Probe Data and Transportation Statistics: Definitions, Uses, and an Approach to Classification

U.S. Department of Transportation
Office of the Secretary of Transportation

Bureau of Transportation Statistics

About the Bureau of Transportation Statistics

Leadership

Rolf R. Schmitt, *Deputy Director*, *Office of Director*

Publication Management

Ed Strocko, *Director*, *Office of Spatial Analysis and Visualization*

About This Report

Project Manager

Kyle Titlow, *Spatial Statistician*, *Office of Spatial Analysis and Visualization*

Major Contributors

BTS: Kyle Titlow, Mindy Liu, Ed Strocko, Rolf Schmitt, and Devon Lechtenberg
MacroSys: Stacey Bricka
WRA: Byron Chigoy and Krishan Viswanathan

Editor

Mikki Stacey (MacroSys)

Visualization Contributors

N/A

Other Contributors

N/A

Report DOI

10.21949/2nq2-sn30

Publication Date

June 2026

Title

Probe Data and Transportation Statistics: Definitions, Uses, and an Approach to Classification

Key Words

Probe data; taxonomy; location-based services; LBS; Global Positioning System; GPS; connected vehicles; CVs; electronic logging device; ELD; cellular data; call data record; CDR; point of interest; POI

Abstract

Probe data are increasingly being adopted by transportation professionals in the United States as a means to better measure the mobility of Americans across and within the national transportation system. Centered on new, rapidly changing forms of technology that allow mobile devices to actively determine and share their locations with analysts as a novel form of big data, probe data offer exciting advantages for measuring mobility, but they rely on technologies and methods that are not transparent to their users or members of the public. This report offers a statistics-centered definition of probe data, contrasts them with nonprobe data, notes common data subtypes, discusses their advantages and limitations, summarizes their history, and proposes ways to differentiate noted subtypes. This report accompanies a set of six user guides on specific probe data types.

Recommended Citation

United States Department of Transportation, Bureau of Transportation Statistics. *Probe Data and Transportation Statistics: Definitions, Uses, and an Approach to Classification*. Washington, DC: 2026. https://doi.org/10.21949/2nq2-sn30.

BTS information service contact information:

Ask-A-Librarian https://transportation.libanswers.com/

Phone 202-366-DATA (3282)

Quality Assurance Statement

The Bureau of Transportation Statistics (BTS) provides high quality information to serve government, industry, and the public in a manner that promotes public understanding. Standards and policies are used to ensure and maximize the quality, objectivity, utility, and integrity of its information. BTS reviews quality issues on a regular basis and adjusts its programs and processes to ensure continuous quality improvement.

Notice

Table of Contents

List of Figures

List of Tables

1. Introduction

To meet its mandate to provide timely, accurate, and credible transportation data on the movement of people and goods in the United States, the Bureau of Transportation Statistics (BTS) promotes innovative methods of travel-data collection, analysis, visualization, and dissemination [BTS 2019].

One source of data that has been at the forefront of many such recent innovations is probes, or the spatially intelligent phones, vehicles, and gadgets that have become ubiquitous among the American population in the last decade. Spurred by technological advances that have made it possible for a wide array of devices to actively determine and emit their own location, probes produce a constant stream of location-rich sightings that illustrate the movements and transportation choices of their holders, information traditional data-collection methods cannot capture to a similar extent. In fact, the pervasiveness of probes, especially smart phones, in daily life has not only transformed the ways people live, work, shop, and recreate but also spurred an entirely new source of data on observed mobility, including new patterns of movement that may not have happened without their availability.

In response to these evolutions, this report evaluates the use and quality of these probe data as a source of mobility information, especially in the context of generating national-level transportation statistics. Probe data collection is ultimately a core area of interest in the U.S. Department of Transportation's (USDOT's) and BTS' efforts to embrace innovative methods that can help improve the ways government assesses, measures, manages, and operates the Nation's transportation systems. Given the significant and distinct advantages probe data offer, especially their ability to measure facets of observed mobility that are invisible to traditional transportation-measurement tools, BTS seeks to evaluate their usability for transportation statistics.

Defining Probe Data: Probe data are generated by technological devices (i.e., probes) that are either carried by individuals (e.g., cell phones) or located on or in vehicles, vessels, or other conveyances (e.g., Global Positioning Systems and Electronic Logging Devices). The devices that produce these data actively determine and emit information about their locations. That information (i.e., probe data)—when monitored, collected, stored, and analyzed—can reveal the movement of each person or conveyance carrying a probe through space and across time even though probe data sightings do not inherently contain identifying information about the device or its holder.

Departmental interest in using probe data has increased in the last 5–10 years due to the following key advantages they offer:

- Relatively inexpensive to procure
- Available for extended periods across time
- Available at a wide range of spatial and temporal resolutions, including many smaller areas of interest not otherwise measurable
- Scalable to meet analysis needs in a short time horizon

A probe, especially one that is mobile and has a device-specific identifier to trace its movement, can produce a trail of location data points that, when collected with consent, reveals much about the mobility and transportation of the device and, therefore, the person or entity to whom it belongs. Tracing such movements, including those on identifiable modes of transport, is possible even when the probe is not explicitly generating data for a transportation-specific purpose; in fact, only half of the probe types listed in Exhibit 1 were explicitly created for transportation use cases or intended for collection and ex-post-facto analysis or pattern detection.

Beyond proffering a definition of probe data, this document also juxtaposes them against nonprobe data-collection methods, such as those that require individual recall (e.g., surveys) and are separated from the flow of vehicles or entities being measured (e.g., highway traffic counters, traffic counters, and satellite imagery). This comparison is necessary because nonprobe data-collection methods, especially surveys, have long histories as trusted sources of transportation data. Increasing transparent understanding of probe data is necessary to be certain that they can truly enhance or replace any traditional methods. This exercise is imperative if the resulting mobility metrics are to meaningfully inform policy decisions and research.

A secondary reason for contrasting probe and nonprobe data is the growing recognition that they are not mutually exclusive and can be fused together to get a wholistic understanding of mobility. Perhaps a surprise given their apparent differences, especially comparing high-tech, Global Positioning System [GPS]-based probes against paper surveys sent through the mail, their respective advantages can be complementary. For example, probes quickly offer much larger sampling frames than high-cost, low-response surveys while needing the results of those surveys to ground-truth any patterns detected in their data. Taking full advantage of these synergies is only possible with a greater understanding of probes and the location-rich data they produce.

In summary, this document defines and summarizes probe data—providing use-case examples and a brief history of the data—and then proposes an approach for categorizing known probe data types. This content is intended to help BTS and other federal agencies better understand probe data sources and assess their applicability to meet each agency's data needs.

The remainder of this chapter explains the definitions of probe and nonprobe data. Chapter 2 provides example use cases and explores data gaps. Chapter 3 summarizes existing barriers to the adoption of probe data for use in transportation statistics. Chapter 4 summarizes the history of probe data. Finally, Chapter 5 defines key terms and considerations for a probe data taxonomy.

1.1. BTS' DEFINITION OF PROBE DATA

Establishing a clear definition of "probe data" and its subclassifications was a necessary first step to determine the scope of the research needed to assess data quality and use cases. BTS thoroughly evaluated the many characteristics of probe data to decide what to include or exclude from its definition.

BTS' definition of probe data hinges on active location detection because this characteristic is necessary and sufficient to distinguish probe data from other forms of remotely detected transportation data, the location information of which is wholly determined by the technology embedded in a receiver rather than through interplay between the target and the receiver. In other words, probe data are defined by an in-situ device that receives the trigger and does all the work to estimate the target's location, whether that be assigning it its own location (i.e., the coordinates of a traffic camera, license plate reader, or toll tag transponder) or using some characteristic of the triggering event to guess the target's relative location. Other elements were carefully excluded from its definition. While all probe data types necessarily involve some element of remote detection, for example, this characteristic is not sufficient to define probe data because multiple nonprobe data types (e.g., license plate readers and Bluetooth receivers) involve passive remote detection. Another consideration was the intent of data collection vis-à-vis transportation. While all probe data are inherently collected with some intention, they do not need to be collected with transportation in mind. Receivers for some types of probe data may be installed for transportation purposes (e.g., infrastructure-based receivers for connected vehicles [CVs]), but often, they are not (e.g., cell phone towers, 5G receivers, GPS satellite constellations). Thus, the intent of data collection is not necessary to define the types of probe data relevant to the generation of transportation statistics.

BTS avoids using the terms "active data" or "passive data" as wholesale synonyms for "probe data" because other features of probes and probe data, beyond location detection, can also be described as "active" and "passive." For example, the mechanisms that trigger a probe to actively find its location can have active or passive elements. A cell phone application (app) or CV, for example, can produce a location sighting while in active use by its user or as a regularly scheduled, passive background message; the probe actively determined its location in both cases, even though the triggers were wholly different. Likewise, probe data can be actively or passively received. In the context of data collection, this distinction is dictated by whether receivers actively look for probe messages or passively receive them. Messages from probes, which include the location information the device actively determined already, can be received either way so long as that method is not central to the source's location detection. Accordingly, BTS chose not to use the term "active data" or "passive data" to define probe data to avoid confounding their various characteristics that could also be described with this language.

