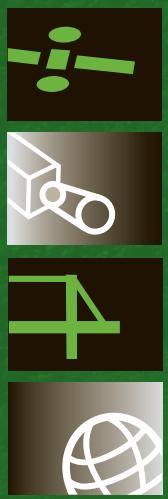


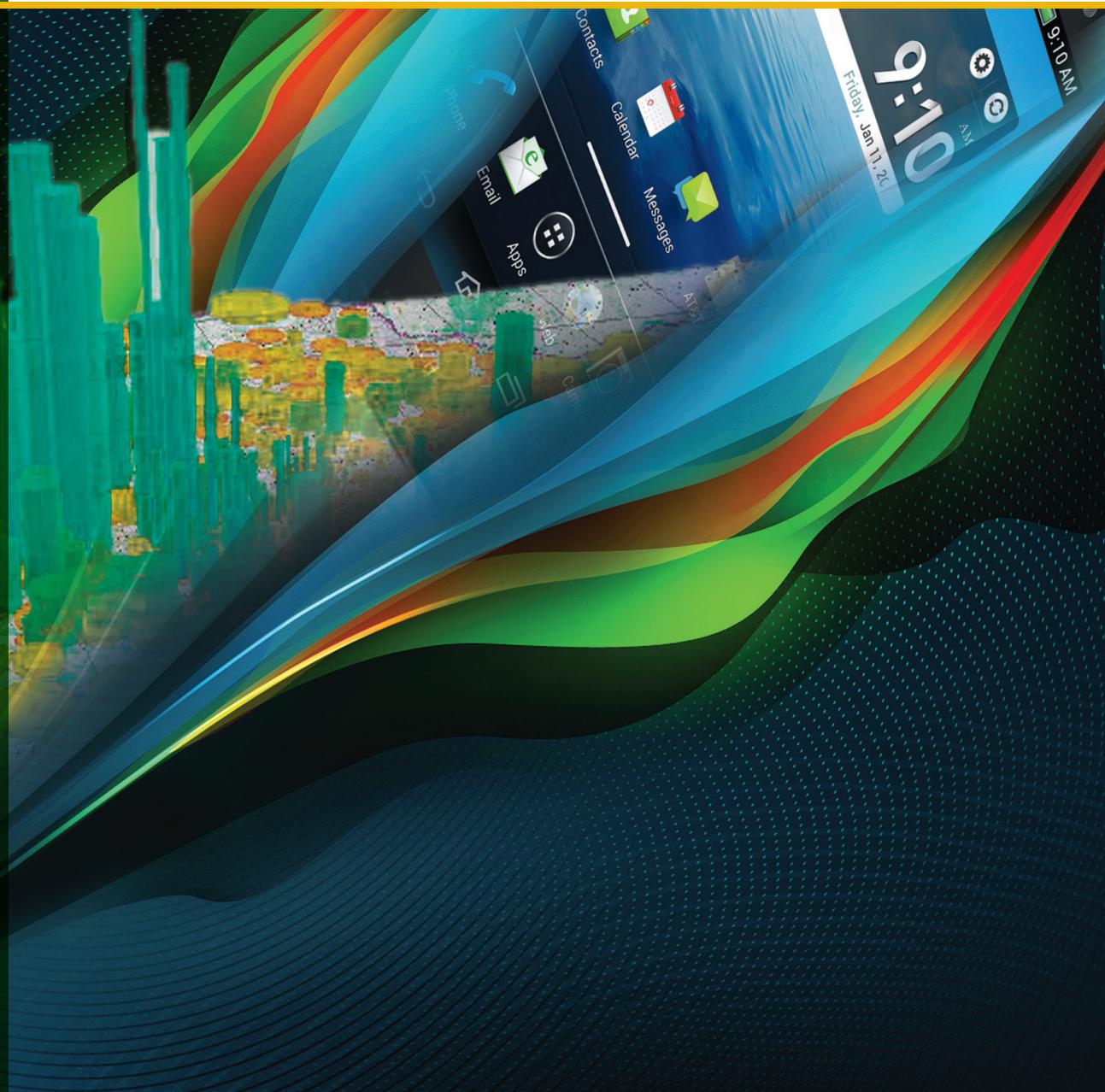
The Exploratory Advanced Research Program



EXPLORATORY ADVANCED RESEARCH

# Cell Phone Data and Travel Behavior Research

SYMPOSIUM SUMMARY REPORT • February 12, 2014



U.S. Department  
of Transportation

Federal Highway  
Administration

## **Notice**

This document is disseminated under the sponsorship of the U.S. Department of Transportation in the interest of information exchange. The U.S. Government assumes no liability for the use of the information contained in this document.

The U.S. Government does not endorse products or manufacturers. Trademarks or manufacturers' names appear in this report only because they are considered essential to the objective of the document.

## **Quality Assurance Statement**

The Federal Highway Administration (FHWA) provides high-quality information to serve Government, industry, and the public in a manner that promotes public understanding. Standards and policies are used to ensure and maximize the quality, objectivity, utility, and integrity of its information. FHWA periodically reviews quality issues and adjusts its programs and processes to ensure continuous quality improvement.

