BTS TRANSPORTATION PROBE DATA GUIDE: CONNECTED VEHICLE DATA

Connected vehicle (CV) data are generated, collected, and transmitted by vehicles equipped with onboard systems and sensors that enable them to exchange data with other vehicles, infrastructure, cloud services, and mobile devices. These communications produce data with records of events that occur while the car is in use (e.g., tire pressure warnings, vehicles on/off events, airbag inflations, impacts, lane deviation warnings).

Car manufacturers play a pivotal role in the CV data ecosystem. They install small computers in vehicles that collect and send out data about the vehicle in standardized formats. The data in these messages are then ingested and interpreted by (1) other vehicles and (2) devices embedded in roadside infrastructure (i.e., roadside units [RSUs] attached to infrastructure, such as traffic signals) [USDOT 2020].

The data produced through these interactions, in turn, contain information that can be used to better understand and illustrate many important facets of the transportation system (vehicle trajectory; vehicle operations; make and model; and safety-related items, such as tire pressure warnings and hard braking events) that are largely unavailable from other data sources at the same scale and precision.

When a CV is on and operating, for example, it continuously generates data packets containing the following information [ITE n.d.a]:

- Vehicle dynamics (speed, heading, acceleration)
- Vehicle safety status (brake status, stability control) [USDOT 2016]
- Precise positioning information
- System operational parameters [ITE n.d.b]

The contents of these data packets are easily digestible and, therefore, transparently understood by planners and statisticians with access to the data, an outcome of standardized CV data architectures and methodologies developed and documented jointly by U.S. government and private companies [Larkin et al 2016].

While CV data have existed since the first vehicles equipped with the requisite technology went into operation in 2015, they remain an emerging probe data source, especially because only one known vehicle maker provides the disaggregate data packets to vendors, agencies, and research organizations for development into public-facing products. While alternative CV data-collection methods exist for technology developers with access to proprietary devices installed in CVs, these sources are not as readily accessible to practitioners. Likewise, CVs are not a representative sample of all vehicles in every region. Overrepresentation occurs in regions with a greater preference for CV-enabled vehicle makes and models.

1. CAPABILITIES

CV data offer several distinct advantages for generating transportation statistics:

- These data are a high-quality source of safety, Intelligent Transportation Systems (ITS), and operational data from road vehicles.
- These high-quality data are highly standardized, the result of (1) extensive negotiations and discussions between government and industry and (2) vehicle-to-infrastructure (V2I) linkages that require and produce the highly standardized data needed for communication between vehicles and roadside devices, like traffic signals and detection equipment at crosswalks.
- CVs provide accurate positioning data, especially compared to location-based services, and precise information on vehicle operations and systems due to the array of sensors and technologies needed to ensure accurate interactions with roadway infrastructure.

2. LIMITATIONS

The following are known limitations of CV data in the development of transportation statistics:

- Data availability is limited by the market penetration of CV makes and models and number of original equipment manufacturers (OEMs) reselling CV data [GVR n.d.].
- Because of the sheer amount of data produced by the myriad sensors on a CV, files require terabytes and petabytes of storage, depending on their time frame and geographical extents, along with cloud computation infrastructure to process and analyze.
- Trip determination requires less guesswork than other sources as the start and stop locations of a vehicle are based on its key-on and key-off events. This strength may also be a weakness in determining long-distance trips as it reduces users' ability to set their own trip-breaking criteria.
- For external data users, creating vehicle tours and long-distance analytics is prohibitive as each CV trip is anonymized from the previous trip and does not have a persistent vehicle identifier (ID).

3. VENDORS AND AGGREGATORS

CV data are available from private vendors in their disaggregate and aggregate forms. Vendors selling disaggregate data require exclusive agreements with an OEM.

Disaggregate vendors ingest and process CV data on behalf of public customers; typical outputs include speed statistics, origin–destination (OD) matrices and trip tables, and other general vehicle activity metrics. Some telematics and ITS technology vendors use CV data for analyses related to signalized intersections and safety—often for insurance integration use cases. These data are commonly used in Signal and SmartCities and Smart Corridors applications.