1.1.1. Types of Probe Data

Exhibit 1 lists major forms of probe data that are relevant to transportation. Each source has a short description linking it to the components of BTS' definition of probe data.

Probe data can be subclassified based on how they are produced and detected. For example, the transportation sector might classify probe data based on whether they are explicitly designed to measure the transportation system (as opposed to probe data that incidentally reveal travel patterns). Probes deliberately implanted into cars or other conveyances are different from cell phones, which incidentally detect transportation vis-à-vis the owner's mobility preferences. Additionally, probes that require user interaction to generate a sighting (e.g., someone making a call or using an app on a phone) are distinct from probes that produce sightings automatically (e.g., vehicle location data that must be emitted over a predetermined unit of time or distance).

Using these definitions and classifications of probe data, BTS established the taxonomy provided in Chapter 5.

Exhibit 1. Common Probe Data Types

GPS-based data emanate from probes embedded with technology that communicates with a constellation of satellites orbiting the earth. Probe data types that rely on GPS include commercial vehicle fleet tracking and in-vehicle navigation.

LBS, or Location-Based Services, provide spatial data derived from technologies embedded in cell phones, which produce data points when users interact with apps. Location is determined using a mix of technologies: tracking device location from a phone's GPS unit or spatial information inferred from cellular networks.

CV data include location, vehicle operational, and safety information emitted by personal and nonpersonal vehicles with embedded telematics technology that can interact with the built environment (vehicle-to-everything technology); location data come from the GPS component of their telematics devices.

ELDs, or Electronic Logging Devices, are attached to commercial vehicles to track their operational hours and general compliance with other regulatory requirements; they source their locations from GPS technology embedded in the devices.

Cellular CDRs, or Call Data Records, are produced from the interaction between a cellular device and consecutive cellular base stations; cell phones are embedded with technology that logs information that can be used to estimate the device's location.

Cellular Location data are derived from a device's location relative to cellular base stations, relying on technology in the phone that logs location-aware data points. In contrast to CDRs, these data are based on the device's actual location rather than the locations of base stations.

POI, or Point of Interest, data are derived, with user consent, from location information emitted from the GPS functionality embedded in smart phones and generated as a user interacts with apps on their device. The output data are aggregated to specific building footprints or business locations, typically for use in understanding activity patterns of businesses.

1.1.2. Contrast With Nonprobe Data

BTS recognizes that many data types share probe and nonprobe characteristics and that they are not mutually exclusive. Smart phones, for example, are probes that produce more than one data type; they are capable of producing some data specifically to measure a mode of transportation (e.g., traffic monitoring, transit route wayfinding, walking or fitness tracking) while producing other types that incidentally capture transportation trends (e.g., any location-aware smart phone app that regularly logs coordinates and locations detected during calls or texts using signals from adjacent cell phone towers). Relatedly, a vehicle can emit its location information to road-adjacent receivers, by which the vehicle acts a data-emitting probe, while passing a license-plate reader or a person counting vehicles at an intersection, both of which would note the vehicle's presence without it playing an active role (i.e., as a form of nonprobe data).

Examples of nonprobe data used for transportation purposes include the following:

- Survey data, which are obtained by reaching out to a sample of respondents to solicit their inputs on how, when, where, why, and with whom they traveled or what goods they transported, are nonprobe data because the location information they generate, along with any other transportation and nontransportation information, is based on recall rather than from a device associated with the respondent. Examples of transportation surveys include the National Household Travel Survey (NHTS), the American Community Survey, Census Transportation Planning Products, and the Commodity Flow Survey, among others. Surveys are sometimes paired with cell phone apps that record the respondents' locations to supplement their responses to a core questionnaire. These apps are a form of probe data comparable to location-based services (LBS).

- Synthetic data are formed when multiple datasets are mathematically joined or modeled to create new datasets. These data are primarily nonprobe since no device-based location detection is involved. They also have no human-intervention component or remote-detection of data points. Examples of synthetic data include the public Longitudinal Employer-Household Dynamics, Replica (a commercial modeled data product), and other sources. Nevertheless, probe-based datasets can be inputs for these products.
- Sensor data include traffic counts, weigh-in-motion data, pedestrian counters, and other data from intelligent transportation systems that are explicitly designed to capture transportation-system characteristics or travel behavior. Even if they have a remote-detection component, these data are not probe data because the primary location-detection element is not done by a device moving with a person, vehicle, or conveyance. Location is primarily determined by the sensor, not the object being sensed.

1.2. OTHER DEFINITIONS OF PROBE DATA

This section discusses other ways probe data have been defined and classified in the literature, offering alternate perspectives on their applicability to the transportation domain.

1.2.1. Federal Highway Administration

The Federal Highway Administration (FHWA) defines probe data as follows:

> Probe data is defined as data that is generated by monitoring the position of individual vehicles (i.e., probes) over space and time rather than measuring characteristics of vehicles or groups of vehicles at a specific place and time. Probe data systems can be further characterized according to whether (a) they require additional roadside or vehicular infrastructure to support them or take advantage of existing infrastructure in innovative ways, (b) they produce detailed tracks of vehicle movements or only identify vehicle arrival at select points, and (c) they provide data in near real-time or only with significant delay. [Mudge et al. 2013]

This definition is similar to the one adopted here and is inclusive of data types involving GPS units, including LBS data, while excluding those that entail the installation of sensors that measure characteristics of vehicles at a specific place or time (e.g., Bluetooth detectors, toll tags, license plate readers).

1.2.2. Texas A&M Transportation Institute

Another complication for classifying probe data types is the different ways the data sources are packaged and disseminated, especially since so many of them are derived from source data not originally intended for transportation use cases.

The Texas A&M Transportation Institute, under the FHWA Pooled Fund Study: Support for Urban Mobility Analysis, offers one potential two-dimension classification: data processing and data delivery [Singh, Sivaraman, Hard 2022].

The first dimension centers on the degree to which the underlying probe data have been processed to make the data useful for transportation applications and studies:

- **Trip trace, or raw, data** are the most granular data (i.e., device level) on the market. Depending on the data-source technology, this type of data could yield the location(s) (trip end) of an activity, an entire trajectory, or a bread crumb trail of data points across time and space at multiple locations.
- **Transformed data** are aggregated trip trace data that are expanded to the population. These data are often sampled, cleaned, and aggregated to produce data products that can be immediately used for further analysis (while some individual trace data points are useful as received, they often require processing to create something useful at scale). In turn, these data products are aggregated by either time or space (e.g., origin–destination [OD] and point-of-interest [POI] data).
- **Modeled data** are trip trace data that are enriched with other nonpassive data sources, such as demographic, land use, and credit card data, and then processed to develop synthetic travel diaries that replicate actual individual trajectories in an area. These datasets, which avoid pitfalls related to data confidentiality that inhibit the creation of public-facing statistical products using raw trip traces, are subsequently used to model a region's daily travel patterns, analogous to an activity-based model and assigned to the regional transportation network.

The second dimension regards the ways probe data are delivered (and the business models associated with each format):

- **Data provider platforms** from data providers offer one-time purchases and periodic data aggregations from their platforms as part of a subscription. Users get access to trip trace data and transformed data, which they can download and use on their own. These platforms are a good alternative for users without the time, technical skills, or cloud infrastructure to handle raw trip trace data but are accessed from platforms that do not have sophisticated or proprietary analysis tools.
- **Service provider platforms** offer access to probe data transformed by the provider using their proprietary methods. The data, which a user will access through a designated platform, are either modeled or include additional inferences (i.e., about modes of travel or demographic attributes) using data-layering or machine-learning techniques. Some of these platforms allow users to upload aggregation zones or choose road segments from the interactive map on the platform to conduct select link or select zone analysis [Singh, Sivaraman, Hard 2022]. Analyses against the data are done within the platform, rather than be downloaded as one-time purchases or aggregations.
- **Third-party platforms** can host many big data sources alongside each other, allowing users to visualize, summarize, and compare big probe datasets against others with which the user has access and is able to upload or plug into the platform. Most of these platforms offer a data downloader feature that allows bulk data downloads of raw trace data that can subsequently be analyzed in-house.

1.2.3. National Cooperative Highway Research Program

Lastly, the National Cooperative Highway Research Program published *Use of Vehicle Probe and Cellular GPS Data by State Departments of Transportation*, which defines a class of probe data as “timestamped location information collected from cell phones, other mobile devices/tablets, or embedded devices in vehicles (like navigation systems or telematics)” [NCHRP 2021].

The report classifies these probe data sources into two categories:

- **Probe-based speed data**—Speed data associated with specific vehicles (passenger car or truck) on roadways
- **Probe-based OD and/or trajectory data**—Data about the start and stop locations of trips made by vehicles on the roadway network, including the route used for the trip

2. Probe Data Advantages

Probe data, especially when they are used in conjunction with nonprobe data, can help fill in gaps and validate existing datasets.

