Report 1—State of the Practice: Collecting Multimodal National and Metropolitan Behavior Data

PL-18-016

March 2018
Updated July 2019

Developing and Implementing Sustainable Strategies and Solutions for Multimodal National Comprehensive Travel Behavior Data and Information

DTFH61-13-D-00020/0210

Federal Highway Administration
Office of Highway Policy Information

U.S. Department of Transportation
Federal Highway Administration

FOREWORD

Since the 1960s, the U.S. Department of Transportation and the Federal Highway Administration (FHWA) within it have explored a wide range of approaches to collecting national travel behavior data covering both local and long-distance travel. FHWA's Nationwide Personal Transportation Survey and National Household Travel Survey as well as the Bureau of Transportation Statistics' American Travel Survey were used to collect such data and information.(1,2) However, changes in technologies, demographics, and personal mobility over the past few decades have increased the cost of traditional approaches that rely entirely on probabilistic sample surveys. Also, given society's overall disinterest in responding to surveys, as evidenced by the continuously declining response rate (which further drives up the cost of traditional approaches), it is critical to evaluate and develop alternatives in gathering such national travel behavior data.

This report provides a synthesis of current approaches in gathering multimodal national and metropolitan travel behavior data as well as an introduction to more advanced and emerging data collection technologies and associated terminology. The project objectives are to explore opportunities for sustainable and comprehensive strategies and solutions for gathering national travel behavior data and information. Topics covered in this report include a summary of the state of the practice, an introduction to emerging data sources from passive location and socioeconomic data sources, and opportunities for moving the practice forward. Transportation agencies and practitioners seeking to improve travel behavior data collection processes will benefit from this report.

TECHNICAL REPORT DOCUMENTATION PAGE

<table>
<tr><td>1. Report No.
PL-18-016</td><td>2. Government Accession No.</td><td colspan="2">3. Recipient's Catalog No.</td></tr>
<tr><td colspan="2" rowspan="2">4. Title and Subtitle
Report 1—State of the Practice: Collecting Multimodal National and Metropolitan Travel Behavior Data</td><td colspan="2">5. Report Date
March 2018, updated June 2019</td></tr>
<tr><td colspan="2">6. Performing Organization Code:</td></tr>
<tr><td colspan="2">7. Author(s)
Dr. Stacey Bricka, Dr. Josie Kressner, Dr. Greg MacFarlane, Dr. Lei Tang, Dr. Jeffrey LaMondia, and Julie Dunbar</td><td colspan="2">8. Performing Organization Report No.</td></tr>
<tr><td colspan="2" rowspan="2">9. Performing Organization Name and Address
MacroSys, LLC
1901 N. Moore Street, Suite 602
Arlington, VA 22209

Transport Foundry
2870 Peachtree Rd NW #915-7318
Atlanta, GA 30305

Auburn University
238 Harbert Engineering Center
Auburn, AL 36849

Dunbar Transportation Consulting
24 Laurel Wood Drive
Bloomington, IL 61704</td><td colspan="2">10. Work Unit No.</td></tr>
<tr><td colspan="2">11. Contract or Grant No.
DTFH61-13-D-00020/0210</td></tr>
<tr><td colspan="2" rowspan="2">12. Sponsoring Agency Name and Address
Office of Highway Policy Information
Federal Highway Administration
200 New Jersey Avenue, SE
Washington, DC 20590</td><td colspan="2">13. Type of Report and Period Covered
Technical Report; July 2017–March 2018</td></tr>
<tr><td colspan="2">14. Sponsoring Agency Code</td></tr>
<tr><td colspan="4">15. Supplementary Notes
The Federal Highway Administration Task Monitor for this project was Daniel Jenkins, P.E.</td></tr>
<tr><td colspan="4">16. Abstract
This report provides a synthesis of current approaches in gathering multimodal national and metropolitan travel behavior data as well as an introduction to the more advanced and emerging data collection technologies and associated terminology. The project objectives are to explore opportunities for sustainable and comprehensive strategies and solutions for gathering national travel behavior data and information. Topics covered in this report include a summary of the state of the practice, an introduction to emerging data sources from passive location and socioeconomic data sources, and opportunities for moving the practice forward.</td></tr>
<tr><td colspan="2">17. Key Words
Travel behavior, multimodal data, emerging technologies, location-based services, NHTS, NextGen NHTS</td><td colspan="2">18. Distribution Statement
No restrictions. This document is available to the public through the National Technical Information Service, Springfield, VA 22161.
http://www.ntis.gov</td></tr>
<tr><td>19. Security Classif. (of this report)
Unclassified</td><td>20. Security Classif. (of this page)
Unclassified</td><td>21. No. of Pages
79</td><td>22. Price
N/A</td></tr>
</table>

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized.

SI* (MODERN METRIC) CONVERSION FACTORS

APPROXIMATE CONVERSIONS TO SI UNITS

Symbol	When You Know	Multiply By	To Find	Symbol
		LENGTH		
in	inches	25.4	millimeters	mm
ft	feet	0.305	meters	m
yd	yards	0.914	meters	m
mi	miles	1.61	kilometers	km
		AREA		
in^2	square inches	645.2	square millimeters	mm^2
ft^2	square feet	0.093	square meters	m^2
yd^2	square yard	0.836	square meters	m^2
ac	acres	0.405	hectares	ha
mi^2	square miles	2.59	square kilometers	km^2
		VOLUME		
fl oz	fluid ounces	29.57	milliliters	mL
gal	gallons	3.785	liters	L
ft^3	cubic feet	0.028	cubic meters	m^3
yd^3	cubic yards	0.765	cubic meters	m^3
		NOTE: volumes greater than 1000 L shall be shown in m^3		
		MASS		
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")
		TEMPERATURE (exact degrees)		
^oF	Fahrenheit	5 (F-32)/9 or (F-32)/1.8	Celsius	^oC
		ILLUMINATION		
fc	foot-candles	10.76	lux	lx
fl	foot-Lamberts	3.426	candela/m^2	cd/m^2
		FORCE and PRESSURE or STRESS		
lbf	poundforce	4.45	newtons	N
lbf/in^2	poundforce per square inch	6.89	kilopascals	kPa

APPROXIMATE CONVERSIONS FROM SI UNITS

Symbol	When You Know	Multiply By	To Find	Symbol
		LENGTH		
mm	millimeters	0.039	inches	in
m	meters	3.28	feet	ft
m	meters	1.09	yards	yd
km	kilometers	0.621	miles	mi
		AREA		
mm^2	square millimeters	0.0016	square inches	in^2
m^2	square meters	10.764	square feet	ft^2
m^2	square meters	1.195	square yards	yd^2
ha	hectares	2.47	acres	ac
km^2	square kilometers	0.386	square miles	mi^2
		VOLUME		
mL	milliliters	0.034	fluid ounces	fl oz
L	liters	0.264	gallons	gal
m^3	cubic meters	35.314	cubic feet	ft^3
m^3	cubic meters	1.307	cubic yards	yd^3
		MASS		
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
		TEMPERATURE (exact degrees)		
^oC	Celsius	1.8C+32	Fahrenheit	^oF
		ILLUMINATION		
lx	lux	0.0929	foot-candles	fc
cd/m^2	candela/m^2	0.2919	foot-Lamberts	fl
		FORCE and PRESSURE or STRESS		
N	newtons	0.225	poundforce	lbf
kPa	kilopascals	0.145	poundforce per square inch	lbf/in^2

*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

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AAT	Google™'s Aggregated and Anonymized Trips
ACS	American Community Survey
API	application programming interface
ATRI	American Transportation Research Institute
ATS	American Travel Survey
ATUS	American Time Use Survey
AV	autonomous vehicle
BLS	Bureau of Labor Statistics
BTS	Bureau of Transportation Statistics
Caltrans	California Department of Transportation
CFS	Commodity Flow Survey
CTPP	Census Transportation Planning Products
DB1B	Airline Origin and Destination Survey
DES	discrete event simulation
DOE	U.S. Department of Energy
EAR	Exploratory Advanced Research
E-E	external-to-external
E-I	external-to-internal
FAF	Freight Analysis Framework
FHWA	Federal Highway Administration
FRA	Federal Railroad Administration
FTA	Federal Transit Administration
GIS	Geographic Information System
GLH	Google™ location history
GPS	Global Positioning System
GTFS-ride	General Transit Feed Specification Ride
HBW	home-based work
H-GAC	Houston-Galveston Area Council
HOFM	Office of Freight Management and Operations
HPMS	Highway Performance Monitoring System
HPPI	Office of Highway Policy Information

I-E	internal-to-external
KSI	fatality or serious injury
LBS	location-based service
LEHD	Longitudinal Employer-Household Dynamics
LOS	level of service
MAG	Maricopa Association of Governments
MORPC	Mid-Ohio Regional Planning Commission
MPO	metropolitan planning organization
MTO	Ministry of Transportation Ontario
MWCOG	Metropolitan Washington Council of Governments
NCDOT	North Carolina Department of Transportation
NCHRP	National Cooperative Highway Research Program
NHO	non-home-based other
NHTS	National Household Travel Survey
NHTSA	National Highway Traffic Safety Administration
NPMRDS	National Performance Management Research Data Set
NTD	National Transit Database
NTS	*National Transportation Statistics*
O-D	origin-destination
ODOT	Oregon Department of Transportation
SEMCOG	Southeast Michigan Council of Governments
SFCTA	San Francisco County Transportation Authority
SHRRP2	Second Strategic Highway Research Program
TAF	Transportation Analysis Framework
TAZ	traffic analysis zone
TBI	Travel Behavior Inventory
TNC	transportation network company
TPO	Transportation Planning Organization
TRB	Transportation Research Board
TxDOT	Texas Department of Transportation
USCB	U.S. Census Bureau
USDOT	U.S. Department of Transportation
UST	underground storage tank

UZA	urbanized area
VDOT	Virginia Department of Transportation
VMT	vehicle miles traveled
WPS	Wi-Fi Positioning System

CHAPTER 1. INTRODUCTION

The United States is at a critical juncture with respect to how to best document travel behavior patterns to support transportation planning and policymaking efforts. Over the past few decades, the cost of conducting traditional probabilistic sample surveys has increased due to decreased response rates, changes in population demographics, the proliferation of personal technologies, and increased personal mobility. With the next decennial census scheduled for 2020, and many State and regional transportation agencies currently developing data collection programs to coincide with this event, many planners and engineers are struggling to generate feasible, comprehensive, and representative travel behavior data programs. In fact, there is much debate on the best approaches to address rising costs, declining response rates, and engaging hard-to-reach populations, such as younger adults, larger households, and minority respondents. Equally important to these agencies is gaining an understanding of where and how big data can complement, supplement, or potentially replace traditional household travel surveys. Many agencies also seek to improve freight and commercial vehicle travel forecasts; however, there are challenges not only in working with existing commercial vehicle and goods movement data but also in exploring third-party big data products.

Travel behavior data are fundamental to inform policy research at the Federal, State, and local levels. Trend data to support policy analysis include travel patterns, motor fuel costs and usage, motor vehicle registrations, drivers' licensure rates, highway user taxation, highway mileage, and highway finance. In addition, these data are increasingly used to understand the connection between health and active transportation, measure the impact of investments in bicycle/pedestrian infrastructure improvements, and document the current fleet composition and related emissions.

The path forward appears to be one that includes both a fusion of ideas and, more importantly, a fusion of data. With the ubiquitous use of cellular phones and their location technologies comes the compilation of location-based service (LBS) data. There is also an increasing volume of passively collected data available from in-vehicle devices (i.e., personal, commercial, occupancy, and fuel efficiency), which may continue to grow as autonomous and connected vehicles become part of the vehicle fleet. Researchers have made significant progress transcribing passive data into trips with varying success in the development and calibration of algorithms to detect trip purpose and travel mode. These passive data sources show promise in being combined with customer data (available at the residential address level) to create synthetic travel datasets and in forming the basis for a new generation of passive data travel demand models.

While the travel behavior community is optimistic of these new opportunities, the reality is that many technical, operational, and institutional details still need to be ironed out. At this point in time, the technology and available data are being used to automate or simplify current processes and practices in response to the current question of "how can these data improve what I am currently doing?" The travel behavior data community is rapidly approaching the point where these technologies and the resultant data provide an opportunity to revolutionize the industry, changing the question to "where can these data take us?"

For purposes of this report, "passive data" refer to information collected without explicit or overt noticeable interaction from a person. There are two types of passive data. The first (referred to as "type 1") is data collected through a device to augment responses from probabilistic surveys. For example, sampled respondents use a smartphone or web application that collects information on trip times and routes in addition to any information the respondents provide directly. A surveyor still must design the survey and recruit respondents, but there is a lower response burden, and the information may be more clean, accurate, and complete. Type 1 passive data are considered to be state of the practice and are discussed in chapter 2.

The second category of passive data (referred to as "type 2") consists of information collected for purposes unrelated to travel behavior research but that nevertheless may be useful in understanding certain aspects of travel behavior. In a way, the data are found by the researcher rather than collected. For example, a mobile application may request to know where its users are so that it can provide more useful information about the surrounding area. A consequence of this is that a developer may know where its users are located at different times of the day and what they may be doing at those times. This information can, in turn, allow the developer to improve its processes and offerings, monetize its data through targeted advertisements, or package and resell the data it receives. Type 2 data are discussed further in chapter 3.

Understanding the new and emerging technologies of type 1 data can help researchers creatively tackle some problems associated with surveys (e.g., high costs and low response rates), but these data do not represent the methodological shift in travel behavior research that type 2 data require. Table 1 compares type 1 and 2 data. One major difference is that type 2 data populations tend to be larger than type 1 data populations, although the extent to which the population differs from the universal population can vary by technology and data provider. Additionally, type 2 data are collected constantly, which contrasts with the episodic nature of surveys augmented with type 1 data. Furthermore, with type 1 surveys, researchers can craft questions that answer their precise needs; however, type 2 data typically do not contain all possible pieces of information, requiring the researcher to find and link other data to gain insight into travel behavior.

Table 1. Typology of passive data.

Characteristic	Type 1 (Augmented)	Type 2 (Found)
Population universe	Random sample	Complete subpopulation
Scope of behavioral understanding	Custom design	Limited to data contents
Collection frequency	One time	Ongoing
Responder burden	Medium to heavy	None
Response rate	Low; the trend is getting lower	Not applicable
Researcher role	Survey design	Data discovery and processing
Types of bias	Nonresponse, coverage, low sample size	Coverage, imputation

Through a synthesis of the state of the practice and emerging passive data applications, the goal of this report is to define a common foundation of understanding on which a new generation of multimodal travel behavior data and models can be built. At the same time, discussions with the travel behavior community are essential to identifying key criteria that can help agencies evaluate the fit of emerging data opportunities with respect to data needs. It is important to note that any mention of private company names and products within this report is not an endorsement by FHWA and are included by the authors for reference purposes only.

This report is structured as follows:

- **Chapter 1. Introduction:** Provides a summary of the report contents.
- **Chapter 2. State of the Practice:** Summarizes current methods for collecting travel behavior data.
- **Chapter 3. Passive Data Collection Technologies and Data Sources:** Introduces emerging methods and sources for passive data collection.
- **Chapter 4. Opportunities and Implications:** Identifies areas for consideration in moving the travel behavior data collection practice forward.
- **Appendix A. Travel Behavior Data Products from Federal Programs:** Details travel behavior data collected and compiled at the Federal level.
- **Appendix B. Recent Projects Using Positional Data**: A short compendium of recent conference presentations that highlight the use of positional data as a complement or substitute for traditional survey data.
- **Appendix C. Private Sector Travel Behavior Data Products:** Summarizes and further details the data products introduced in chapter 3.

CHAPTER 2. STATE OF THE PRACTICE

This chapter summarizes the state of practice with respect to multimodal travel behavior data initiatives at the Federal and State/regional levels for both daily and long-distance travel. A list of travel behavior data collected or compiled across relevant Federal agencies is included in appendix A.

STATE OF THE PRACTICE IN COLLECTING MULTIMODAL DAILY TRAVEL BEHAVIOR DATA

Regional and statewide travel surveys are conducted primarily to serve as an input to travel demand models. Travel demand models are used to support long-range transportation plan development and evaluate transportation system improvements at the corridor and sub-area levels. Increasingly, the data are also used to answer policy questions, inform economic development decisions, and serve as inputs for health-related planning tools. While the main surveys conducted across the United States are household and transit on-board surveys, agencies also look to establishment and special generator intercept surveys to better understand employee and visitor daily travel patterns and campus-wide travel surveys to capture student travel both on and off campus. External station surveys have mostly been replaced with passive data alternatives.

Despite the difficulties and costs of collecting travel behavior data, today's travel demand models are more data hungry than ever. The traditional trip-based models have become more disaggregated through the addition of time-of-day and destination choice components. Advanced activity-based models are now in full operation in several metropolitan areas and require data from all household members and preferably for multiple days. These advanced models put pressure on the already strained travel surveys by demanding additional data from a diminishing population willing to participate. The state of the practice with respect to these core travel surveys is as follows:

- **Household surveys:** One of the biggest challenges with household surveys is low response rates, which erode the representativeness of the survey results. Current design practices entail the use of enhanced address-based samples, with web forms and smartphones as the dominant data collection modes. From a survey administration viewpoint, State and regional transportation agencies are beginning to conduct smaller sample surveys more frequently (i.e., continuous cross-sectional design) instead of performing more traditional large sample studies every 8–10 yr. The benefits of these improvements include: (1) more frequent data collection enables agencies to monitor, track, and report on emerging trends; (2) address-based sampling targets households regardless of the type of home telephone (landline or cellular phone); and (3) agencies gain the ability to oversample geographic and demographic characteristics associated with lower incidence travel modes (e.g., transit, bicycle) or emerging travel modes (e.g., ride hailing).

  The best known household travel survey is the National Household Travel Survey (NHTS), which has been conducted every 5–8 yr since 1969 and was most recently completed in 2017.[1] The data are used to support Federal policy initiatives as well as State- and regional-level planning efforts. Since 1990, the Federal Highway Administration's (FHWA's) NHTS Program has included a pooled fund opportunity where agencies can buy into the survey through the purchase of additional samples. This option is used by State and regional agencies. Agencies not participating in the NHTS pooled fund conduct their own travel surveys with varying designs and questions, or they "borrow" data from similar regions. All household travel surveys in the United States document trip purpose, mode of travel, time of day of travel, and locations (i.e., origin-destination (O-D)) of travel.

The use of smartphone applications that collect type 1 passive data and prompt respondents for specific details about each trip when necessary is generally accepted as state of the practice. Most smartphone applications also include real-time validation of the passively captured details. The data collected through these applications can be exported and shared and are roughly the equivalent of a travel diary, although they are missing information on travel purpose and travel party. The Maricopa Association of Governments (MAG) and the San Diego Association of Governments both fielded 100 percent smartphone household travel surveys in 2016–2017.[3,4] Additionally, the Ohio Department of Transportation Household Travel Survey extends this practice by collecting both local and long-distance travel as well as running the project as a continuous survey (over a 5-yr period).[5] These hybrid methods reduce respondent burden significantly, as the applications ask fewer questions, but provide more detailed travel diary data.

- **Workplace, visitor, and special generator surveys:** Integrated establishment survey designs that capture the workplace, visitor, commercial vehicle, and count details at one time for more efficient data collection are now the standard. Significantly improved methods include using tablets for in-person interviewing, databases of companies to serve as sampling frames, and passive technology for detailed person and vehicle counts. These technologies help to lower the cost of fielding the surveys, resulting in more efficient use of labor. In addition, the administration of surveys using tablets provides for real-time mapping of locations, thus improving the quality of O-D data.

  Establishment surveys are not conducted as regularly as household travel surveys. Those performed recently include those by MAG (2015–2017), New York Metropolitan Council of Governments (2014–2016), and the Oklahoma–Kentucky–Indiana Council of Governments (2015–2016) efforts.[3,6,7] These studies were conducted as intercept interviews of employees and patrons of local businesses selected to fill stratified sampling goals based on industry, employee size, and geographic location within the region.

- **Commercial vehicle and freight surveys:** These surveys include multi-day Global Positioning System (GPS) data collection with minimal reliance on driver logs. Some agencies are exploring the use of passive data as the primary source of O-D flows, with smaller traditional survey data relegated to the validation role. The increased reliance on GPS technology and move to passive data are thought to provide higher-quality and more complete data with reduced burden for vehicle drivers.

The recently completed MAG Establishment Survey included three different commercial vehicle/freight components: (1) a phone survey of businesses to establish vehicle ownership and usage patterns, (2) a travel survey where businesses were asked to equip vehicles with GPS devices, and (3) the purchase of StreetLight Data™ for the region.[(3)] A similar study was conducted for metropolitan planning organizations (MPOs) located in Colorado's Front Range region (Denver, Colorado Springs, Pueblo, and North Front Range), which also had three components: (1) a phone survey of businesses to establish vehicle ownership and usage patterns, (2) the collection of truck GPS and (3) the assessment of American Transportation Research Institute (ATRI) truck data.[(8,9)]

- **Transit surveys:** Transit surveys use boarding and alighting data, GPS, bar code scanning, and other passenger-related data to form the basis of transit flow data. For surveys, tablets with real-time geocoding are used to improve the level of O-D geocoding rates as well as the overall quality of the data. Most major metropolitan regions have an on-board survey completed within the past 5–7 yr.

  Service performance data for transit agencies across the United States are available via the National Transit Database (NTD), which is a primary source of data and statistics on national transit systems.[(10)] To meet legislative requirements, around 850 transit agencies in urbanized areas (UZAs) that receive Federal Transit Administration (FTA) funds submit data to the NTD. NTD performance data are used to apportion FTA funds to transit agencies in UZAs. The information provided by the NTD include data on transit profiles, national transit summaries and trends, time–series data on transit systems, monthly ridership, and safety.

