active transport users: Based on emerging work into the development of air quality sensors for smartphones (or linked to smartphones). Specific areas of air quality concern (or "hot spots") could be measured in conjunction with the number of users who would be impacted by them, the duration during which they would pass through them, and likely times of exposure. Such information would give better measures of environmental safety based on more thorough accounting.
- Trending topics of interest: By mining the contents of such social network shares as "Tweets" and status updates sent from mobile phones, in addition to reviewing the history of internet searches, it is possible to track and quantify the spread of topics of interest (such as illnesses, events, places, or climate conditions depending upon the spatial resolution). Such information can provide jurisdictions with measurable information regarding both transportation needs (for example, if a particular area is currently gaining more popularity and will need additional transit services) and, potentially, health or human services needs (for example, the example of Google searches being used to track the spread of the flu (http://www.google.org/flutrends/)).

These examples provide some suggestions of the types of indicators that may be developed over time given new and emerging technologies. This is only a small example of three fairly pragmatic indicators that may be used to better determine and measure sustainability and efforts towards sustainability within an urban area; however, they reveal the types of inputs that planners may have access to with minimal resource usage.

PRIVACY CONCERNS

Although the technologies discussed here have great potential to be of value in the search for practical, accurate and valuable sustainability indicators, various concerns with their application will also be encountered. In particular, and perhaps most notably, is the overarching concern for privacy that is often raised when examining the use of technologies tied to a specific person or vehicle (such as a cellular or smart phones, or in-vehicle sensors). Due to the rapidity of technological innovation and dissemination, it has been difficult for privacy policies and practices to keep pace with the evolving world of data collection (as described in (Levine 2011)) As such, privacy protections, including in the United States, tend to be scattershot, often industry-specific and disaggregated. In addition, consumers often have little awareness about the potential uses of their collected data, which may engender additional privacy concerns. For public agencies involved in the collection and use of data for the development of sustainability indicators, there are often overarching guidelines (described in (Cottrill 2011)) that may impact the ability of these agencies to collect, mine or present data on an individual level, or without proper disclosure and consent. However, there are few comprehensive guidelines relevant to private firms related to these matters, which may have a larger impact on data sets that are purchased by public agencies or that are used for the development of sustainability indicators in the private sector. According to (Ozer et al. 2010), "When many different companies hold copies of valuable information about consumers, the privacy protection afforded to consumers is only as strong as the weakest link. Unfortunately, in a climate lacking both clear legal protections and strong privacy safeguards provided by LBS [location-based services] themselves, consumers are starting to recognize and experience privacy failures." Many of the data sources outlined above, particularly those associated with mobile phones, have the capacity to provide detailed data about the movements of data subjects over time, which may, in turn, reveal far more about individuals than they would perhaps feel comfortable with.

The recognition and discussion of these concerns have been growing, particularly with recent privacy infringements by such firms as Google and Facebook, and the growing availability of LBS applications on smartphones and tablets. Suggestions regarding how to address these concerns have ranged from those predominantly focused on the use of overarching or broadly-applicable privacy policies, such as the European Data Protection Directive, which regulates the processing of personal data within the European Union; to more voluntary practices that leave the responsibility for his privacy protection largely to the consumer himself. The latter approach, which often takes the form of opt-out clauses, is viewed by most privacy advocates as counter to the interests of the consumer, but is looked upon favorably by many industry groups. Additional methods that have been suggested include the masking or aggregation of data before release by the collecting agency to the public or third parties, context-sensitive controls that would give the consumer the ability to determine personally her level of data disclosure dependent upon her comfort with the specific situation, or use of technological methods of privacy protection (such as secure servers and regular deletion of data). While each of these methods presents a degree of privacy protection for the types of data discussed above, it is clear that privacy protection in the realm of emerging technologies, in particular those related to location and transportation, is an open question, and one that deserves closer attention.

DATA QUALITY AND MANAGEMENT

Related to the issue of privacy discussed above is that of data quality and management. Massive data sets such as those described above, particularly if collected in real-time, require careful oversight and treatment if they are to be used in ways that are both useful and minimally invasive. This issue is particularly critical as the rapid evolution of technology and resulting data have shifted the paradigms under which transport and sustainability professionals operate, requiring more knowledge and understanding of such matters as data fusion (Faouzi *et al.,* 2011), interoperability, and data protection.