2.1. APPLICATIONS OF PROBE DATA

Probe data are altogether capable of measuring a wide range of transportation system characteristics, even if they were not originally intended for such a purpose. They also help measure changes in mobility caused by the very technology that defines them as probe data (i.e., smart phones).

Table 1 summarizes common travel metrics and whether they are discernible using probe data. Some measurements are still only available from nonprobe sources.

Table 1. Travel Metrics by Data Sources

Travel metric	Data source
Number of trips and origin or destination (i.e., start or stop location)	Surveys, LBS, GPS, CVs
Number of trips by mode and distance	GPS, CVs, ELDs (car and truck), transit and active transportation (survey)
Vehicle speeds and travel times	GPS,* LBS, Bluetooth (corridor section only), CVs*
Demographic composition of travel and characteristics of the traveler or vehicle (e.g., vehicle classification)	Survey
Regional commodity flows	Survey
Traffic counts (and/or volumes)	Radar, loop detectors, tube traffic counts, probe data**
Regional activity patterns and changes	Surveys, LBS, GPS, CVs

*FHWA provides the National Performance Management Research Data Set (Operations Performance Measurement—FHWA Operations [n.d.]) to state departments of transportation (DOTs) and Metropolitan Planning Organizations for use in developing performance measures [FHWA 2020a].
**Several state DOTs have utilized traffic counts obtained by the algorithmic processing of probe data [Turner, Tsapakis, Koeneman 2020].
ELD = electronic logging device.

Refer to Section 4.3 for a list of programs through which probe data are already in use and making an impact in the transportation sector.

2.2. RELATIONSHIP BETWEEN PROBE AND NONPROBE DATA

Even though the value of probe data is typically framed via the advantages they offer against traditional, nonprobe data sources, the relationship between the two is more complex and increasingly complementary and bilateral. For example, probe data typically cannot provide a complete picture of mobility on their own and must be fused with alternative sources of travel-behavior data, especially because characteristics of the user and vehicle are deliberately scrubbed from probe data before they are made public to ensure confidentiality. Information on reasons and modalities for travel and whether a mode is being taken by an individual or group is not inherent to probe data. Thus, probe data require supplemental data to produce mobility estimates that are consistent, reliable, sustainable, contextual, and maximally useful to the public. In fact, transportation professionals often explicitly combine probe data with nonprobe data to provide more context on travel trends or validate the patterns and insights derived from probe data alone.

Traffic-count data, for example, is a form of nonprobe data that the transportation sector commonly uses to calibrate probe-based estimates. These data come from traffic-counting devices, such as tube counters (Figure 1). These tubes measure traffic by counting the number of axles that cross them. They are implemented throughout the United States for a variety of uses, including federal reporting via FHWA’s Highway Performance Monitoring System as well as local traffic studies. Counts from a tube at a known location can be matched with probe data observed at the same site to expand the latter data to the full population of vehicles.

Figure 1. Pneumatic Traffic-Counting Device Installation

Source: ADOT 2011.

Traffic and other nonprobe counting methods obtain only snapshots of travel at specific locations. These counting methods do not reliably capture more expansive patterns away from where they are installed, missing the frequency of travel by persons and households; trade and commodity flows; and near-real-time, macro-level travel demand and trends. To capture data on these broader elements of the transport system, the industry has traditionally relied on semiregularly conducted household travel surveys, but it is increasingly eager to replace these surveys with probe data as a more efficient and accurate means to fill these critical gaps.

2.3. GAPS ACROSS PROBE AND NONPROBE DATA

While the interplay of probe and nonprobe data can give a broad understanding of the transportation system in a specified region, some gaps will remain—especially regarding elements of the mobility landscape that are poorly understood. Two such elements are international and long-distance travel.

For international travel, surveying foreigners at the port of entry is difficult and expensive, and—even if successful—the information captured is limited. Further, not all forms of probe data will be available on multiple sides of the border (especially those involving cell phone data, the availability of which may be affected by national laws, incompatible data standards, or cell phone providers with different technology requirements).

Long-distance travel is similarly inconsistently captured across transportation data-collection modes. Surveys, for example, rarely capture long-distance travel information, while longitudinal studies of probes are complicated by the tendency of probe data vendors to reset the unique identifiers associated with their devices every time the device or associated vehicle is turned on and off. Observing the same device over a longer tour, especially one where a device might be turned on and off several times before it returns to its home location, is impossible.

More research is needed on the usability of probe data for these two topic areas to determine if they are entirely incompatible.

3. Probe Data Limitations

Though probe data may offer significant advantages for generating national-level transportation statistics, they cannot remedy all the limitations of existing transportation-measurement tools.

Before the advent of probe data, transportation agencies were traditionally in complete control of their data-collection programs, including sample sizes, data-collection methods, and analysis approaches. With this amount of control, agency statisticians and planners could pick and choose their own methods and assumptions, leaving staff with a level of understanding such that they could be certain their data-collection strategies were appropriate for their agency's institutional needs. In contrast, probe data—which often entail advanced, nontransparent technologies from third parties that often require proprietary knowledge to understand—do not offer a similar level of data program quality control despite the insights they are uniquely able to provide when compared to traditional collection methods.

Examples of data-quality loss between traditional and probe data include the following:

- Probe data sources, like GPS, LBS, and CVs, do not collect all the same mobility characteristics as traditional surveys. For example, they do not independently capture trip purpose, travel mode used, or travel party composition.
- Probe data, as a form of big data,[1] require distinct forms of technical expertise, including staff members well versed in computation, data science, and data engineering. These skills, in turn, are not typically taught in planning and engineering schools, nor are these skills common among transportation agency staff. As such, big data analytics are outsourced to third-party data analytics companies or data vendors, who are not obligated to share information about the data or models they use to convert the underlying data into useable products. Thus, their methods cannot be checked or emulated by anyone on staff with the necessary skillset.
- Surveys require agency sponsors to balance high costs and obtaining an appropriate number of responses for statistical validity [FHWA 2023]. While probe data may have lower upfront costs, the technical staff who use them often express significant concerns about data validity and their inability to match or compare them with corresponding metrics from survey data. For example, establishment and household surveys are used to develop trip generation rates, but probe data cannot be used for this purpose without significant changes to the way these rates are calculated and implemented.

A recent study by Ugurel et al. [2024] surveyed over 50 Metropolitan Planning Organization (MPO) staff members, who confirmed these tensions around probe data as data users. Their research found that, while big data products offer promising tools and new perspectives for transportation planning, adoption faces several key challenges. The primary challenges are the lack of transparency in data-collection and -processing methods and questions about the accuracy and validity of data products (i.e., underdeveloped ground truthing and data-validation methods). Another challenge is evolving regulatory standards for data, including data-privacy

---

1 "Big data" refers to massive, complex datasets that traditional data-management systems cannot handle. Refer to the following link for more information: https://www.ibm.com/think/topics/big-data [Badman, Kosinski n.d.].

concerns, opaque rules by platforms like iOS and Android, ever-changing data applications, and government actions to preserve more consumer privacy and restrict access to the data.

Ugurel et al. [2024] also found that MPO planners and managers have different concerns about the use of probe datasets. Planners' concerns center on the usefulness of the data, especially given poor data coverage of low-volume roads and inability to capture hard-to-reach segments of the population (i.e., those without cell phones), while their managers' qualms often focus on their opaque cost structures and their relative quality (i.e., whether the data meet explicit and implicit standards and guidelines). While these concerns about probe data cannot be completely ameliorated with better knowledge alone, providing more public information on probe data would ensure decisions to use them are better informed.

The multitude of probe data vendors, product offerings, data sources, and technologies have increased confusion in the marketplace. Recent volatility in passive data sources (i.e., the bankruptcy of several LBS data providers) illustrates this trend.

Other barriers to probe data adoption center on agency staff. Without personnel fully dedicated to using and understanding probe data, institutions cannot appreciably comprehend or communicate the shortcomings of the data, let alone breakdown institutional barriers to their adoption or build the in-house capacity to directly collect, process, analyze, and glean insights from large datasets. Insufficient resources to adequately train staff on handling and processing these massive datasets, coupled with broader market forces keeping agencies from hiring the right staff, mean most agencies cannot keep pace with all the changes happening in the probe data marketplace.

Nevertheless, agency interest and awareness of big probe data remain high (Ugurel et al. [2024] found 83 percent of MPOs have experimented with or used big data sources through trial subscriptions, purchases, or consultants), making the ability to build capacity to better understand data offerings and the nuances that differentiate available data products imperative. An aim of this document as well as its six guides on specific data types is to provide resource-limited, data-curious staff with an easy-to-follow digest on the background and uses of probe data vis-à-vis public planning and the generation of transportation statistics [Titlow et al. 2026a–2026f].

4. Probe Data History

Section 4.1 provides a timeline of the evolution of probe data, Section 4.2 shows the changing landscape of probe data, and Section 4.3 summarizes existing data programs.