- **External station surveys:** External station surveys are no longer conducted by directly stopping drivers to obtain travel details. Instead, the current practice relies on fusing Bluetooth™ and third-party cellular and passive GPS data streams that are interpreted through custom algorithms.

  Most smartphones and other smart devices include Bluetooth™ chips. When device owners enable their Bluetooth™, the devices locate antennas within their broadcast range. These antennas receive unique identification numbers from devices even if they do not connect to their network. Bluetooth™ antenna operators can locate and track devices within their signal range. If an operator builds a network of antennas, it is possible to track devices throughout an area or within a network.

  For transportation planning, data collection efforts using Bluetooth™ data often involve setting up antennas and recording the device IDs that ping the Bluetooth™ antennas. This approach is useful for establishing a cordon or screen line to replace intercept surveys. This kind of data collection approach has also been used with Wi-Fi technology to collect information about indoor movements inside large facilities (e.g., inside stadiums or shopping malls). Placing Bluetooth™ antennas along a corridor can allow researchers to measure travel times. Bluetooth™ data tend to not be sold by data aggregators but rather directly collected with devices and systems available from commercial hardware developers for specific project applications.

STATE OF THE PRACTICE IN COLLECTING MULTIMODAL LONG-DISTANCE SURVEYS

Long-distance travel is critical to regional economic development, interstate commerce, freight system operations, and quality of life. In fact, the U.S. Travel Association reported that in 2016, Americans completed more than 1.7 billion long-distance trips for leisure purposes and more than 457.4 million long-distance trips for business purposes.[(11)] This resulted in over $990.3 billion in direct spending from long-distance travel, which translates to 2.7 percent of the 2016 national gross domestic product.[(11)] The scope of long-distance travel continues to expand each year, and many anticipate an even greater role for this travel as megaregions continue to develop across the United States, causing greater volumes of travel to extend beyond traditional metropolitan planning region limits.[(12)] Not surprisingly, many States and MPOs now seek to incorporate these often-ignored long-distance trips into their long-range transportation plans to proactively address potential economic, congestion, and growth issues.

Passenger

The definition of a long-distance trip varies across studies. For example, the 2013 Longitudinal Survey of Overnight Travel captured tours that included at least a one-night stay in a location away from the respondent's home over a 1-yr period.[(13)] In contrast, the 2009 Michigan Department of Transportation's Long-Distance Travel Survey focused on tours of 50 mi or more (one way) from an individual's home that occurred over a 3-mo period.[(14)] A review of similar surveys also used the same definition for a long-distance trip of 50 mi or more one way.[(15)]

Nationally, there are two sources of long-distance travel associated with travel surveys. The first is the 1995 American Travel Survey (ATS), which was sponsored by the U.S. Department of Transportation's (USDOT's) Bureau of Transportation Statistics (BTS).[(2)] In this year-long effort, households were contacted quarterly to collect information about trips that were 100 mi or more (one way). In 2001, ATS was merged into the NHTS design.[(1)] This combined study collected the typical daily travel for a 1-yr period and also included details about all trips 50 mi (one way) from home that were completed in the 4 weeks prior to the assigned travel day. Long-distance trip details included tour characteristics, access/egress modes, and overnight stops. If no trips were reported for the 4-week period, participants were queried about the last long-distance trip completed.

The 2009 NHTS did not collect any data about long-distance travel.[(1)] For the recently completed 2017 NHTS, there were no national questions about long-distance travel.[(1)] However, seven add-on agencies asked their constituents for some long-distance details. These agencies used varying definitions of long-distance travel but centered about the theme of trips made either 50 or 75 mi from home.

Long-distance travel is significantly different from daily travel, and data are often difficult to collect simultaneously with traditional daily travel surveys, given respondent burden and the cost associated with a longer survey. Many regions either capture these trips in their external counts or borrow parameters from other regions.[(15)] Complicating data collection efforts is the fact that many agencies and organizations often require different sets of long-distance travel characteristics to address a varied range of transportation and economic decisions. Additionally, long-distance travel patterns differ significantly from daily travel, meaning that survey instruments and collection efforts utilized must be different from those implemented for daily travel surveys.

FHWA funded an Exploratory Advanced Research (EAR) project to evaluate approaches for collecting long-distance travel data.(16) It was designed to identify novel, innovative, and cost-effective data collection alternatives to collecting long-distance travel. Research included a series of smaller studies to investigate alternative candidate approaches for future long-distance surveys. These smaller studies included the following:

- Integrating existing survey data with the results of the new survey.
- Designing a core probabilistic sample by using microdata from the 1995 ATS and the 2001 NHTS.
- Performing data fusion and imputation research on combining probabilistic and non-probabilistic data.
- Exploring post-processing methods used with advanced travel survey methods (e.g., GPS, smartphone, and social media) in order to impute trip information.
- Developing smartphone applications and Facebook™ surveys that combine survey data with passive location tracking to improve data quality and minimize bias.(16,17)

Another source of long-distance travel is the Transportation Analysis Framework (TAF), which was FHWA's first effort to estimate long-distance passenger movements, defined in this effort as trips with a distance greater than 100 mi.(18) It consists of a set of trip tables that provide information on person trip flows at the county-to-county (or equivalent-to-equivalent) level for base year 2008 and future year 2040 for different passenger modes (e.g., auto, bus, air, and rail). A trip is defined as a one-way trip. The TAF includes the following three items:

- Long-distance passenger O-D tables.
- Long-distance passenger travel demand model.
- Intercity bus ridership data for the top 200 markets.

Freight

On the freight side, the Commodity Flow Survey (CFS) is the primary data source on goods movement in the United States at the national and State levels.(19) It is a shipper-based survey that is conducted by a partnership between BTS and the U.S. Department of Commerce's U.S. Census Bureau (USCB). The CFS has been conducted every 5 yr since 1993. It is the only publicly available goods movement data source for highways. Data from the CFS and other related sources are combined to create the Freight Analysis Framework (FAF).(20) Produced by BTS and FHWA, FAF is a comprehensive data source that provides information on freight movement of all transportation modes among States and major metropolitan areas.

CHAPTER 3. PASSIVE DATA COLLECTION TECHNOLOGIES AND DATA SOURCES

Passive data collection technologies are being used to generate new sources of travel behavior data. These new data sources are used in a variety of contexts within the U.S. travel behavior data industry. Improvements to technology and services running on that technology are responsible for the increasing availability of passive data, and it is impossible to separate the usefulness of passive data from the technology used to collect, process, and distribute it.

The chapter begins with a brief introduction to the types of passive data technologies that generate data useful for transportation planning (i.e., type 2 data as defined in chapter 1). This is followed by a more detailed discussion with examples of how the data are beginning to be used in practice. Examples are used to illustrate specific applications. This chapter also includes examples of where the data are being used as well as which applications show promise, or, at a minimum, point out limitations in the new data opportunities. Applications were selected to illustrate specific or unique aspects of the technologies or applications and are not intended to be all inclusive. It is important to note that the investigation and application of passive data in transportation planning is undergoing rapid testing and adoption across the United States, as evidenced by the abstracts of recent conference presentations included in appendix B. As a result, this chapter is not intended to serve as an exhaustive inventory of relevant projects.

POSITIONAL DATA

The first category of type 2 passive data is positional data. Positional data provide records that primarily contain the spatial location of observations. Table 1 includes a list of companies known to collect positional data from their customers or clients. It is organized by the technology used to collect the data. A summary of data products by vendor is provided in appendix C.

Table 1. Positional data providers and data collection technologies.

Positional Data Provider	Cell Tower Triangulation	GPS (Vehicle)	GPS (Mobile)	Wi-Fi Positioning System (WPS)/ LBS
Actively Selling/Available				
AirSage™*	—	Other	—	Primary
StreetLight Data™ (partners with Cuebiq™ and INRIX™)	—	Primary	—	Primary
HERE™	—	Primary	—	—
INRIX™	—	Primary	—	—
ATRI	—	Primary	—	—
SkyHook™	—	—	—	Primary
Cuebiq™	—	—	—	Primary
SafeGraph™	—	—	—	Primary
Twitter™**	—	—	—	Primary
Not Selling				
Google™/Android™	Primary	—	Primary	Primary
AT&T™	Primary	—	—	—
Apple™	Primary	—	Primary	Primary
Facebook™	—	—	—	Other

—Not applicable.
*As of early 2019, AirSage™ is selling historical products from cell tower triangulation covering dates up to 2017 from either Verizon™ or Sprint™. Starting in 2018, they only sell products with LBS-type sources.
**Twitter™ does not sell its data but does have an application programming interface (API) that researchers can query against to collect their own data.

Cell Tower Triangulation

Cellular phones operate by receiving radio waves from physical towers and sending them back to the towers in response. When a device connects to multiple towers, it is possible to triangulate the location of the device within some level of tolerance. Some of the major cellular carriers do their own triangulation to balance their networks and understand usage better. Others were selling their tower data to firms who cleaned, triangulated, and resold the data to transportation planners, typically as aggregate O-D trip flow tables. Beginning in 2018, these data seem to no longer be readily available on the market.

The trip flow tables from cell tower data can be segmented by time of day and day of the week. The resellers were imputing likely home and work locations and, consequently, work and home trip purposes, by observing a device's location during the day and overnight across multiple days. The carriers did not release demographic information on the phone's user to the resellers in general, but the resellers may have attempted to infer individual and household attributes by matching the device's home location to aggregate U.S. Census data.

Cellular phone data tend to have large sample sizes, device persistence, and deep market penetration. A total of 95 percent of individuals in the United States own a cellular phone, and 98 percent of those users are served by just four large national carriers.[21,22] Resellers that contract with only one or two of these carriers can therefore access data from a large share of a region's population. Because cellular phone users typically carry their phone with them at all times, these data measure travel behavior irrespective of the travel mode.

Travel modelers were increasingly using O-D matrices derived from cellular triangulation, typically resold by AirSage™, in model development for external trip models and calibration/validation efforts. The availability, sample size, cost, and easy distribution of these matrices brought their use effectively into standard practice for the aforementioned modeling activities within the last few years.

External Trip Models

Travel models must account for trips made by individuals who live outside the study region but travel into or through the region. These trips will be missed in a household survey sampling frame that draws only from residents within the study region, so other methods to capture trips into or through the region are necessary. Historically, many agencies have intercepted traffic on the edge of the region and surveyed the stopped traffic; these types of studies are expensive and politically unpopular. An alternative has been to apply gravity models to traffic counts at the external stations.

O-D matrices developed from cellular triangulation data provide a compelling data source for external trip models. This process is exemplified in work performed by the North Carolina Department of Transportation (NCDOT).[23] In this study, researchers established districts both within the model region and beyond its boundaries at each entrance station and purchased cellular O-D matrix data representing 1 mo of travel between those geographies. Quality control on the data included a comparison to results from a household travel survey conducted in roughly the same period captured with the cellular data. Model results were validated against traffic counts at the relevant external stations. The modelers then used the cellular data to establish trips made by residents and non-residents as well as trips by work and non-work purposes. Observed through-trip tables were generated, and external trip models were successfully developed. The modelers concluded that, "mobile phones are a useful source of data for the development and estimation of external trip models that represent observed local traffic patterns" (p. 31).[23]

Validation/Calibration

Modelers also used cellular O-D matrices to calibrate district-to-district flows within a region. Regions of all sizes have done this, including Washington, DC; Tyler, TX; and the aforementioned work performed by NCDOT.[23–25] The data have also been used in statewide models in Idaho.[26] In many of these cases, researchers had a contemporaneous household travel survey to compare the O-D matrices against, and the cellular matrices compared favorably.

In Washington, DC, the Metropolitan Washington Council of Governments (MWCOG) purchased O-D data to assess its viability in updating forecasts of external, through, and visitor travel.[24] In particular, agency staff sought to understand how the trips compared to modeled trips and activities as well as potential biases or data limitations. The data were processed to convert O-D to production/attraction format. In addition, MWCOG staff aggregated vendors' trip purposes and traveler type for more direct comparisons to the current model structure. Their assessment showed good matches with respect to motorized person trips. Regarding trip purpose, the purchased data contained more home-based work (HBW) trips and fewer non-home-based other (NHO) trips as compared to the model output. In the aggregate, total time and distances were in agreement, but average trip times were different for the HBW and NHO trips.

Validation was conducted in the NCDOT study by assigning cellular matrices directly to the highway network and comparing the resulting volumes against observed highway traffic counts.(23) Screen line and cordon analyses could reveal systemic bias in the cellular data for specific populations or geographies, allowing the planners or the data provider to take corrective action. This type of validation must consider that corridors with high transit ridership may have high O-D flows in the cellular data but lower observed traffic counts.

Related Research

In recent years, two research efforts have provided insight and a better understanding of how cellular data might benefit travel behavior analysis. The first was the Cell Phone Data and Travel Behavior Research Symposium that was sponsored by FHWA on February 12, 2014.(27) The topic of this symposium was using cellular location data for national travel behavior research. Participants included a wide range of data providers, researchers, and professionals who shared their experiences in using cell data. At that time, the main issues regarding cell data were the availability and application of cellular data, the merging of cellular data with other data sources (including travel survey data), and the validation of cellular data. Several researchers from public agencies, private firms, and academic institutes also discussed their experiences in using cellular data.

The second research effort was funded by the National Cooperative Highway Research Program (NCHRP) entitled "NCHRP 08-95: Cell Phone Location Data for Travel Behavior Analysis."(28) The objectives of this research were to explore the accuracy of using cellular phone data for travel behavior analysis and to provide guidelines on using these data for such an analysis. The evaluation centered about interviews with transportation agencies that had experience in purchasing and using cellular phone location data for travel behavior study. The guidelines address the issues under the topics of cellular data acquisition, data quality, and data applications.

Limitations

Their common use aside, cellular O-D matrices are not without limitations and caveats. Virtually all who have used such data warn that the records appear less reliable at smaller levels of geography, which conforms to the technical limitations of cellular triangulation. Acquiring aggregate data over longer periods can reduce errors. Similarly, the algorithms to determine likely home and work locations can be less accurate in places where there is shift work or large numbers of overnight workers, such as factories and hospitals. Further, there are unknown and thus unexplored biases related to carrier choices of different classes of users. If a particular cellular carrier is disproportionately popular with a certain segment of the population, then the resold data could be biased in the same way.

GPS

GPS is a satellite-based coordinate determination regimen developed and operated by the U.S. Air Force and available for civilian use globally. GPS receivers in mobile phones and vehicles determine their position on Earth by triangulating signals from a constellation of satellites. Many companies now provide navigation and other location-aware services to users through GPS-enabled devices and resell the information they collect in various forms. GPS accuracy is generally within 4 meters.

The Maryland State Highway Administration purchased 4 mo of GPS trajectory data and worked with the University of Maryland's Center for Advanced Transportation Technology lab to assess the data.(29) The data captured approximately 20 million trips. A visualization of the data was created that showed O-D and waypoints associated with each trip over time lapse, illustrating the network usage over time.(29) The trips were analyzed to detect trip duration and time between waypoints as well as summarized in O-D matrices showing travel at different geographic levels. The data also allowed for an analysis of trips along I-95 by time of day, day of week, and travel time as well as turning movements along the route. Researchers are working on tools to help agencies leverage the value of this type of data.

A drawback to GPS traces relative to cellular triangulation data can be selection bias. Not all vehicles are equipped with GPS devices, and some users may only turn on their devices when they need directions to an unknown location. Also, a large share of in-vehicle GPS devices are likely commercial fleet vehicles that have markedly different use patterns from the general population. On the other hand, this fleet-focused data may be valuable in studying freight movements or other populations not typically sampled in local transportation surveys. Compared with cellular triangulation data, GPS devices are traced both more frequently and precisely. As a result, GPS data can be used to construct routes and travel speeds in addition to a subpopulation O-D pattern.

Freight

Given the gaps of GPS data in capturing personal trips, some of the pioneering passive data research has aimed at understanding freight movements and commercial vehicle behavior.

ATRI is an analysis organization funded by a consortium of trucking companies and is thus a primary vendor of truck GPS data. ATRI places GPS receivers in a participating sample of trucks and resells the trace data to transportation planners; the resold data are precise, detailed, and deep in their representation of the trucking network.(30) Researchers have used ATRI GPS data to study truck parking, travel time reliability, and truck highway routing.(31–33) Users who purchase the GPS data generally develop their own tools to clean and process the data for their use case.

In addition, StreetLight Data™ offers a commercial product, which contains data from heavy vehicles and commercial fleets, primarily coming from INRIX™. StreetLight Data™ processes and resells these data as O-D matrices, including selected link matrices.(34)

GPS data have also been used to develop external truck trip models for a regional travel model and to study congestion at major ports.(35,36) In the former, a research team evaluated data obtained from AirSage™, StreetLight Data™, and ATRI to determine whether the data could provide a low-cost and effective replacement to external survey methods in Allen County, OH.[35] The O-D estimation included extracting trajectories from the device data, scaling that data to the population, and comparing the data to historical data for the same geography. The data were shown to approximate through traffic but did not perform as well with respect to replicating the internal–external and external–internal traffic flows. In addition, preliminary conclusions suggest that methods are needed to reconcile the scale of the data across different data sources.

In the second example (a study of congestion at major ports), GPS data were used to analyze trends and indicators for six major ports in the United States.[36] The research focused on creating metrics using GPS data. These metrics were then used to compare the activity at each port. Results of the study suggest that the creation of metrics from GPS data was a straightforward task and that comparisons between ports was useful and informative. The authors suggested that this approach could be applied to describe and compare other areas characterized by heavy truck activity.

There remain many freight-related questions that existing passive data products do not address. GPS data available through ATRI or StreetLight Data™ do not include any information on freight loads, including the commodity the vehicle is carrying or even if the vehicle is loaded or empty. Shipping companies often regard such data as proprietary, and small-sample surveys can suffer from high variance in shipment data.

Highway Networks

The precision of GPS data allows service providers to generate detailed routable networks with fine-resolution speed profiles. Given that commercial vehicles face the same networks and speeds as passenger vehicles, network products derived from GPS data may not be biased in the way that GPS-derived demand data may be. HERE™ sells routable networks as shapefiles that have been used in numerous products and projects as well as speed data by hour and day type. An important step in model development is to obtain a correct and complete highway network and determine free-flow and congested speeds for each link; the HERE™ data include this information for virtually all highway facilities in the United States.

A recent presentation included in the proceedings for a 2014 Transportation Research Board (TRB) conference illustrates the use of GPS to identify truck networks.[30] Examples included mapping average weekday afternoon peak period speeds on the interstate system for a specific time period and also the distribution of truck flows from Miami, FL, to destinations across the United States over a 7-d period. The data can also be used to illustrate before and after traffic flows along the network due to accidents and natural disasters such as a rock slide closing a portion of the Interstate system.

Limitations

As noted previously, two major limitations to GPS data include the potential for selection bias associated with which vehicles are equipped with in-vehicle GPS and also a “capture bias” in the sense that not all travelers use their GPS regularly.

WPS and LBS

Modern smartphones have multiple technology options to locate their positions, including cellular triangulation, GPS, and WPS. To develop WPS, device manufacturers (e.g., Google™, Apple™, and Microsoft™) and LBS providers operating within device applications (e.g., SafeGraph™, SkyHook™, and Cuebiq™) periodically ping each device available to them to check the devices' locations using GPS and the closest set of cell IDs that identify base transceiver stations in the cellular network. When pinged, phones send back the location data from these other technologies along with publicly broadcast Wi-Fi access points by their service set identifier and media access control data. These data collectively build up-to-date databases about where Wi-Fi networks are located, and location database vendors use the databases to predict device locations when GPS is not available.

Device applications heavily use WPS for better location accuracy. Several types of applications on smartphones require location information to properly function, including mapping and navigation, ride hailing and ride sharing, augmented reality, etc. The applications receive a device's location through GPS or WPS, depending on the device settings, and pass the location to the application's developer. Such LBSs are therefore an amalgamation of cellular triangulation, GPS, and WPS. Developers use these data to improve their products or monetize the data through ad service or reselling.

Companies currently selling this kind of data in the market today are location database vendors who provide LBS and application developers. Because there are many database vendors providing location services to app developers, the locations of individual users tend to be less persistent than those in cell tower triangulation, meaning individuals are observed for a smaller number of days. For example, two or three LBS providers might each contain pieces of the full location history of a single device in a day, whereas one cellular provider will observe all the location history of that same device, in theory. As such, the number of observations in LBS location data can be quite high, but depending on the question asked of the data, the number of "usable" records is usually much lower. In particular, it is difficult to understand schedule-based behavior such as tours throughout a day with WPS data coming from vendors.

However, device manufacturers have access to the devices' operating systems and have large market penetration. Accordingly, the WPS data, along with the GPS data they collect, could estimate population movements well. To date, none of the device manufacturers are releasing LBS data for external uses under any mechanism.[1]

The potential exists to use different types of social media data to model or examine parts of metropolitan travel. For example, Foursquare™ is a local search-and-discovery application that helps its users find retail establishments near them. Users "check in" at establishments, alerting their connections as to their whereabouts and giving establishments an idea of the types of people who use their services. An example of the use of social media data to estimate time-of-arrival at commercial establishments was performed by comparing Foursquare™ check-in data from Austin, TX, in 2010 with calibrated O-D and time-of-day data from the region's model.[(37)] The results showed promise in modeling time-of-day zonal arrival pattern estimation.

[1]Based on publicly available information, Sidewalk Labs™ does not have access to Google™/Android™ location data for their products.

Data from Twitter™'s API have been used to examine different patterns in travel behavior by many researchers, but they acknowledge the self-selection bias.(38,39) One study evaluated the usefulness of Twitter™ data in a statewide travel demand model estimation.(40) The authors compared geo-tagged tweets to data from a statewide household travel survey and found that the data showed similar trip lengths, spatial distributions, and differences in trip durations.