For purposes of quality determination of individual data sets, both development of appropriate metadata (such as method of creation and collection and collecting agency or organization), as well as systematic quality checks (addressing such issues as timeliness and detected accuracy) are necessary to ensure appropriate data maintenance. Particularly for data sets where the quality of data may be mixed (for example, location detection for smartphones may come from GPS, GSM, Wi-Fi or accelerometer data, with varying degrees of accuracy), it will be necessary to ensure that collected data are appropriately "tagged" to allow for accurate algorithms to be developed. Moreover, when mining linked datasets, it is necessary to ensure that data are appropriately matched and linked to ensure that accurate results can be reported. Such issues as disparate identifiers, differences in the temporal structure of collected data, or alternate methods of location determination may require that additional data manipulation or modification take place. This, in turn, requires that metadata detailing this manipulation be created in order that such linked data sets may be used appropriately. Ensuring that such issues are addressed in the initial stages of data collection will better allow for analysis and manipulation at scale.

Such issues, while they have always been of concern in data management, must be highlighted as they will come more to the forefront of indicator development as more diverse data collection agents are able to share and use disparate data sets.

CONCLUSION

In the 21st century, making transportation more sustainable has become a priority in the planning community. Measuring sustainability, however, presents substantial challenges. Not only do the indicators that we use need to adequately reflect principles of sustainability, the data used to compile these indicators needs to be accurate and complete, and at the moment we seem to fall short on both accounts. Nonetheless, the 21st century is also the century of Big Data being collected from new and ubiquitous technologies. The main goal of this paper was to review the current practice with regards to sustainable mobility indicators and highlight the potentials from new technologies (from infrastructure to smartphones).

First, we discussed the need of indicators and highlighted several areas where indicators are commonly used. Second, we presented a list of currently indicators and identified problems with present data collection technics. Third, we introduced and described several new technologies that are relevant for data collection. Finally, we discussed how and where these new technologies could be applied. Clearly, the first gain resides in higher accuracy, which is greatly needed at the moment.

We subsequently illustrated some of the benefits by going a through a thorough scenario using the GPS traces of one of the authors.

Overall, new technologies present substantial benefits to better measure sustainability in the realm of transportation, and while managing a vast amount of data will not be trivial, the transportation community has much to gain. One specific potential is to combine various data sources such as smartphones, social networks and infrastructure sensors for model calibration and real time operations. New technologies seem to offer endless opportunities both for developed and developing countries in our overarching goal to develop better indicators and hence plan more sustainable transport systems.

REFERENCES

Bohlin, Anders, Harald Gruber, and Pantelis Koutroumpis. 2010. “Diffusion of New Technology Generations in Mobile Communications.” *Information Economics and Policy* 22 (1) (March): 51–60. doi:10.1016/j.infoecopol.2009.11.001.

Bongardt, Daniel, Dominik Schmid, Cornie Huizenga, and Todd Litman. 2011. “Sustainable Transport Evaluation: Developing Practical Tools for Evaluation in the Context of the CSD Process.” In Eschborn, Germany: United Nations Department Of Economic And Social Affairs.

Borken, Jens. 2003. “Indicators for Sustainable Mobility: a Policy Oriented Approach.” In *Transport & Environment - 1st Symposium*. Actes INRETS. Arcueil, FRANCE: Institut national de recherche sur les transports et leur sécurité.

Burke, Jeffrey A, D Estrin, Mark Hansen, Andrew Parker, Nithya Ramanathan, Sasank Reddy, and Mani B Srivastava. 2006. “Participatory Sensing.” http://www.escholarship.org/uc/item/19h777qd.

CIA-Asia, and EMBARQ. 2007. “Sustainable Urban Transport in Asia: Making the Vision a Reality”. Partnership for Sustainable Urban Transport in Asia (PSUTA): Asian Development Bank (ADB), Clean Air Initiative for Asian Cities (CIA-Asia), Swedish International Development Cooperation Agency (Sida), World Resources Institute (EMBARQ-WRI). http://www.wri.org/publication/sustainable-urban-transport-asia.

Cobb, Clifford, and Craig Rixford. 1998. *Lessons Learned from the History of Social Indicators*. San Francisco, CA: Redefining Progress.

Cottrill, Caitlin. 2011. “Location Privacy: Who Protects?” *URISA Journal* 23 (2): 49–59.

Derrible, S., S. Saneinejad, L. Sugar, and C. Kennedy. 2010. “Macroscopic Greenhouse-gases Emissions Model of Urban Transportation for Municipalities.” *Transportation Research Record* 2191: 174–181.

Dictionary.com. 2012. “Smartphone.” *Collins English Dictionary - Complete & Unabridged 10th Edition.*

Economist. 2009. “Beyond Voice: New Uses for Mobile Phones Could Launch Another Wave of Development.” *Special Report: Telecoms in Emerging Markets*, September 24. http://www.economist.com/node/14483848.