4.1. EVOLUTION OF PROBE DATA

Travel-behavior data began to significantly diverge from data derived from traditional collection methods around the year 2000, corresponding to the arrival and public proliferation of the modern, commercially available location-detection and data transmittal technologies that underlie the core probe types noted in Exhibit 1 [White House 2000].

Table 2 provides a timeline of innovations in probe data along with sample references for milestones. The rapid evolution of probe data technologies summarized here illustrate the changing landscape researchers are navigating to understand and best utilize these data.

Table 2. Probe Data Evolution Timeline

Era	Period	Milestone	Technology evolution	References
Traditional data collection	Pre-2000	Implementation of NHTS, regional travel surveys, truck surveys, Vehicle Inventory and Use Survey, floating car speed studies, and cordon intercept studies	Manual surveys, trip diaries, cordon line intercept studies, floating car speed studies	FHWA 2022
Early GPS adoption	Pre-2000	Placement of on-board GPS devices on trucks and passenger vehicles at selective availability (i.e., vehicle location revealed at a coarse geographic level and not yet with specific latitude–longitude coordinates)	Limited-accuracy GPS due to selective availability	Yalamanchili et al. 1999; Battelle 1999
	May 2000	Implementation of President Clinton's Executive Order to allow civilians to use full-precision GPS (previously it was only available at selective availability)	Full-precision GPS access for civilian use	White House 2000
	2000s	Introduction of in-vehicle navigation and CVs	Early in-vehicle GPS-based navigation systems	California State Senate Committee on Transportation and Housing and Senate Committee on Judiciary 2016

Era	Period	Milestone	Technology evolution	References
Data-collection innovation	2000–2013	Various implementations of truck fleet GPS, in-vehicle, Personal Digital Assistant GPS for travel behavior studies, including travel survey enhancements	GPS-enhanced travel surveys (respondents given GPS unit to track their movements to supplement survey responses with precise spatial data)	Li et al. 2004; Forrest, Pearson 2005
	2008–2009	Emergence of cell phone probe data	Cell phone location tracking for traffic analysis	Gur et al. 2009; Liu et al. 2008
	2008–2010	Emergence of Bluetooth device capture for traffic, corridor, and external data collection	Bluetooth detectors for travel time	Barcelö et al. 2010; Ma, Van Zuylen, Van Dalen 2012; Carpenter, Fowler, Adler 2012
Big data revolution	2012	Implementation of Moving Ahead for Progress in the 21st Century Act federal requirements for the adoption of ELDs in commercial motor vehicles	Mandated ELD tracking for commercial vehicles	FMCSA 2024
	2013	Development of big data analytical techniques, enabled by better computing and data storage tools and methods, making it feasible to archive and analyze GPS data—now possible to use data ubiquitously collected from GPS units carried by the population at large	Large-scale data aggregation and analytics	Desharnais, Chapleau 2013
	2014–2017	Commercial availability of cellular call-detection record data	Cellular call data records providing triangulated location and timestamps of devices, allowing analysts to determine individuals' speeds, travel times, and OD travel patterns	Iqbal et al. 2014; Hard et al. 2017; Huntsinger 2017; Çolak et al. 2019; Huntsinger, Ward 2015

Era	Period	Milestone	Technology evolution	References
LBS and CVs	2015–present	Ubiquitous integration of CV technology into auto manufacturing (vehicle to vehicle, vehicle to infrastructure, and V2X)	V2X communications technology providing traces that can be archived and analyzed to study travel patterns of CVs and their occupants	UMD, MDOT 2025
	2016–2018	Commercial availability of LBS data	Applications for LBS data in transportation planning developed (i.e., studying the location information and timestamps from sightings to determine person travel characteristics from observed movements)	Sadeghvaziri, Rojas, Jin 2016; Lemp et al. 2019
	2019	Commercial/public availability of CV data from third-party sources	Cloud-connected vehicle data services, which provide vehicle data (rather than person data); aggregate products show trips and other mobility patterns observed from actual vehicle movements	—
	2019	Commercial availability of ELD data	Commercial vehicle ELDs providing hours of service and coarse OD tables based on observed commercial vehicle movements	—

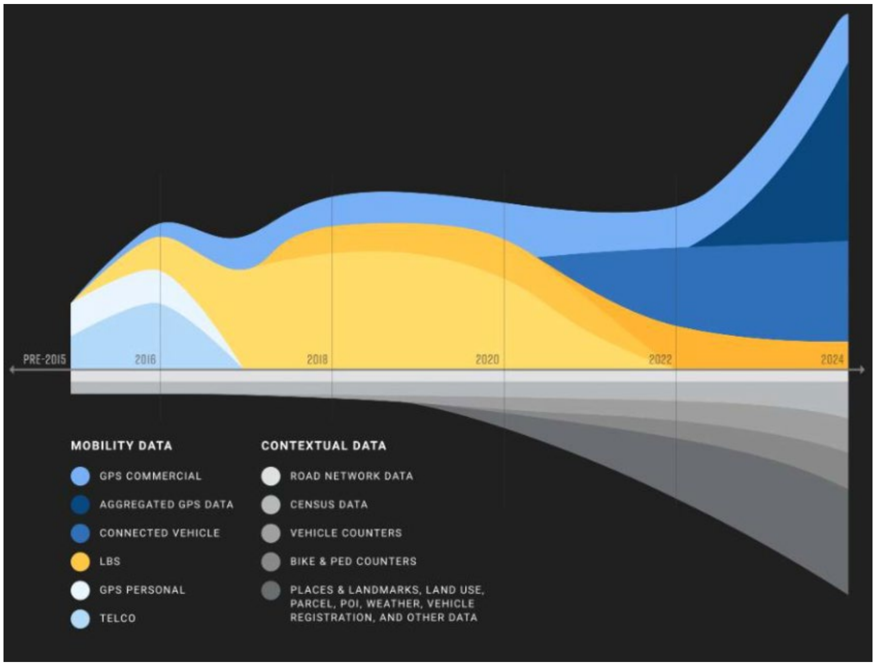
—Not applicable.
ELD = electronic logging device; vehicle to everything = V2X.

4.2. PROBE DATA—A CHANGING LANDSCAPE

The variation in classifications of probe and nonprobe data is partially a reflection of the rapid innovations and deprecations in different forms of technology and standards of data quality and assurance over the past 20 years (Table 2).

Early in the evolution of these sources, disaggregate probe data (i.e., raw location sightings) were the most common [Chigoy, Gerke 2019]. Over time, however, they become less common, driven by changes in the data markets resulting from data governance changes imposed by data providers and platforms due to growing concerns about the risks to privacy associated with probe data [Apple 2021]. Adding to this decline, the composition of probe data sources has changed since 2019 with the addition of new sources, like CVs. Figure 2 is from a commercial data product developer and illustrates how these trends in the data market have played out since 2015.

Figure 2. Commercial Data Provider Data Foundation



Source: StreetLight Data 2024.

4.3. EXISTING NATIONAL AND STATE DATA PROGRAMS

A major advantage of probe data is that they can supplement and enhance traditional, national-level sources of transportation data by reducing survey-respondent burden, enabling the release of mobility data in a timelier, more cost-efficient manner.

Nevertheless, the relationship between probe and nonprobe data is not unidirectional. Traditional national-level data sources, like the NHTS, also support sensitivity analyses that help validate probe data, ensuring results and statistics drawn from them are valid and consistent with other measurements of the same phenomena. In fact, multiple state and national programs responsible for producing statistics on larger geographic areas use probe data sources to ensure complete coverage of their study areas, something that would not otherwise have been feasible with traditional sources alone.

Examples of such state and national programs include:

- **FHWA’s National Performance Management Research Data Set (NPMRDS)**—This program publishes travel time and speed data collected anonymously from a fleet of probe vehicles—passenger cars and freight trucks—equipped with GPS-enabled cell phones. Time and location information generated by these probes are used to calculate and aggregate speed and travel-time data into 5-minute, 15-minute, and 1-hour increments along segments (technically referred to as “Traffic Message Channels”) of nationally significant highways in the United States [Bricka, Zhang, Schroeder 2020].

- **BTS Daily Mobility Statistics (DMS)**—Discontinued in April 2024, the DMS program used LBS data sourced from smart phone apps to estimate the number of Americans staying at home during the onset of COVID-19 as well as the temporal and spatial trip patterns (at national, state, and county levels) of the same traveling public over the subsequent years (January 2019–April 2024) of the pandemic [BTS 2025].
- **FHWA's Next Generation NHTS National OD Data Products**—This program includes national, multimodal OD tables reflecting passenger and truck travel in the United States. The passenger OD data are sourced from passively collected cell phone location data (i.e., LBS), and the truck data are sourced from truck fleet GPS providers. The OD table is aggregated to 583 geographic zones: 447 based on state-specific metropolitan statistical areas, micropolitan statistical areas, and some counties within a state for geography continuity and 136 created from the remaining areas within each state [FHWA 2022]. This product harnesses the ubiquity of probe data to provide a national travel monitoring product more than once every 5 years.
- **Statewide commercial probe data products (multiple)**—Many state departments of transportation, MPOs, counties, and cities subscribe to or purchase probe data products for internal transportation planning. To help these entities with these procurements, the Eastern Transportation Coalition (TETC) provides its members with probe data acquisition resources through its Transportation Data Marketplace [TETC 2022].
- **Smart Cities and transportation systems management and operations (TSMOs)**—these programs, which broadly seek to optimize the safe, efficient, and reliable use of existing and planned transportation infrastructure for all modes and geographical areas through the multijurisdictional coordination, require and entail many types of probe data, especially those that can be used across all the different facets of the transportation system they are seeking to improve: transit management, traffic signal coordination, traffic incident management, CV and automated vehicle deployment, multimodalism, and managing travel demand. TSMOs favor probe data from CVs, marrying them with in-situ nonprobe data systems, like pedestrian cameras at crosswalks [FHWA 2020b].