Transit Fare Collection Systems

Many transit agencies use electronic fare collection systems, which can be designed to collect route-level ridership and O-D data to help with service planning. Sometimes, card-based fare systems do not require riders to tap out of the system, but destinations can be modeled. Mobile ticketing systems, unlike card-based systems, can collect location data from devices using the application to collect information on travel behavior.

Travel behavior specific to transit is often collected with targeted efforts because it is difficult to observe enough transit trips in a random sample at the regional level. It is standard to conduct on-board transit surveys to estimate ridership and O-D flows of passengers. More recently, data collection and modeling efforts have focused on using data that agencies already gather with their fare collection systems and real-time tracking systems. The Massachusetts Bay Transportation Authority pieced together data from their fare collection system and automated counters to develop a ridership model that estimates when people are traveling, where they are getting on trains and buses, where they are transferring and, using the origin of their subsequent trip, inferring where they are ending their journeys.(41)

These kinds of data harvesting or mining could become easier with a data standard. Through a partnership between the Oregon Department of Transportation (ODOT) and Oregon State University, researchers have been developing the General Transit Feed Specification Ride (GTFS-ride).(42) It is an open, fixed-route transit ridership data standard that allows for improved ridership data collection, storing, sharing, reporting, and analysis. With GTFS-ride, mobile applications that take advantage of GTFS-ride real-time data, such as Citymapper™, Transit App™, and One Bus Away™, as well as applications that individual transit agencies publish, could aggregate the observed LBS data and report route-level O-D data in GTFS-ride format. As more agencies build out mobile ticketing systems similar to what the Metropolitan Atlanta Rapid Transit Authority is doing, GTFS-ride data could become more robust.(43)

SOCIOECONOMIC DATA

A second source of passive data that has the potential to support travel behavior studies is socioeconomic data, which focus on the traveler but can reveal details about the travel as well. Examples of this type of data include targeted marketing lists and social media data.

Targeted Marketing Lists: Businesses (Points of Interest) and Households

Commercial marketers who attempt to target likely audiences for their products purchase lists of consumers at the household and address level. These types of databases are compiled from many different sources (i.e., loyalty programs, customer lists, property tax records, purchasing behavior, product surveys, etc.). Parallel products exist for marketers who target businesses and commercial enterprises rather than households. Though the primary audience of these data products are marketers, they contain many pieces of data useful to transportation planners that come in a form similar to the American Community Survey (ACS) Public User Microdata Sample with address.

A benefit of targeted marketing records is their large sample size relative to traditional surveys. Targeted marketing firms strive to provide a detailed picture of the full population. Because the data are inexpensive, researchers can obtain sampling rates of 10 or 25 percent affordably.(44,45) The completeness of specific attributes varies, but important items, such as household income and presence of children, typically have high rates of availability. There is some concern about the degree to which targeted marketing data are accurate or correct; some addresses can take a while to update after households move, for instance, or individuals living in a household with different last names are classified as separate households.

Data from marketing lists can also be used as a microscopic population in disaggregate modeling. For example, InfoUSA™ firm listings were used to model how firms in Philadelphia, PA, react to local traffic congestion.(46) Additionally, household listings have been used to study home prices in Atlanta, GA.(47) Household lists can also be used as a sampling frame in generating synthetic populations.(48) Point-based data from marketing lists on households and businesses can be aggregated within arbitrary zones to create socioeconomic files for travel demand models. NCDOT purchased InfoUSA™ firm listings to provide the base year employment data for their statewide and MPO models.(49)

Related to targeted marketing data are data from industry observers. A number of commercial firms collect data on particular industries, which they then resell to industry analysts, investors, and researchers. Examples of such data that have been used in transportation planning include Edmunds™ (vehicle sales) and FW Dodge™ (construction). ODOT purchased construction data from FW Dodge™ to provide initial year floor space inventory and calibrate the land development module within its statewide integrated travel and land use model.(50)

Social Media

Many individuals participate in online social networks where they interact with other individuals and firms. The applications that enable individuals to participate in their networks collect data on interpersonal relationships, geospatial locations, and attitudes. In some cases, the applications offer LBSs for users and advertisers. The data available from social networks vary widely among networks based on the network's audience, capabilities, and privacy policies. Social networks are promising data sources for researchers because they provide a trove of digitized micro data that is large and accessible relative to custom surveys or experimental results. The sampling frame of some networks is large as well, with 79 percent of individuals operating a Facebook™ account in 2016.(51)

A major drawback to virtually all social network data is that they are highly self-selective; users select which services they use as well as the types of interactions they wish to have on the networks. Another potential source of bias is that users tend to engage in signaling; they post activities or travel that they feel will be interesting to their connections rather than a complete diary of their activities. Similarly, establishments that appear popular on social media may have become so through marketing rather than through genuine popularity, and locations that generate high volumes of real traffic may have minimal social media presences. Finally, some social media applications like Waze™ collect extremely relevant travel behavior data and may share incident data with transportation agencies but do not otherwise release their data for general research purposes.

Social networks provide a barely tapped reservoir of sociodemographic information for a large, if potentially biased, sample of the population. But beyond the individual-level data such networks provide, it is the data on interpersonal relationships that provide a promising avenue for new types of travel models. For example, in a study of undergraduate students' air travel, researchers asked respondents for permission to query their Facebook™ friends' lists.[52] The researchers were able to reconstruct the social relationships among the survey respondents and establish a link to their travel behavior.

HYBRID AND ADVANCED DATA APPLICATIONS

A central challenge in using type 2 passive data in travel behavior studies is that they are typically incomplete. Cellular O-D matrices provide information on flows between districts but not information on trip frequency, tour chaining, traveler characteristics, trip purposes, and other attributes typically acquired in household surveys used for modeling or planning. Targeted marketing lists contain detailed demographic information on specific households but almost nothing on their travel behaviors. Some characteristics can be inferred by matching probable home locations with U.S. Census geographies, but such aggregated information cannot match the behavioral specificity available from a household travel survey.

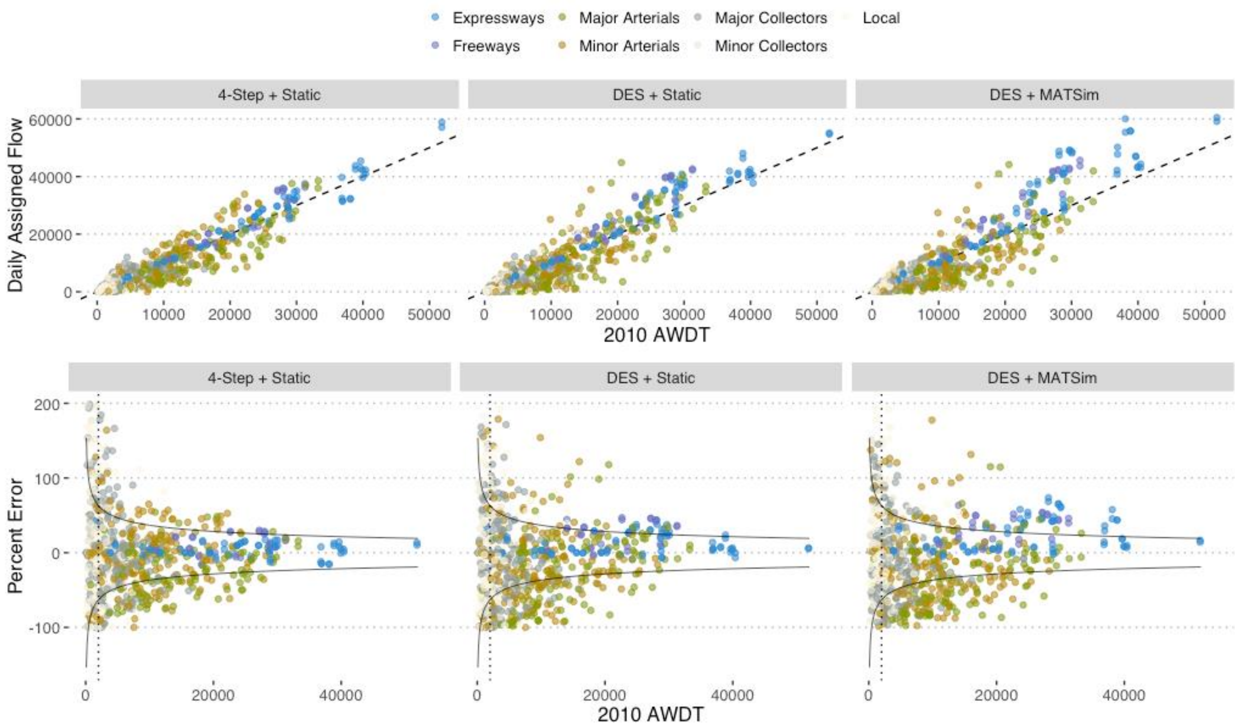
There are examples where researchers have used type 2 passive data to infer a great deal of information about travel behavior in the absence of a local household travel survey. They include novel applications of cellular phone data and GPS traces as well as attempts to statistically combine positional and socioeconomic passive data. Specific examples of constructing detailed travel diaries with trips and tours directly from cellular or GPS data include the following:

- A recent study in Maryland applied various elementary machine-learning techniques to raw GPS traces to look at a number of issues beyond simple O-D flows, including estimating weigh-in-motion avoidance and calculating travel time sheds.[29] In this case, researchers faced two primary obstacles common to GPS data: (1) they suspected that the trace data may over-represent commercial vehicles, and (2) they determined that ad-hoc heuristics are required to identify trips and activities precisely.

- A more comprehensive attempt occurred in the San Francisco, CA, Bay Area, where researchers partnered with a major cellular phone carrier and received raw call record data.[53] The researchers processed the data to create diaries of each phone's travels, as is commonly done to generate O-D matrices; however, in this case, they fed the data into a series of machine-learning models that generated trips and tours, classifying activities from land use data and the call data records themselves. The result was a highly predictive activity-based travel model that is sensitive to changes in population and land use inputs in the same manner as might have been created from econometric models estimated on survey data.

An important aspect of transportation represented in modern activity-based models concerns joint travel behavior, and many surveys accordingly request information on which household members participated in trips or activities. Passive data in the forms many people encounter do not explicitly inform this type of model; however, in a study in Spain, researchers obtained call detail records from a major telecoms firm.(54) This allowed them to both triangulate the devices' positions over time and also build a social network from calls between devices. The researchers observed that mobile devices that called each other tended to travel to the same destinations at the same times, an intuitive conclusion that nevertheless has profound implications for activity location and destination choice modeling. Classical understanding of travel behavior suggests that people base their discretionary activity locations primarily on destination attractiveness and travel costs, with other unobserved components relegated to random error; however, it may be that selecting common destinations with acquaintances is at least as predictive. Large-scale passive data can uncover this relationship in a way that is impossible with household surveys, as recruiting all members of all respondents' social networks is infeasible.

Previous studies show what is possible when researchers obtain access to a large amount of passive data directly from a service provider.(29,53,54) A similar attempt has been made to explore what may be possible with passive data products currently available under commercial contracts.(48,55) In this effort, researchers started with archetypal daily patterns, including tour departure times and activity durations, processed from the NHTS. They then fed these patterns into a discrete event simulation (DES), which probabilistically joined aggregate O-D matrices and targeted marketing data from commercial providers for a few cities. The result of this process was a synthetic population with synthetic travel diaries that matched the true population in at least two validation measures. The validation for this study was performed in three cities: Seattle, WA; Atlanta, GA; and Asheville, NC. In the first two cities, the synthetic diaries compared favorably with the results of recent household travel surveys on multiple dimensions, indicating that the synthetic diary was a plausible substitution of the household survey for at least some measures. In the third city (Asheville, NC), the trip diaries were assigned to a highway by both static user equilibrium assignment and microsimulation (using MATSim™) and the forecasted traffic volumes were compared to a recently calibrated four-step travel model for the region. The results of this validation exercise are given in figure 1.(48) The synthetic DES assignments are within acceptable error margins, and the assignment could be improved with elementary calibration techniques.



© Transport Foundry.

Figure 1. Graphs. Assignment validation by four-step model and synthetic DES.[48]

In fall 2017, FHWA launched an EAR effort with the University of Maryland, employing data fusion techniques to generate O-D tables at the national and metropolitan levels.[56] Three sources of mobile device data are being used: cellular phone, GPS, and smartphone application data. The research plan calls for the data products to be segregated by mode, purpose, time of day, socioeconomic, and demographic variables for both the 2016 base year and future year scenarios. The results from each data source are being generated at the national level (with flows between metropolitan statistical areas) and at the traffic analysis zone (TAZ) level within the Baltimore, MD, metropolitan area and will then be compared and validated.

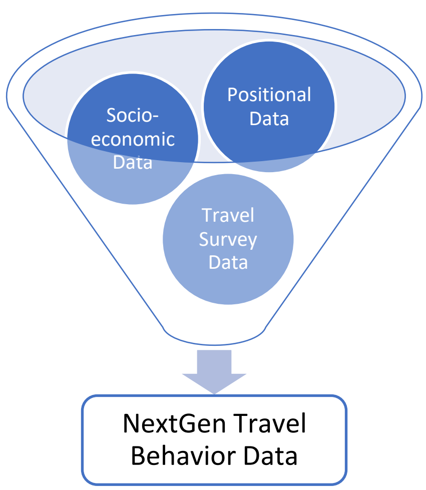
It is important to note that for each of these hybrid examples, the passive data were not simply passable substitutes for household surveys; rather, the use of passive data allowed researchers to make insights or predictions that would not be possible with a survey. This fact is clear in the case of using call data records to build a complete social network in addition to activity paths. In the other cases, the large quantities of local observations obviate the lengthy estimation and calibration periods in survey-based model development. Given this, it is not unreasonable to think that planning organizations could obtain new training data and therefore a new base year model each year. Further, as the training data are available throughout the country, the models can be more easily transferred between regions. This transferability can help disseminate best practices and lower development costs for agencies.

CHAPTER 4. OPPORTUNITIES AND IMPLICATIONS

With the increasing availability and access to passively collected location data, the travel behavior community is approaching a time of unparalleled opportunities for harnessing new data sources. Financially pressured agencies are exploring the viability of leveraging these new data sources and technologies to complement or even replace expensive travel survey data (especially in terms of O-D data), while academics and practitioners alike explore not only how the data can be used in travel behavior applications but also the implications of doing so. This project seeks to capture both the current exploratory efforts regarding these emerging data sources and relevant resources for agencies needing to make decisions about their own data programs.

The data themselves vary in terms of content and volume. Passive data contain large volumes of O-D data that can be mined to identify trips and processed to determine trip purpose, mode, length, and duration characteristics with varying levels of success; these data tell researchers what is taking place on the transportation facilities in a region. The socioeconomic data describe the residents at specific addresses, complementing census data that can also be summarized for specific geographies. Combined, these data inform researchers on who is traveling. What is missing from both data stories is the "why" of travel, those motivating factors and needs that result in the daily travel and underlying choice of mode, destination, and time of day of travel. The passive data are deep in terms of observations but shallow in terms of variables. The traditional survey data are strong in terms of variables but shallow in terms of sample size (i.e., number of observations). Data fusion efforts suggest that some combination of all three sources is possible, and research underway is focusing on how fusion can take place, when it is appropriate, and what mechanisms are needed to create next generation travel survey data products for transportation planning and policy applications, as illustrated in source: fhwa.

figure 2.



Source: FHWA.

Figure 2. Illustration. Potential framework for next generation travel behavior data.

This exploration of data fusion is motivated by increased survey costs resulting from declining response rates but also tempered by the unknowns associated with potentially migrating to the rapidly emerging passively collected data sources available on the market. While the new data sources themselves provide a much higher volume of O-D data, they are shallow in terms of the specific details needed to support the data-hungry travel demand models that are in wide use today. While the travel data are traditionally considered the primary source of details, the volume of passively collected data may lead to an inversion of the practice where the passive data are the primary source of details and the survey data are used to validate, calibrate, and fill in the gaps missing from the passive data (to provide depth to the shallower passive data).

In truth, the “perfect” travel dataset is quickly becoming an element of fantasy (if it was ever attainable in the first place). Data collection costs continue to increase while survey response rates continue to decline. Instead, passive datasets may be leveraged by transportation researchers and social scientists of all types, with other available products to construct the needed information and travel pattern summaries. In a way, researchers in the field regularly do this already when they join aggregate demographic information from U.S. Census geographies to their data in order to bring in additional variables. The challenge is to do this with disaggregate data that may not have convenient join fields. For example, cellular O-D data can be joined with household and firm marketing data to create synthetic travel diaries. A recent TRB Innovations Deserving Exploratory Analysis project did this with a mix of simulation and probabilistic linking, satisfactorily replicating many aspects of traditional household travel surveys.(55)

While passive data can provide a wealth of information on travel patterns, there are two main considerations inherent to this type of data. First, passively collected data are incredibly siloed. Each source focuses on a single mode, purpose, timeframe, or activity. While this is good because it allows researchers to characterize trends that relate to that topic, it also poses challenges in that it may be hard to relate the different sources and datasets. Second, passive data do not follow a predetermined sampling procedure, instead capturing broad snapshots of travel patterns and behavior. While it is known that these sources provide a large volume of data, it is unknown how representative the data are and what biases may exist in the data. For example, the data are largely collected using technology that include smartphones, credit cards, and social media that, while widespread, are not fully integrated into every American home.

Perhaps most important is understanding how passive data will influence how travel forecasts and decisionmaking may change in the future. The travel demand models used today to inform transportation planning are built on probabilistic sample surveys. Fundamental changes to travel behavior data collection will necessitate changes to travel demand models. Agencies have invested significantly in their travel demand models as well as the data programs to support those models, so identifying cost efficiencies are key to any type of migration of data and introduction of new data tools. Researchers have begun adapting analysis methods for passive datasets so that they can accommodate datasets larger than a computer can put into memory. (Traditional data processing methods in travel modeling tend to hold all data in memory.) These advances can inform new approaches for transportation planning.

This report provides an introduction to the abundance of passive data opportunities currently available in the United States; these sources are expected to continue to grow. Federal, State, and metropolitan transportation decisionmakers and planners may be able to leverage these passive data to meet increased data demands, provide more information across wider timeframes, and understand the mobility (and needs) of more travelers. However, there is still much to study and learn, including the following:

- How to best fuse the needed traveler characteristic data (such as income, trip purpose, specific stop locations along a tour, etc.) with the passively collected data.
- Whether it is possible to build models based on the passive data and validate/calibrate using survey data.
- How well the travel behavior data obtained through these new sources trend when compared to historic travel behavior data (e.g., U.S. Census, NHTS, and ACS) over time.
- What level of training and guidance is needed by agency staff obtaining, protecting, and applying passive data.

This project seeks to investigate answers through the sharing of ideas and experiences. Many of these topics are too complex for a single agency or individual to fully investigate, but a shared effort, such as a pooled fund approach, can provide a common platform from which agencies and individuals can explore and gain a better understanding of the new data approaches discussed in this report. Adoption of a pooled fund platform could lead to an organized and cost-effective approach to designing, collecting, purchasing, and using current and emerging multimodal travel behavior data. Where agencies and staff may find the introduction of new data approaches daunting given their historic investment in more traditional methods, joining a new data pooled fund effort could provide the tools and support necessary to make such a transition a feasible and even cost effective alternative.

APPENDIX A. TRAVEL BEHAVIOR DATA PRODUCTS FROM FEDERAL PROGRAMS

This appendix provides a summary of some of the national multimodal travel behavior datasets published by various Federal agencies. It is organized by Federal agency, with first an introduction and overview of the agency activities followed by a summary of specific travel behavior reports and datasets.

FHWA'S OFFICE OF HIGHWAY POLICY INFORMATION (HPPI)

The main purpose of FHWA's HPPI is to provide national data for surface transportation for use in policy and planning efforts at the national, State, and metropolitan levels. The primary multi-modal travel behavior data provided by HPPI include the following:

- **TAF (parts 1–3):** TAF is FHWA's first effort to estimate long-distance (i.e., trips greater than 100 mi) passenger movements.(17) It consists of a set of trip tables that provide information on person trip flows at the county-to-county (or equivalent-to-equivalent) level for base year 2008 and future year 2040 for different passenger modes (e.g., auto, bus, air, and rail). A trip in the trip table is defined as a one-way trip.

- **NHTS:** NHTS has been conducted every 5–8 yr since 1969 (specifically, in 1969, 1977, 1983, 1990, 1995, 2001, 2009, and 2017).(1) The recently completed 2017 NHTS collected data from April 2016 to April 2017. The datasets include demographic details about the participating households, their members, and their vehicles and details about a 24-h assigned travel period for household members ages 5 and older. The data are used to support policy research, travel demand modeling, and special topical studies in the areas of health, electric vehicles, ride sharing, and travel by various subpopulation groups such as baby boomers and millennials.

- ***Highway Statistics*:** The *Highway Statistics* series are annual reports that have been published since 1945.(57) The reports provide analyzed statistical information on motor fuel, motor vehicle registrations, drivers' licenses, highway user taxation, highway mileage, travel, and highway finance. Tables and charts are used to present these statistics. The data used to compile these reports are mostly from highway data submitted by each State to FHWA.

- **Highway Performance Monitoring System (HPMS):** HPMS is an information system and database for the Nation's highways.(58) It provides data on the extent, condition, performance, use, and operating characteristics of national highways. It has been modified several times since it was first developed in 1978 to reflect the changes of the highway systems, legislation, and national priorities as well as new technology. The modifications have also aided in consolidating or streamlining reporting requirements.

- **Motor fuel data:** Data regarding motor fuel is compiled by FHWA based on monthly reported data on the number of gallons of fuel taxed by each State.(59) The previous year's data are then used to attribute Federal revenue to the States.

FHWA'S EAR PROGRAM

FHWA's EAR Program was established by legislation to address research with the potential to transform transportation systems.[60] This program has funded various projects related to multimodal travel behavior research, two of which are described further in the following subsections.