4. MARKETS

The primary market for CV data is within the transportation sector. Common demand areas within the field include the following:

- ITS programs along CV corridors and in SmartCities, where sections of roadways, intersections, and signalized intersections are optimized for vehicle-to-vehicle (V2V) and V2I connectivity
- Safety analyses
- Operations, such as bottleneck analysis and congestion management
- Planning applications, such as OD studies

CV data's secondary markets retain tangential connections to transportation. These domains include the following use cases:

- Telematics[1] applications, which use the vehicle's operations and trajectory data for insurance and driver alerts
- Electric vehicle (EV) planning to understand where EV activity is highest and, thus, aid in determining charger placement, where most EVs are CVs

5. SCALE OF TRANSPORTATION USES

These data are available for national state, and local use, mostly from vendors and transportation research agencies. Currently, they only support nationwide analyses related to national speed data and traffic-volume datasets.

6. GENERAL DETAILS

The following sections detail background information on how CV data are generated, why they exist as a data type, and the pieces of information generated and contained in their data points.

6.1. How Are Data Captured

The following sections detail elements of the technology underlying the generation, capture, conveyance, and protection of CV data points.

6.1.1. Collection Frequency and Event Trigger Mechanisms

CV data points are generated under the following circumstances:

- CV data are collected by purpose-built onboard units (OBUs), which are installed in vehicles by their OEMs.
- CV data can be collected as follows:
  - Actively collected after a specific trigger: vehicle ignition (on or off), safety events (hard braking, crash detection, etc.)

---

[1] Numerous companies outside traditional planning and engineering are developing technologies that use these data to monitor driving habits: https://www.cmtelematics.com/ [Cambridge Mobile Telematics 2025].

- Passively collected from infrastructure communications (i.e., cross walk warnings of the presence of pedestrians, among others) [FHWA 2019]
  - Continuously collected:
    - Safety-critical Basic Safety Messages (BSMs), sent 10 times per second (every 100 ms) [Chen et al. 2021][2]
    - Regular vehicle position messages, sent every 1–5 seconds
    - Event-specific messages, sent at the exact time and location of key events, i.e., ignition turned on or off, seatbelt utilization, airbag going on or off, abrupt changes in or around the vehicle, or a collision
- Communications occur between the vehicle and infrastructure when it is in range of RSUs. An RSU is a transportation infrastructure communications device that is a part of a Cooperative ITS transportation environment. RSUs are independent of the OEMs (OEM-neutral in design) and typically installed by the department of transportation (city, county, state, or toll authority).
- Sightings are generally regular and predictable, especially for V2V safety messages and V2I event messages.

6.1.2. Location Determination Technology

The location elements of CV data points are generated and defined under these conditions:

- Coordinates are determined through dedicated Global Positioning System (GPS) receivers, integrated with CV OBUs.
- CVs have higher precision than consumer devices, often augmented with differential GPS and interactions with detailed mapping data for enhanced precision and other technologies [USDOT 2019].
- CV-based position accuracy typically achieves 1–3 feet in ideal conditions.
- Augmented positioning using V2I (e.g., RSUs) can maintain accuracy in environments that can block or disrupt GPS reception (tunnels, urban canyons).

6.1.3. Data Transmission

CV data points are transmitted from the vehicle through the following procedures:

- Data are transmitted via dedicated communication channels, such as Dedicated Short-Range Communications at 5.9 GHz.
  - Data are initially stored on devices attached to the infrastructure as well as stored by the OEM.
  - Infrastructure streams data to traffic-management centers (TMCs) and cloud data centers via cellular and fiber-optic networks, although the precise technology and nature of these connections is not publicly known.
- Transmission security uses specialized encryption and certificate-management systems.
- Data prioritization occurs automatically, with safety messages being given highest priority.
- Communications typically range between 850–1,650 feet for direct V2V and V2I.

---

[2] The Minnesota Department of Transportation defines BSMs: containing data about the vehicle's position, speed, and acceleration. Roadside receivers, RSUs, can capture BSM broadcasts and translate them into information about traffic conditions. If every vehicle is equipped with awareness, BSMs can be combined to calculate traffic flows, speeds, and densities [Chen et al. 2021].