BTS is not entirely new to the realm of probe data; it has been actively exploring and publishing experimental statistics using these data, including several products explicitly pulled from nonhighway modes of transportation:

- Using GPS-enabled automatic information system (AIS) messages from ships to study ports and other freight and passenger vessel movement trends [BTS 2026a]
- Harnessing radar- and GPS-based air traffic control position messages relayed from aircraft via the Federal Aviation Administration's System Wide Information Management platform to measure the national aviation system
- Utilizing GPS data from fleets of freight trucks to study the movement and implications of freight movement around North America, producing public-facing statistics through a program known as the BTS Freight Mobility Initiative [BTS 2026b]

Other USDOT modal offices and administrations have also embraced probe data:

- FHWA's NPMRDS program, which began in 2013, offers NPMRDS, at no charge to local, state, and federal transportation agencies, for monitoring and reporting transportation-system performance measures and setting and meeting mobility objectives and targets [FHWA 2020a]. NPMRDS plays a major role in ensuring agencies meet Moving Ahead for Progress in the 21st Century Act regulations and improving the quality of data reported in other FHWA products.

- FHWA's NHTS OD data products include multimodal passenger and truck travel OD tables at the national and zonal levels from passively collected data sources, and they contain raw cell phone location data from multiple vendors [FHWA 2022].
- The Federal Motor Carrier Safety Administration's electronic logging device (ELD) requirement started on December 16, 2019, mandating the use of ELDs by all carriers and drivers subject to the ELD rule. An ELD is a probe that automatically records a driver's driving time and other hours of service (HOS) data to allow for easier, more accurate HOS recordkeeping [FMCSA 2017]. An ELD monitors a vehicle's engine to capture data on whether the engine is running, whether the vehicle is moving, miles driven, and duration of engine operation (engine hours). With an ELD, law enforcement can review a driver's HOS by viewing the ELD's display screen, a printout from the ELD, or retrieving data electronically from the ELD. The raw ELD messages, however, cannot be accessed by members of the public, even for statistical generation purposes.
- The Maritime Administration (MARAD) uses GPS-assisted AIS probe data to monitor maritime operations and logistics. Data obtained from the Sea-web™ Movements Database allows MARAD to track ship movements to monitor maritime traffic, port operations, and supply chain logistics [MARAD 2023].
- The Volpe National Transportation Center maintains SeaVision, another platform for analyzing GPS-assisted probe data from vessels. The web-based tool is primarily oriented toward users with operational and security needs for real-time maritime information, although it does have functionalities that permit statistical generation and analytical study.
- The Federal Transit Administration published a 2023 report that studied the feasibility of LBS probe data for transit operations and planning, calling attention to their advantages for studying riders' mobility while noting drawbacks, namely privacy concerns and their inability to capture Americans without smart phones [FTA 2023].

Federal agencies other than USDOT also utilize the extensive coverage probe data offer. These agencies use them to develop more effective strategies to address safety and infrastructure challenges on a national scale. Examples of other federal agencies using probe data include the following:

- The Centers for Disease Control and Prevention (CDC) uses LBS data for public-health research, COVID-19 movement analysis, and epidemic forecasting. In 2020, CDC published a report that assessed the impact of stay-at-home orders on population movement using publicly available device data from SafeGraph [SafeGraph n.d.].
- The Environmental Protection Agency conducted a pilot study using a commercially available dataset from AirSage, Inc., that contained counts from cellular call data records (CDRs) and GPS data to study visits to water recreational areas. The purpose of this research was to improve management of public lands and address social inequities in recreational access [Nichols et al. 2023].

5. Categorizing Probe Data Types

To help the transportation community gain a better, more transparent understanding of probe data, BTS proposes a taxonomy for categorizing and distinguishing probe data's constituent data types (Section 5.2). BTS has two primary goals for this categorization:

1. Use the collected information to plan future statistical programs based on probe data
2. Provide the findings to the public, including fellow transportation planners and those involved in the generation of statistical products, to promote greater transparency on these critical data types

To accomplish these goals, this taxonomy shall explain relationships between data types, identify commonalities between data formats, and provide a standard vocabulary for referring to probe data types and their core elements. Hopefully, this initial taxonomy will spark conversations about building a singular, standardized taxonomy of probe data types in the transportation field.

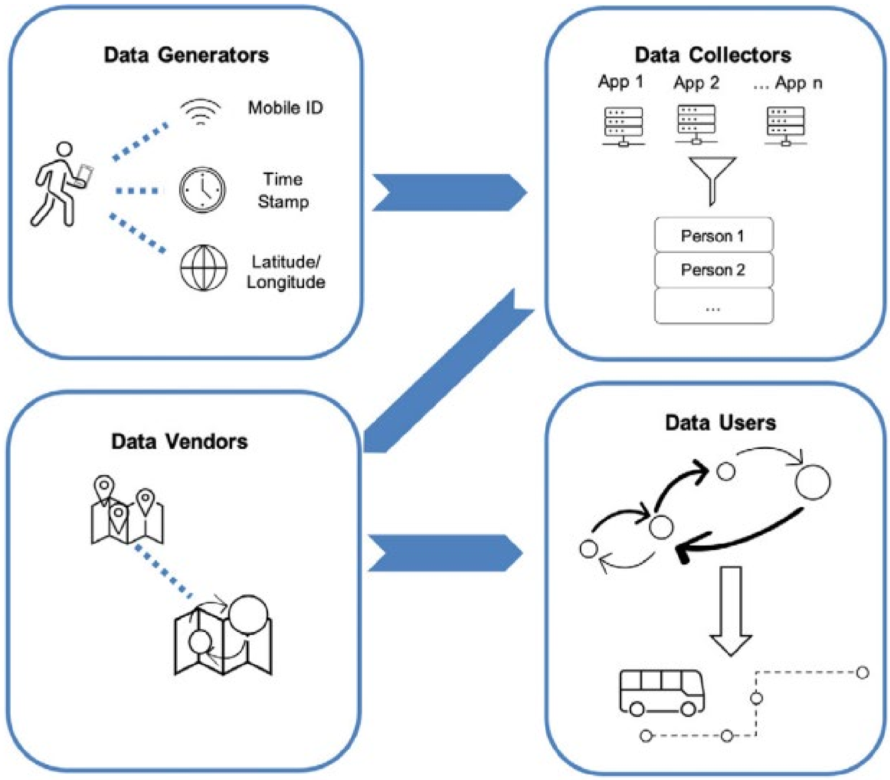
The following sections identify the entities involved in the probe data marketplace (Section 5.1) and propose dimensions for a taxonomy that could be adopted by planners and statisticians across the industry to assess and specify future data products and purchases (Section 5.2).

5.1. PROBE DATA MARKETPLACE

An awareness of the probe data marketplace is essential for end users to better understand the data they have purchased (Subsection 5.2.2.2). For the purposes of this taxonomy, BTS defines the four main actors in this marketplace as follows (Figure 3):

- **Data generators:** These entities are responsible for the devices and apps that generate raw probe data through GPS-enabled devices, smartphone apps, CV computers, or any other location-detection and -emitting technology.
- **Data collectors (or aggregators):** These companies contract with data generators (e.g., mobile-device app developers and CV original equipment manufacturers) to receive panels of their data for re-sale. In some instances, a data generator also collects and reaggregates its data for re-sale.
- **Data vendors:** These companies purchase anonymized, disaggregate data from aggregators and add value, commonly running the data through further privacy filters or adding industry-specific metrics or attributes to the raw data. Depending on the level of aggregation they perform, they preprocess data on behalf of customers who lack the in-house technical expertise to work with disaggregate, raw probe data. Nevertheless, some vendors (depending on the data type) will sell minimally processed, disaggregate probe data that more sophisticated customers can use for their own big data analyses.
- **Data users:** These entities use the data for a variety of applications. They can be in-house staff to an aggregator or vendor, or they can be a third-party consumer (i.e., local government planner, academic researcher, or private sector consultant) who purchases a data product. For the latter entities, their use of the data will be restricted by the terms of their purchase agreement with their vendor or aggregator and, depending on the data type and format, any aggregation or anonymization done before the data are received.

Figure 3. Probe Data Marketplace



Source: Shaw, Flint, Englin 2023.

Transportation agencies most commonly procure probe data from vendors, not aggregators.