## Technical Report Documentation Page

|   |  |  |                                  |
|---|--|--|----------------------------------|
| 1. Report No.<br>FHWA-HRT-14-060  | 2. Government Accession No.                          | 3. Recipient's Catalog No.   |                                  |
| 4. Title and Subtitle<br><br>Cell Phone Data and Travel Behavior Research:<br>Symposium Summary Report  |  | 5. Report Date<br>July 2014  | 6. Performing Organization Code: |
|   |  |  |                                  |
| 7. Author(s)<br>Gregory Bucci and Tom Morton  |  | 8. Performing Organization Report No.  |                                  |
| 9. Performing Organization Name and Address<br>U.S. Department of Transportation Volpe Center<br>55 Broadway, Cambridge, MA 02142<br>Woodward Communications<br>1420 N Street, NW., Suite 102, Washington, DC 20005   |  | 10. Work Unit No.  |                                  |
|   |  | 11. Contract or Grant No.<br><b>Contract DTFH61-10-V-00031</b><br><b>Contract DTFH61-09-F-00027</b>  |                                  |
| 12. Sponsoring Agency Name and Address<br>Federal Highway Administration<br>Office of Policy Information, 1200 New Jersey Ave., SE<br>Washington, DC 20590<br>Office of Corporate Research, Technology, and Innovation<br>Management, 6300 Georgetown Pike<br>McLean, VA 22101-2296   |  | 13. Type of Report and Period Covered<br>Symposium Summary Report, February 2014   |                                  |
|   |  | 14. Sponsoring Agency Code<br>HRTM-30  |                                  |
| 15. Supplementary Notes<br>FHWA's Contracting Officer's Task Manager (COTM): Zachary Ellis, HRTM-30<br>Technical Contact: Brad Gudzinas   |  |  |                                  |
| 16. Abstract<br><br>This report summarizes the key themes from a symposium held on February 12, 2014, to discuss opportunities and challenges using cellular location data for national travel behavior analysis. Participants discussed the availability of cellular data and the common types of licensing agreements; applications of cellular data and how it can be leveraged; fusion of cellular data in terms of merging it with other data sources; and validation of cellular data to determine accurate and meaningful results. Particular focal points included applications and limitations of land-use models and data, and using surveys in conjunction with cellular location data to facilitate accuracy and precision. |  |  |                                  |
| 17. Key Words<br>Cell phone, cellular data, national travel behavior, data analysis, data validation, data sources, travel monitoring, data availability, licensing agreements, mobile phones, acoustic sensors.  |  | 18. Distribution Statement<br>No restrictions. This document is available to the public through the National Technical Information Service, Springfield, VA 22161. |                                  |
| 19. Security Classif. (of this report)<br>Unclassified  | 20. Security Classif. (of this page)<br>Unclassified | 21. No. of Pages<br>28   | 22. Price<br>N/A                 |

| SI* (MODERN METRIC) CONVERSION FACTORS                             |                             |                             |                             |                     |
|--|-----------------------------|-----------------------------|-----------------------------|---------------------|
| APPROXIMATE CONVERSIONS TO SI UNITS                                |                             |                             |                             |                     |
| Symbol   | When You Know               | Multiply By                 | To Find                     | Symbol              |
| <b>LENGTH</b>  |                             |                             |                             |                     |
| in   | inches                      | 25.4                        | millimeters                 | mm                  |
| ft   | feet                        | 0.305                       | meters                      | m                   |
| yd   | yards                       | 0.914                       | meters                      | m                   |
| mi   | miles                       | 1.61                        | kilometers                  | km                  |
| <b>AREA</b>  |                             |                             |                             |                     |
| in <sup>2</sup>  | square inches               | 645.2                       | square millimeters          | mm <sup>2</sup>     |
| ft <sup>2</sup>  | square feet                 | 0.093                       | square meters               | m <sup>2</sup>      |
| yd <sup>2</sup>  | square yard                 | 0.836                       | square meters               | m <sup>2</sup>      |
| ac   | acres                       | 0.405                       | hectares                    | ha                  |
| mi <sup>2</sup>  | square miles                | 2.59                        | square kilometers           | km <sup>2</sup>     |
| <b>VOLUME</b>  |                             |                             |                             |                     |
| fl oz  | fluid ounces                | 29.57                       | milliliters                 | mL                  |
| gal  | gallons                     | 3.785                       | liters                      | L                   |
| ft <sup>3</sup>  | cubic feet                  | 0.028                       | cubic meters                | m <sup>3</sup>      |
| yd <sup>3</sup>  | cubic yards                 | 0.765                       | cubic meters                | m <sup>3</sup>      |
| NOTE: volumes greater than 1000 L shall be shown in m <sup>3</sup> |                             |                             |                             |                     |
| <b>MASS</b>  |                             |                             |                             |                     |
| oz   | ounces                      | 28.35                       | grams                       | g                   |
| lb   | pounds                      | 0.454                       | kilograms                   | kg                  |
| T  | short tons (2000 lb)        | 0.907                       | megagrams (or "metric ton") | Mg (or "t")         |
| <b>TEMPERATURE (exact degrees)</b>                                 |                             |                             |                             |                     |
| °F   | Fahrenheit                  | 5 (F-32)/9<br>or (F-32)/1.8 | Celsius                     | °C                  |
| <b>ILLUMINATION</b>  |                             |                             |                             |                     |
| fc   | foot-candles                | 10.76                       | lux                         | lx                  |
| fl   | foot-Lamberts               | 3.426                       | candela/m <sup>2</sup>      | cd/m <sup>2</sup>   |
| <b>FORCE and PRESSURE or STRESS</b>                                |                             |                             |                             |                     |
| lbf  | poundforce                  | 4.45                        | newtons                     | N                   |
| lbf/in <sup>2</sup>  | poundforce per square inch  | 6.89                        | kilopascals                 | kPa                 |
| APPROXIMATE CONVERSIONS FROM SI UNITS                              |                             |                             |                             |                     |
| Symbol   | When You Know               | Multiply By                 | To Find                     | Symbol              |
| <b>LENGTH</b>  |                             |                             |                             |                     |
| mm   | millimeters                 | 0.039                       | inches                      | in                  |
| m  | meters                      | 3.28                        | feet                        | ft                  |
| m  | meters                      | 1.09                        | yards                       | yd                  |
| km   | kilometers                  | 0.621                       | miles                       | mi                  |
| <b>AREA</b>  |                             |                             |                             |                     |
| mm <sup>2</sup>  | square millimeters          | 0.0016                      | square inches               | in <sup>2</sup>     |
| m <sup>2</sup>   | square meters               | 10.764                      | square feet                 | ft <sup>2</sup>     |
| m <sup>2</sup>   | square meters               | 1.195                       | square yards                | yd <sup>2</sup>     |
| ha   | hectares                    | 2.47                        | acres                       | ac                  |
| km <sup>2</sup>  | square kilometers           | 0.386                       | square miles                | mi <sup>2</sup>     |
| <b>VOLUME</b>  |                             |                             |                             |                     |
| ml   | milliliters                 | 0.034                       | fluid ounces                | fl oz               |
| L  | liters                      | 0.264                       | gallons                     | gal                 |
| m <sup>3</sup>   | cubic meters                | 35.314                      | cubic feet                  | ft <sup>3</sup>     |
| m <sup>3</sup>   | cubic meters                | 1.307                       | cubic yards                 | yd <sup>3</sup>     |
| <b>MASS</b>  |                             |                             |                             |                     |
| g  | grams                       | 0.035                       | ounces                      | oz                  |
| kg   | kilograms                   | 2.202                       | pounds                      | lb                  |
| Mg (or "t")  | megagrams (or "metric ton") | 1.103                       | short tons (2000 lb)        | T                   |
| <b>TEMPERATURE (exact degrees)</b>                                 |                             |                             |                             |                     |
| °C   | Celsius                     | 1.8C+32                     | Fahrenheit                  | °F                  |
| <b>ILLUMINATION</b>  |                             |                             |                             |                     |
| lx   | lux                         | 0.0929                      | foot-candles                | fc                  |
| cd/m <sup>2</sup>  | candela/m <sup>2</sup>      | 0.2919                      | foot-Lamberts               | fl                  |
| <b>FORCE and PRESSURE or STRESS</b>                                |                             |                             |                             |                     |
| N  | newtons                     | 0.225                       | poundforce                  | lbf                 |
| kPa  | kilopascals                 | 0.145                       | poundforce per square inch  | lbf/in <sup>2</sup> |