6.1.4. Identifiers and Privacy

Identifying information on the owner of an LBS data-generating device is protected by the following procedures:

- The OEMs collect and store all data received from the CV, a subset of which is passed onward to consumers.
- The base unit of commercially available CV data is a trip, which is defined as the GPS location pings and vehicle events between the vehicle's key-on event and the vehicle's subsequent key-off event.
- Trips and their positional GPS and event data are segmented and assigned a unique hashed ID to prevent trip-chain reconstruction (i.e., it is not possible to know which trips preceded or followed a given trip).
- Trips may include a Squish Vehicle Identification Number (VIN), which is a shortened VIN and allows for the identification of the vehicle model and engine type. Squish VIN tables, however, are available online for public use [ESP 2025].
- Additional security measures, known as Security Credential Management Systems (SCMSs), are implemented for V2I and V2V interactions (SCMSs are similar to secure sockets layer and transport layer security encryption).

6.1.5. Observation Unit

The unit of analysis for CV data is the vehicle whose onboard computers produce the core data packets.

In turn, these data packets come in two flavors:

- The coordinates and timestamps of the vehicle's trip (i.e., a breadcrumb trace of sightings)
- The labeled events and their GPS sightings and timestamps that occurred during the trip (e.g., hard braking, tire-pressure warning, seatbelt usage, and collisions)

The multiple sensors that provide vehicle state data, reflected in both data packet varieties, can also be a source of event data, such as passenger airbag deactivation.

6.2. Data Source Background

The following sections provide background on the history, development, and application of CV data, especially vis-à-vis the generation of transportation statistics.

6.2.1. Original Purpose

The emergence and development of CV data have been centered on the following goals:

- CV technology was developed to improve transportation safety and efficiency within the framework of ITS, which encompasses a broad range of technologies and systems designed to improve transportation safety, efficiency, and overall performance of the transportation system.

- The basic technology was created through government–industry partnerships and data standardization. Nevertheless, each OEM may have additional technologies and features that deviate from others, and they may not all have a standardized set of sensors contributing to the standardized set and format of messages the vehicles send.
- Core applications include the following:
  - Collision detection and avoidance
  - Intersection safety and performance
  - CV corridors, which use CV technology to optimize vehicle flows either along a highway or across signalized intersections and have typically been developed as public–private partnerships, with some level of government installation of RSUs

6.2.2. Transportation Focus

As CV data have developed as a source, their transportation-centricity has been characterized as follows:

- Whether collected and used internally by OEMs or made available to analytics companies and public entities, CV data have been primarily used for transportation research. Documented applications within transportation include:
  - Safety studies
  - OD studies
  - Congestion-management studies
  - Speed studies
  - Route studies
  - Special events studies
- Direct integration via fiber-optic and cloud infrastructure with TMCs and signalized intersections improve and measure system performance.

6.3. Data Contents

The following sections note the pieces of information contained in CV data points, including elements generated at inception as well as attributes that must be estimated.

6.3.1. Core Attributes

The following elements are available for all CV data points and are generated by default by an onboard computer:

- Timestamp
- Latitude–longitude coordinates
- Squish VIN to identify vehicle model, body style, and engine classification
- Direction of travel (i.e., heading)
- Speed and acceleration, via speedometer
- Key-on and key-off events (plus coordinates and timestamp)

6.3.2. Non-Universal Attributes

Data elements that may be available for a CV data point—outside the core attributes but still generated by an onboard computer—include the following:[3]

- Hard braking and antilock brake activation
- Seatbelt and airbag status
- Lane-departure warning
- Exterior lights status (brake lights, turn signals)
- Wiper status
- Vehicle automation level and mode

6.3.3. Derived Attributes

Data elements that are not inherently generated by the CV's onboard computer, and therefore must be derived or estimated by context, include the following:

- Map or route location may include a derived road network name although the source of that road network data is not documented.
- Traffic signal phase and timing information may be included if using data from connected intersections.

6.3.4. Sample Characteristics

The sample penetration of CV data vis-à-vis the traveling American public is determined as follows:

- CV data are currently limited to equipped vehicles (market penetration began around 2015), but they are implemented almost universally across OEMs that produce CVs.
    - Not all OEMs resell CV data.
    - Some OEMs work exclusively with a data analytics company to package data for commercial purposes.
- Samples are reduced in regions with older, non-CV vehicle models or where CVs have greater or lesser market presence due to individual make and model preferences.
- More concentrated data are available in specific corridors with infrastructure deployment.
- Government fleets and transit vehicles are frequently equipped.