Users should be aware that aggregators and vendors cannot publicly reveal many details, if any, on how they source and compile their data products. Data aggregator websites do not typically reveal the sources of their datasets, as they are restricted by nondisclosure agreements (NDAs) and competitive market forces; aside from not being able to name the data generators (or fellow aggregators) from whom they receive their data, they are very unlikely to disclose methods to convert their source data into transformed outputs or the provisions they take to protect data user privacy. Data vendors are similarly bound to NDAs that restrict their ability to name their respective data aggregator partner(s). Nevertheless, some vendors may provide—either as a white paper, website section, or direct ask—higher level information about certain elements of their methodologies, namely data validation techniques, the type(s) of probe data they ingest, and in some cases, data-expansion procedures.

5.2. PROPOSED DATA TAXONOMY

The purpose of this section is to propose a taxonomy for better understanding and assessing probe data, especially in the context of generating transportation statistics.

To start, several dimensions for characterizing and differentiating known probe data types are proposed (Subsection 5.2.1). Once these categorization heuristics are detailed, the dimensions of a probe data taxonomy are proposed (Subsection 5.2.2.1). The chapter concludes with proposed standardized terminology and classifications (Subsection 5.2.2.2).

Note that some concepts discussed in previous chapters are purposefully presented again within this new context. The repetition of these details is necessary in some places to ensure the foundation for the proposed taxonomy appears all in one place without sending the reader back and forth within the report.

5.2.1. Dimensions for Differentiating and Evaluating Probe Data Types

To help users differentiate the myriad forms of probe data (while distinguishing them from traditional, nonprobe data types), the contract team provided a set of dimensions on which each form can be evaluated. Each dimension highlights the ways probe data can vary under the umbrella definition that they be generated by a technological device located on or with a person, vehicle, vessel, or conveyance that actively determines and emits its location at a given moment in time.

The following dimensions cover various aspects of data acquisition, quality, utility, and market considerations; in fact, they all speak to questions commonly asked of probe data:

- **Volatility assessment:** How stable and reliable is the data source over time?
- **Resource requirements:** What monetary, personnel, skills, and technology resources are needed to acquire and utilize the data?
- **Public and legal policy considerations:** Are there privacy concerns or regulatory hurdles?
- **Geographic coverage and limitations:** What is the spatial extent of the data (i.e., local, regional, and/or national), and are there areas of missing coverage (e.g., rural areas often missing granular data)?
- **Infrastructure modification requirements:** Does using the data require new infrastructure?
- **Data sharing and selling willingness:** How accessible are the data from vendors?
- **Subject-matter expert (SME), community, and institutional acceptance levels:** How well are the data trusted and adopted by the transportation community?
- **Extract, transform, load (ETL) complexity evaluation:** How difficult is it to prepare the data for use?
- **Immediacy or real-time availability:** How quickly are the data available after an event?
- **Scale of data:** How big is the sample or what is the sampling rate?

For the probe types listed in Exhibit 1, Table 3 summarizes how these sources match with the dimensions listed in Section 5.2.2 along with sample vendors.

Table 3. Probe Data Dimensions Versus Probe Data Types

Probe data type	Volatility	Resource requirements	Public and legal policy considerations	Geographic coverage and limitations	Infrastructure modification requirements	Data sharing and selling willingness	SME, community, and institutional acceptance level	ETL complexity	Immediacy or real-time availability	Scale of data
LBS	High	Low	Privacy protected by various agreements	National	Need state of practice tech stack	High	High	Low	Varies by vendor (but most can turn around in a day to a quarter—depending on use case)	Big
CV	High	Low	Privacy protected by various agreements	National, where detection infrastructure is installed	Need state of practice tech stack	High	High	Low		Big
Cellular (CDR and passive signaling)	Stable	Low	Privacy protected by various agreements	National, where cell tower network is sufficiently dense	Need state of practice tech stack	High	High	Low		Big
LBS_ POI	High	Low	Privacy protected by various agreements	National	Need state of practice tech stack	High	Mid	Low		Big
Fleet GPS	Stable	Low	Privacy protected by various agreements	Global, anywhere with adequate GPS detection from satellites	Need state of practice tech stack	Low	Mid	Low		Big
ELD	Stable	Low	Privacy protected by various agreements	National, anywhere with adequate GPS detection from satellites	Need state of practice tech stack	Low	Mid	Low		Big

5.2.2. Proposed Data Taxonomy Primary Dimensions

The following subsections describe proposed dimensions for data characteristics (Subsection 5.2.2.1) and data processing and delivery (Subsection 5.2.2.2).

When considering what type of probe data is appropriate for a given location or use case, practitioners should ponder these dimensions to help decide. No form of probe data will be a perfect fit for all the data and statistical needs at a transportation-oriented agency, so trade-offs will need to be made between options.

5.2.2.1. Data Characteristics Dimensions

The following dimensions characterize the content and attributes of the data type:

- **Location data:** What are the origin, quality, and precision of their core spatial information (e.g., GPS coordinates and cellular pings)?
- **Frequency of collection:** How often are the data collected (e.g., continuous, hourly, or daily)?
- **Transportation modes:** Which modes of transportation (e.g., passenger, commercial, rail, bike, and/or pedestrian) are innately covered (versus requiring estimation)?
- **Operational characteristics:** Do the data innately include vehicle operating characteristics, like speed, acceleration, or engine hours, or do these characteristics have to be estimated?
- **Travel versus nontravel activity:** Do the data innately differentiate between actual travel and other activities, or does this information have to be estimated (e.g., parking, and POI activity)?
- **Scalability potential:** Can the data be scaled from the local to the national level?
- **Sample penetration:** What proportion of the target population is captured by the data?
- **Known biases and limitations:** What inherent biases or gaps might exist in the data, especially vis-à-vis its sample of the population being measured?
- **Temporal consistency for trend analysis:** Are the data consistent over time for trend or archival analysis?

The following dimensions characterize applications of the data type:

- **Use cases:** What specific applications do the data support (e.g., assessing travel frequency or system performance)?
- **Activity types:** What specific activities are innately captured, and what activities must be estimated (e.g., business visits and commuting)?
- **Transportation exclusivity:** Was the data product created specifically for the transportation market, or is transportation a secondary use?
- **Opportunity cost analysis:** What are the trade-offs in choosing this data source, especially compared to alternate forms of probe or nonprobe data?
- **Economic indicators:** Can the data inform economic activity or trade conditions?

The following dimensions characterize the usability of the data type:

- **Data expertise:** What are the specific skills required to work with the data?
- **Data costs:** What are the financial implications for using these data (e.g., acquisition costs, predictability of pricing, and pricing model)?

- **Licensing and use restrictions:** What (if any) legal limitations exist on data use?
- **Documentation:** What is the availability and quality of explanatory materials?
- **Transparency in data-selection processes:** How open are vendors about their methodologies? How responsive and informative are the vendors to questions?
- **Shelf-life of data:** For how long are the data useful or relevant? Would they be useful archived? What is the stability and longevity of the data source?
- **Data formats:** In what file format are the data structured and delivered?
- **Data fusion needs:** What other data sources (if any) are needed for normalization or expansion?
- **Data quality and assurance:** What measures for accuracy and reliability exist?
- **Privacy:** How is the privacy of the device producing the data protected (e.g., anonymization and filters)?
- **Vendors:** Who are the providers, and what is their reputation?
- **User community:** Who are the primary users of the data and what are their specific needs? Would it be easy to identify and contact fellow users?

5.2.2.2. Proposed Standardized Terminology and Classifications

As users seek to better understand and differentiate transportation probe data types going forward, the following ideas should be centered in the discussion:

- **Provide a clear definition of "probe data":** Promote a singular definition of probe data, distinguishing them from traditional nonprobe data, like surveys or traffic counters.
- **Acknowledge the gray areas:** Be transparent that certain data sources exist in a gray area between probe and nonprobe.
- **Standardize data type names and core definitions:** Adopt consistent terminology for primary probe data types (e.g., GPS, LBS, ELD, CV, cellular CDRs, and POI), and provide concise definitions as outlined in Subsection 5.2.2.1. Use the same terms when referring to the individual data points produced from each source.
- **Emphasize data-processing levels:** Make users aware of the distinction between raw (trip trace), transformed, and modeled data outputs, helping users understand that degrees of aggregation, cleaning, and inference directly impact data granularity and interpretability.
- **Explain data-delivery models:** Clarify the ways probe data products are accessed (e.g., data provider platforms, third-party platforms, and service providers), which helps users identify suitable procurement and analysis methods appropriate for their infrastructure, resources, personnel, and technical capabilities.
- **Utilize a common set of evaluation dimensions:** Promote standard evaluation factors for all probe data sources. Once a collective set of core acceptance criteria are developed, perhaps using the dimensions listed here, they should be at the core of any selection process (e.g., geographic coverage, temporal consistency, privacy, data quality, cost, and expertise needed).
- **Identify marketplace actors:** Remain aware of the probe data marketplace (e.g., data generators, collectors and aggregators, vendors, and data users). This awareness ensures buyers understand the data supply chain and, therefore, why aggregators and vendors cannot reveal their raw data sources or methodologies (due to NDAs).