European Commission. 2011. “Roadmap to a Single European Transport Area – Towards a Competitive and Resource Efficient Transport System”. White Paper COM(2011) 144 final. Brussels, Belgium: European Commission. http://ec.europa.eu/transport/strategies/2011_white_paper_en.htm.

Fabish, Lisa, and Peter Haas. 2011. “Measuring the Performance of Livability Programs.” *Transportation Research Record: Journal of the Transportation Research Board* 2242 (-1) (December 1): 45–54. doi:10.3141/2242-06.

Faouzi, Nour-Eddin El, Henry Leung, and Ajeesh Kurian. 2011. "Data Fusion in Intelligent Transportation Systems: Progress and Challenges – A Survey." *Information Fusion* 12 (1) (January): 4–10. doi:10.1016/j.inffus.2010.06.001.

Gartner. 2012. "Gartner Says Worldwide Sales of Mobile Phones Declined 2 Percent in First Quarter of 2012; Previous Year-over-Year Decline Occurred in Second Quarter of 2009." *Gartner Newsroom*. May 16. http://www.gartner.com/it/page.jsp?id=2017015.

Gonzalez, Marta C., Cesar A. Hidalgo, and Albert-Laszlo Barabasi. 2008. "Understanding Individual Human Mobility Patterns." *Nature* 453 (7196) (June 5): 779–782. doi:10.1038/nature06958.

Grob, G.R. 2010. "Future Transportation with Smart Grids & Sustainable Energy." *International Journal of COMADEM* 13 (4): 2–7.

Gudmundsson, Henrik. 2003. "Making Concepts Matter: Sustainable Mobility and Indicator Systems in Transport Policy*." *International Social Science Journal* 55 (176): 199–217. doi:10.1111/j.1468-2451.2003.05502003.x.

ITU. 2012. "Key Global Telecom Indicators for the World Telecommunication Service Sector." *International Telecommunications Union*. http://www.itu.int/ITU-D/ict/statistics/at_glance/keytelecom.html.

Keirstead, James, and Matt Leach. 2008. "Bridging the Gaps Between Theory and Practice: a Service Niche Approach to Urban Sustainability Indicators." *Sustainable Development* 16 (5): 329–340. doi:10.1002/sd.349.

Krauth, Elfriede, Hans Moonen, Viara Popova, and Martijn Schut. 2005. "Performance Indicators in Logistics Service Provision and Warehouse Management – a Literature Review and Framework." In *Euroma*. Budapest, Hungary. http://www.cs.vu.nl/~schut/pubs/Krauth/2005a.pdf.

Lautso, Kari, Klaus Spiekermann, and Michael Wegener. 2002. "Modelling Policies for Urban Sustainability." In *42nd Congress of the European Regional Science Association (ERSA)*. Dortmund, Germany: European Regional Science Association. http://www.spiekermann-wegener.com/pub/pdf/PROPOLIS_ERSA2002.pdf.

Levine, Danielle. 2011. "Facebook and Social Networks: The Government's Newest Playground for Information and the Laws That Haven't Quite Kept Pace." *Hastings Communications and Entertainment Law Journal (COMM-ENT)* 33 (3) (March 22): 481–498.

Litman, Todd. 2007. "Developing Indicators for Comprehensive and Sustainable Transport Planning." *Transportation Research Record: Journal of the Transportation Research Board* 2017 (1) (December 1): 10–15. doi:10.3141/2017-02.

———. 2012. "Well Measured: Developing Indicators for Sustainable and Livable Transport Planning". Victoria Transport Policy Institute.

Di Lorenzo, G., and F. Calabrese. 2011. "Identifying Human Spatio-temporal Activity Patterns from Mobile-phone Traces." In *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC*, 1069–1074. http://www.scopus.com/inward/record.url?eid=2-s2.0-83755228639&partnerID=40&md5=f2d6fb1d0e49779946e29b6ba64521aa.

Lynch, Kevin. 1960. *The Image of the City*. Cambridge, MA: MIT Press. http://www.amazon.ca/exec/obidos/redirect?tag=citeulike09-20&path=ASIN/0262620014.

Millett, Lynette I., Deborah L. Estrin, Committee on Computing Research for Environmental and Societal Sustainability, Computer Science and Telecommunications Board, and Division on Engineering and Physical Sciences. 2012. *Computing Research for Sustainability*. Washington D.C.: National Academies Press. http://www.nap.edu/openbook.php?record_id=13415.

De Montjoye, Yves-Alexandre, César A. Hidalgo, Michel Verleysen, and Vincent D. Blondel. 2013. "Unique in the Crowd: The Privacy Bounds of Human Mobility." *Scientific Reports* 3 (March 25). doi:10.1038/srep01376. http://www.nature.com/doifinder/10.1038/srep01376.