Cellular Phone Data and Travel Behavior Research Symposium

FHWA's EAR Program, in coordination with HPPI, sponsored a symposium in 2014 on using cellular location data for national travel behavior research and how to combine cellular data with traditional survey methods.[27] The major issues discussed included the availability and application of cellular data, merging of cellular data with other data sources, and validation of cellular data. Participants were from both the public and private sectors, including data providers, researchers, and professionals. Several researchers from public agencies, private firms, and academic institutes discussed their experiences in using cell data.

"Design of a Completely New Approach for a National Household-Based Long-Distance Travel Survey Instrument" Project[16]

The EAR Program also funded a project between 2011 and 2013 to identify novel, innovative, and cost effective data collection alternatives to collect long-distance travel.[16] The project included a series of smaller studies to investigate alternative candidate approaches for future long-distance surveys. They included integrating existing survey data with the results of the new survey, designing a core probabilistic sample, performing data fusion and imputation research on combining probabilistic and non-probabilistic data, exploring post-processing methods used with advanced travel survey methods (e.g., GPS, smartphone, and social media) to impute trip information, and developing smartphone applications and Facebook™ surveys that combine survey data with passive location tracking to improve data quality and minimize bias.

FHWA'S OFFICE OF FREIGHT MANAGEMENT AND OPERATIONS (HOFM)

The purpose of FHWA's HOFM is to promote smooth and secure freight flows on the national transportation system and across national borders.[20] The main activities of HOFM include conducting research, developing analysis tools, and providing data for freight transportation. Further, HOFM provides training to transportation professionals, provides guidance and funding on freight programs, and conducts other freight-related activities to achieve its purpose. The primary travel behavior data sources used by HOFM include CFS and FAF, which are described further in the following subsections.

CFS

CFS is the primary data source on goods movement in the United States at the national and State levels.[19] It is a shipper-based survey that is conducted through a partnership between USDOT's BTS and the U.S. Department of Commerce's USCB. It has been conducted every 5 yr since 1993 and is the only publicly available goods movement data source for the highway mode.

FAF

FAF is produced by BTS and FHWA.[20] It is a comprehensive data source that provides information on freight movement of all transportation modes among States and major metropolitan areas. FAF is comprised of data from different data sources that represent multiple business sectors.

USDOT'S BTS

BTS is an independent statistical agency within USDOT that provides statistics on commercial aviation, multimodal freight, and transportation economics.[61] The main purpose of BTS is to promote intermodal transportation and provide direction to national transportation policymaking. The primary travel behavior data produced by BTS includes two products in particular (i.e., *National Transportation Statistics* (NTS) and the Airline Origin and Destination Survey (DB1B)) that focus on travel behavior in more detail.[62] These are described further in the following subsections.

NTS

NTS, which is published by BTS, is a large online database that contains over 260 tables that are updated annually (with on average of 50 tables being updated each quarter of the year).[62] NTS is a reproduction of data from a variety of data sources.

DB1B

Each quarter, the U.S. Office of Airline Information records a 10 percent sample of airline tickets from reporting carriers' records of the number of passengers, fare classes, coupon types, distances, and travel times for each air trip between every O-D in the airline network.[63] This includes detailed information on carriers, airports, and service (e.g., delays, capacity issues, and connections for different itineraries). Information can be used to model air traffic patterns, market shares, and passenger flows.

USDOT'S NATIONAL HIGHWAY TRAFFIC SAFETY ADMINISTRATION (NHTSA)

NHTSA is a safety and crash prevention agency responsible for developing and implementing a variety of educational, engineering, and enforcement programs to reduce the number of deaths, injuries, and economic costs due to highway vehicle crashes.[64] One such program is the 2012 National Survey of Bicyclist and Pedestrian Attitudes and Behaviors.[65] The purpose of this survey was to promote safety of bicyclists and pedestrians. Researchers interviewed 7,509 respondents who were 16 yr or older from a national representative sample. The survey provided information on overall behavior, trip characteristics, habits, satisfaction, and safety of bicyclists and pedestrians.

USDOT'S FTA

FTA's main responsibility is to provide technical and financial support to local transit systems. NTD is the primary source of multi-model travel behavior data and statistics provided by FTA.[10] To meet legislative requirements, around 850 transit agencies in UZAs receive FTA funds and therefore submit data to NTD. NTD performance data are used to apportion FTA funds to transit agencies in UZAs. The information provided by NTD include data on transit profiles, national transit summaries and trends, time-series data on transit systems, monthly ridership, and safety.

USDOT'S FEDERAL RAILROAD ADMINISTRATION (FRA)

FRA's mission is to enable safe, efficient, and reliable railroad transportation.(66) The primary data provided by FRA to promote railroad safety is the safety analysis data, which are available to the public through FRA's Office of Safety Analysis website.(66) The statistical information provided on the website include railroad inventory data, accident and incidents, and highway–rail crossing information.

U.S. DEPARTMENT OF ENERGY (DOE)

The primary data provided by DOE that are related to multimodal travel behavior include the following:

- Alternative fuels and advanced vehicle data.(67)
- *Monthly Energy Review* (Table 1.8 summarizes motor vehicle mileage, fuel consumption and fuel economy by vehicle type).(68)

USCB

USCB is a principle Federal statistical agency that serves as the leading data source for the U.S. people and economy. The primary responsibility of USCB is to conduct the U.S. Census. Further, USCB also conducts the economic census and many other surveys or programs. The primary data from USCB that are useful for multimodal travel behavior studies include the following:

- ACS is an ongoing survey conducted by USCB.(69) It collects detailed information about people and the workforce in the United States. The subjects included in the survey are social, housing, economic, and demographic. Travel behavior information is collected under the "Journey to Work" section of the survey.(70)
- Census Transportation Planning Products (CTPP) is a set of special tabulations created from large sample survey data collected by USCB.(71) Currently, the survey used for CTPP is the ACS.(69) There are three parts in the CTPP tabulations: parts 1 and 2 provide information on the residence-based household characteristics and workplace-based worker characteristics, respectively, and part 3 provides information on travel flow between workplace and home.

U.S. DEPARTMENT OF LABOR'S BUREAU OF LABOR STATISTICS (BLS)

BLS is an independent Federal statistical agency. Its primary responsibilities focus on the labor market and measure activity, working conditions, and related economic indicators. BLS collects, analyzes, and publishes economic information to serve its users.(72) BLS's American Time Use Survey (ATUS) is one of the data sources that can be used for travel behavior research.(73)

Although not widely used, ATUS is an alternative source of travel details that can be used to complement and supplement traditional travel behavior studies. It is described in further detail in the following subsection.

ATUS

ATUS is a continuous survey used to collect information on how survey respondents spend their time on various activities the day before their interview day.(73) A random sample of individuals ages 15 and older is selected from the sample households, which are chosen from BLS's Current Population Survey. Information on the amount of time spent on travel activities is collected in the survey along with other activities, such as working, socializing, etc.

NCHRP

NCHRP, which is sponsored by FHWA and State transportation departments, is a transportation research forum administered by TRB.(74) One recently completed NCHRP project (project 08-95) addresses issues important to the question of the future of travel behavior data.(28) Starting in 2014, NCHRP sponsored a research project on cellular phone location data for travel behavior. The objectives of this research were to explore the accuracy of using cellular phone data for travel behavior analysis and to provide guidelines on using these data for such analysis. The evaluation was conducted by interviewing transportation agencies that have experience in purchasing and using cellular phone location data for travel behavior study. The main outcome of the study is guidelines that address the issues under the topics of cell data acquisition, data quality, and data applications.

APPENDIX B. RECENT PROJECTS USING POSITIONAL DATA

This section summarizes recent projects using positional data, as presented in Table 1 in chapter 3 in this report (i.e., the first category of type 2 passive data defined in chapter 1). Positional data provide records that primarily contain the spatial location of observations. Based on data collection technologies, there are five main types of positional data as follows:

- Cell tower triangulation.
- GPS (vehicle).
- GPS (mobile).
- WPS/LBS.
- Other.

The summary of recent projects using positional data are obtained from an online literature search and organized by primary type of positional data. Recent studies published or presented at the following transportation conferences were reviewed:

- TRB Tools of the Trade conferences (see http://www.trbtoolsofthetrade.org/).
- TRB 2019 Annual Meeting (see http://amonline.trb.org/?qr=1).
- August 2018 NHTS workshop (see http://onlinepubs.trb.org/onlinepubs/Conferences/2018/NHTS/Program.pdf).
- TRB Innovations in Travel Modeling conferences (see http://itmconference.com/).
- TRB Planning Application conferences (see https://www.trbappcon.org/).

CELL TOWER TRIANGULATION

Project 1: 21st Century Transportation Planning in Lake Tahoe Using Cellular Mobility Data Analytics[75]

- **Data provider:** Unknown.
- **Sponsor:** Stantec.
- **Description:** Stantec used big data from wireless devices to identify travel patterns of various groups of people living, working, and visiting Lake Tahoe. This information was vertically integrated into ArcGIS10.4 with attributes that included land use, transit stops, multimodal infrastructure locations, public parking, traffic counts, crashes/accident locations, social demographics, U.S. Census Longitudinal Employer-Household Dynamics (LEHD) data, ownership, recreation sites, and popular entertainment

destinations to analyze, synthesize, and visualize the findings and recommendations. Infographics were created that resonated with the stakeholders, elected officials, and the public. In summary, the researchers presented the magnitude of the problem and provided a detailed plan with costs, infrastructure requirements, and approximately how many auto trips would be removed with each shift in mode share.

Project 2: "Data-Driven Prediction System of Dynamic People-Flow in Large Urban Network Using Cellular Probe Data"[(76)]

- Data provider: Unknown.
- **Sponsor:** Ford Motor Company.
- **Description:** Cellular probe data, which are collected by cellular network operators, have emerged as a critical data source for human-trace inference in large-scale urban areas. However, because cellular probe data of individual mobile phone users are temporally and spatially sparse (unlike GPS data), few studies predicted people flow using cellular probe data in real time. In addition, it is hard to validate the prediction method at a large scale. This paper proposed a data-driven method for dynamic people-flow prediction, which contains four models. The first model is a cellular probe data preprocessing module, which removes the inaccurate and duplicated records of cellular data. The second module is a grid-based data transformation and data integration module, which is proposed to integrate multiple data sources, including transportation network data, point-of-interest data, and people movement inferred from real-time cellular probe data. The third module is a trip-chain–based human daily trajectory generation module, which provides the base dataset for data-driven model validation. The fourth module is for dynamic people-flow prediction, which is developed based on an online inferring machine-learning model (random forest). The feasibility of dynamic people-flow prediction using real-time cellular probe data is investigated. The experimental results show that the proposed people-flow prediction system could provide prediction precision of 76.8 and 70 percent for outbound and inbound people, respectively. This is much higher than the single-feature model, which provides prediction precision around 50 percent.

Project 3: "Trip-Chain-Based Travel-Mode-Shares-Driven Framework Using Cellular Signaling Data and Web-Based Mapping Service Data"[(77)]

- **Data provider:** Unknown.
- **Sponsor:** Ford Motor Company.
- **Description:** The signaling data of cellular phones, as a passively generated, real-time, wide-coverage, low-cost data source, have been widely used in recent studies to understand human activity and model urban travel demand. However, in contrast with the GPS data, cellular phone signaling data are sparsely distributed in time and space, which makes travel mode inference a challenge. Recent studies presented methods of deriving users' home and work locations, O-D trips, and other activities. Very few provided a

complete and feasible framework for travel mode derivation with effective validation methods. This paper provides a real time travel mode derivation framework using signaling data and a web-based mapping service. A trip chain model is proposed to detect individual activity patterns and derive the trips of mobile phone users. Then, the travel mode of each trip is identified by a Fuzzy K-Means model, which is trained and validated by the point-to-point travel time from a web-based mapping service. Finally, the travel mode shares are aggregated and scaled to the whole population of the study area. The framework is demonstrated using cellular signaling data from 1.9 million users in Shanghai, China, for 7 days and citywide point-to-point travel times from a web-based mapping service for 3 of those 7 days. Comparing the modeled travel mode shares with travel survey data and transportation hub statistics demonstrates the plausibility and efficiency of using a large data source (mobile trace data and web-based mapping) to accurately assess the travel modes of people in a big city using the proposed framework.

Project 4: Utilization of Mobile Device Data for Model Validation and Development(78)

- **Data provider:** AirSage™.
- **Sponsor:** AirSage™.
- **Description:** Combining their patented data collection and WiSE analysis technology engine, AirSage™ extracts actionable information from the geo-location of mobile devices, leveraging the largest base of high-quality location data in the transportation industry. This presentation presents several examples on the application of geo-location data in trip matrix (O-D) development and validation.

Project 5: Modeling Automated Vehicles with a Passive Data Model(79)

- **Data provider:** AirSage™.
- **Sponsor:** Transport Foundry.
- **Description:** The current Government and large industry leaders are investing in policies and technologies to make AVs a reality. In August 2016, Uber announced that it would make self-driving cars available for hailing in the next few months. Even though AVs seem to be right around the corner, few travel models are built to study the effects of AVs, leaving planners and policy makers unsure how to prepare. This presentation shares a study of AVs in the Asheville, NC, region using a pattern-based demand model built from passive large-scale data and MATSim™. This setup provides a microscopic framework from which to analyze short-term responses to AVs assuming shared fleets, privately owned fleets, or a mix of the two. Although AVs present an array of questions—notably the interaction of AVs with transit and the changes to trip-making behavior—this application focuses on measuring the sensitivity of total vehicle miles traveled (VMT) and average commute time to differing assumptions of AV adoption and use.

Project 6: Empirical Analysis of External Travel and State Highway 130 in Austin, Texas Using Cellular Data[(80)]

- **Data provider:** Unknown.

- **Sponsor:** Texas A&M Transportation Institute.

- **Description:** Assigning trip routes to polygon-based cellular O-D data is a challenge. A select link analysis, based on shortest distance or time path, can be used to assign these trips to roadways (a form of map matching). Whether it be external-to-external (E-E) trips passing through a study area, external-to-internal (E-I) trips entering the area, or internal-to-external (I-E) trips leaving the area, different roadway types, such as toll roads, can have varying effects on the different types of external travel.

  This presentation describes the process of mapping the most likely route for polygon-based cellular O-D data under multiple scenarios to investigate the impact of external travel on the toll road State Highway 130 in Austin, TX.

  The process of mapping the route for polygon-based O-D pairs centers on the shortest time path between O-D pairs. The network was assembled as a combination of the Austin travel demand model and the Texas Statewide Analysis Model network. Using TRANSCAD (software used in developing travel demand models), a select link analysis based on shortest time path was used to obtain the O-Ds of traffic-crossing–specified roadways (i.e., selected links). The initial select link analysis included travel times based on speed limit inputs from the original model. From this analysis, it was determined that Interstate Highway 35 and State Highway 130 were two of the primary locations of external travel through the study area.

GPS (VEHICLE)

Project 1: Revealing Freight Vehicle Tours and Tour Patterns from GPS Vehicle Tracking and Driver Survey Data[(81)]

- **Data provider:** Unknown.

- **Sponsor:** Massachusetts Institute of Technology.

- **Description:** This study attempted to fill the research gap identified regarding post-processing methods specific for freight GPS data (except for stop detection). Two GPS vehicle tracking and driver surveys were conducted in Singapore and Boston, MA, for this purpose. The research findings were as follows:

  - Tour identification algorithm led to sensible results on several tour-level indicators.

  - There were no major differences between labeling stop chains from an individual tour or whole day perspective.

- Whole day perspective allowed for further insight into operations: non-negligible percentage of vehicle/days (one-half Singapore and one-third Boston) demonstrated variable pickup locations, challenging assumptions of pickup location(s) as base.
- Algorithm development must incorporate robustness to several data issues (e.g., wrong/nonsensical validation, missing stops, etc.).

Project 2: Practical Application of Archived Probe Vehicle Data[82]

- **Data provider:** HERE™.
- **Sponsor:** Oregon DOT.
- **Description:** Third-party information vendors using big data are expanding rapidly in number and products. It is very challenging to know which product best meets project needs affordably. Access to independent evaluation and use of products is hard to come by, yet the transportation planning industry is beginning to use these products.

  FHWA purchased the National Performance Management Research Data Set (NPMRDS) to support *Fixing America's Surface Transportation Act* reporting requirements, providing this dataset to States and MPOs at no charge. AASHTO organized a pooled fund program to provide analytics using the NPMRDS data for any organization interested in joining, but there is a cost. In order to gain more detailed data and direct access to the analytics, ODOT purchased HERE™ probe-vehicle data along with the IPeMS^{R} web analytics tool.

  This presentation illustrates several examples of how ODOT has applied probe-vehicle data in the planning world, showcasing input data development, data analysis and visualization of results. All three examples can be done using HERE™ or NPMRDS data provided by FHWA.

Project 3: I-295 Truck Corridor Forecasts Development: Richmond, VA[83]

- **Data provider:** StreetLight Data™.
- **Sponsor:** Richmond Regional Transportation Planning Organization.
- **Description:** This study explored the use of Streetlight Data™ O-D data for subarea corridor traffic forecasting and performed truck and auto forecasts on the I-295 corridor. Lessons learned from this study include the following:
  - Big data can be used to develop O-D seed matrices for corridor studies.
  - StreetLight Data™ provides promising O-D distribution for corridor studies.
  - Truck percentages should not be obtained from StreetLight Data™ (auto and truck indices should be obtained using separate processes).

- The procedure illustrated in the presentation successfully adjusted for any errors in base year validation.
- One limitation of this procedure is that it did not address the model's future growth uncertainties.

***Project 4: Combining NHTS and Passive OD Data for Charleston, SC*[84]**

- **Data provider:** AirSage™ and ATRI.
- **Sponsor:** Resource Systems Group.
- **Description:** AirSage™ pioneered the transformation of wireless network signaling data into powerful location intelligence information. Early research led to unique developments, patents, and methodologies. This presentation highlights the project experience and findings of Charleston, SC, using passive data from AirSage™ 870 × 870 matrices and by residents and visitors and ATRI data for over 37,000 trucks, over 150,000 truck trips, and 30 days of data for O-D estimation.

***Project 5: TNCs Today: A. Profile on San Francisco Transportation Network Company Activity*[85]**

- **Data provider:** Lyft and Uber APIs.
- **Sponsor:** Northeastern University and San Francisco County Transportation Authority (SFCTA).
- **Description:** This study demonstrates how SFCTA partnered with Northeastern University and used collected GPS data from Lyft/Uber APIs to develop a profile of San Francisco transportation network company (TNC) activity. The research addressed the following questions regarding TNCs:
    - Number of vehicles.
    - Number of trips.
    - When and where.
    - VMT.
    - Geographic coverage.

The limitations of the research were as follows:

- Intra-San Francisco only.
- Trip details are imputed.
- No information is available about travelers, chosen product, or vehicle occupancy.

Project 6: A Suite of Model Updating and Validation Procedures Using Third Party Origin-Destination Data[86]

- **Data provider:** StreetLight Data™ and INRIX™.
- **Sponsor:** Ohio Department of Transportation.
- **Description:** The presentation discussed Ohio Department of Transportation's model updating and validation procedures using INRIX™/StreetLight Data™ passive data. It developed a set of procedures to extract and analyze these third-party data. The department also updated model validation procedures to incorporate the data.

Project 7: Expanding the Uses of Truck GPS Data in Transportation Planning and Analysis[87]

- **Data provider:** ATRI and StreetLight Data™
- **Sponsor:** MAG.
- **Description:** Big data sources that became available in the past few years, especially truck GPS data from commercial data sources, have formed strong foundations for the freight modeling and regional truck movement forecasts. In the development and maintenance of truck models, the most common uses of truck GPS data are the extraction of trip matrices and estimation of tour models. As GPS data contain a large amount of spatial–temporal information, to get the most out of it, additional data mining and visual analytics applications are needed to transform data into meaningful insights about truck activities.

  As one of Second Strategic Highway Research Program (SHRP2) C20 Freight Model Grant recipients, MAG acquired GPS data for light, medium, and heavy-duty trucks from third-party data providers (i.e., ATRI and StreetLight Data™) to facilitate the development of tour-based behavioral truck models. Beyond the standard trip matrices extraction and tour model estimation, MAG analyzed the GPS data in various ways to benefit transportation planning, freight study, and traffic study, including validation between GPS sample and vehicle classification counts, truck tour characteristic analysis, truck trajectories reconstruction on the network using an innovative map matching algorithm, and visualization of the truck movement and route choice patterns.

  This presentation mainly focuses on (1) an overview of sample data characteristics, (2) a brief discussion of GPS data processing, and (3) showcase of truck GPS data analysis and visual analytics applications for transportation planning and analysis.

Project 8: Tour-Based Truck Travel Models Using Truck GPS Data[88]

- **Data provider:** StreetLight Data™ and ATRI.
- **Sponsor:** MAG.
- **Description:** Part of the SHRP2 C20 Freight Model Grant, which MAG successfully secured, calls for the development of a next generation freight demand model that involves synthesizing firms, linking suppliers with buyers, creating supply chains, estimating truck tours by industry sector, and integrating it with regional activity-based model (ABM). In order to develop truck tour-based models, MAG acquired and processed truck GPS data from two vendors: StreetLight Data™ and ATRI. The processed data from StreetLight Data™ yielded a database of 266,832 tours and 1,216,754 trips from over 17,000 single-unit trucks. On the other hand, the processed ATRI data resulted in about 81,090 tours from 39,080 combination-unit or heavy trucks. These two truck tour databases formed a strong foundation to estimate robust tour-based models for various industry sectors for three truck types (i.e., light, medium, and heavy).