\*SI is the symbol for the International System of Units. Appropriate rounding should be made to comply with Section 4 of ASTM E380.  
(Revised March 2003)

## Table of Contents

|  |          |
|--|----------|
| <b>Introduction .....</b>  | <b>1</b> |
| <b>Opening Remarks .....</b>   | <b>3</b> |
| <b>Data Coverage, Resolution, and Characteristics .....</b>                      | <b>4</b> |
| <b>Presentations .....</b>   | <b>4</b> |
| Bill King, <i>AirSage</i> .....  | 4        |
| Daniel Rolf and Keith Hangland, <i>HERE</i> .....                                | 4        |
| Rick Schuman, <i>INRIX</i> .....   | 4        |
| Nick Cohn, <i>TomTom</i> .....   | 5        |
| <b>Major Themes Discussed .....</b>  | <b>5</b> |
| <b>Key Takeaways .....</b>   | <b>5</b> |
| <b>Agency and Academic Experience Using Cell Data .....</b>                      | <b>6</b> |
| <b>Presentations .....</b>   | <b>6</b> |
| Cynthia Chen, <i>University of Washington</i> .....                              | 6        |
| Xiaowei Xu, <i>University of Arkansas</i> .....                                  | 6        |
| Elizabeth Birriel, <i>Florida Department of Transportation</i> .....             | 6        |
| Camille Kamga and Anil Yazici, <i>City University of New York</i> .....          | 7        |
| Andrew Rohne, <i>Ohio-Kentucky-Indiana Regional Council of Governments</i> ..... | 7        |
| Krishnan Viswanathan, <i>CDM Smith</i> .....                                     | 7        |
| <b>Major Themes Discussed .....</b>  | <b>8</b> |
| <b>Key Takeaways .....</b>   | <b>8</b> |

|  |           |
|--|-----------|
| <b>Linking Trip Information to Behavior Data .....</b>   | <b>9</b>  |
| <b>Presentations .....</b>   | <b>9</b>  |
| <i>Lei Zhang, University of Maryland .....</i>   | <i>9</i>  |
| <i>Billy Bachman, Westat .....</i>   | <i>9</i>  |
| <i>Shan Jiang, Massachusetts Institute of Technology.....</i>  | <i>9</i>  |
| <b>Major Themes Discussed .....</b>  | <b>10</b> |
| <b>Key Takeaways .....</b>   | <b>10</b> |
| <br>   |           |
| <b>Relating Census Block and American Community Survey ..... 11<br/>County Data, Demographic, and Economic Data to<br/>Derived Trip Data</b> |           |
| <b>Presentation .....</b>  | <b>11</b> |
| <i>Ben Pierce, Battelle .....</i>  | <i>11</i> |
| <b>Major Themes Discussed .....</b>  | <b>11</b> |
| <b>Key Takeaways .....</b>   | <b>11</b> |
| <br>   |           |
| <b>Panel Summary and Next Steps .....</b>  | <b>12</b> |
| <b>Major Themes Discussed .....</b>  | <b>12</b> |
| <b>Key Takeaways .....</b>   | <b>12</b> |
| <br>   |           |
| <b>Appendix A: Agenda .....</b>  | <b>15</b> |
| <b>Appendix B: Attendees .....</b>   | <b>17</b> |

## List of Figures

|   |    |
|---|----|
| Figure 1. AirSage pinpoints a user's location. ....   | 4  |
| Figure 2. Cellular analytics illustrate where a region's visitors .....<br>originated and highlights areas of demand. | 5  |
| Figure 3. Clustering algorithm developed to represent individuals .....<br>in a complex system.                       | 6  |
| Figure 4. Taxi data used to determine demand for taxis during .....<br>inclement weather.                             | 7  |
| Figure 5. AirSage cellphone data sample. ....   | 8  |
| Figure 6. Road usage based on cell phone data linked with census .....<br>tracts to illustrate congestion exposure.   | 9  |
| Figure 7. Example of a large national survey with sample sizes in the .....<br>tens of thousands.                     | 11 |

## List of Acronyms and Abbreviations

### **General Terms**

|      |                                      |
|------|--------------------------------------|
| EAR  | Exploratory Advanced Research        |
| FHWA | Federal Highway Administration       |
| FDOT | Florida Department of Transportation |
| GPS  | Global Positioning System            |
| OD   | origin-destination                   |
| NHTS | National Household Travel Survey     |
| TRB  | Transportation Research Board        |





# Introduction

**O**n February 12, 2014, the Federal Highway Administration's (FHWA's) Office of Highway Policy Information, with the Exploratory Advanced Research (EAR) Program, sponsored a Cell Phone Data and Travel Behavior Research Symposium. The symposium brought together data providers in private industry, researchers in academic fields, and other professionals from both public agencies and private businesses and institutes to explore opportunities and challenges using cellular location data for national travel behavior analysis. The authors of this document summarize the key themes discussed during the symposium held at the Transportation Research Board's (TRB's) Keck Center in Washington, DC.

Discussions centered on the availability of cellular data and the common types of

licensing agreements, applications of cellular data and how it can be leveraged, fusion of cellular data in terms of merging it with other data sources, and validation of cellular data to determine accurate and meaningful results. Particular focal points included applications and limitations of land-use models and data and using surveys in conjunction with cellular location data to facilitate accuracy and precision.

FHWA sponsored this symposium in cooperation with the TRB Travel Survey Methods Committee, Special Task Force on Data for Decisions and Performance Measures, and Task Force on Understanding New Directions for the National Household Travel Survey (NHTS).