- **Focus on use cases and data gaps:** Connect probe data types to specific, appropriate transportation use cases (e.g., travel frequency, system performance, and economic impact), which include those facets of the transport system that probe data do not best equip researchers to study (e.g., capturing international and long-distance travel and providing context on how and why travel occurs). This focus will temper user expectations and cue them into situations when and where probe data need be fused with traditional and/or nonprobe data sources.
- **Stress transparency and validation:** Emphasize the importance of transparency in data-collection and -processing methods as well as making it clear what methods were used to validate probe-based data products. Users should ask vendors for as much information as possible about methodologies, quality-assurance and quality-control procedures, expansion factors, and data sources, including the spatial and temporal coverage and quality of their data.
- **Acknowledge the changing landscape:** Know the probe data market is highly volatile and rapidly evolving, with new sources emerging and others changing or deprecating in the face of rapidly changing data privacy laws and practices. Continuous learning and adaptation are required to keep pace with changes, something practitioners should consider when incorporating probe data into their workflows.

By providing clear dimensions for their evaluation and promoting the development of a singular vocabulary for working with probe data, users can make informed, maximally transparent decisions to use these innovative data sources for transportation analysis and policy.

5.3. INDIVIDUAL DATA TYPE GUIDES

Accompanying this broad overview of probe data is a set of user guides on six data types commonly used or requested by the transportation statistics and planning communities:

1. **LBS** data are passive probe data derived from cell phone apps that produce location information with user consent. Location is determined using a mix of technologies that are collected by different methods based on the device's location (via either satellite-based GPS technology or cellular or Wi-Fi network triangulation) [Titlow et al. 2026a].
2. **POI and aggregated LBS** obtain location data from cell phone apps (with user consent) and aggregate them to specific building footprints. They are typically used for understanding activity patterns at businesses [Titlow et al. 2026b].
3. **Cellular data (CDRs)** result from interactions between cellular devices and base stations with known locations. Cellular location data are passive probe data that are associated with a device's location relative to base stations or cell towers [Titlow et al. 2026c].
4. **CVs** provide location, vehicle operational, and safety data collected from personal and nonpersonal vehicles with embedded telematics technology that can interact with purpose-installed receivers in the built environment (vehicle-to-everything technology) [Titlow et al. 2026d].
5. **Fleet GPS** provide location data derived from various satellite-based commercial navigation technologies, such as CVs, commercial vehicle fleet tracking, and in-vehicle navigation (passenger or freight) [Titlow et al. 2026e].
6. **ELDs** are required in commercial vehicles and track HOS, parking, and other regulatory requirements and produce messages with GPS-based location information [Titlow et al. 2026f].

Each guide is structured with an identical set of sections, mapped onto the data dimensions summarized herein. This structure aims to ease comparisons across data types, especially for those trying to determine the suitability of each for their use cases or agency needs.

6. References

ADOT. 2011. *Transportation decisions driven by traffic counts*. Phoenix, AZ: Arizona Department of Transportation. https://azdot.gov/blog-article/transportation-decisions-driven-traffic-counts. Last accessed March 3, 2025.

Apple. 2021. “Upcoming AppTrackingTransparency requirements,” *News*. Cupertino, CA: Apple. https://developer.apple.com/news/?id=ecvrtzt2. Last accessed November 20, 2024.

Badman, Annie, and Matthew Kosinski. n.d. *What is big data?* Armonk, NY: IBM. https://www.ibm.com/think/topics/big-data. Last accessed March 13, 2025.

Barcelö, Jaume, Lidin Montero, Laura Marqués, and Carlos Carmona. 2010. “Travel Time Forecasting and Dynamic Origin-Destination Estimation for Freeways Based on Bluetooth Traffic Monitoring,” *Transportation Research Record*, 2175, no. 1. Thousand Oaks, CA: Sage Journals. https://doi.org/10.3141/2175-03. Last accessed March 13, 2025.

Battelle. 1999. *Heavy-Duty Truck Activity Data*. Washington, DC: U.S. Department of Transportation. https://rosap.ntl.bts.gov/view/dot/4679. Last accessed March 13, 2025.

Bricka, Stacey, Lei Zhang, and Paul Schroeder. 2020. *Assessing Changes in Travel Behavior Data Collection*. Washington, DC: U.S. Department of Transportation. https://nhts.ornl.gov/assets/MacroSys_FHWAAssessChangesReport_01212020.pdf. Last accessed March 13, 2025.

BTS. 2019. *About the Bureau of transportation Statistics*. Washington, DC: Bureau of Transportation Statistics. https://www.bts.gov/about-BTS. Last accessed March 13, 2025.

———2025. *Daily Mobility Statistics*. Washington, DC: U.S. Department of Transportation. https://www.bts.gov/daily-travel. Last accessed March 13, 2025.

———2026a. *Port Performance Freight Statistics Program*. Washington, DC: Bureau of Transportation Statistics. https://www.bts.gov/ports. Last accessed May 18, 2026.

———2026b. *Trucking movements: BTS Freight Mobility Initiative (FMI)*. Washington, DC: Bureau of Transportation Statistics. https://www.bts.gov/explore-topics-and-geography/topics/freight-transportation/trucking-movements-bts-freight-mobility. Last accessed May 18, 2026.

California State Senate Committee on Transportation and Housing and Senate Committee on Judiciary. 2016. *Telematics 101: How Much Your Car Knows About You*. Sacramento, CA: California State Senate. https://sjud.senate.ca.gov/sites/sjud.senate.ca.gov/files/telematics_hearing_background_paper.pdf. Last accessed March 13, 2025.

Carpenter, Charles, Mark Fowler, and Thomas J. Adler. 2012. “Generating Route-Specific Origin–Destination Tables Using Bluetooth Technology,” *Transportation Research Record*, 2308, no. 1. Thousand Oaks, CA: Sage Journals. https://doi.org/10.3141/2308-10. Last accessed March 13, 2025.

Chigoy, Byron, and Michael Gerke. 2019. *2018 Alamo Area MPO External Study Using LBS and GPS Data Technical Summary*. College Station, TX: Texas A&M Transportation Institute, Texas Department of Transportation Travel Survey Program.

Çolak, Serdar, Lauren P. Alexander, Bernardo G. Alvim, Shomik R. Mehndiratta, and Marta C. González. 2019. “Analyzing Cell Phone Location Data for Urban Travel: Current Methods, Limitations, and Opportunities,” *Transportation Research Record*, 2526, no. 1. Thousand Oaks, CA: Sage Journals. https://doi.org/10.3141/2526-14. Last accessed March 13, 2025.

Desharnais, Marie-Christine, and Robert Chapleau. 2013. *Scrutinizing Weekly Travel Behavior Patterns in Paratransit: Results of a Big Data Experiment*. Paper No. 13-3790. Washington, DC: Transportation Research Board 92nd Annual Meeting.

FHWA, 2020a. *The National Performance Management Research Data Set (NPMRDS) and Application for Work Zone Performance Measurement*. Washington, DC: U.S. Department of Transportation. https://ops.fhwa.dot.gov/publications/fhwahop20028/index.htm. Last accessed March 13, 2025.

———2020b. *Integrated Corridor Management and the Smart Cities Revolution: Leveraging Synergies*. Washington, DC: U.S. Department of Transportation. https://ops.fhwa.dot.gov/publications/fhwahop16075/ch2.htm. Last accessed March 13, 2025.

———2022. *2022 NextGen NHTS National Passenger OD Data*. Washington, DC: U.S. Department of Transportation. https://nhts.ornl.gov/od/. Last accessed March 13, 2025.

———2023. *FHWA NHTS Report: Travel Survey State of the Practice*. Washington, DC: U.S. Department of Transportation. https://nhts.ornl.gov/assets/NextGen%20NHTS_State%20of%20Practice_032423.pdf. Last accessed March 13, 2025.

———n.d. *Operations Performance Measurement Program*. Washington, DC: U.S. Department of Transportation. https://ops.fhwa.dot.gov/perf_measurement/. Last accessed March 24, 2026.

FMCSA. 2017. *Electronic Logging Device (ELD) Rule: Facts for the Commercial Motor Vehicle Industry*. Washington, DC: U.S. Department of Transportation. https://www.fmcsa.dot.gov/sites/fmcsa.dot.gov/files/docs/regulations/hours-service/elds/81576/usdot-fmcsaeld-brochureen.pdf. Last accessed March 13, 2025.

———2024. *MAP-21 - Moving Ahead for Progress in the 21st Century Act*. Washington, DC: U.S. Department of Transportation. https://www.fmcsa.dot.gov/mission/policy/map-21-moving-ahead-progress-21st-century-act. Last accessed March 14, 2025.

Forrest, Timothy L., and David F. Pearson. 2005. “Comparison of Trip Determination Methods in Household Travel Surveys Enhanced by a Global Positioning System,” *Transportation Research Record*, 1917, no.1. Thousand Oaks, CA: Sage Journals. https://doi.org/10.1177/0361198105191700108. Last accessed March 13, 2025.