  The objective of the truck tour model is to develop truck trip chains by industry sector and truck type. These truck trip chains are then grouped into the major linkages based on land uses the trucks make stops at and the probability of making another stop based on the number of previous stops. The tour-based model generates the number of stops by industry sector, number of stops on a tour, stop purposes, and the location and time of day of stops.

  All the tour model components were coded in R, and each component was individually assessed and calibrated. The reasonability of explanatory variables was determined by their magnitude, *t*-statistic, and their relation to the dependent variable. The individual model outputs were also compared against the truck GPS data to assess the model performance. These comparisons indicated that the model components are predicting very closely to the observed data.

  This presentation focuses on two aspects: (1) processing of truck GPS data from two different sources, and (2) developing tour-based models by truck type. This presentation also discusses the calibration and validation of these models.

Project 9: Identifying Truck Stop and Service Station Locations to Improve Passively Collected Commercial Truck GPS Data[89]

- **Data provider:** Unknown.
- **Sponsor:** Resource Systems Group.
- **Description:** Passively collected data can greatly enhance traditional data sources used as inputs to travel demand models. However, passive datasets often lack sufficient context to fully characterize trips without integrating additional data sources. This project demonstrates a new method to identify the spatial location of truck stops and service

stations where long- and short-distance commercial trucks may dwell between trip O-Ds that constitute work-related activity locations.

Owners of underground storage tanks (USTs) are required to register with a State agency, and these agencies are required to maintain a registry of facilities pursuant to Federal regulations. While UST registries vary State to State, they commonly contain information regarding the name, owner, and address of the facility as well as the number, size, contents, and operational status of USTs contained within each facility. A novel method was developed to identity the location of truck stops using State-level UST administrative records. First, State UST records were filtered to identify facilities that have USTs containing diesel a cumulative volume of at least 10,000 gal and have USTs containing gasoline with a cumulative volume of at least 10,000 gal (to filter out depots/maintenance facilities). Text analytics were then performed on the facility name and owner fields to refine the selection. Facility addresses were then geocoded, mapped, and verified using satellite imagery. Geofences were then drawn around each identified facility, including attached parking lots. These truck stop/service station geofences were then used to flag GPS pings, which occurred while commercial vehicles were stationary at truck stops, thereby identifying situations in which commercial vehicles made intermediate stops and/or diverted to refuel, rest, or comply with hours-of-service regulations at truck stops. Identified intermediate stops/diversions can then be accounted for when deriving O-D matrices from passively collected data.

In a case study application in Nebraska, 2,258 unique facilities were identified from the State UST registry. These facilities were filtered to a final list of 272 potential truck stops/services stations, and 108 of these locations were confirmed with satellite imagery. The resulting geofences contained 734,660 of 5,921,709 (12.4 percent) total daily pings from a passive commercial vehicle GPS dataset and, when combined with criteria to determine when commercial vehicles were stopped, identified at least one intermediate stop for 18.8 percent of the commercial vehicles present in the passive data set for that day.

Project 10: Validating Trip Distribution in Southeast Michigan Using GPS Data[90]

- **Data provider:** StreetLight Data™.
- **Sponsor:** Southeast Michigan Council of Governments (SEMCOG).
- **Description:** It is critically important that a travel demand model accurately predicts how trips are distributed throughout a region. Traditionally, datasets such as household travel surveys, CTPP, and LEHD data have been used to support calibration of trip distribution models. More recently, big data obtained from cell phone and GPS devices have been used to supplement information about trip distribution patterns. This presentation explores the use of StreetLight Data™ to improve the regional trip distribution model for SEMCOG. This data source provides information for both personal and commercial vehicles, allowing analysis to be conducted separately for each type of travel.

For personal travel, StreetLight Data™ was compared to the CTPP worker flow data and two household travel surveys conducted in 2005 and 2015. The household travel surveys were weighted to account for variability in sampling rates for different demographic groups, while the StreetLight Data™ was not weighted. Evaluation of the unweighted data suggested that demographic characteristics affect sampling rates in the StreetLight Data™. The presentation shows an evaluation of various expansion approaches for their ability to reduce sampling bias while preserving rather than obfuscating trip patterns present in the original data. Opportunities and limitations encountered using StreetLight Data™ as a supplement to traditional data sources are then be identified. In particular, the StreetLight Data™ provides information about external trips that are difficult to observe as well as information on difficult to reach traveler segments, including visitors to the region, by providing data on trips starting at the airport and at external stations.

StreetLight Data™ also serves as a source of commercial trip data. The StreetLight Data™ commercial vehicle data were more difficult to evaluate than personal travel data due to a lack of other sources of observed data in the SEMCOG region. For this analysis, commercial vehicle trip patterns in the StreetLight Data™ were compared to trip patterns obtained through survey efforts at the Canada–United States border conducted by Transport Canada. This comparison led to adjustments to data processing and analysis methodology.

Project 11: Opportunities for Regional DTA Validation with Mobile Data(91)

- **Data provider:** StreetLight Data™.
- **Sponsor:** Caliper.
- **Description:** This report discusses using mobile data in validation of regional and wide-area dynamic traffic assignment (DTA) models. It first examines the state of practice in model calibration and validation. Then, a MAG bottleneck study is used as an example to validate vehicle trajectory data using the Time-Lapse Aerial Photography Survey. Finally, the report summarizes how mobile data can assist with DTA validation.

Project 12: Approaches to Evaluating VMT using GPS-based Probe Vehicle Data(92)

- **Data provider:** StreetLight Data™.
- **Sponsor:** StreetLight Data™.
- **Description:** Traffic flow, often measured as level of service (LOS), has traditionally been used as a primary indicator of performance and basis for project evaluation but has significant recognized limitations. VMT reduction has been recognized as an important complementary measure for evaluating improvements, as measures that reduce VMT can provide more permanent congestion relief, reduce greenhouse and criteria gas emissions, and decrease wear and tear on roadways. VMT is typically estimated using methods, such as petroleum use and surveys that are too general, and that no longer reflect today’s changing vehicle technology mix. In contrast, real-world GPS data enable planners to

estimate the total volume and length of trips in virtually any size geography—from small municipalities to States and entire regions. To use VMT as a guide for policy and infrastructure decisions, more nuance and precision is required. Planners must be able to assess the causes behind changes in VMT and to see how patterns change by geography, income, urban form, weather and more.

This presentation explores how VMT has become an increasingly viable, useful metric for transportation project prioritization. It focuses on the ways that big data from GPS-enabled mobile devices and connected cars can be used to evaluate VMT accurately and efficiently. It uses real-world example case studies to illustrate the types of questions that VMT evaluations based on big data can answer for planners.

By analyzing VMT in three different types of zones (a TAZ, a region, and a State) using both traditional methods and big data-based methods, the presentation demonstrates how planners can use GPS data to truly understand VMT, including assessing the causes behind changes in VMT. To provide transportation planners with a roadmap for using VMT calculations to evaluate projects in their own communities, the presentation evaluates sample projects for these regions using a big data-based VMT estimate and a traditional VMT estimate and then contrasts the results. It also explores changes in regional VMT over time; the contributions of particular areas, land use, and development types to regional VMT; the contributions of short personal and commercial trips to VMT, and the respective impact of internal trips and pass-through trips on VMT.

Project 13: Analysis of GPS Fleet Tracking Data to Infer Commercial Vehicle Travel Patterns in Ontario[93]

- **Data provider:** Unknown.
- **Sponsor:** Ministry of Transportation of Ontario (MTO).
- **Description:** WSP | Parsons Brinckerhoff is preparing two large-scale multimodal transportation forecasting models for MTO, both of which include tour-based models of commercial truck travel. Currently, estimation and calibration of disaggregate commercial vehicle travel models are hindered due to a lack of high-quality data describing truck travel behavior. GPS fleet-tracking data are becoming more widely available and show large promise for this application, as they provide travel information over extended time periods with highly accurate spatial and temporal resolution. The primary difficulty using this data source is that the only available data are the position and time history of the observed vehicles.

  This paper presents a novel GPS processing to convert the raw GPS data into a travel diary of trips, stops, and tours, with an emphasis on observing the behavior of urban truck travel. Particular emphasis is placed on a multi-step stop identification procedure and on a novel depot identification algorithm that uses clustering techniques to identify locations with many long stops. Depot identification is particularly important, as it identifies home locations, which must be accurately identified as they are key to tour formation.

Finally, this paper presents an overview of the observed results, including summary distributions of stop dwell times, trip and tour distances and durations, number of trips per tour, tour departure times, and the relationship between tour departure time and duration.

Project 14: Using Passive GPS Data to Understand Truck Flows in Nebraska[94]

- **Data provider:** ATRI.

- **Sponsor:** Resource Systems Group.

- **Description:** ATRI now collects over 2 billion truck positions in North America every day, including medium duty single-unit trucks as well as semi-tractor trailers. Over the past 8 yr, since the original incorporation of ATRI's truck GPS data in the Indiana statewide model, many States have followed suit. Nebraska is among the latest States to harness the tremendous power of these data to both understand truck flows in their State and incorporate them in their travel model.

  Given the size of ATRI's database, it was helpful to draw a sub-sample of their data. The sub-sample for Nebraska was drawn from 8 weeks in 2017 spread over all four quarters and contained nearly 300 million truck position records. This yielded data on over 260,000 individual trucks making over 1 million trips. The data were carefully processed to filter out anomalies in the GPS data, such as loss of signal, as well as to distinguish intermediate stops at rest stops, truck stops, and similar locations from true truck O-Ds as they would be represented in FAF flows and for most modeling purposes. The data were then expanded to truck counts using a three-step procedure. First, iterative screenline fitting was used to factor the data. Second, parametric expansion factors as a function of trip distance were estimated, and finally, O-D matrix estimation from counts was used with careful constraints on how much it could adjust the matrix. The resulting dataset, expanded to represent all truck trips in Nebraska correcting for the systematic duration bias present in all passive data, was analyzed both to produce visualizations and summaries for understanding truck flows and to estimate parameters for the truck component of the Nebraska statewide model.

GPS (MOBILE)

Project 1: A Multi-Resolution Approach in Investigating the Impacts of Pre-Planned Road Capacity Reduction Based on Smartphone Trajectory Data: A Case Study of Lane Closure Event on Mopac Expressway, Austin TX[95]

- **Data provider:** Metropia.

- **Sponsor:** Metropia.

- **Description:** Pre-planned events, such as constructions or special events, lead to road capacity reductions and create bottlenecks in the traffic network. The traffic impact of such events goes beyond local, as informed drivers may detour to alternative corridors for faster travel speed, and, subsequently, the traffic congestion may propagate to the entire region. Traditional traffic impact analyses are typically based on simulation models, fixed-location sensor data, or survey data, which have various shortcomings respectively. In this research, the researchers propose the use of real trajectory data collected via smartphone GPS module that are capable of keeping track of individual drivers' behavior change before and after road capacity reduction, combined with system-wide dynamic traffic condition and roadway geometry network to investigate the impacts of pre-planned events in a multi-resolution manner. First, the traffic impact of such events at network level is analyzed, indicating how traffic may propagate to alternative corridors from the system perspective. Second, at the individual driver level, behavior changes and corresponding outcomes are examined by the comparison of before and after travel behavior. Finally, regression models are used to explain drivers' detour behavior choice with spatial and temporal features of interest. A case study based on a lane closure event on Mopac expressway in Austin, TX, is used as an example in this research, which shows a local freeway capacity reduction has a significant impact on other freeways in Austin and that drivers' detour behavior exhibits three major patterns and highly depends on spatial features such as trip length, distance to freeway entrance and to other alternative freeways, in addition to the time of the day this trip happens.

Project 2: Visualization of Truck GPS Origin-Destination[96]

- **Data provider:** Unknown.

- **Sponsor:** Houston-Galveston Area Council (H-GAC).

- **Description:** H-GAC is currently developing a tour-based truck model using observed GPS data purchased from a third-party vendor. The truck GPS data are used for model estimation and validation. There are no other data to verify the accuracy of the purchased truck GPS data. H-GAC visualizes the GPS O-D in Geographic Information System (GIS) as a mean of high-level reasonable check.

  H-GAC purchased observed truck GPS data from a third-party vendor. The truck GPS data are separated into heavy, medium, and light trucks. H-GAC visualizes the truck O-D locations on GIS. The O-D are aggregated into a TAZ level. The following checks are performed:

  1. Comparison of the truck O-D to the employment locations.

  2. Comparison of the truck O-D to the general perception of truck activity.

  3. Investigation of selected individual TAZs where numbers of trucks O-D could not be explained by number of employments

  4. Investigation into whether areas of special truck interests, such as cargo seaports, are producing and attracting a lot of truck trips relatively to the rest of the region.

  The visualization suggests that the observed truck GPS O-D data conform to the commonly known truck activity regions. However, there are TAZs where the level of observed truck O-D could not be well explained by known employment. These inconsistency between the observed truck O-D and known employment distribution could suggest issues on GPS data sampling, accuracy of known employment location, and potential uncaptured local factors.

Project 3: "Use of a Smartphone GPS Application for Recurrent Travel Behavior Data Collection"[97]

- **Data provider:** rMove™.
- **Sponsor:** Metropolitan Council.
- **Description:** The Metropolitan Council's Travel Behavior Inventory (TBI) has been conducted approximately every 10 yr since 1949 in the greater Twin Cities region to collect household travel survey data for the regional travel demand model and regional planning purposes. In 2018, the council transitioned to a recurrent data collection program using a smartphone-based GPS application, rMove™, as the primary means of data collection to obtain more current, accurate, and detailed spatial, temporal, and survey-specific data. The recurrent TBI program collects data biennially beginning with a starter wave of 7,500 complete households. A pilot study of 407 complete households was conducted in May 2018 to test two study designs, measure response rates, and evaluate the resulting data. To increase response and reduce respondent burden, a split sample was used to evaluate two study designs in the pilot. Household travel surveys have historically collected data through a two-part survey process where households provide demographic data in part 1 and travel diary data in part 2. For the pilot, an all-in-one design was tested where households were invited to recruit directly into the smartphone application eliminating the two-part process. For the traditional two-part design, households were invited to recruit either online or over the telephone before downloading the smartphone application for the diary. The pilot results provide a direct comparison of the two-part and all-in-one designs at the household, person, and trip levels. The council will use the pilot results to determine survey methodology for future waves.

Project 4: Getting the Most out of Your Data: Applying Passively Collected Travel Demand Management Data to Transportation Planning in El Paso[98]

- **Data provider:** Metropia.
- **Sponsor:** Texas Department of Transportation (TxDOT).
- **Description:** The generation of data in today's digital environment provides engineers and planners with a wealth of information that in the past was too inaccessible to be of use. One of the greatest challenges currently facing transportation agencies is how to determine the best way to interpret and use these emerging data sources. Data developed primarily for Transportation System Management & Operations (TSM&O) may also have applications for transportation planning. Identifying additional uses of data developed for other purposes can help planning agencies to make better use of existing

resources and greatly enhance the benefit to the public of a given data source. To that extent, the El Paso District of the TxDOT, the Camino Real Regional Mobility Authority (CRRMA) and the El Paso MPO decided to implement Metropia Synergy (Metropia), an incentive-based Active Traffic and Mobility Management platform to allow travelers to discover and engage available mobility options. Data acquired via the platform would be used to provide regional travel patterns, mobility and reliability performance measures as well as support the traffic management center, TransVista's, operations.

This presentation discusses how the Metropia Synergy data were assessed, how suitability was determined, and the development of guidelines for the proper use and interpretation of the data. Factors addressed include defining data density thresholds required for using the data to develop volume delay curves, O-D seed matrices, and time segmentation factors, identifying sufficient market penetration of the app to ensure the proper representation of the traveling public, methods of comparison between the newer Metropia data and other more traditional data sources to establish suitability, and assessing the support for other planning and modeling related activities. The presentation addresses how travel demand management and traffic operations data generated by this source such as travel time and speed estimates, reliability metrics, travel patterns for construction zones, O-D seed matrices, and arterial LOS can aid in the development of travel demand and operation models

Project 5: Driving Activity Locations Inferred from Smartphone Data [99]

- **Data provider:** Google™ Location History.
- **Sponsor:** Florida International University.
- **Description:** In order to support efficient transportation planning decisions, household travel survey data with high levels of accuracy are essential. Due to a number of issues associated with conventional household travel surveys, including high cost, low response rate, trip misreporting, and respondents' self-reporting bias, Government and private agencies are desperately searching for alternative data collection methods. Recent advancements in smart phones and GPS technologies present new opportunities to track travelers' trips. Considering the high penetration rate of smartphones, it seems reasonable to use smartphone data as a reliable source of individual travel diary. The Google™ location history (GLH) data provide an opportunity to explore the potential of these data. One month of GLH data are obtained from 50 participants. This presentation describes the data processing methods in deriving travel information, including trip ends, modes, activity types, etc. GIS tools are also employed to facilitate the data processing. The results show great promise of using GLH data as a supplement or complement to conventional travel diary data. It shows that GLH provides sufficient high resolution data that can be used to study people's movement without respondent burden, and potentially it can be applied to a large scale study easily. These data provide the opportunity to facilitate the investigation of various issues, such as less frequent long-distance travel, daily variations in travel behavior, and human mobility pattern in large spatio-temporal scale.

Project 6: Using Multiple Years of Truck GPS Data for Freight Model Development and Validation[(100)]

- **Data provider:** Unknown.
- **Sponsor:** Resource Systems Group.
- **Description:** Passively collected truck GPS trace data are now available to travel demand model developers as a data source for the estimation, calibration, and validation of truck and freight demand models. Their usage is becoming increasingly common and approaches using a single recent sample of data or, for example, four slices of data from the most recent year are typical. With the availability now of several years of past data from truck GPS trace data vendors, it is possible to develop a multi-year GPS dataset of the region being modeled and make comparisons of changes over time. This offers the opportunity to understand trends in the region and to test whether the model being developed is capable of responding to input changes through short term forecasting sensitivity tests where 1 yr of truck GPS traces are used for estimation or calibration of a base year model and a second year of truck GPS traces are used for forecast (or back-cast) validation.

  The application presented is the development of a truck model for an east coast metropolitan area for a private client. Two years (2015 and 2017) of passively collected truck GPS trace data were procured for the project for the use of designing, estimating, and calibrating the model and then evaluating the forecast sensitivity of the model. The truck GPS trace data were processed using consistent methods for the 2 yr into trips through stop identification methods and then grouped into tours. The data are being expanded using truck counts to represent the full population of trucks moving in and through the region. The data are being compared with each other and with other local data such as land use data to understand the differences in truck travel behavior between the years and to embody the model with the ability to forecast those changes.

WPS/LBS

Project 1: Identifying Truck Bias in LBS Data[(101)]

- **Data provider:** Unknown.
- **Sponsor:** TxDOT and Texas A&M Transportation Institute.
- **Description:** LBS and cellular data provide unprecedented scope of the daily travel and activity of millions of people. However, the mode of these travel activities is mostly unknown. One question for practitioners is to what degree are commercial vehicles present in LBS data? After all, truck drivers carry cell phones too—for both work and personal use.

  This presentation describes findings to address this question using results of external travel analyses in Texas using LBS, cellular, and GPS data sources. These external analyses are of model boundaries as well as an extended buffer area surrounding these regions and include over 100 million trips (inclusive of internal trips). This assessment utilizes LandSat imagery, LEHD data, Census data, and GPS in relation to TAZs designed to segregate industrial, commercial, and residential areas.

Project 2: Insights from Big and Small Data: Which Trips and Travelers are Captured by Location-Based Services Data?[102]

- **Data provider:** Unknown.
- **Sponsor:** Ohio Department of Transportation.
- **Description:** This study presented a case study that evaluated similarities and differences between smartphone data collected for a household travel survey using Resource System Group's rMove™ app and LBS data collected passively from smartphone apps for FHWA's TMIP.

  The study concluded that passive LBS and rMove™ data together allow the transportation community to understand travel better than either alone.

Project 3: Use of Big Data to Calibrate and Validate Travel to Special Travel Destinations[103]

- **Data provider:** Unknown.
- **Sponsor:** San Diego Association of Governments.
- **Description:** Special travel destinations are unique with respect to magnitude of travel; type of traveler/purpose (non-residents, recreational, health related, etc.); and spatial, temporal, modal distribution of travel to/from location. Special travel destinations are often under-represented in travel surveys. Special destinations in San Diego County include beaches, major shopping centers, hospitals, and parks.

  This study used passively collected LBS smartphone app data to better understand San Diego travel patterns and calibrate the region's activity-based model. This study compared the following:

  - Big data to intercept survey data collected at beaches.
  - Big data and intercept survey data to model results.
  - Model results to cordon traffic counts.

The conclusions of this study were as follows:

  - LBS data are useful for understanding travel to special destinations.
  - Model comparisons to survey revealed some useful insights, including the following:

    - Sites measured by acres of active space (e.g., beaches and parks) are the most challenging to represent accurately in terms of magnitude of travel; models of non-resident travel help.

    - Major shopping centers not necessarily a special market.

    - Hospital-related travel may require special treatment to match real-world constraints (less onerous accessibility terms in destination choice). Caution should be used in how land-use data classifies employment, particularly medical (sometimes coded as university or government) and entertainment.

Project 4: Using Big Data to Understand Travel Behavior to Parks(104)

- **Data provider:** StreetLight Data™.

- **Sponsor:** StreetLight Data™.

- **Description:** According to the National Park Service, there were nearly 331 million visits to national parks in 2017 alone. With such a high volume of visitors and limited facilities with little options for expansion, how do transportation planners start to take action and tackle this problem? How do they understand exactly where visitors come from to visit national, State, and local parks before they even start to consider effective alternative modes of transportation?