## Opening Remarks

The symposium began with representatives of the sponsoring offices providing a brief introduction to their interests in cellular data. David Winter, Director of FHWA's Office of Highway Policy Information, began by stating that cellular data is a very important emerging field. He noted that cellular data can help provide reliable data for national decisionmaking and that the critical next steps include how best to integrate and utilize cellular data with current methods. Winter explained that the Office of Highway Policy Information has worked with TRB and others on several projects, such as the NHTS task force. According to Winter, this type of collaboration will be useful moving forward as researchers explore emerging areas, such as cellular data.

David Kuehn, EAR Program Manager, explained that the EAR Program seeks out advances in science to test possible highway applications. He noted that the program works to reach out to the

community, not only to determine what areas are on the cutting edge, but also to improve research outcomes through peer and panel reviews. Kuehn also highlighted that methods for applying cellular data to highway transportation can be viewed as high-risk, high-reward propositions. Cellular data aligns with many of the EAR Program focus areas and can contribute to the long-term goal of better system investment and better system operations.

Tianjia Tang, Travel Monitoring and Surveys Team Leader in the Office of Highway Policy Information, explained that research on applying cell data will assist with determining why, when, and how people travel in a timely and comprehensive manner, which is scientifically and statistically significant. Tang stated that a primary goal would be to determine individual trips, roadway usage, and the reason for the trip on a nationwide scale.

# Data Coverage, Resolution, and Characteristics

In this session, industry representatives provided a brief overview of their companies and the types of data they collect and are able to provide. Symposium participants then discussed challenges relating to privacy and modes.

## Presentations

### Bill King, AirSage

Bill King of AirSage explained that the company is a data provider with contracts in place with two major cell phone carriers. King highlighted how AirSage is able to pinpoint a user's location by triangulating between multiple towers, as shown in figure 1. He noted users are anonymous, but AirSage is able to link to demographics at the census tract level. AirSage is also able to make inferences about users' home and work locations

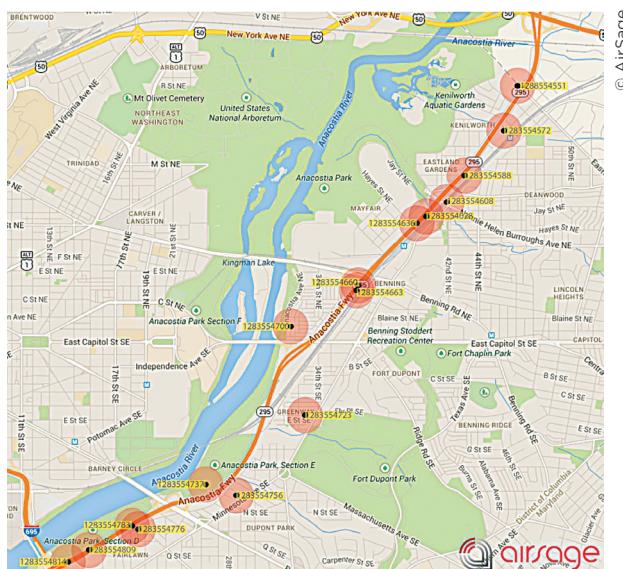


Figure 1. AirSage pinpoints a user's location.

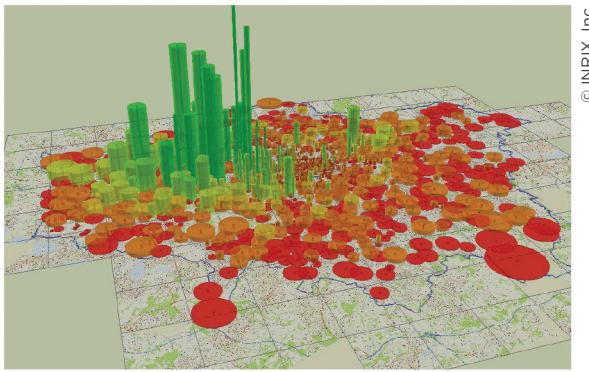
within a dataset of extrapolated person trips. King stated that the data can be licensed to end users for a specific purpose.

### Daniel Rolf and Keith Hangland, HERE

Daniel Rolf and Keith Hangland of HERE informed symposium participants that, although originally a mapping company, HERE now uses probe data collected from different devices by using applications or cell phones. Rolf and Hangland noted that HERE is not only collecting and selling probe data but is also using it. They explained that HERE data power freight and trucking industries and are also beginning to be used for government planning activities. Rolf and Hangland told participants that data can also be used to analyze traffic movement and efficiency to solve engineering and planning issues.

### Rick Schuman, INRIX

Rick Schuman of INRIX stated that the primary business for INRIX is real-time congestion or traffic movement data based on 25–30 percent of travelers, as shown in figure 2. Schuman noted that this can be obtained in real time or through a record or catalog. He explained that INRIX provides processed data, which is more usable for most buyers than raw data. Schuman told symposium participants that there is also an important distinction between metro-level data (generally associated with a project) and national-level data (generally associated with a product).



© INRIX, Inc.

Figure 2. Cellular analytics illustrate where a region's visitors originated and highlights areas of demand.

### Nick Cohn, TomTom

Nick Cohn of TomTom informed symposium participants that one feature of TomTom's data is origin-destination (OD). Cohn noted that, like other providers, available data are a sample of the total population. He explained that sample sizes and local data availability vary greatly, making it difficult to standardize a national product. Cohn highlighted that TomTom gathers Global Positioning System (GPS) data from vehicle movements used for both speed and OD. He stated that TomTom uses OD data, along with mapping and shape files, and makes data available through licenses. TomTom also creates customized datasets for specific customers and their needs.

### Major Themes Discussed

The symposium participants discussed several major themes during this session, as follows:

- **Connecting data and behavior.** Although land-use data are helpful and available at a city level, in terms of making connections between the location data and user behavior, there is currently no such nationwide map.

- **Identifying needs.** The financial and advertising industries already use OD analytics but have different needs in terms of precision and bias. For the transportation industry, it should be established what margin of error is acceptable in terms of deriving behavior.

- **Confirming data usage.** Licensing differs for educational facilities, researchers, government, and commercial entities who want to use available data.

- **Making data available.** There are privacy issues in terms of which aspects of data are available.

- **Determining mode.** The ability to determine the mode (e.g., car, bus, cycling, and walking) varies depending on location and context as well as the acceptable margin of error.

