New Data Inventory Matrix

The quality and relevance of data are crucial for accurate and reliable forecasts in transportation planning and modeling. While many data sets used by the forecasting and modeling community are publicly available, additional specialized, privately sourced data are often necessary to enhance the modeling process. In the absence of a household travel survey, new data vendors offer similar information using proprietary data collection methods. With an increasing number of data vendors in the market, it can be challenging to determine the most suitable sources for the analysis at hand. This document provides a data matrix designed to give transportation planners, modelers, and other professionals a comprehensive overview of the types of data collected by some of the most prevalent vendors in the field. The matrix categorizes key data attributes, enabling professionals to easily assess which data sources best align with their specific needs.

The matrix provides the following publicly available information:

Data Type – Refers to the nature of the data; that is, the specific kind of information collected (e.g., trip data, O-D flows, safety data, truck movement data, etc.).

Data Source – Defines where or how the data is collected. Common sources include GPS, LBS, CV data, and sensor-based detections.

Data Release Cadence – Indicates how frequently new data is made available (e.g., real-time, weekly, monthly, quarterly). Some vendors, like TomTom, offer real-time data, whereas others, like LOCUS and StreetLight, provide monthly, weekly, or quarterly updates.

Data Latency – Refers to the time periods reflected in the data (e.g., real-time, daily, weekly).

Data Resolution – Indicates the spatial or temporal granularity level of the data (e.g., hourly, 15 mins OR longitude/latitude, census block, etc.).

Data Accessibility – Indicates how users can access the data (e.g., self-serve platform, API, etc.).

Data QA/QC – Describes the validation methods or protocols used to ensure and check the data's accuracy, reliability, and overall quality. For example, data from vendors like AirSage and INRIX are independently validated, while Geotab uses a patented algorithm to validate their data.

Travel Modes – Outlines the types of transportation modes captured in the data (e.g., auto, transit, bicycle, pedestrian, and trucks).

Data Can Be Expanded To – Gives specific details on what additional data types are used to expand or enrich the primary data set (e.g., traffic data, population data, synthetic data, etc.).

Expansion Data Sources – Indicates additional data sources that can be used to enhance or expand the primary data (e.g., ACS, Census, etc.).

Resources – Includes references to vendor websites or reports that provide further information on relevant vendor details.

	AirSage	ATRI	Flow Labs	Geotab	HERE	INRIX	LOCUS	Replica	StreetLight	Streetlytics	TomTom
Data Type	O-D flows, Safety data	Truck movement data	Signal data, Detection data, CV data	Trip data, Vehicle trajectory data (e.g., harsh braking)	Real-time and historical traffic data	Traffic data, Trip data, Parking data, Safety data	O-D flows	O-D flows, Traveler demographics, AADT, Speed profiles, Turning movement counts	Trip data, O-D flows, Routing, AADT, Turning movement counts, VMT, VHD	O-D flows, Traffic counts, Traveler demographics, Mobile data	O-D flows, Traffic data
Data Source	GPS, LBS, CV	GPS	CV, Data from sensors & detectors	GPS, CV	GPS	GPS	LBS	LBS, CV	GPS, LBS, CV	GPS, Cellular	GPS
Data Release Cadence	-	Monthly	-	Real-time	Two days	Real-time	Quarterly	Weekly	Weekly, Monthly, Semi-Monthly, Annual	-	Real-time
Data Latency	Hourly	Real-time	Real-time	Real-time	5 mins	15 mins	Hourly	Real-time	15 mins	-	Real-time
Data Resolution	LBS: census block, GPS: lat./long.	Lat./long.	-	Lat./long., Geohash	50 cm	Lat./long., 10m -100m roadway sections	Census block	Census block	Lat./long., Census block	Lat./long.	30-50 cm
Data Accessibility	Self-serve platform	CSV format	AI-powered self-serve platform	Self-serve platform, API	Self-serve platform, API	Self-serve platform, API	Self-serve platform	Self-serve platform	Self-serve platform, API	-	Self-serve platform, API
Data QA/QC	Independent validation by the University of Maryland; FHWA, NHTS	-	Data validated at 94.4% accuracy	Validated using curve logic algorithm (a patented algorithm)	Travel times validated against Google Maps	Independent validation of private sector traffic data	Validated against counts, household & transit surveys, farecard & toll data	Validated against travel/on-board surveys, counts, and customer-provided data	Validated against permanent counter data and external sources (surveys, sensors, etc.)	Validated against nationwide survey of Mobility in Germany (MiD)	Internal quality checks (e.g., anomalies in data patterns, missing data, etc.)
Travel Modes	Vehicles, Bicycle, Pedestrian	Heavy-duty (commercial) trucks	Private auto, Trucks	Private auto, Trucks	Vehicles, Trucks	Private auto, Trucks	Motorized, Pedestrian, Bicycle	Private auto, Transit, Taxi/TNC, Bicycle, Pedestrian, Medium and heavy trucks	Private auto, Transit, Bicycle, Pedestrian, Trucks	Vehicles, Transit, Pedestrian	Private auto, Trucks
Data Can Be Expanded To	Survey data	Traffic data Survey data	Freight data, Pedestrian push button data	Traffic data, GIS data	GIS data	-	Population, School & employment	Synthetic data set	Census data, Traffic counts	Population data	GIS
Expansion Data Sources	NHTS, FHWA	FHWA	Geotab	-	-	-	ACS	Census (PUMS)	-	ACS	-
Resources	AirSage	NOCoE; ATRI Data	Flow Labs; Announcement	Geotab	HERE Traffic; Imagery Contract	INRIX	Locus	Replica; FTA Report	StreetLight	CITILABS; Streetlytics	TomTom