  This presentation explains how big data from mobile devices, such as smart phones and connected vehicles, can reveal valuable insights for public transit to parks where parking for cars is limited. Using Boulder County, CO, as a case study, attendees learn the process of using geospatial records from mobile devices to create O-D matrices. The presentation also describes the process for locating where visitors are coming from and at what volume, what routes visitors are taking to go to popular state parks, and the demographics of these visitors. It demonstrates how cities, counties, and park operators can use these data to identify the best potential locations for transit onto parks, which would improve congestion and mitigate the lack of parking. Finally, the presentation covers how addressing these challenges will also reduce the number of VMT around and better preserve these natural habitats by expanding on this case study and sharing how VMT can be estimated using this same source of data.

Project 5: Using Travel Time and Origin-Destination Data in Transportation Planning: A Metropolitan Planning Organization's Example(105)

- **Data provider:** Unknown.

- **Sponsor:** Mid-Ohio Regional Planning Commission (MORPC).

- **Description:** Under the Ohio Department of Transportation’s recent subscription to big data services, all public agencies in Ohio gained access to travel time and O-D data collected by GPS devices, mobile phones, and location-based apps on smart phones/tablets. MORPC, the MPO for the Columbus, OH, metropolitan area, has become one of the most active users of such data and successfully applied them in various transportation planning activities and studies via analytical statistic measures and visualization.

This presentation describes how MORPC applied travel time data to produce meaningful summary measures in both an annual report card and project evaluation process for the Metropolitan Transportation Plan and Transportation Improvement Program. Also, an innovative way to visualize travel time data over a long period of time will be proposed for individual roadway segments so that congestion patterns and anomalies can be easily identified. The measures and visualizations were also applied in several sub-area traffic studies.

MOPRC investigated the big data available through the Ohio Department of Transportation to analyze and visualize the O-D and route choice patterns for both passenger cars and trucks in a comprehensive study for the Rickenbacker Area in Columbus, a vibrant logistics hub that has grown into an important engine in central Ohio. In the study, the O-D data extracted from the consultant service were analyzed to show the overall travel patterns in the area and its interaction with the other parts of central Ohio region. Seasonal patterns were examined, especially for holiday seasons near distribution warehouses. Major freight highway routes to the Rickenbacker area were identified through an in-depth analysis of the data.

Lessons learned from the above experiences were shared. Potential future applications of such data are currently under investigation, such as combining the O-D data with other powerful data (e.g., LEHD O-D employment statistics) and producing meaningful near-real-time congestion measures to inform SmartColumbus initiatives.

Project 6: Using Google's Aggregated and Anonymized Trip Data to Estimate Dynamic Origin-Destination Matrices for San Francisco(106)

- **Data provider:** Google™'s Aggregated and Anonymized Trips (AAT).
- **Sponsor:** SFCTA.
- **Description:** Historically, O-D data could only be collected using time-consuming and expensive vehicle intercept or license plate surveys. Recent technological advancements have resulted in use of Bluetooth™ detectors and cell phone call detail records to passively collect data that can be used to estimate O-D demand. In this presentation, a new passive data source, Google™'s AAT, is presented. Google™'s Better Cities program, which seeks to minimize traffic congestion, speed up journeys, improve safety, and reduce the amount of money spent on infrastructure, has partnered with SFCTA to make available aggregated and anonymized O-D flow information from location reports. This dataset is derived from users who have chosen to store their location information from Google™-enabled devices on Google™ servers. The aggregate data are anonymized by applying differential privacy algorithm. Since the dataset accounts for only a sample of travelers, Google™ provides relative trips rather than absolute counts.

Aggregate flow data (in terms of relative trips) to and from approximately 90 districts covering the 9-county San Francisco Bay Area in 1-h increments were provided to SFCTA to support the San Francisco Freeway Corridor Management Study. The primary objective of this study is to assess strategies for improving the performance of the US-101 and I-280 corridor, which connects San Francisco and Silicon Valley. Google™ also provided O-D flows for four primary freeway segments on the US-101 and I-280 corridor in San Francisco for these same districts and time periods. A total of 6 mo of daily and hourly data were provided, covering 3 mo each from spring and fall 2015.

The presentation focuses on the methods used and models estimated to convert the weights provided in the AAT dataset to actual person trips resulting in hourly O-D matrices. In addition, a comparison of the estimated O-D demand to that indicated by regional travel demand model and household travel survey is presented.

Project 7: Activating Big Data for Active Transportation with a Statewide Data Platform[107]

- **Data provider:** StreetLight Data™.
- **Sponsor:** StreetLight Data™.
- **Description:** Big data analytics derived from mobile device data can meet planners' increasing demands for complex and comprehensive bike and pedestrian data collection and accelerate project prioritization. This presentation aims to demonstrate how by sharing the shared data platform created for the California Department of Transportation (Caltrans) to develop unique bike and pedestrian travel analytics, such as volumes of pedestrian and bike trips for every road segment and TAZ in the State, demographics of bikers and pedestrians, popular O-Ds, and more.

  This presentation review and explains the process for developing these metrics, sharing how the teams trained machine learning algorithms on a truth dataset provided by an external third party to ensure the appropriate disaggregation by mode for bicyclists, pedestrians, and vehicular trips. Validation work conducted using data from Caltrans and other third parties is also presented.

  Finally, the presentation shares how Caltrans is leveraging this groundbreaking dataset by diving into the scoring criteria (developed by StreetLight Data™ in consultation with Caltrans) that indicate the likelihood of locations having a high return on investment from new biking and pedestrian infrastructure.

Project 8: Understanding Travel Using Location Based Services Data[108]

- **Data provider**: StreetLight Data™.
- **Sponsor:** Cambridge Systematics.
- **Description:** LBS data collected from cell phone apps can provide valuable insights into travel behavior. A dataset representing device movement within New Jersey was obtained for a period of 6 mo in 2017 and analyzed to gain relevant insights.

The data were used to better understand trip-making behavior, such as travel across months and times of day, frequency of visits to locations of interest, and even in-State versus out-of-State visitors. The results from three distinct locations (Hoboken station, Metlife stadium, and Liberty State Park) are presented and discussed as part of this application.

OTHER

Project 1: Finding Uses for Temporal Elements of Big Data[109]

- **Data provider:** Unknown.
- **Sponsor:** Whitman, Requardt, and Associates.
- **Description:** The use of big data in the development of travel demand models and use as part of projects is no longer considered new or innovative. Many agencies have adopted the use of one of the data sources for use in developing external models or for validation of regional travel patterns. At the project level, the use of big data is commonly used for understanding travel patterns and draws to a given corridor or facility. The industries use of big data for these applications will become more common place, especially with the release of several upcoming reports formalizing the acceptable use of these sources.

  With the standard applications in mind, new methods or applications of big data are being sought. This presentation looks at two such examples where the temporal nature of big data can be realized rather than the static average day. The first is within the context of regional travel demand modeling, and the second is related to project level traffic forecasting.

  The Hampton Roads region in Virginia has always been considered a challenging region to model for several reasons, including the significant number of water crossings, prevalence of toll facilities in the region, and unique proximity to the urbanized portions of Northern Virginia and access to tourist destinations to the south. Whitman, Requardt, and Associates is working on an update to the Hampton Roads regional model that, with the use of big data, will improve the model's sensitivity and reliance on ad-hoc adjustments necessary to make the model validate. Big data will be used to help calibrate the model's external demand and distribution patterns accounting for the water crossings and barriers (both physical and perceived) to travel that a traditional model is not able to see. In addition, the temporal range of information available from the big data will allow for the understanding of travel behavior outside of the typical average, which has not been possible.

Project 2: Using Big Data in Small and Medium Sized Regions: Three Case Studies and Lessons Learned[110]

- **Data provider:** AirSage™ and StreetLight Data™.
- **Sponsor:** Whitman Requardt, and Associates.

- **Description:** Conclusions are as follows:
    - Case Study 1:
        - Provided evidence of validity of the regional model and provided credibility to the forecast process.
        - Showed value of data to the sponsoring agency.
    - Case Study 2
        - Produced a model that met the Virginia Department of Transportation (VDOT) validation criteria with small data investment.
        - Able to account for the travel patterns observed in the region with the use of the big data while maintaining transparency in the model and forecast sensitivity.
    - Case Study 3
        - Challenges in defining geography within the Streetlight framework to support the internal cordon survey and to support the external model development.
        - Effective at both applications but requires forethought in developing methodology.

Project 3: An Agile, Data-Driven Ensemble Modeling Framework for the Illinois Statewide Travel Demand Model (111)

- **Data provider:** Unknown.
- **Sponsor:** Illinois Department of Transportation.
- **Description:** Travel modeling and forecasting are facing two simultaneous significant challenges. On the one hand, modelers are being asked by planners to provide answers to questions about how transformative technologies such as AVs might reshape transportation. On the other hand, data vendors and tech-savvy executives are asking modelers how they are incorporating streams of passive data and insights from machine learning and advances in artificial intelligences. The thesis of this presentation is that the latter challenge is actually a key part of the solution to the former and that passive data-driven ensemble modeling can help planners understand how technology is transforming transportation as it happens.

  In the face of the transformative changes in the realm of transportation, such as the emergence of new modes and technologies, the State of Illinois has undertaken the development of a statewide travel forecasting model to support its transportation planning, programming, and policy making. The design of the new model leverages the availability of passive data together with insights from machine learning in a modular, data-driven framework designed to support ensemble modeling. The framework synthesizes a number of recent advances in travel modeling and is presented here as an example of best practice.

Both the data-driven nature and ensemble methods of the new model design promise improved forecasting validity based on proven applications in other industries as well as limited applications to travel forecasting. The ensemble approach also provides built-in cross-validation of component models, as differences in their forecasts should be plausible and logically related to the differences in their assumptions and methodologies. The approach is also efficient in not wasting preliminary model development efforts, accumulating models over time rather than discarding earlier tools in favor of later, more complex ones. Moreover, this supports ongoing development of new components alongside and without interfering with a working model. Finally, this ability to develop new components and produce alternative forecasts without requiring the redevelopment of the entire modeling system aims to provide an agile and nimble modeling platform to support the incorporation of new data and address the emergence of new trends in coming years in shorter and on-going model development cycles.

Project 4: Overview of Methods for Validation and Expansion of Passive Origin-Destination Data[112]

- **Data provider:** AirSage™ and ATRI.
- **Sponsor:** Resource Systems Group.
- **Description:** This presentation reviewed different types/sources of passive O-D data and concluded the following:
    - All suffer from systematic biases.
    - Biases can be corrected through analysis together with other data sources.
    - Ensemble expansion methods are best for now.
    - Count-based methods are necessary for now.
    - Smartphone travel survey data are especially promising in correcting passive data at the disaggregate level.

Project 5: Incorporating Big Data in an Activity-Based Travel Model: The Chattanooga Case Study[113]

- **Data provider:** AirSage™ and ATRI.
- **Sponsor:** Chattanooga-Hamilton County Planning Agency.

- **Description:** This presentation compared traditional data (survey, LEHD CTPP/ACS, etc.) with passive data for model calibration. It then discussed how LEHD and expanded passive data are used in DaySim destination choice calibration.

Project 6: GPS and Cell Data for Medium and Small Cities[114]

- **Data provider:** AirSage™ and INRIX™.

- **Sponsor:** Texas A&M Transportation Institute.

- **Description:** This presentation showcased experiences using cell data and GPS data for five medium and small Texas MPOs.

Project 7: "Forecasting Current and Next Trip Purpose with Social Media Data and Google Places"[115]

- **Data provider:** Google™ Places API and Twitter™ API.

- **Sponsor:** University at Buffalo and The State University of New York.

- **Description:** Trip purpose is crucial to travel behavior modeling and travel demand estimation for transportation planning and investment decisions. However, the spatial-temporal complexity of human activities makes the prediction of trip purpose a challenging problem. This research addresses the problem of predicting both current and next trip purposes with both Google™ Places and social media data. First, this paper implements a new approach to match points of interest from the Google™ Places API with historical Twitter™ data. Therefore, the popularity of each point of interest can be obtained. Additionally, a Bayesian neural network is employed to model the trip dependence on each individual's daily trip chain and infer the trip purpose. Compared with traditional models, it is found that Google™ Places and Twitter™ information can greatly improve the overall accuracy of prediction for certain activities, including "EatOut," "Personal," "Recreation," and "Shopping" but not for "Education" and "Transportation." In addition, trip duration is found to be an important factor in inferring activity/trip purposes. Further, to address the computational challenge in the Bayesian neural network, an elastic net is implemented for feature selection before the classification task. This research can lead to three types of possible applications: activity-based travel demand modeling, survey labeling assistance, and online recommendations.

Project 8: Using Big Data in Freeway Corridor Studies[117]

- **Data provider:** HERE™ and StreetLight Data™.

- **Sponsor:** Oregon DOT.

- **Description:** The purpose of this freeway corridor study is to analyze alternatives for improving operations on a segment of I-5 South in Wilsonville, OR, such as adding a southbound auxiliary lane on I-5 from the Wilsonville Road interchange (Exit 283) on-ramp to the Canby/Hubbard interchange (Exit 282A) off-ramp. The addition of a southbound auxiliary lane is expected to reduce merging conflicts and relieve the traffic bottleneck that occurs at the high-volume Wilsonville Road southbound on-ramp merge.

The presentation focuses on the operational analysis of existing conditions and future conditions of proposed alternatives and present the final plan. The congestion on I-5 South through the Wilsonville area has been increasing during the recent years, causing substantial delay and adding unreliability, especially in peak commuting periods. ODOT and local agencies partnered to study options for improving traffic operations and safety. The consultant developed an approach to apply big data for the analysis of existing corridor operations and performances of future alternatives.

An issue of concern along this corridor is the high percentage of vehicles using the I-5 segment to make short interchange-to-interchange trips across the Willamette River. To help understand the travel patterns, the consultant team analyzed a large set of O-D data provided by StreetLight Data™ and applied the results in freeway operational analysis.

Regarding travel time reliability, the HERE™ traffic data were analyzed, which provides travel time and speed information collected from sample devices, to help the project team understand the variations of congestions and travel time differences across typical weekdays.

Project 9: Using Big Data to Estimate Weekend VMT for Crash Analysis [(118)]

- **Data provider:** ATRI and StreetLight Data™.
- **Sponsor:** Delaware Valley Regional Planning Commission.
- **Description:** One of the goals of the Indianapolis MPO recent long-range transportation plan update was “Make Safe.” One of the criteria for project selection involved proximity to high-crash areas. To identify these areas, the region was divided into a 1-km grid, and crash locations containing a fatality or serious injury (KSI) from ATRI were placed in grid cells. KSI in each grid cell were then normalized by VMT, which was taken from a base year travel demand model run. When the results were initially evaluated, it was found that there was a very high KSI per VMT rate near a large sports complex in the Indianapolis suburbs that hosts numerous events during weekends. It was then realized that the crash data were for the full week, but the VMT from the model was for a weekday, resulting in a KSI per VMT measure that was biased towards areas with higher weekend traffic.

  In the absence of a weekend travel demand model, it was decided to use passive data to estimate weekend travel. The region was divided into a 2-km grid, and data from Streetlight Data™ were used to estimate the ratio of weekend trips to weekday trips in each grid cell. The weekday VMT in each 1-km grid cell was then multiplied by the observed weekend-weekday travel ratio from its associated 2-km grid cell to obtain an estimate of the weekend VMT for that 1-km grid cell. The weekday and weekend VMT values were then combined to obtain an estimate of the weekly VMT, of which the KSI were normalized by, resulting in a less biased identification of the region's most dangerous areas.

Project 10: Passive Data Modeling: A Method for Fusing Passive Data and Survey [119]

- **Data provider:** Unknown (cell tower triangulation, GPS, LBS, etc.).
- **Sponsor:** Transport Foundry.
- **Description:** This presentation provided an example of how the NextGen NHTS data products could be used for planning purposes using a data-driven (or passive data) model.

Project 11: Using Big Data to Explore Long-Distance Freight Travel on Non-interstate Corridors in Texas [120]

- **Data provider:** Unknown.
- **Sponsor:** Jacobs.
- **Description:** The availability of big location data provides a number of opportunities and challenges for transportation agencies and policymakers. While a number of suppliers of big transportation data exist, the application of these data and the common data products to a range of planning contexts and scales has not been determined. There has been little exploration of the use of these data sources to evaluate the performance of statewide freight travel corridors.

  This presentation describes an exploratory comparison of pre-processed big transportation data on commercial vehicle movements with aggregate regional freight flow data and details the development of a network assignment model that seeks to maximize the benefits of each dataset. These data sources are evaluated in the context of evaluating long-distance freight travel on several non-Interstate corridors in Texas.

  Big transportation data can offer unprecedented detail that is difficult to capture in traditional travel data. However, many of the methods used to mine information from big data have been developed for urban and regional contexts where trips are relatively short and stops are limited. Modeling freight vehicle travel, on the other hand, requires accurate information about the movement of commodities from O-D, which frequently spans regions and even States. Common assumptions used to mine information on trip distance and trip ends from commercial vehicle big data may not be appropriate for evaluating long-distance commercial freight travel, where rest stops and refueling breaks may be misidentified as trip ends. However, long-distance freight models frequently make simplistic assumptions about route choice behavior. These models may not accurately depict long distance travel where several non-interstate highway corridors are available. Working with a big transportation data platform, this presentation attempts to cross-validate a number of indices of interregional and interstate commercial travel against region-to-region freight flow data from FAF4.

It begins by comparing O-D movements between Texas regions and between Texas and other States not served by an interstate route from end to end. It then compares synthetic freight volumes on non-interstate corridors as provided by the big data platform and as available in the FAF4 database. Finally, a probabilistic long-distance commercial network assignment model is proposed that maximizes the granular information on roadway use from big data while accurately depicting long-distance commercial travel. With this approach, freight flows are accurately routed across a number of alternative non-Interstate travel routes.

Project 12: Model Validation Using "Novel" Big Data [(122)]

- **Data provider:** AirSage™, INRIX™, StreetLight Data™, Streetlytics, and Teralytics.
- **Sponsor:** Atlanta Regional Commission.
- **Description:** The presentation discussed the following model validation examples using novel big data:
    - Activity-based model DTA Integration (SHRP2 C10) with INRIX™ data.
    - I-85 Bridge Collapse Travel Patterns with Streetlytics Data.
    - Externals Model with AirSage™ Data.
    - I-285/GA-400 Interchange Reconstruction Commute Options with Streetlight Data™.
    - Regional Origin-Destination Analysis with Teralytics Data.
    - Volume-Delay-Reliability Functions (SHRP2 L04) with NPMRDS Data.

Project 13: Understanding Regional Travel Patterns with Big Data [(123)]

- **Data provider:** StreetLight Data™ (GPS vehicle and LBS).
- **Sponsor:** Northeastern Indiana Regional Coordinating Council.
- **Description:** The Northeastern Indiana Regional Coordinating Council invested in passively collected big data to better understand the movements of both people and truck freight into, out of, through, and within Northeastern Indiana. Key findings include:
    - Traditional surveys cannot provide a picture of the O-D trip matrix at the same level of zones or even moderately disaggregate districts.
    - Traditional surveys typically contain observations for 3 percent or less of the cells in the O-D matrix.

- Passive O-D data typically provide observations for a quarter to a third of the cells in a regional O-D matrix.

Project 14: Alternate Methodologies for Origin-Destination Data Collection [124]

- **Data provider:** Unknown.
- **Sponsor:** Polk County Transportation Planning Organization (TPO).
- **Description:** There is a wide range of O-D data sources, approaches, technologies, and techniques that did not exist until recently. Many of these are passive data extraction techniques that use devices with GPS. Anonymous tracking of GPS signals provides cost savings in data collection and allows for larger sample sizes than traditional O-D survey techniques.

  There are a multitude of considerations in evaluating and selecting approaches to collecting data on trip O-Ds. The most obvious of these is cost, although this is a difficult consideration to quantify. While some data vendors have a standard cost template covering a variety of factors, the competitive nature of data acquisition also means there is some flexibility on the part of vendors to remain competitive. Other criteria in selecting an O-D data collection approach include study area size and geography, information needs, trip purposes, and transportation modes. Geographic considerations are crucial in selecting the best approach as different travel patterns might dominate within a single transportation corridor or subarea versus an entire region. Information needs can also vary; for example, a data source for trip O-Ds may differ from a study needing information on auto occupancy. Collecting information on trip purpose necessitates different methodologies than studies of general traffic. Focus on specific travel modes is another consideration as different data collection methodologies can provide data on autos, trucks, or transit vehicles.

  This presentation covers a range of data sources and considerations in selecting methodologies, including vendor/product name, approach, sampling unit, survey periods, relative vintage, pros/benefits, cons/disadvantages, and relative cost. This information was obtained through Stantec experience using these alternative methodologies in toll corridor feasibility studies, demonstrations and discussions with vendors, and a project for the Polk County TPO on comparing alternative methodologies.

  This is an update to a recent top 5 presentation at the 2016 TRB Tools of the Trade Conference to reflect findings from an ongoing phase 2 of the Polk TPO project and more recent advancements in the mining of passive data, an area where technology is continuously moving forward.

Project 15: Big Data "Triage" Before Modeling[125]Data provider: StreetLight Data™.

- **Sponsor:** StreetLight Data™.
- **Description:** Modeling is time and money consuming. This presentation demonstrated how big data can help improve the process as follows:
  - Use big data to identify the highest modeling priorities.
  - Drill down on causes of (and solutions to) congestion on specific roadways.
  - Combine big data analytics to identify high-potential project opportunities.
  - Use empirical data to help decide which models to build first.

Project 16: Vehicle Emissions and Air Quality: How Big Data Can Help[126]

- **Data provider:** StreetLight Data™.
- **Sponsor:** StreetLight Data™.
- **Description:** The U.S. National Emissions Inventory relies on modeled inputs at the county level. In this presentation, the CRC A-100 case study provides improved data compared to defaults and shows the benefits of big data to be:
  - Better spatial/temporal resolution.
  - Differences unique to individual cities.
  - Truck trends not seen in other datasets.