### Key Takeaways

The symposium participants noted several key takeaways during the discussion, as follows:

- Data are available from providers through specific licenses and agreements.
- OD trips can be determined; however, the reason for a trip and mode are less certain and have wider margins for error.
- Privacy issues weigh upon data availability to some degree.

# Agency and Academic Experience Using Cell Data

In this next session, researchers from various agencies and academic institutions provided descriptions of the work they have been conducting and their uses of cellular data. Symposium participants then discussed challenges relating to sampling and dataset extrapolation.

## Presentations

### Cynthia Chen, *University of Washington*

Cynthia Chen identified certain challenges and limitations associated with the use of cell data. These include difficulties with penetration rates and sample sizes (e.g., in some markets users have multiple phones), as well as the varying market shares of different carriers. Another challenge is oscillation between two cell towers and therefore being unable to determine an exact location. In a preliminary dataset of 8,000 people, Chen was able to determine 90 percent of homes and workplaces within a certain area. Current key challenges that she and her team are working on include validation of trips and population expansion.

### Xiaowei Xu, *University of Arkansas*

Xiaowei Xu conducted research that focuses on determining data clusters and associating individuals with groups of people or communities. He accomplished this through an algorithm that connects individuals' social networking activities, as shown in figure 3. One specific example

where Xu applied this algorithm was at college football games. Xu noted that challenges arise when cleaning and using data, as well as when finding a dynamic algorithm to enhance the result. Xu told the symposium participants that a major area to improve upon is to establish new methods for validating travel behavior by using cell data.

### Elizabeth Birriel, *Florida Department of Transportation*

Elizabeth Birriel explained that the Florida Department of Transportation (FDOT) has a desire to use cellular data to improve information used in 511 systems. These systems provide information about travel systems and travel routes. This information improvement was already achieved by using 2010 data, but it would be helpful to

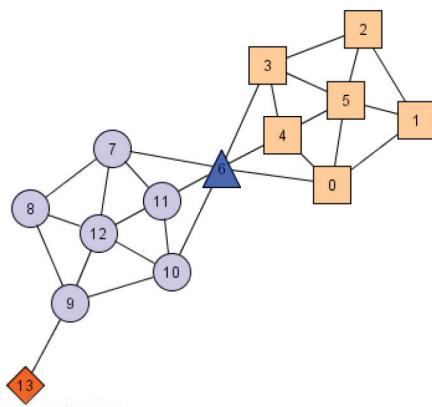


Figure 3. Clustering algorithm developed to represent individuals in a complex system.

continue to refine these systems through the use of more data as coverage was strong for the interstate highway system but spotty for arterial roadways. FDOT is also looking at how cellular data and automated vehicles can be leveraged with each other moving forward.

#### **Camille Kamga and Anil Yazici, City University of New York**

Camille Kamga and Anil Yazici noted that they have used cellular data for mobility studies, specifically within New York City. One study involved using taxi data and determining that there is higher demand for taxis during inclement weather, as shown in figure 4. Data were also used to determine, as others have found, that the morning and afternoon peak periods are not symmetric. Further challenges and areas of research include determining the reason for these findings, for example, if changes in mobility are based on trip cancellations or mode shifts.

#### **Andrew Rohne, Ohio-Kentucky-Indiana Regional Council of Governments**

Andrew Rohne used data from AirSage to attempt to model trips in his region of interest (greater Cincinnati, OH). In doing so, some challenges were identified regarding counts and capacities. Rohne compared his results using AirSage data to Ohio Department of Transportation and NHTS data to see how results matched and what the discrepancies were. Lessons learned and areas of future research include thinking through which data are available and what time of year the data would be most desirable for research purposes. Rohne also noted the possibility of incorporating cellular data in an NHTS expansion.

#### **Krishnan Viswanathan, CDM Smith**

Krishnan Viswanathan explained that he used data from AirSage to attempt to model trips in his region of interest (Wilmington, NC), as shown in figure 5. He derived various trips and examined

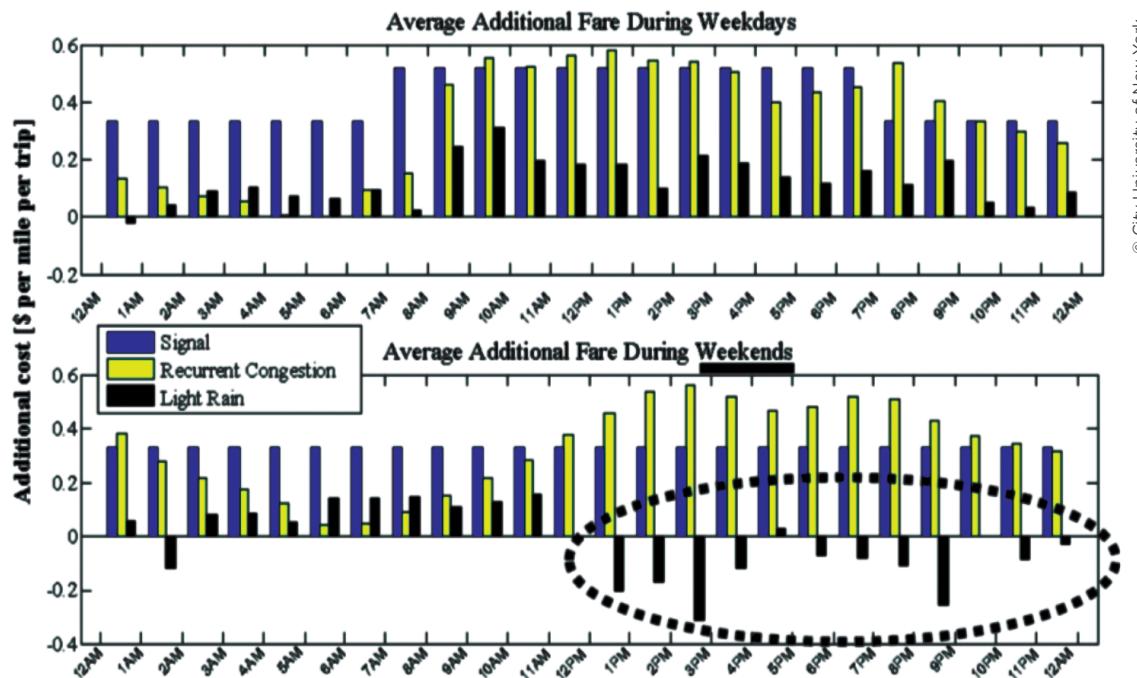


Figure 4. Taxi data used to determine demand for taxis during inclement weather.