Munda, G. 2005. "'Measuring Sustainability': A Multi-criterion Framework." *Environment, Development and Sustainability* 7 (1): 117–134.

Newman, Peter, and Jeffrey R. Kenworthy. 1989. *Cities and Automobile Dependence : a Sourcebook*. Brookfield, Vt.: Gower Technical.

Ozer, Nicole, Chris Conley, David Hari O'Connell, Tamar R. Gubins, and Ellen Ginsburg. 2010. "Location-Based Services: Time for a Privacy Check-In." *SSRN eLibrary* (November 29). http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1732269.

Pei, Yi, Adjo Amekudzi, Michael Meyer, Elise Barrella, and Catherine Ross. 2010. "Performance Measurement Frameworks and Development of Effective Sustainable Transport Strategies and Indicators." *Transportation Research Record: Journal of the Transportation Research Board* 2163 (1) (December 1): 73–80. doi:10.3141/2163-08.

Samberg, Stuart, Alon Bassok, and Shawna Holman. 2011. "Method for Evaluation of Sustainable Transportation." *Transportation Research Record: Journal of the Transportation Research Board* 2242 (-1) (December 1): 1–8. doi:10.3141/2242-01.

Sandblad, B., A.W. Andersson, A. Kauppi, and G. Isaksson-Lutteman. 2010. "Development and Implementation of New Principles and Systems for Train Traffic Control in Sweden." In *WIT Transactions on the Built Environment*, 114:441–450. http://www.scopus.com/inward/record.url?eid=2-s2.0-78649418596&partnerID=40&md5=9474312b41a8c9858acbfaf1c5ab5faf.

Schneider, Adam. 2012. "Make a Google Map from a GPS File." *GPSVisualizer*. http://www.gpsvisualizer.com/map_input.

Sharmeen, Fariya, Theo Arentze, and Harry Timmermans. 2010. "Modelling the Dynamics Between Social Networks and Activity-travel behavior:Literature Review and Research Agenda." In *12th World Conference on Transport Research*. Lisbon, Portugal.

Silvis, Julia, Deb Niemeier, and Raissa D'Souza. 2006. "Social Networks and Travel Behavior: Report from an Integrated Travel Diary." In *11th International Conference on Travel Behaviour Research*. Kyoto, Japan.

Simini, Filippo, Marta C. Gonzalez, Amos Maritan, and Albert-Laszlo Barabasi. 2012. "A Universal Model for Mobility and Migration Patterns." *Nature* 484 (7392) (April 5): 96–100. doi:10.1038/nature10856.

Steenbruggen, J., M.T. Borzacchiello, P. Nijkamp, and H. Scholten. 2011. "Mobile Phone Data from GSM Networks for Traffic Parameter and Urban Spatial Pattern Assessment: a Review of Applications and Opportunities." http://www.scopus.com/inward/record.url?eid=2-s2.0-79955505348&partnerID=40&md5=6cc2cfd7d536ae4559e0aef45e50d63f.

UNCDR. 2010. "Bangkok Declaration for 2020 – Sustainable Transport Goals for 2010-2020." In Bangkok, Thailand: United Nations Centre for Regional Developmennt. http://www.uncrd.or.jp/env/5th-regional-est-forum/doc/bangkok_declaration.pdf.

Wolf, Jean, Randall Guensler, and William Bachman. 2001. "Elimination of the Travel Diary: Experiment to Derive Trip Purpose from Global Positioning System Travel Data." *Transportation Research Record: Journal of the Transportation Research Board* 1768 (-1) (January 1): 125–134. doi:10.3141/1768-15.

Wolf, Jean, and Michelle Lee. 2008. "Synthesis of and Statistics for Recent GPS-enhanced Travel Surveys." In *8th International Conference on Survey Methods in Transport: Harmonisation and Data Comparability*, 17. Annecy, France.

Woody, Todd. 2010. "iPhone App Lets Agencies Crowdsource Environmental Monitoring." *Grist*. November 12. http://grist.org/article/2010-11-11-iphone-app-lets-agencies-crowdsource-environmental-monitoring/.

Zegras, P. Christopher. 2005. *Sustainable Urban Mobility: Exploring the Role of the Built Environment*. Ph.D. Dissertation. Cambridge, MA: Massachusetts Institute of Technology. http://web.mit.edu/czegras/www/Final%20VersionV3.pdf.

Zmud, Johanna, and Jean Wolf. 2003. "Identifying the Correlates of Trip Misreporting - Results from the California Statewide Household Travel Survey GPS Study." In *Moving Through Nets: The Physical and Social Dimensions of Travel*, 17. Lucerne, Switzerland.