FTA. 2023. *Location-Based Services Data for Transit Agency Planning and Operations: Market Scan and Feasibility Analysis.* https://www.transit.dot.gov/sites/fta.dot.gov/files/2023-05/FTA0245-Research-Report-Summary.pdf. Last accessed October 16, 2025.

Gur, Yehuda J., Shlomo Bekhor, Charles Solomon, and Leonid Kheifits. 2009. “Intercity Person Trip Tables for Nationwide Transportation Planning in Israel Obtained from Massive Cell Phone Data,” *Transportation Research Record*, 2121, no. 1. Thousand Oaks, CA: Sage Journals. https://doi.org/10.3141/2121-16. Last accessed March 13, 2025.

Hard, Edwin N., Byron T. Chigoy, Praprut Songchitruksa, et al. 2017. *Comparison of Cell, GPS, and Bluetooth Derived External O-D Data – Results from the 2014 Tyler, Texas Study*. Washington, DC: Transportation Research Board 96th Annual Meeting.

Huntsinger, Leta F. 2017. *The Lure of Big Data: Evaluating the Efficacy of Mobile Phone Data for Travel Model Validation*. Washington, DC: Transportation Research Board 96th Annual Meeting.

Huntsinger, Leta F., and Kyle Ward. 2015. “Using Mobile Phone Location Data to Develop External Trip Models,” *Transportation Research Record*, 2499, no.1. Thousand Oaks, CA: Sage Journals. https://doi.org/10.3141/2499-04. Last accessed March 13, 2025.

Iqbal, Md Shahadat, Charisma F. Choudhury, Marta C. González, and Pu Wang. 2014. *Development of Origin-Destination Matrices Using Mobile Phone Call Data: A Simulation Based Approach*. Washington, DC: Transportation Research Board 93rd Annual Meeting.

Lemp, Jason, Jingjing Zang, Xiao Yun Chang, et al. 2019. *Assessing the Impact of Mobile Data Processing Algorithms and Expansion Methods on the Inference of Travel Behavior Metrics*. Paper No. 19-05687.Washington, DC: Transportation Research Board 98th Annual Meeting.

Li, Hainan, Randall Guensler, Jennifer Ogle, and Jun Wang. 2004. “Using Global Positioning System Data to Understand Day-to-Day Dynamics of Morning Commute Behavior,” *Transportation Research Record*, 1895, no. 1. Thousand Oaks, CA: Sage Journals. https://doi.org/10.3141/1895-11. Last accessed March 13, 2025.

Liu, Henry X., Adam Danczyk, Rashmi Brewer, and Ray Starr. 2008. “Evaluation of Cell Phone Traffic Data in Minnesota,” *Transportation Research Record*, 2086, no. 1. Thousand Oaks, CA: Sage Journals. https://doi.org/10.3141/2086-01. Last accessed March 13, 2025.

Ma, Yinyi, Henk J. Van Zuylen, and Jan Van Dalen. 2012. *Freight Origin-Destination Matrix Estimation Using Multiple Data Sources*. Paper No. 12-0693. Washington, DC: Transportation Research Board 91st Annual Meeting.

MARAD. 2023. *Data and Reports*. Washington, DC: U.S. Department of Transportation. https://www.maritime.dot.gov/data-reports. Last accessed March 12, 2025.

Mudge, Richard, Hani Mahmassani, Robert Haas, Alireza Talebpour, and Lindsey Carroll. 2013. *Work Zone Performance Measurement Using Probe Data*. Washington, DC: U.S. Department of Transportation. https://ops.fhwa.dot.gov/wz/resources/publications/fhwahop13043/fhwahop13043.pdf. Last accessed March 13, 2025.

NCHRP. 2021. NCHRP Synthesis 561: *Use of Vehicle Probe and Cellular GPS Data by State Departments of Transportation*. Washington, DC: The National Academies Press. https://doi.org/10.17226/26094. Last accessed August 28, 2025.

Nichols, Elizabeth, Angela O'Brien, Shelley Feuer, and Jennifer Childs. 2023. *Use of Mobile Phone Location Data in Official Statistics, Social, Demographic and Health Studies*. Washington, DC: U.S. Census Bureau. https://www2.census.gov/adrm/CBSM/rsm2023-03.pdf. Last accessed March 13, 2025.

Sadeghvaziri, Eazaz, Mario B. Rojas IV, and Xia Jin. 2016. "Exploring the Potential of Mobile Phone Data in Travel Pattern Analysis," *Transportation Research Record*, 2594, no. 1. Thousand Oaks, CA: Sage Journals. https://doi.org/10.3141/2594-04. Last accessed March 13, 2025.

SafeGraph. n.d. *Social Distancing Metrics*. Denver, CO: SafeGraph. https://docs.safegraph.com/docs/social-distancing-metrics. Last accessed March 13, 2025.

Shaw, Jingsi, Lafcadio Flint, and Eric Englin. 2023. *Location-Based Service Data for Transit Agency Planning and Operations: Market Scan and Feasibility Analysis*. Washington, DC: U.S. Department of Transportation. https://www.transit.dot.gov/sites/fta.dot.gov/files/2023-05/FTA-Report-No-0245.pdf. Last accessed January 17, 2025.

Singh, Gargi, Vijayaraghavan Sivaraman, and Ed Hard. 2022. *Technical Memorandum: State of Emerging Mobility Big Data Sources and its Applications, Task 1: Evaluate Mobility Datasets*. College Station, TX: Texas A&M University. https://static.tti.tamu.edu/tti.tamu.edu/documents/TTI-2022-8.pdf. Last accessed August 12, 2025.

StreetLight Data. 2024. *StreetLight Data Sources & Methodology White Paper*. San Francisco, CA: StreetLight. https://learn.streetlightdata.com/methodology-data-sources-white-paper. Last accessed March 13, 2025.

TETC. 2022. *Transportation Data Marketplace*. College Park, MD: The Eastern Transportation Coalition. https://tetcoalition.org/transportation-data-marketplace/. Last accessed March 13, 2025.

Titlow, Kyle, Mindy Liu, Ed Strocko, et al. 2026a. *BTS Transportation Probe Data Guide: Location Based Services (LBS) Data*. Washington, DC: Bureau of Transportation Statistics. https://doi.org/10.21949/hst8-0343. Last accessed May 18, 2026.

———2026b. *BTS Transportation Probe Data Guide: Aggregate Location Based–Services (LBS), Navigation (LBS or GPS), and Point-of-Interest Data*. Washington, DC: Bureau of Transportation Statistics. https://doi.org/10.21949/f1w2-dh59. Last accessed May 18, 2026.

———2026c. *BTS Transportation Probe Data Guide: Non-LBS Cellular Probe Data*. Washington, DC: Bureau of Transportation Statistics. https://doi.org/10.21949/8q9s-gw77. Last accessed May 18, 2026.

———2026d. *BTS Transportation Probe Data Guide: Connected Vehicle Data*. Washington, DC: Bureau of Transportation Statistics. https://doi.org/10.21949/sd55-6x72. Last accessed May 18, 2026.

———2026e. *BTS Transportation Probe Data Guide: Fleet GPS Data*. Washington, DC: Bureau of Transportation Statistics. https://doi.org/10.21949/t34r-w727. Last accessed May 18, 2026.

———2026f. *BTS Transportation Probe Data Guide: Electronic Logging Device Data*. Washington, DC: Bureau of Transportation Statistics. https://doi.org/10.21949/2tza-cq27. Last accessed May 18, 2026.

Turner, Shawn M., Ioannis Tsapakis, and Pete Koeneman. 2020. *Evaluation of StreetLight Data's Traffic Count Estimates from Mobile Device Data*. St. Paul, MN: Minnesota Department of Transportation Office of Research & Innovation. https://rosap.ntl.bts.gov/view/dot/57948. Last accessed March 9, 2025.

Ugurel, Ekin, Xinhua Wu, Ryan Wang, Brian H.Y. Lee, and Cynthia Chen. 2024. *Metropolitan Planning Organizations' Uses of and Needs for Big Data*. Sydney, Australia: Findings. https://doi.org/10.32866/001c.127143. Last accessed March 13, 2025.

UMD and MDOT. 2025. *Advanced Transportation Technologies & Applied Practices (ATTAP)*. College Park, MD: University of Maryland.

White House. 2000. *President Clinton: Improving the Civilian Global Positioning System (GPS)*. Washington, DC: The White House. https://clintonwhitehouse4.archives.gov/WH/EOP/OSTP/html/0053_4.html. Last accessed March 13, 2025.

Yalamanchili, Lalit, Ram M. Pendyala, N. Prabaharan, and Pramodh Chakravarthy. 1999. "Analysis of Global Positioning System-Based Data Collection Methods for Capturing Multistop Trip-Chaining Behavior," *Transportation Research Record*, 1660, no. 1. Thousand Oaks, CA: Sage Journals. https://doi.org/10.3141/1660-08. Last accessed March 9, 2025.