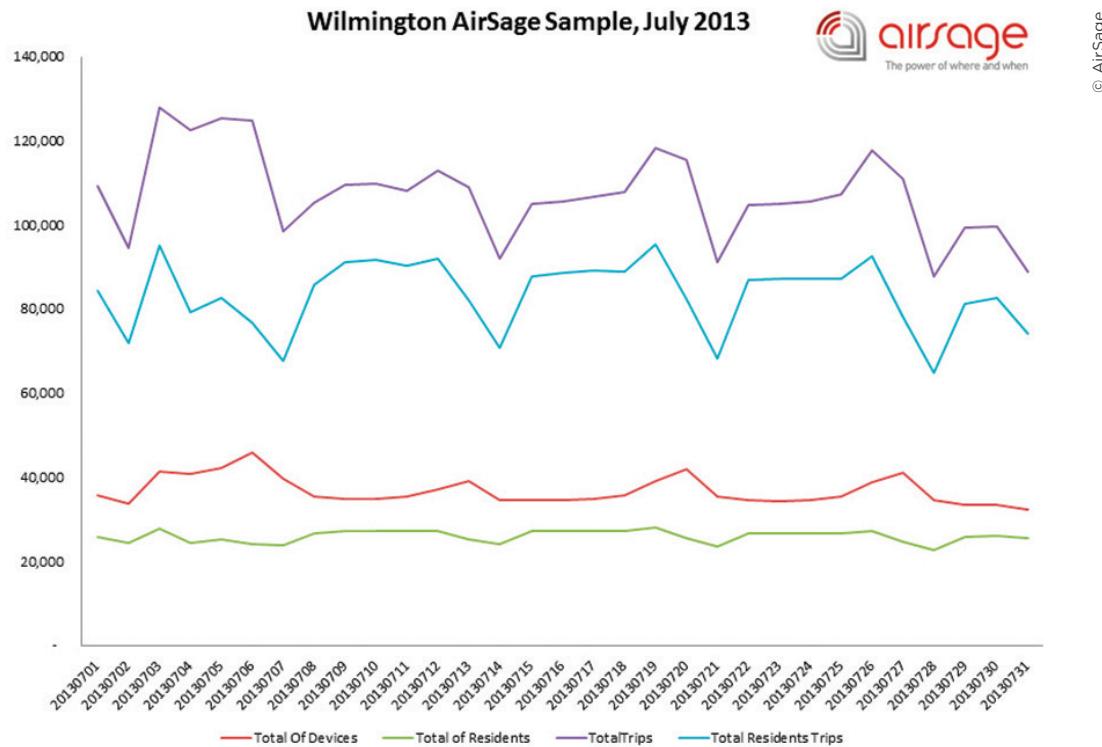


Figure 5. AirSage cellphone data sample.

different studies on non-home, residential, and home-to-home-based trips. To better understand movement and trips, Viswanathan noted that a possible next step would be to recruit users to willingly provide location data and demographic information.

### Major Themes Discussed

The symposium participants discussed several major themes during this session, as follows:

- **Determining a timeframe.** It is important while conducting research of this nature to determine what periods of the year will be most useful to analyze data, as data are typically purchased in monthly blocks.
- **Focusing on long-distance travel.** Although it is possible to determine long-distance travel to some degree, it is generally not the focus of regional level analysis as the possible long-distance traveler simply moves through the region without stopping.

- **Accessing different data sets.** Data consists of the movement of people rather than trips in the traditional sense. It is possible to access data that are extrapolated at the census tract level or to have unextrapolated data sets, such as device movements and a mobile survey.

### Key Takeaways

The symposium participants noted several key takeaways during the discussion, as follows:

- Validating results is key to determining the usefulness and meaningfulness of data.
- An important challenge moving forward is linking data to behavior through demographic information.
- Although it is possible to determine long-distance travel and mode based on duration, location, and land use, it remains inexact.

# Linking Trip Information to Behavior Data

In this session, researchers from various agencies and academic institutions provided descriptions of the work that they have been conducting and their uses of linking cellular and other mobile data with behavior and environment data. Symposium participants then discussed challenges relating to land use and margins of error.

## Presentations

### Lei Zhang, University of Maryland

Lei Zhang's focus has been to research trip purpose by using passively collected data and by attaining varying trip characteristics and details depending on the model. Zhang linked the 1995 American Travel Survey and cell data for urban trips. Zhang found that the most difficult types of trips to categorize were personal in nature, because there is less of a pattern. His model utilized land use, time of day, and duration and works fairly well across regions. Zhang noted that a useful next step would be linking the data in terms of a time series.

### Billy Bachman, Westat

Billy Bachman focused on what can be learned from GPS data in conjunction with travel surveys. He highlighted that developing the link to behavior is important for building forecasts and determining future movement. A challenge that Bachman noted included ensuring that the data stream is uninterrupted or does not contain missing data. This is a specific area in which

cellular data could prove to be beneficial. Integrating cellular data into traditional survey approaches could be particularly beneficial in filling in gaps and answering lingering questions.

### Shan Jiang, Massachusetts Institute of Technology

Shan Jiang worked on extracting data from cellular devices to link trips to activity. This included looking at possible nodes for daily travel, which typically consisted of home and work. In addition, Jiang linked the data to census tract data with the goal of determining the most heavily impacted roadways in terms of traffic to target congestion reduction, as shown in figure 6. Jiang noted that a key challenge is validating cell phone inferred results with survey and GPS data.

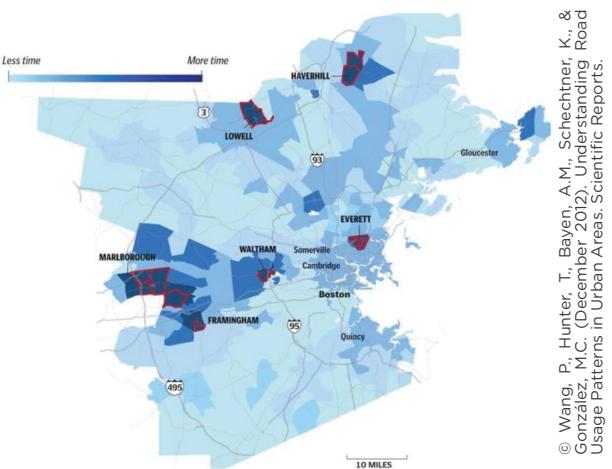


Figure 6. Road usage based on cell phone data linked with census tracts to illustrate congestion exposure.

