Traffic Operations

November 2025

Project Title:
Big Data Deployment Requirements for Transportation Use Cases

Task Number: 4583

Start Date: April 1, 2026

Completion Date: March 31, 2028

Task Manager:
Abdullah Faiyaz
Transportation Engineer, Electrical
Abdullah.Faiyaz@dot.ca.gov

DRISI provides solutions and knowledge that improves California's transportation system.

Big Data Deployment Requirements for Transportation Use Cases

The development of practical deployment requirements for big data to better support current and future transportation use cases. The deployment requirements will include the required inputs of traffic measurements and the minimum technical specifications necessary for sampling rate, penetration rate, latency, data fusion, and aggregation approaches to ensure accurate and reliable estimates of traffic measurements.

WHAT IS THE NEED?

Big data has been utilized to support various transportation use cases, including managed lane performance assessment, route estimation, and traffic demand modeling. However, with recent rapid development in sensing technologies, network communications, and artificial intelligence (AI), the increasing amount of data available from various heterogeneous sources creates a number of challenges that hinder transportation agencies from applying big data to support their current practices of transportation planning, traffic monitoring, and management. For example, speed measures calculated from the data retrieved from probe vehicles or smartphones can be biased or even error-prone due to low penetration rates. Vehicle/pedestrian route inference using GPS traces can be wrong due to low sampling rates and limited GPS accuracy. Travel demand estimation using data fusion from multiple data sources, such as travel surveys, probe traces, and traffic sensor measurements, can be significantly inaccurate due to variations in sampling methods and potential demographic biases in the data. The deployment of AI-powered video/ Light Detection and Ranging (LiDAR) sensors at road intersections can replace the conventional approach to assessing intersection performance, such as delay and LOS, but requires the deployment of a new Intelligent Transportation Systems (ITS) architecture to support the data management and communication needs. With the rapid adoption of Generative AI (GenAI) by the public, it is becoming increasingly challenging to assess the veracity of data retrieved from social media. Considering the above

challenges, particularly the concerns regarding data velocity, variety, and veracity, there is an urgent need to develop deployment requirements for big data that are tailored to meet the needs of various transportation use cases.

WHAT ARE WE DOING?

This research will conduct a comprehensive review of existing transportation use cases and big data sources. Develop algorithms to replicate the identified big data sources and transportation use cases, and leverage the synthetic "ground-truth" data generated from microsimulations in an extensive corridor network to assess the performance of these use cases under various common data issues. Develop data normalization standards. Develop deployment requirements for big data to support better operations of the transportation use cases.

M1 (2 M): Review results of transportation use cases
M2 (3 M): Review results of big data sources
M3 (11 M): Technical memo on the performance assessment
M4 (16 M): Technical memo on the big data deployment requirements, including data normalization requirements
M5 (24 M): Final report and workshops

Resources: Dr. Qijian Gan & Postdoc Researcher. Data from the I-210 corridor & One Aimsun software license.

WHAT IS OUR GOAL?

The objective of this research is to develop practical deployment requirements for big data that support current and future transportation use cases more effectively. The deployment requirements should include, but are not limited to, the required inputs of traffic measurements and the minimum technical specifications necessary for sampling rate, penetration rate, latency, data fusion, and aggregation approaches to ensure accurate and reliable estimates of traffic measurements.

This research is a timely effort to tackle the challenges (e.g., data velocity, variety, and veracity concerns) in managing and properly using big data from various heterogeneous sources, especially from those where emerging technologies like advanced infrastructure sensors, artificial intelligence (especially GenAI), Vehicle-to-Everything (V2X) communications, and connected and automated vehicles (CAVs) are involved.

This research is the first deployment-oriented effort that oversees data issues at a system level and systematically develops deployment requirements to provide Caltrans guidance on the procurement and usage of big data to support various transportation use cases.

WHAT IS THE BENEFIT?

Caltrans can leverage the big data deployment requirements as a guideline to make informed decisions on the procurement of third-party data to support its transportation use cases.

The big data deployment requirements developed in this study will help state agencies like Caltrans better understand the data needs to support a particular transportation use case, the available big data sources, and the corresponding minimum technical requirements. This will help state agencies, such as Caltrans, make informed decisions when procuring the right big data sources to meet their needs.

A deeper understanding of big data and the technical requirements to support various transportation use cases can help foster public-private partnerships in data exchange and sharing. This can increase the effectiveness, efficiency, and resiliency in managing our statewide transportation system, while also reducing costs by eliminating ad hoc and redundant efforts.

WHAT IS THE PROGRESS TO DATE?

The research contract is currently being reviewed

by the Department of Procurement and Contracts
(DPAC) and has not been executed.