6.3.5. Data Quality

Documented concerns about the quality and usability of CV data include the following:

- A limited number of CVs are on the roadway relative to the full population of non-CVs.
- CV data are limited to passenger vehicles.[4]

---

[3] Attribute data from CVs are continually evolving, and older vehicles, even if they are CVs, may not have the same attributes as newer vehicles.

[4] CVs can also be trucks as this technology is important for automated driving [Abrams 2025].

- Trips can be round-trip if the vehicle is not keyed off during intermediate stops (such as drive-throughs or school drop-offs), resulting in high circuity ratios that reflect total distance traveled rather than direct origin-to-destination trips.
- Trips are not linked and require special processing[5] to develop tours or identify long-distance trips.

7. TEMPORAL AND SPATIAL SCALES OF THE DATA

The following sections characterize the historical, temporal, and spatial coverage, availability, and precision of CV data.

7.1. Temporal Coverage

Details on the historical penetration and availability of CV data in the United States include the following:

- Limited historical data due to emerging technology but increasing detail as the technology evolves
- Most deployments operational from 2015 onward
- Data volume increasing with CV vehicle-market penetration

7.2. Temporal Resolution

Details on the precision of temporal information generated with CV data points include the following:

- BSMs typically 10 times per second (100ms), measuring factors like speed and braking
- Signal phase and timing messages usually 10 times per second [FHWA 2018]
- RSU broadcasts 1–10 times per second
- Event-based messages triggered immediately if detected (e.g., airbag inflation, collision)

7.3. Spatial Coverage

Details on the geographic coverage and availability of CV data in the United States include the following:

- Coverage is available throughout the United States but varies due to the following:
    - Spatial variation in penetration of newer vehicle models
    - Spatial variation in market preferences for certain vehicle manufacturers
    - Granularity and spatial distribution of testing corridors and within Smart-Cities programs [USDOT 2025]
- Coverage is growing with infrastructure investment.

[5] CV datasets are very large (terabytes to petabytes) and thus require large compute and storage for analysis.

7.4. Spatial Resolution

The spatial information generated by a CV is very high-precision, GPS-based position data, especially while in optimal conditions for satellite detection. Location data points are precise enough to identify lane-specific vehicle positions.

During these cases, when GPS is not available, the other sensors and receivers involved in producing and receiving CV data can step in to ensure high-precision location data.

8. USE CASES OF THE DATA

The following sections note documented use cases and applications of CV data, with specific attention to those related to transportation and transportation statistics.

8.1. Transportation Use Cases

Examples of transportation-centric applications of CV data include the following:

- Transportation-planning applications:
  - OD studies, which use the key-on and key-off events and coordinates to determine OD as well as trajectories to determine intermediate ODs, such as geographical borders
  - Travel time reliability analysis, which uses speed data [StreetLight 2024]
  - Special event and route studies, which match the data points to specific routes
- Traffic-engineering applications:
  - Signal timing optimization based on Automated Traffic Signal Performance Measures
  - Bottleneck identification and mitigation
  - Safety analysis to identify locations prone to hard braking and abrupt lane changes
  - Corridor and network performance
- Travel statistics generation by transportation departments and planners:
  - Peak period and rush hour traffic studies and seasonal traffic trend analyses
  - Travel Time Index and Planning Time Index analysis
  - National Performance Management Measure reporting, including level of travel time reliability
- Incident and emergency management:
  - Automatic incident detection and verification
  - Emergency response route planning
  - Evacuation effectiveness monitoring
  - Post-incident analysis and system recovery assessment
  - Secondary crash-risk prediction and prevention

8.2. Nontransportation Use Cases

Examples of nontransportation-centric applications of CV data include the following:

- Emergency response coordination among government agencies
- Evacuation management
- Insurance telematics and risk assessment
- Vehicle emissions monitoring (CVs can be both electric and gas powered)

REFERENCES

Abrams, Cameron. 2025. "Self-Driving Semi-Truck Company Launches Operations in Texas," *The Texan*. Austin, TX: The Texan. https://thetexan.news/issues/transportation/self-driving-semi-truck-company-launches-operations-in-texas/article_2f7fa446-62e3-40d0-a88c-18159d59b8ae.html. Last accessed September 18, 2025.