© Wang, P., Hunter, T., Bayen, A.M., Schechtnar, K., & González, M.C. (December 2012). Understanding Road Usage Patterns in Urban Areas. Scientific Reports.

### **Major Themes Discussed**

The symposium participants discussed several major themes during this session, as follows:

- **Improving accuracy.** Cellular data are generally less accurate than GPS-tracking data in terms of position. Cellular data focuses mainly on travel among larger zones and monitors the handoff among different cell coverage areas, whereas GPS data can pinpoint specific OD locations.
- **Validating results.** Multiple data along the same paths assist in validating results by using cellular data.
- **Processing issues.** Pre-processing of cell data by the telecommunication companies that collect the data and post-processing by data providers can leave artifacts that limit use for understanding travel behavior.

- **Developing a nationwide standard.** Land-use models and data are extremely helpful, but specificity and configuration varies from region to region, and there is no nationwide standard land-use model.

### **Key Takeaways**

The symposium participants noted several key takeaways during the discussion, as follows:

- Land-use modelling is an important aspect of using cellular data but is inconsistent across regions or nationwide.
- A key challenge moving forward will be linking data from cellular devices to make it more meaningful.
- Validating results and determining the margin of error that one is willing to accept is critical.

# Relating Census Block and American Community Survey County Data, Demographic, and Economic Data to Derived Trip Data

In this session, Ben Pierce from Battelle presented on reducing bias and variance in sampling by linking cellular data with travel surveys and other datasets. Symposium participants then discussed challenges relating to changing methods based on emerging technologies.

## Presentation

### Ben Pierce, Battelle

Ben Pierce of Battelle began his presentation with an exercise to demonstrate the law of large numbers. This law indicates that as a group, it is possible to get close to the right answer, and as the size of the group increases, the closer to the answer the group will get. Pierce explained that statistics can measure variance in large surveys, as shown in figure 7; however, bias also needs to be considered. Pierce recommended taking an unbiased but imprecise survey and using precise cellular

data to calibrate it. He noted that, even if cellular data are biased, it may still be necessary to accept a certain level of bias to reduce the mean squared error.

## Major Themes Discussed

The symposium participants discussed the following major themes after the presentation:

- **Reconciling improvements.** A key challenge moving forward is attempting to reconcile technological improvements with changing methods and how information is collected and analyzed.
- **Reducing uncertainty.** It is possible to reduce the level of uncertainty within the results.

## Key Takeaways

During the discussion, the symposium participants agreed that leveraging and reconciling surveys and cellular data will be helpful in determining a broader picture of what is happening.

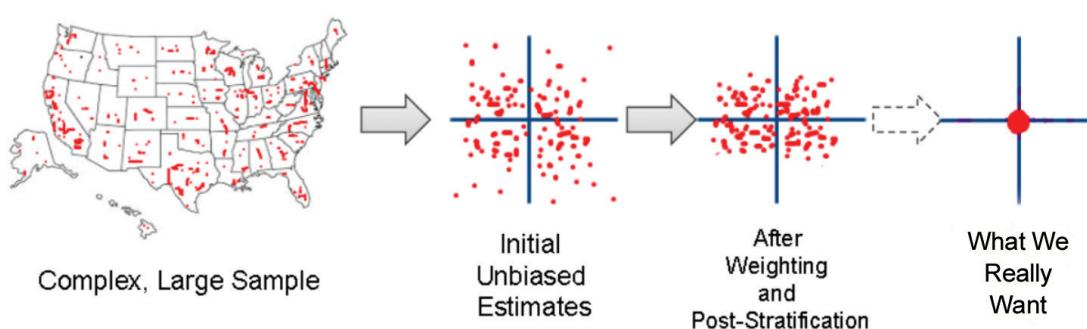


Figure 7. Example of a large national survey with sample sizes in the tens of thousands.

## Panel Summary and Next Steps

This session involved general discussion regarding the key trends touched upon throughout the day.

### Major Themes Discussed

The participants discussed the following overall themes:

- **Utilizing data.** Symposium participants noted that there remains tremendous value in using data despite the limitations. Several models and algorithms are transferable between GPS and cellular location data, which allows for some ease of use.
- **Finding a balance.** Some type of balance between traditionally collected data and emerging data is the best way to focus research efforts moving forward to answer national policy questions.
- **Developing standards.** It would be helpful to develop a common or standard language when discussing emerging issues and when differentiating among the various types of data streams (e.g., GPS or cellular, phone on or off).
- **Working with coverage differences.** Coverage in rural areas is available as long as there is service, but it may be less precise than in urban areas. Sampling issues and discrepancies occur not only in rural and urban areas, but also in terms of

different types of carriers and areas of higher or lower income.

### Key Takeaways

The participants noted several final takeaways from the symposium, as follows:

- Participants noted that different data are better. As researchers continue to gain a better understanding of mobility, traffic patterns, congestion, and behavior, a variety of data can increase understanding with accuracy and precision.
- For cell data to support highway policy information at the national level, standard product outputs can increase the value over ad hoc outputs designed for multiple projects.
- Although cellular data are available from various providers, acquiring data can be costly, and the licensing agreements can limit the use of the data. Fusing the cellular data with demographic information, census tract data, travel surveys, or other data will be a critical next step in determining travel behavior and developing applications for cellular data. Although there are certain privacy issues and other challenges that will need to be addressed over time, this data fusion will help to unlock further advantages of cellular data.