Project 17: Development of an Agile & Data-Driven Model Framework[127]

- **Data provider**: Unknown.
- **Sponsor:** Resource Systems Group.

  Description: This presentation discusses ensemble modeling and forecasting that uses different types of data including passive data. Details discussed include:
  - How to combine multiple forecasts into a "consensus" forecast via binding function.
  - The passive data works best with different (uncorrelated) models.
  - The use of passive data has led to greater predictive validity (e.g., in meteorology).

Project 18: Applying Big Data to Small Projects(128)

- **Data provider:** StreetLight Data™.
- **Sponsor:** StreetLight Data™.
- **Description:** This presentation discusses the following two case studies:
    - **Case Study 1: Scanning Roadways for TDM Opportunities in Northern VA:**
        - **Challenge:** Northern Virginia's severe congestion cannot be addressed by highway expansion. VDOT needs to reduce travel in single occupancy vehicles.
        - **Data-driven solution:** VDOT compared every single regional corridor for mode-shift potential creating, which led to far more effective cost/benefit process and expenditures of infrastructure dollars.
    - **Case Study 2: Understanding the Impact of Expressways on Congestion in Fredericksburg, VA:**
        - **Challenge:** Fredericksburg experiences intense congestion hotspots along their stretch of I-95 between mileposts 125-145. There is a need to understand travel patterns to, from, and within this region.
        - **Data-driven solution:** Big data was used to identify, screen, and evaluate common O-Ds using I-95 to shift trips to other modes and routes.

Project 19: Data Analytics and Modeling Methods for Tracking and Predicting O-D Travel Trends Based on Mobile Device Data(129)

- **Data provider:** Unknown (cell tower triangulation, GPS, and LBS).
- **Sponsor:** University of Maryland.
- **Description**: This EAR project supports the development of the NextGen O-D component. A key goal of this project, led by the Maryland Transportation Institute at the University of Maryland, is to produce national- and MPO-level O-D matrices from different big data sources (e.g., cell phone, GPS, and LBS) that are segregated by mode, trip purpose, time-of-day, month-of-year, and socio-demographics. All O-D products will available in the public domain. In addition, the institute will develop open-source algorithms for accurately imputing missing information and for sample expansion in the production of O-D and other travel behavior statistics from passively collected data sources.

APPENDIX C. PRIVATE SECTOR TRAVEL BEHAVIOR DATA PRODUCTS

This appendix compiles the data products discussed in chapter 3 by provider. The contents of this appendix are intended only to summarize product offerings based on publicly available details at the time of this report and for the products referenced in chapter 3. It is not intended to be a comprehensive list of offerings and should not be construed as any type of product or vendor endorsement. Furthermore, any mention of private company names or products is not an endorsement by FHWA and are included by the authors for reference purposes only.

Table 2. Examples of private sector travel behavior data products.

Company	Primary Data Provided	Travel Behavior Data Product Examples	Application Examples
AirSage™(130)	GPS and LBS data	• Trip matrix (trip patterns). • Target location analysis (points of interest). • Audience insights (audience exposure to outdoor mediums).	O-D matrices used in external trip models and calibration/validation efforts.
INRIX™(131)	GPS data	• Analytics (movement patterns of roadways and populations). • Parking (finding, comparing, and paying for parking). • Traffic (traffic information for major road types).	O-D trajectories and waypoints, network usage over time, inputs to calculate travel speeds, trip duration, travel time, and turning movements along routes.
ATRI(9)	GPS data (commercial vehicles)	• Trucking industry datasets (operational, finance, performance, and safety). • Trucking industry GPS datasets.	Truck parking, travel time reliability, and truck highway routing.
StreetLight Data™(34)	GPS and LBS data	• O-D matrices. • Select link analysis. • 2016 average annual daily traffic. • Trip purpose. • Average travel times and distributions. • Commercial and personal travel vehicle comparisons.	O-D matrices, including selected link metrics.
HERE™(132)	GPS data	• HERE™ traffic (combine precision mapping, real-time big data, and historical insights). • HERE™ positioning (global positioning with cross-platform compatibility and online APIs	Detailed routable networks with fine resolution speed profiles by hour and day type, which can be used to determine free-flow and congested speeds for each link.

Company	Primary Data Provided	Travel Behavior Data Product Examples	Application Examples
		for use with any compatible connected device). • HERE™ tracking (hardware agnostic, global, reliable, and complete Internet of things location toolkit).	
InfoUSA™[133]	Consumer and firm data	Targeted lists compiled across multiple directory and event-driven sources.	Generate synthetic populations, socioeconomic files for use in travel demand models, and base year employment data.

REFERENCES

1. Federal Highway Administration. (2017). "National Household Travel Survey." (website) U.S. Department of Transportation, Washington, DC. Available online: http://nhts.ornl.gov/index.shtml, last accessed December 19, 2017.

2. Bureau of Transportation Statistics. (2017). "The 1995 American Travel Survey (ATS)." (website) U.S. Department of Transportation, Washington, DC. Available online: https://www.transtats.bts.gov/DatabaseInfo.asp?DB_ID=505&Link=0, last accessed December 19, 2017.

3. Livshits, V., Brown, E., Copperman, R., DeBoer, K., Hekele, A., Komanduri, A., and Zhang, W. (2017). *MAG Large-Scale Regional Travel Surveys and Bottleneck Survey*, Travel Model Improvement Program Webinar, Federal Highway Administration, Washington, DC. Available online: https://www.fhwa.dot.gov/planning/tmip/community/webinars/summaries/20171018/, last accessed December 19, 2017.

4. San Diego Association of Governments. (2017). "San Diego Regional Transportation Study." (website) San Diego Association of Governments, San Diego, CA. Available online: http://www.sandag.org/index.asp?classid=13&projectid=540&fuseaction=projects.detail, last accessed December 19, 2017.

5. Ohio Department of Transportation. (2017). "Ohio Moves Transportation Study." (website) Ohio Department of Transportation, Columbus, OH. Available online: https://ohiomoves.org/web/pages/home?locale=en-US, last accessed December 19, 2017.

6. New York Metropolitan Transportation Council. (2017). "Regional Establishment Survey (RES)." (website) New York Metropolitan Transportation Council, New York, NY. Available online: https://www.nymtc.org/Data-and-Modeling/Travel-Surveys/Regional-Establishment-Survey, last accessed December 19, 2017.

7. OKI Regional Council of Governments. (2017). "OKI Regional Council of Governments Home Page." (website) OKI Regional Council of Governments, Cincinnati, OH. Available online: http://www.oki.org, last accessed December 19, 2017.

8. Denver Regional Council of Governments. (2014). "Freight and Goods Movement." (website) Denver Regional Council of Governments, Denver, CO. Available online: https://drcog.org/programs/transportation-planning/freight-and-goods-movement, last accessed January 2, 2018.

9. American Transportation Research Institute. (2012). "ATRI Home Page." (website) American Transportation Research Institute, Arlington, VA. Available online: http://atri-online.org/, last accessed December 29, 2017.

10. Federal Transit Administration. (2017). "The National Transit Database (NTD)." (website) U.S. Department of Transportation, Washington, DC. Available online: https://www.transit.dot.gov/ntd, last accessed December 19, 2017.

11. U.S. Travel Association. (2017). *U.S. Travel Answer Sheet*, U.S. Travel Association, Washington, DC. Available online: https://www.ustravel.org/system/files/media_root/document/Research_Fact-Sheet_US-Travel-Answer-Sheet.pdf, last accessed December 19, 2017.

12. U.S. Department of Transportation. (2017). *Beyond Traffic: 2045 Final Report*, U.S. Department of Transportation, Washington, DC. Available online: https://www.transportation.gov/sites/dot.gov/files/docs/BeyondTraffic_tagged_508_final.pdf, last accessed December 19, 2017.

13. Aultman-Hall, L., Harvey, C., LaMondia, J.J., and Ritter, C. (2015). "Design and Response Quality in a One-Year Longitudinal Survey of Overnight and Long-Distance Travel," *Transportation Research Procedia, 11*, pp. 136–153, Transportation Research Board, Washington, DC. Available online: http://www.sciencedirect.com/science/article/pii/S2352146515003038, last accessed December 19, 2017.

14. Michigan Department of Transportation. (2017). "MI Travel Counts Household Travel Survey." (website) Michigan Department of Transportation, Lansing, MI. Available online: http://www.michigan.gov/mdot/0,1607,7-151-9615_51690---,00.html, last accessed December 19, 2017.

15. Schiffer, R. et al. (2012). *Long-Distance and Rural Travel Transferable Parameters for Statewide Travel Forecasting Models*, NCHRP Report 735, National Cooperative Highway Research Program, Washington, DC. Available online: http://www.trb.org/Publications/Blurbs/168389.aspx, last accessed December 19, 2017.

16. Federal Highway Administration. (2017). "Design of a Completely New Approach for a National Household-Based Long-Distance Travel Survey Instrument." (website) U.S. Department of Transportation, Washington, DC. Available online: https://highways.dot.gov/design-completely-new-approach-national-household-based-long-distance-travel-survey-instrument, last accessed December 19, 2017.

17. Federal Highway Administration. (2013). *The Exploratory Advanced Research Program Fact Sheet: Designing a New National Household Travel Survey*, Publication No. FHWA-HRT-13-081, U.S. Department of Transportation, Washington, DC. Available online: https://www.fhwa.dot.gov/advancedresearch/pubs/13081/, last accessed December 19, 2017.

18. Office of Highway Policy Information. (2016). "Traveler Analysis Framework." (website) Federal Highway Administration, Washington, DC. Available online: https://www.fhwa.dot.gov/policyinformation/analysisframework/, last accessed December 29, 2017.

19. Bureau of Transportation Statistics. (n.d). "Commodity Flow Survey." (website) U.S. Department of Transportation, Washington, DC. Available online: https://www.bts.gov/cfs/, last accessed July 26, 2019.

20. Freight Management and Operations. (2018). "Freight Analysis Framework." (website) U.S. Department of Transportation, Washington, DC. Available online: https://ops.fhwa.dot.gov/freight/freight_analysis/faf/, last Accessed December 29, 2017.

21. Pew Research Center. (2017). "Mobile Fact Sheet." (website) Pew Research Center, Washington, DC. Available online: http://www.pewinternet.org/fact-sheet/mobile/, last accessed December 29, 2017.

22. Statista. (n.d.). "Wireless subscriptions market share by carrier in the U.S. from 1st quarter 2011 to 3rd quarter 2017." (website) Statista, Hamburg, Germany. Available online: https://www.statista.com/statistics/199359/market-share-of-wireless-carriers-in-the-us-by-subscriptions/, last accessed December 29, 2017.

23. Huntsinger, LF. and Ward, K. (2015). "Using Mobile Phone Location Data to Develop External Trip Models," *Transportation Research Record 2499*, pp. 25–32, Transportation Research Board, Washington, DC. Available online: https://journals.sagepub.com/doi/10.3141/2499-04, last accessed July 26, 2019.

24. Metropolitan Washington Council of Governments. (2014). *Status Report: Ongoing review of O-D cellular data for The TPB modeled area*, Metropolitan Washington Council of Governments, Washington, DC. Available online: https://www.mwcog.org/file.aspx?&A=On2SiF8%2b7kDhhDF8cAkGJmlcjT2f4zgV05SSM7QeISU%3d, last accessed December 29, 2017.

25. Hard, E.N., Chigoy, B.T., Songchitruksa, P., Farnsworth, S.P., Borchardt, D.W., and Green, L.L. (2016). *Comparison of Cell, GPS, and Bluetooth Derived External O-D Data—Results from the 2014 Tyler, Texas Study*, Transportation Research Board, Washington, DC. Available online: https://trid.trb.org/view/1439253, last accessed December 29, 2017.

26. Stabler, B. (2014). "Development of the Idaho Statewide Travel Demand Model Trip Matrices Using Cell Phone OD Data and Origin Destination Matrix Estimation," *TREC Friday Seminar Series*, 30, Portland State University, Portland, OR. Available online: http://pdxscholar.library.pdx.edu/cgi/viewcontent.cgi?article=1029&context=trec_seminar, last accessed December 29, 2017.

27. Bucci, G. and Morton, T. (2014). *Cell Phone Data and Travel Behavior Research: Symposium Summary Report*, Report No. FHWA-HRT-14-060, Federal Highway Administration, Washington, DC. Available online: https://www.fhwa.dot.gov/advancedresearch/pubs/14060/14060.pdf, last accessed December 29, 2017.

28. Transportation Research Board. (2016). "NCHRP 08-95: Cell Phone Location Data for Travel Behavior Analysis." (website) Transportation Research Board, Washington, DC. Available online: http://apps.trb.org/cmsfeed/TRBNetProjectDisplay.asp?ProjectID=3637, last accessed December 29, 2017.

29. Markovic, N., Sekula, P., Laan, Z.V., Andrienko, G., and Adrienko, N. (2017). *Applications of Trajectory Data in Transportation: Literature Review and Maryland Case*

Study, Cornell University Library, Cornell, NY. Available online: https://arxiv.org/abs/1708.07193, last accessed December 29, 2017.

30. Turnbull, K.F. (2014). "Developing Freight Fluidity Performance Measures: Supply Chain Perspective on Freight System Performance," *Transportation Research Circular E-C187*, Transportation Research Board, Washington, DC. Available online: http://onlinepubs.trb.org/onlinepubs/circulars/ec187.pdf, last Accessed December 29, 2017.

31. Torrey, W.F. and Murray, D. (2017). *Managing Critical Truck Parking Tech Memo #2: Minnesota Case Study—Utilizing Truck GPS Data to Assess Parking Supply and Demand*, American Transportation Research Institute, Arlington, VA. Available online: http://atri-online.org/wp-content/uploads/2017/02/Managing-Critical-Truck-Parking-Tech-Memo-2-02-2017-1.pdf, last Accessed December 29, 2017.

32. Wang, Z., Goodchild, A., and McCormack, E. (2017). "A methodology for forecasting freeway travel time reliability using GPS data," *Transportation Research Procedia*, *25*, pp. 842–852, Elsevier, Oxford, United Kingdom. Available online: http://www.sciencedirect.com/science/article/pii/S2352146517307688?via%3Dihub/, last accessed December 29, 2017.

33. Camargo, P., Hong, S., and Livshits, V. (2017). "Expanding the Users of Truck GPS Data in Freight Modeling and Planning Activities," *Transportation Research Record 2646*, pp. 68–76, Transportation Research Board, Washington DC. Available online: https://journals.sagepub.com/doi/10.3141/2646-08 , last accessed July 26, 2019.

34. StreetLight Data™. (2017). "StreetLight Data." (website) StreetLight Data™, San Francisco, CA. Available online: online: https://www.streetlightdata.com/, last accessed December 29, 2017.

35. Miller, H.J., O'Kelly, M.E., Jaegal, Y., Bachman, W., Huntsinger, L., and Macfarlane, G. (2016). *Progress Report: Estimating External Travel Using Purchased Third-Party Data*, State Job #134877, Ohio State University, Columbus, OH. Available online: http://www.otdmug.org/wordpress/wp-content/uploads/2016/12/Dec2016_OSU_Archived_Data_v3.pdf, last accessed December 29, 2017.

36. Komanduri, A. et al. (2017). *Using Big Data to Develop Comparative Commercial Vehicle Metrics for Port Traffic at Major Ports in the U.S.*, Presented at the 96th Transportation Research Board Meeting, Washington, DC.

37. Hu, W. and Jin, P.J. (2017). "An Adaptive Hawkes Process Formulation for Estimating Time-of-Day Zonal Trip Arrivals with Location-Based Social Networking Check-in Data," *Transportation Research Part C: Emerging Technologies*, *79*, pp. 136–155, Elsevier, Oxford, United Kingdom. Available online: http://www.sciencedirect.com/science/article/pii/S0968090X17300426, last accessed December 29, 2017.

38. Chaniotakis, E., Antoniou, C., Aifadopoulou, G., and Dimitriou, L. (2017). "Inferring Activities from Social Media Data," *Transportation Research Record 2666*,

pp. 29–37, Transportation Research Board, Washington, DC. Available online: http://pubsindex.trb.org/view/2017/C/1438596, last accessed December 29, 2017.

39. Zhang, Z., He, Q., and Zhu, S. (2017). *Exploring Travel Behavior with Social Media: An Empirical Study of Abnormal Movements Using High-Resolution Tweet Trajectory Data*, Presented at the 96th Annual Transportation Research Board Meeting, Washington, DC. Available online: http://pubsindex.trb.org/view/2017/C/1437752, last accessed December 29, 2017.

40. Lee, J.H., Davis, A., McBride, E., and Goulias, K.G. (2017). *Exploring Social Media Data for Travel Demand Analysis: A Comparison of Twitter, Household Travel Survey, and Synthetic Population Data in California*, Presented at the 96th Annual Transportation Research Board Meeting, Washington, DC. Available online: http://pubsindex.trb.org/view/2017/C/1437326, last accessed December 29, 2017.

41. Miller, S. (2017). "Inside Boston's Bus Revamp," *CityLab*, Atlantic Monthly Group, Washington, DC. Available online: https://www.citylab.com/transportation/2017/04/inside-bostons-bus-revamp/522726/, last accessed December 29, 2017.

42. Oregon State University. (2018). "Partnering for better transit planning." (website) Oregon State University, Corvallis, OR. Available online: http://mime.oregonstate.edu/partnering-better-transit-planning, last accessed February 16, 2018.

43. Bowling, C. (2016). "MARTA's smartphone app would replace Breeze card," *The Atlanta Journal-Constitution*, The Cox Media Group, Atlanta, GA. Available online: http://www.myajc.com/news/transportation/marta-smartphone-app-would-replace-breeze-card/8wGUfLCefoHP3RzqvV8ToM/, last accessed December 29, 2017.

44. Kressner, J. and Garrow, L. (2014). "Lifestyle Segmentation Variables as Predictors of Home-Based Trips for Atlanta, Georgia, Airport," *Transportation Research Record 2266*, pp. 20–30, Transportation Research Board, Washington, DC. Available online: https://doi.org/10.3141/2266-03, last accessed December 29, 2017.

45. Binder, S., Macfarlane, G., Garrow, L., and Bierlaire, M. (2014). "Associations among household characteristics and emissions failures: An application of targeted marketing data," *Transportation Research Part A: Policy and Practice, 59*, pp. 122–133, Elsevier, Oxford, United Kingdom. Available online: https://doi.org/10.1016/j.tra.2013.11.005, last accessed December 29, 2017.

46. Sweet, M.N. (2014). "Do firms flee traffic congestion?," *Journal of Transport Geography*, *35*, pp. 40–49. Elsevier, Oxford, United Kingdom. Available online: http://www.sciencedirect.com/science/article/pii/S0966692314000064, last accessed December 29, 2017.

47. Macfarlane, G., Garrow, L., and Moreno-Cruz, J. (2015). "Do Atlanta residents value MARTA? Selecting an autoregressive model to recover willingness to pay," *Transportation Research Part A: Policy and Practice*, *78*, pp. 214–230, Elsevier,

Oxford, United Kingdom. Available online: http://www.sciencedirect.com/science/article/pii/S096585641500138X, last accessed December 29, 2017.

48. Kressner, J. (2017). *Synthetic Household Travel Data Using Consumer and Mobile Phone Data*, NCHRP IDEA Project 184, Transportation Research Board, Washington, DC. Available online: http://onlinepubs.trb.org/onlinepubs/IDEA/FinalReports/Highway/NCHRP184_Final_Report.pdf, last accessed December 29, 2017.

49. North Carolina Statewide Model Users Group. (2006). *North Carolina Statewide Model Users Group Meeting April 26, 2006 in Raleigh North Carolina*, North Carolina Department of Transportation, Raleigh, NC. Available online: https://connect.ncdot.gov/projects/planning/TPB%20Model%20User%20Groups/Group%20Meeting%20-%20April%2026,%202006%20Notes.pdf, last accessed December 29, 2017.

50. WSP Parsons Brinkerhoff, Inc., ECONorthwest, HBA Specto Incorporated, and Resource Systems Group, Inc. (2017). *SWIM Version 2.5 Model Development Report*, Donnelly, R. (ed.), Oregon Department of Transportation, Salem, OR. Available online: http://www.oregon.gov/ODOT/Planning/Documents/Statewide-Integrated-Model-Vers2-5.pdf, last accessed December 29, 2017.

51. Greenwood, S., Perrin, A., and Duggan, M. (2016). "Social Media Update 2016." (website) Pew Research Center, Washington, DC. Available online: http://www.pewinternet.org/2016/11/11/social-media-update-2016/, last accessed December 29, 2017.

52. Zhang, B. (2014). *Development of a Social Weights Matrix to Consider Friendship Influences on Air Travel*, Master's Thesis, Georgia Institute of Technology, Atlanta, GA. Available online: https://smartech.gatech.edu/handle/1853/52312, last accessed December 29, 2017.

53. Yin, M., Sheehan, M., Feygin, S., Paiement, J-F., and Pozdnoukhov, A. (2017). "A Generative Model of Urban Activities from Cellular Data," *Institute of Electrical and Electronics Engineers Transactions on Intelligent Transportation Systems*, *PP*(99), pp. 1–15, IEEE Intelligent Transportation Systems Society, New York, NY. Available online: http://ieeexplore.ieee.org/abstract/document/7932990/?reload=true, last accessed December 29, 2017.

54. Picornell, M., Ruiz, T., Lenormand, M., Ramasco, J.J., Dubernet, T., and Frías-Martínez, E. (2015). "Exploring the potential of phone call data to characterize the relationship between social network and travel behavior," *Transportation*, *42*(4), pp. 647–668, Springer US, New York, NY. Available online: http://dx.doi.org/10.1007/s11116-015-9594-1, last accessed December 29, 2017.

55. Transport Foundry. (n.d.). "Synthetic Travel Diaries from Passive Data." (website) TRB Innovations Deserving Exploratory Analysis Program (NCHRP-184), Transport Foundry, Atlanta, GA. Available online: http://projects.transportfoundry.com/trbidea/, last accessed December 29, 2017.