Cambridge Mobile Telematics. 2025. *Cambridge Mobile Telematics – The world leader in telematics*. Cambridge, MA: Cambridge Mobile Telematics. https://www.cmtelematics.com/. Last accessed September 18, 2025.

Chen, Rongsheng, Michael Levin, John Hourdos, and Melissa Duhn. 2021. *Generating Traffic Information from Connected Vehicle V2V Basic Safety Messages*. St. Paul, MN: Minnesota Department of Transportation. https://mdl.mndot.gov/items/202108. Last accessed September 18, 2025.

ESP. 2025. *VIN Research – SquishVIN Data Tables – ESP Data Solutions*. Wilbraham, MA: ESP Data Solutions. https://www.espdata.com/core-products/vinpower-squish-vin-flat-data-tables/. Last accessed September 18, 2025.

FHWA. 2018. "Smart Vehicles, Smart Signals, Smart Cities: How Cyber Physical Systems Can Foster Traffic Synergy." Washington, DC: Federal Highway Administration. https://highways.dot.gov/media/326. Last accessed September 18, 2025.

———2019. *Effects on Intelligent Transportation Systems Planning and Deployment in a Connected Vehicle Environment*. Washington, DC: Federal Highway Administration. https://ops.fhwa.dot.gov/publications/fhwahop18014/chap3.htm. Last accessed September 18, 2025.

GVR. n.d. *Connected Car Market Size, Share & Trends Analysis Report By Connectivity Solutions (Embedded, Integrated, Tethered), By Technology, By Application, By Sales Channel (OEM, Aftermarket), By Region, And Segment*. San Francisco, CA: Grand View Research. https://www.grandviewresearch.com/industry-analysis/connected-car-market. Last accessed September 18, 2025.

ITE. n.d.a. "Connected Automated Vehicles/Connected Interoperability (CAV/CTI) Standards," *Resources*. Washington, DC: ITE. https://www.ite.org/technical-resources/topics/standards/cav-cti-standards/. Last accessed September 17, 2025.

———n.d.b. "Connected/Automated Vehicles," *Resources*. Washington, DC: ITE. https://www.ite.org/technical-resources/topics/connected-automated-vehicles/. Last accessed September 17, 2025.

Larkin, James, James O'Hara, Michael McGurrin, and Gene McHale. 2016. *USDOT Guidance Summary for Connected Vehicle Deployments: Application Deployment*. Washington, DC: USDOT. https://rosap.ntl.bts.gov/view/dot/31550. Last accessed September 18, 2025.

StreetLight. 2024. *Speed Validation White Paper*, Version 2.0. San Francisco, CA: StreetLight.

USDOT. 2016. “Connected Vehicle Applications: Safety.” Washington, DC: U.S. Department of Transportation. https://rosap.ntl.bts.gov/view/dot/30928. Last accessed September 17, 2025.

———2019. *Nationwide Differential Global Positioning System (NDGPS) Program*. Washington, DC: U.S. Department of Transportation. https://www.transportation.gov/pnt/nationwide-differential-global-positioning-system-ndgps-program. Last accessed September 19, 2025.

———2020. *Connected Vehicle Deployment Technical Assistance: Roadside Unit (RSU) Lessons Learned and Best Practices*. Washington, DC: U.S. Department of Transportation. https://rosap.ntl.bts.gov/view/dot/50749. Last accessed September 17, 2025.

———2025. *SMART Grants Program*. Washington, DC: U.S. Department of Transportation. https://www.transportation.gov/grants/SMART. Last accessed September 19, 2025.

About This Document

Title:	*BTS Transportation Probe Data Guide: Connected Vehicle Data*
Publication Date:	April 2026
Contributor(s):	BTS: Kyle Titlow, Mindy Liu, Ed Strocko, Rolf Schmitt, and Devon Lechtenberg. MacroSys: Stacey Bricka WRA: Byron Chigoy and Krishnan Viswanathan
DOI:	10.21949/sd55-6x72