"-" indicates unavailable information

Last Updated: October 2024

Acronym Glossary

AADT—*Annual Average Daily Traffic*—represents the average number of vehicles using a specific segment of the roadway (in both directions) per day, calculated over the course of a year.

ACS—*American Community Survey*—is an annual survey conducted by the U.S. Census Bureau to collect detailed social, economic, housing, and population data for the nation, states, counties, and other geographic areas down to the block group level.

AI—*Artificial Intelligence*—is technology that enables computers and machines to simulate several aspects of human intelligence, such as learning, reasoning, problem-solving, creativity, and autonomy.

API—*Application Programming Interface*—is a set of rules or protocols that enables different software applications to communicate and share data.

CSV—*Comma-Separated Values*—is a text file consisting of a data record with fields separated by commas

CV—*Connected Vehicle*—a vehicle that uses wireless communication technology to exchange data with other vehicles, infrastructure, and systems. This provides useful information to the driver, enabling them to make safer and more informed decisions.

FHWA—*Federal Highway Administration*—is an agency within the U.S. Department of Transportation responsible for overseeing and supporting the design, construction, maintenance, and safety of the nation's highways and various federally and tribally owned lands.

GIS—*Geographic Information Systems*—is a computer system for capturing, checking, storing, and visualizing spatial and geographic data.

GPS—*Global Positioning System*—is a global navigation satellite system that provides location and time information of an object anywhere on Earth.

LBS—*Location-Based Services*—refers to services that utilize real-time location data from mobile devices to offer information, navigation, or other services.

NHTS—*National Household Travel Survey*—is a survey conducted by the Federal Highway Administration to collect data on personal and household travel behavior in the U.S.

O-D—*Origin-Destination*—data or matrices that show the starting (origin) and ending (destination) points of trips.

PUMS—*Public Use Microdata Sample*—is a subset of raw data from the U.S. Census and ACS that allows for detailed research at the individual and household level.

QA/QC—*Quality Assurance and Quality Control*—are procedures and protocols used to ensure the accuracy, reliability, and quality of data or processes.

TNC—*Transportation Network Company*—is a company (e.g., Uber or Lyft) that uses app-based platforms to provide on-demand transportation services for passengers.

VHD—*Vehicle Hours of Delay*—measures the degree of congestion on a roadway by calculating the extra time spent in traffic beyond what people would experience if they were traveling at free-flowing speed.

VMT—*Vehicle Miles Travelled*—represents the total number of miles traveled by all vehicles in a specific area over a given period.