56. Center for Geospatial Information Science. (n.d.). “Award from the Federal Highway Administration.” (website) University of Maryland, College Park, MD. Available online: https://geospatial.umd.edu/news/award-federal-highway-administration, last accessed December 29, 2017.

57. Office of Highway Policy Information. (2017). “Highway Statistics Series.” (website) Federal Highway Administration, Washington, DC. Available online: https://www.fhwa.dot.gov/policyinformation/statistics.cfm, last accessed December 29, 2017.

58. Office of Highway Policy Information. “Highway Performance Monitoring System (HPMS).” (website) Federal Highway Administration, Washington, DC. Available online: https://www.fhwa.dot.gov/policyinformation/hpms.cfm, last accessed December 29, 2017.

59. Office of Highway Policy Information. (n.d.). “Motor Fuel Data and the Highway Trust Fund.” (website) Federal Highway Administration, Washington, DC. Available online: https://www.fhwa.dot.gov/policyinformation/motorfueldata.cfm, last accessed December 29, 2017.

60. Exploratory Advanced Research Program. (2016). “About the Exploratory Advanced Research Program.” (website) Federal Highway Administration, Washington, DC. Available online: https://www.fhwa.dot.gov/advancedresearch/about.cfm#pro, last accessed December 29, 2017.

61. U.S. Department of Transportation. (2017). “Understanding the Office of the Assistant Secretary for Research and Technology (OST-R).” (website) U.S. Department of Transportation, Washington, DC. Available online: https://www.transportation.gov/transition/assistant-secretary-for-research-and-technology, last accessed December 29, 2017.

62. Bureau of Transportation Statistics. (n.d.). “National Transportation Statistics.” (website) U.S. Department of Transportation, Washington, DC. Available online: https://www.bts.gov/product/national-transportation-statistics, last accessed July 26, 2019.

63. Bureau of Transportation Statistics. (n.d.). “Airline On-Time Tables—December 2015.” (website) U.S. Department of Transportation, Washington, DC. Available online: https://www.bts.gov/topics/airline-time-tables, last accessed July 26, 2019.

64. National Highway Traffic Safety Administration. (1997). *NHTSA 2020 Report: People Saving People*, U.S. Department of Transportation, Washington, DC. Available online: https://one.nhtsa.gov/nhtsa/whatis/planning/2020Report/2020report.html, last accessed December 29, 2017.

65. Schroeder, P. and Wilbur, M. (2013). *2012 National Survey of Bicyclist and Pedestrian Attitudes and Behavior Volume 1: Summary Report*, Report No. DOT HS 811 841 A, U.S. Department of Transportation, Washington, DC. Available online: https://www.nhtsa.gov/sites/nhtsa.dot.gov/files/811841a.pdf, last accessed December 29, 2017.

66. Federal Railroad Administration. (2018). "Federal Railroad Administration Office of Safety Analysis Home Page." (website) U.S. Department of Transportation, Washington, DC. Available online: http://safetydata.fra.dot.gov/OfficeofSafety/default.aspx, last accessed December 29, 2017.

67. USA.gov. (n.d.). "U.S. Department of Energy." (website) U.S. Department of Energy, Washington, DC. Available online: https://www.usa.gov/Federal-agencies/u-s-department-of-energy, last accessed December 29, 2017.

68. Alternative Fuels Data Center. (2017). "Maps and Data." (website) U.S. Department of Energy, Washington, DC. Available online: http://www.afdc.energy.gov/data, last accessed December 29, 2017.

69. U.S. Census Bureau. (n.d.). "American Community Survey (ACS)." (website) U.S. Census Bureau, Washington, DC. Available online: https://www.census.gov/programs-surveys/acs/, last accessed December 29, 2017.

70. U.S. Census Bureau. (2017). "American Community Survey (ACS): Subjects Included in the Survey." (website) U.S. Census Bureau, Washington, DC. Available online: https://www.census.gov/programs-surveys/acs/guidance/subjects.html, last accessed December 29, 2017.

71. Office of Planning, Environment, and Realty. (2017). "Census Transportation Planning Products." (website) Federal Highway Administration, Washington, DC. Available online: https://www.fhwa.dot.gov/planning/census_issues/ctpp/, last accessed December 29, 2017.

72. Bureau of Labor Statistics. (2018). "About BLS." (website) U.S. Department of Labor, Washington, DC. Available online: https://www.bls.gov/bls/infohome.htm, last accessed December 29, 2017.

73. Bureau of Labor Statistics. (2017). *American Time Use Survey—2016 Results*, News Release USDL-17-0880, U.S. Department of Labor, Washington, DC. Available online: https://www.bls.gov/news.release/pdf/atus.pdf, last accessed December 29, 2017.

74. Transportation Research Board. (2018). "National Highway Cooperative Research Program." (website) Transportation Research Board, Washington, DC. Available online: http://www.trb.org/NCHRP/NCHRPOverview.aspx, last accessed December 29, 2017.

75. Albright, C. (2018). *21st Century Transportation Planning in Lake Tahoe Using Cellular Mobility Data Analytics*, Innovation in Travel Modeling (ITM) Research Brief, Presented at the 7th TRB Innovations in Travel Modeling Conference, Atlanta, GA. Available online: https://www.dropbox.com/s/7idfa41c440catk/_ITM2018_InnovationsInTravelModelingITMResearchBrief.pdf?dl=0, last accessed July 16, 2019.

76. Chen, X., Wan, X., Ding, F., Li, Q., McCarthy, C., Cheng, Y., and Ran, B. (2019). "Data-Driven Prediction System of Dynamic People-Flow in Large Urban Network Using Cellular Probe Data." *Journal of Advanced Transportation*, *2019*(49), pp. 1–12.

77. Chen, X., Wan, X., Li, Q., Ding, F., McCarthy, C., Cheng, Y., and Ran, B. (2019). “Trip-Chain-Based Travel-Mode-Shares-Driven Framework Using Cellular Signaling Data and Web-Based Mapping Service Data.” *Transportation Research Record 2673*, (3), pp. 51–64, Transportation Research Board, Washington, DC. Available online: https://doi.org/10.1177/0361198119834006, last accessed July 16, 2019.

78. King, B. (2018). *Utilization of Mobile Device Data for Model Validation and Development*, Presented at the 7th TRB Innovations in Travel Modeling Conference, Atlanta, GA. Available online: http://onlinepubs.trb.org/onlinepubs/Conferences/2018/ITM/BKing.pdf, last accessed July 16, 2019.

79. Macfarlane, G. (2017). *Modeling Automated Vehicles with a Passive Data Model*, Presented at the 16th TRB Planning Application Conference, Raleigh, NC. Available online: https://www.trbappcon.org/oldsite/2017conf/PresentationDetailscbdc.html?abstractid=248, last accessed July 16, 2019.

80. White, L.D. (2019). *Empirical Analysis of External Travel and State Highway 130 in Austin, Texas Using Cellular Data*, Presented at the 17th TRB Planning Application Conference, Portland, OR.

81. Alho, A., Sakai, T., Bhavathathran B.K., Jeong, K. and Ben-Akiva, M. (2018). *Revealing Freight Vehicle Tours and Tour Patterns from GPS Vehicle Tracking and Driver Survey Data*, Presented at the 7th TRB Innovations in Travel Modeling Conference, Atlanta, GA. Available online: http://onlinepubs.trb.org/onlinepubs/Conferences/2018/ITM/AAlho2.pdf, last accessed July 16, 2019.

82. Arnold, R. (2019). *Practical Application of Archived Probe Vehicle Data*, Presented at the 17th TRB Planning Application Conference, Portland, OR.

83. Aryal, S. and Varanasi, S. (2018). *I-295 Truck Corridor Forecasts Development: Richmond, VA*, Presented at the 7th TRB Innovations in Travel Modeling Conference, Atlanta, GA. Available online: http://onlinepubs.trb.org/onlinepubs/Conferences/2018/ITM/SAryal.pdf, last accessed July 16, 2019.

84. Bernardin, V., Sadrsadat, H., and Chen J. (2018). *Combining NHTS and Passive OD Data for Charleston, SC,* Presented at the 2018 NHTS Workshop, Washington, DC.

85. Cooper, D., Castiglione, J., Wilson, C., and Mislove, A. (2018). *TNCs Today: A. Profile on San Francisco Transportation Network Company Activity*, Presented at the 7th TRB Innovations in Travel Modeling Conference, Atlanta, GA. Available online: http://onlinepubs.trb.org/onlinepubs/Conferences/2018/ITM/DCooper.pdf, last accessed July 16, 2019.

86. Giaimo, G. (2018). *A Suite of Model Updating and Validation Procedures Using Third Party Origin-Destination Data*, Presented at the 7th TRB Innovations in Travel Modeling Conference, Atlanta, GA. Available online: http://onlinepubs.trb.org/onlinepubs/Conferences/2018/ITM/GGiaimo2.pdf, last accessed July 16, 2019.

87. Hong, S. (2017). *Expanding the Uses of Truck GPS Data in Transportation Planning and Analysis*, Presented at the 16th TRB Planning Application Conference, Raleigh, NC. Available online: https://www.trbappcon.org/oldsite/2017conf/PresentationDetailsae44.html?abstractid=284, last accessed July 16, 2019.

88. Kuppam, A. (2017). *Tour-Based Truck Travel Models Using Truck GPS Data*, Presented at the 16th TRB Planning Application Conference, Raleigh, NC. Available online: https://www.trbappcon.org/oldsite/2017conf/PresentationDetails766c.html?abstractid=141, last accessed July 16, 2019.

89. Mansfield, T. (2019). *Identifying Truck Stop and Service Station Locations to Improve Passively Collected Commercial Truck GPS Data*, Presented at the 17th TRB Planning Application Conference, Portland, OR.

90. McAtee, S. (2017). *Validating Trip Distribution in Southeast Michigan Using GPS Data*, Presented at the 16th TRB Planning Application Conference, Raleigh, NC. Available online: https://www.trbappcon.org/oldsite/2017conf/PresentationDetailscc6b.html?abstractid=171, last accessed July 16, 2019.

91. Morgan, D. (2018). *Opportunities for Regional DTA Validation with Mobile Data*, Presented at the 7th TRB Innovations in Travel Modeling Conference, Atlanta, GA. Available online: http://onlinepubs.trb.org/onlinepubs/Conferences/2018/ITM/DMorgan.pdf, last accessed July 16, 2019.

92. Schewel, L. (2017). *Approaches to Evaluating VMT using GPS-based Probe Vehicle Data*, Presented at the 16th TRB Planning Application Conference, Raleigh, NC. Available online: https://www.trbappcon.org/oldsite/2017conf/PresentationDetailse0b1.html?abstractid=339, last accessed July 26, 2019.

93. Sharman, B. (2017). *Analysis of GPS Fleet Tracking Data to Infer Commercial Vehicle Travel Patterns in Ontario*, Presented at the 16th TRB Planning Application Conference, Raleigh, NC. Available online: https://www.trbappcon.org/oldsite/2017conf/PresentationDetails5a15.html?abstractid=213, last accessed July 16, 2019.

94. Trevino, S. (2019). *Using Passive GPS Data to Understand Truck Flows in Nebraska*, Presented at the 17th TRB Planning Application Conference, Portland, OR.

95. Hu, X. (2017). *A Multi-Resolution Approach in Investigating the Impacts of Pre-Planned Road Capacity Reduction Based on Smartphone Trajectory Data: A Case Study of Lane Closure Event on Mopac Expressway, Austin TX*, Presented at the 16th TRB Planning Application Conference, Raleigh, NC. Available online: https://www.trbappcon.org/oldsite/2017conf/PresentationDetails46c3.html?abstractid=258, last accessed July 16, 2019.

96. Lam, C.P. (2019). *Visualization of Truck GPS Origin-Destination*, Presented at the 17th TRB Planning Application Conference, Portland, OR.

97. Lynch, J., Dumont, J., Greene, E., and Ehrlich, J. (2019). "Use of a Smartphone GPS Application for Recurrent Travel Behavior Data Collection." *Transportation Research Record*, Transportation Research Board, Washington, DC. Available online: https://doi.org/10.1177/0361198119848708, last accessed July 16, 2019.

98. Ramirez, T. (2017). *Getting the Most out of Your Data: Applying Passively Collected Travel Demand Management Data to Transportation Planning in El Paso*, Presented at the 16th TRB Planning Application Conference, Raleigh, NC. Available online: https://www.trbappcon.org/oldsite/2017conf/PresentationDetails636d.html?abstractid=139, last accessed July 16, 2019.

99. Sadeghvaziri, E. (*2017). Driving Activity Locations Inferred from Smartphone Data*, Presented at the 16th TRB Planning Application Conference, Raleigh, NC. Available online: https://www.trbappcon.org/oldsite/2017conf/PresentationDetailsafef.html?abstractid=277, last. accessed July 16, 2019.

100. Smith, C. (2019). *Using Multiple Years of Truck GPS Data for Freight Model Development and Validation*, Presented at the 17th TRB Planning Application Conference, Portland, OR.

101. Chigoy, B. (2019). *Identifying Truck Bias in LBS Data*, Presented at the 17th TRB Planning Application Conference, Portland, OR.

102. Flake, L., Bernardin, V., Greene, E., Jeffrey Dumont, J., and Anderson, R. (2018). *Insights from Big and Small Data: Which Trips and Travelers are Captured by Location-Based Services Data?* Presented at the 7th TRB Innovations in Travel Modeling Conference, Atlanta, GA. Available online: http://onlinepubs.trb.org/onlinepubs/Conferences/2018/ITM/LFlake.pdf, last accessed July 16, 2019.

103. Freedman, J., Dhakar, N., Curry, R., and Sun, W. (2018). *Use of Big Data to Calibrate and Validate Travel to Special Travel Destinations*, Presented at the 7th TRB Innovations in Travel Modeling Conference, Atlanta, GA. Available online: http://onlinepubs.trb.org/onlinepubs/Conferences/2018/ITM/WSun.pdf, last accessed July 16, 2019.

104. Manzo, C. (2019). *Using Big Data to Understand Travel Behavior to Parks*, Presented at the 17th TRB Planning Application Conference, Portland, OR.

105. Su, Y. (2019). *Using Travel Time and Origin-Destination Data in Transportation Planning: A Metropolitan Planning Organization's Example*, Presented at the 17th TRB Planning Application Conference, Portland, OR.

106. Sana, B. (2017). *Using Google's Aggregated and Anonymized Trip Data to Estimate Dynamic Origin-Destination Matrices for San Francisco*, Presented at the 16th TRB Planning Application Conference, Raleigh, NC. Available online: https://www.trbappcon.org/oldsite/2017conf/PresentationDetails27b8.html?abstractid=279, last accessed July 26, 2019.

107. Schewel, L. (2019). *Activating Big Data for Active Transportation with a Statewide Data Platform*, Presented at the 17th TRB Planning Application Conference, Portland, OR.

108. Viswanathan, K. (2019). *Understanding Travel Using Location Based Services Data*, Presented at the 17th TRB Planning Application Conference, Portland, OR.

109. Avner, J. (2019). *Finding Uses for Temporal Elements of Big Data*, Presented at the 17th TRB Planning Application Conference, Portland, OR.

110. Avner, J. (2018). *Using Big Data in Small and Medium Sized Regions: Three Case Studies and Lessons Learned*, Presented at the 16th National TRB Tools of the Trade Conference, Kansas City, MO. Available online: http://onlinepubs.trb.org/onlinepubs/Conferences/2018/Tools/JAvner.pdf, last accessed July 16, 2019.

111. Bernardin, V. and Chen, S. (2019). *An Agile, Data-Driven Ensemble Modeling Framework for the Illinois Statewide Travel Demand Model*, Presented at the 17th TRB Planning Application Conference, Portland, OR.

112. Bernardin, V. (2018). *Overview of Methods for Validation and Expansion of Passive Origin-Destination Data*, Presented at the 7th TRB Innovations in Travel Modeling Conference, Atlanta, GA. Available online: http://onlinepubs.trb.org/onlinepubs/Conferences/2018/ITM/VBernardin2.pdf, last accessed July 16, 2019.

113. Chen, J., Bernardin, V., and Lee, Y. (2018). *Incorporating Big Data in an Activity-Based Travel Model: The Chattanooga Case Study*, Presented at the 7th TRB Innovations in Travel Modeling Conference, Atlanta, GA. Available online: http://onlinepubs.trb.org/onlinepubs/Conferences/2018/ITM/JChen.pdf, last accessed July 16, 2019.

114. Chigoy, B. (2018). *GPS and Cell Data for Medium and Small Cities*, Presented at the 16th National TRB Tools of the Trade Conference, Kansas City, MO. Available online: http://onlinepubs.trb.org/onlinepubs/Conferences/2018/Tools/BChigoy-1.pdf, last accessed July 16, 2019.

115. Cui, Y., Meng, C., He, Q., and Gao, J. (2018). "Forecasting Current and Next Trip Purpose with Social Media Data and Google Places." *Transportation Research Part C: Emerging Technologies*, *97*, pp. 159–174, Transportation Research Board, Washington, DC. Available online: https://doi.org/10.1016/j.trc.2018.10.017, last accessed July 16, 2019.

116. Darwiche, A. and Fox, J. (2018). *A Proposed Tool for Transit Fare Structure Evaluation: Integration of Smart Card Data with Regional Demand Models*, Presented at the 7th TRB Innovations in Travel Modeling Conference, Atlanta, GA. Available online: http://onlinepubs.trb.org/onlinepubs/Conferences/2018/ITM/ADarwiche.pdf, last accessed July 16, 2019.

117. Delahanty, R. (2019). *Using Big Data in Freeway Corridor Studies*, Presented at the 17th TRB Planning Application Conference, Portland, OR.

118. Flood, J. (2019). *Using Big Data to Estimate Weekend VMT for Crash Analysis*, Presented at the 17th TRB Planning Application Conference, Portland, OR.

119. Kressner. J. and Macfarlane, G. (2018*). Passive Data Modeling: A Method for Fusing Passive Data and Survey*, Presented at the 7th TRB Innovations in Travel Modeling Conference, Atlanta, GA.

120. Norton, T. (2019). *Using Big Data to Explore Long-Distance Freight Travel on Non-interstate Corridors in Texas*, Presented at the 17th TRB Planning Application Conference, Portland, OR.

121. Padgett, T. (2018). *Big Data – From Corridors to Regions*, Presented at the 16th National TRB Tools of the Trade Conference, Kansas City, MO. Available online: http://onlinepubs.trb.org/onlinepubs/Conferences/2018/Tools/TPadgett.pdf, last accessed July 16, 2019.

122. Rousseau, G. (2018). *Model Validation Using "Novel" Big Data*, Presented at the 7th TRB Innovations in Travel Modeling Conference, Atlanta, GA. Available online: http://onlinepubs.trb.org/onlinepubs/Conferences/2018/ITM/GRousseau2.pdf, last accessed July 16, 2019.

123. Sadrsadat, H., Bernardin, V., and Avery, D. (2018). *Understanding Regional Travel Patterns with Big Data*, Presented at the 7th TRB Innovations in Travel Modeling Conference, Atlanta, GA. Available online: http://onlinepubs.trb.org/onlinepubs/Conferences/2018/ITM/HSadrsadat.pdf, last accessed July 16, 2019.

124. Schiffer, R. (2017). *Alternate Methodologies for Origin-Destination Data Collection*, Presented at the 16th TRB Planning Application Conference, Raleigh, NC. Available online: https://www.trbappcon.org/oldsite/2017conf/PresentationDetailsddb2.html?abstractid=236, last accessed July 16, 2019.

125. Schewel, L. (2018). *Big Data "Triage" Before Modeling*, Presented at the 7th TRB Innovations in Travel Modeling Conference, Atlanta, GA. Available online: http://onlinepubs.trb.org/onlinepubs/Conferences/2018/ITM/LSchewel.pdf, last accessed July 16, 2019.

126. Schewel, L. (2018). *Vehicle Emissions and Air Quality: How Big Data Can Help*, Presented at the 16th National TRB Tools of the Trade Conference, Kansas City, MO. Available online: http://onlinepubs.trb.org/onlinepubs/Conferences/2018/Tools/LSchewel.pdf, last accessed July 16, 2019.

127. Trevino, S. and Bernardin, V. (2018*). Development of an Agile & Data-Driven Model Framework*, Presented at the 7th TRB Innovations in Travel Modeling Conference, Atlanta, GA. Available online: http://onlinepubs.trb.org/onlinepubs/Conferences/2018/ITM/STrevino.pdf, last accessed July 16, 2019.

128. Thye, C. (2018). *Applying Big Data to Small Projects*, Presented at the 16th National TRB Tools of the Trade Conference, Kansas City, MO. Available online: http://onlinepubs.trb.org/onlinepubs/Conferences/2018/Tools/CThye.pdf, last accessed July 16, 2019.

129. Zhang, L. and Ghader, S. (2018). *Data Analytics and Modeling Methods for Tracking and Predicting O-D Travel Trends Based on Mobile Device Data*, Presented at the 7th TRB Innovations in Travel Modeling Conference, Atlanta, GA.

130. AirSage™. (n.d.). "AirSage Home Page." (website) AirSage™, Atlanta, GA. Available online: https://airsage.com, last accessed December 29, 2017.

131. INRIX™. (2017). "INRIX Home Page." (website) INRIX™, Kirkland, WA. Available online: http://inrix.com/, last accessed December 29, 2017.

132. HERE™. (2018). "Here Home Page." (website) HERE™, Chicago, IL. Available online: https://here.com/en, last accessed December 29, 2017.

133. InfoGroup. (2018). "InfoUSA Home Page." (website) InfoGroup, Papillion, NE. Available online: https://www.infousa.com/, last accessed December 29, 2017.