# Appendices

# Appendix A: Agenda

## FHWA Cell Phone Data and Travel Behavior Research Symposium February 12, 2014

|                    |   |
|--------------------|---|
| 8:30-8:45 a.m.     | <b>Welcome and Introductions</b><br>David Winter, <i>Federal Highway Administration, Office of Highway Policy Information</i><br>Tom Palmerlee, <i>Transportation Research Board</i>  |
|                    | <b>Overview of Exploratory Advanced Research Program</b><br>David Kuehn, <i>Federal Highway Administration, Exploratory Advanced Research Program</i>   |
| 8:45-9 a.m.        | <b>Issues to Resolve and Understand</b><br><b>Why Are We Gathering?</b><br>Tianjia Tang, <i>Federal Highway Administration, Travel Monitoring and Surveys Team</i>  |
| 9-10:15 a.m.       | <b>Data Coverage, Resolution, and Characteristics</b><br>Moderator: Brad Gudzinas, <i>Federal Highway Administration, Travel Monitoring and Surveys Team</i> <ul style="list-style-type: none"><li>• Bill King, <i>AirSage</i></li><li>• Daniel Rolf, <i>HERE</i></li><li>• Keith Hangland, <i>HERE</i></li><li>• Rick Schuman, <i>INRIX</i></li><li>• Nick Cohn, <i>TomTom</i></li></ul>   |
|                    | <b>Discussion</b>   |
| 10:15-10:30 a.m.   | <b>Break</b>  |
| 10:30 a.m.-12 p.m. | <b>Agency and Academic Experience Using Cell Data</b><br>Moderator: Guy Rousseau, <i>Atlanta Regional Commission</i> <ul style="list-style-type: none"><li>• Cynthia Chen, <i>University of Washington</i></li><li>• Xiaowei Xu, <i>University of Arkansas</i></li><li>• Elizabeth Birriel, <i>Florida Department of Transportation</i></li><li>• Camille Kamga, <i>City University of New York, Transportation Research Center</i></li><li>• Anil Yazici, <i>City University of New York, Transportation Research Center</i></li><li>• Andrew Rohne, <i>Ohio-Kentucky-Indiana Regional Council of Governments</i></li><li>• Krishnan Viswanathan, <i>CDM Smith</i></li></ul> |
|                    | <b>Discussion</b>   |

|                |   |
|----------------|---|
| 12-1 p.m.      | <b>Lunch</b>  |
| 1-2:15 p.m.    | <b>Linking Trip Information to Behavior Data</b> <ul style="list-style-type: none"><li>• <i>Deriving Trip Data From Cellular Data</i></li><li>• <i>Cellular Data, Social-Demographic, and Economic Aspects of the Data Generators</i></li></ul> Moderator: Cynthia Chen, <i>University of Washington</i> <ul style="list-style-type: none"><li>• Lei Zhang, <i>University of Maryland</i></li><li>• William (Billy) Bachman, <i>National Cooperative Highway Research Program 08-89: Applying GPS Data to Understand Travel Behavior</i></li><li>• Shan Jiang, <i>Massachusetts Institute of Technology</i></li></ul> |
|                | <b>Discussion</b>   |
| 2:15-3:30 p.m. | <b>Relating Census Block and American Community Survey County Data, Demographic, and Economic Data to Derived Trip Data</b> <ul style="list-style-type: none"><li>• <i>Non-Probability-Based Data for Universe Coverage</i></li></ul> Moderator: Stacey Bricka, <i>Texas A&amp;M Transportation Institute</i> <ul style="list-style-type: none"><li>• Ben Pierce, <i>Battelle</i></li></ul>   |
| 3:30 p.m.      | <b>Panel Summary and Next Steps</b><br>Brad Gudzinas, <i>Federal Highway Administration, Travel Monitoring and Surveys Team</i>   |
|                | <b>Adjourn</b>  |

## Appendix B: Attendees

| Name                 | Organization   |
|----------------------|--|
| William Bachman      | Westat   |
| Peter Bang           | Federal Highway Administration                           |
| Elizabeth Birriel    | Florida Department of Transportation                     |
| Stacey Bricka        | Texas A&M Transportation Institute                       |
| Greg Bucci           | U.S. Department of Transportation Volpe Center           |
| Cynthia Chen         | University of Washington                                 |
| Nick Cohn            | TomTom   |
| Gareth Coville       | The MITRE Corporation                                    |
| Vivian Daigler       | Nustats  |
| Elenna Dugundji      | Netherlands e-Science Center and University of Amsterdam |
| Daniel Fisher        | AirSage  |
| Rhett Fussell        | Parsons Brinckerhoff                                     |
| Deepak Gopalakrishna | Battelle   |
| Ryan Gordon          | RTI International  |
| Sarah Griffith       | Nustats  |
| Brad Gudzinas        | Federal Highway Administration                           |
| Keith Hangland       | HERE   |
| Ed Hard              | Texas A&M Transportation Institute                       |
| Ryan Hicks           | Boston Region Metropolitan Planning Organization         |
| Weimin Huang         | Cambridge Systematics                                    |
| Shan Jang            | Massachusetts Institute of Technology                    |

| Name                 | Organization   |
|----------------------|--|
| Camille Kamga        | City University of New York,<br>Transportation Research Center |
| Nicole Katsikides    | Federal Highway Administration                                 |
| Bill King            | AirSage  |
| Anurag Komanduri     | Cambridge Systematics  |
| David Kuehn          | Federal Highway Administration                                 |
| Michelle Lee         | Resource Systems Group   |
| Huan Li              | Mygistics  |
| Yangwen Liu          | Virginia Department of Transportation                          |
| Nancy McGuckin       | Consultant   |
| Jasmy Methipara      | Macrosys   |
| Rolf Moeckel         | University of Maryland   |
| Elaine Murakami      | Federal Highway Administration                                 |
| Dan Murray           | American Transportation Research Institute                     |
| Thomas Palmerlee     | Transportation Research Board                                  |
| Julie Parker         | Bureau of Transportation Statistics                            |
| Ben Pierce           | Battelle   |
| Eric Pihl            | Federal Highway Administration                                 |
| Jim Powers           | U.S. Department of Transportation Volpe Center                 |
| Andrew Rohne         | Ohio-Kentucky-Indiana Regional Council of Governments          |
| Daniel Rolf          | HERE   |
| Guy Rousseau         | Atlanta Regional Commission                                    |
| Hossain Sanjani      | Bureau of Transportation Statistics                            |
| Jean-Daniel Saphores | University of California, Irvine                               |
| Rick Schuman         | INRIX  |

| Name                 | Organization   |
|----------------------|--|
| Tianjia Tang         | Federal Highway Administration                                 |
| Rich Taylor          | Federal Highway Administration                                 |
| Krishnan Viswanathan | CDM Smith  |
| David Winter         | Federal Highway Administration                                 |
| Xiaowei Xu           | University of Arkansas   |
| Anil Yazici          | City University of New York,<br>Transportation Research Center |
| Lei Zhang            | University of Maryland   |

Office of Corporate Research, Technology,  
and Innovation Management  
6300 Georgetown Pike  
McLean, VA 22101-2296

Publication No. FHWA-HRT-14-060  
HRTM-30/07-14(1M)E