

## Research Connection: Hybrid Data Implementation

### Questions & Answers

1. Does Caltrans have a contract or contracts for districts to use third-party data such as INRIX and/or big data aggregation and dissemination platform such as RITIS (Regional Integrated Transportation Information System)? This will be great resources for us to complement actual counts and PeMS data from our count stations. What is the process to request access to the data and/or systems?

No comment on Caltrans contracts, or processes.

2. What types of data are we getting from this third party?

No comment on current Caltrans contracts, please see the final report for a survey of available offerings from leading vendors.

3. Can this also do vehicle classification? What accuracy percentage will this record?

Some vendors maintain separate sets of information based on whether data are derived from consumer or commercial fleets. We are not yet aware of something as detailed as specific vehicle type classification.

4. Was there any analysis (or plan for future analysis) on hybrid or 3rd party data integration for ramp meter operation? Currently D3 uses occupancy from upstream detector stations to develop ramp meter timing plans.

No analysis was performed for integration with real-time traffic control systems. However, this is a great area for future research.

5. What are the parameters defining "quality"?

There are many parameters for defining quality. Some high-level parameters include the following:

- Accuracy—Degree to which the data from data feed matches ground truth
- Accessibility—Relative ease in the use of the data feed for a given traffic application
- Completeness—Degree to which values are present for all the fields of the data feed
- Availability—Degree to which real-time data feeds are online, operating correctly, and providing data
- Coverage—Degree to which the data feed represents the whole of what is to be measured
- Purity—Degree to which the data feed consists of raw data
- Validity—Degree to which the data feed fields satisfy acceptance requirements

6. For real world pilot, what would be recommended way of measuring delays.

For a real-world pilot, it depends on the infrastructure type. The details are in the report, but for mainline freeways, the best method was the hybrid calculation using flow information from fixed-point sensors, such as loops, and speed information from

## Research Connection: Hybrid Data Implementation

### Questions & Answers

vendors. For HOV lanes, third-party data are not yet available. More research is needed to address delay on on-ramps, off-ramps, or arterial roads.

7. Anthony, do you have any suggestions on what Caltrans can do to improve its point detection data quality, especially meta-data?

The way to improve data quality is to make sure that someone or some group at Caltrans has a job description that is specifically data quality. Data quality is something that needs to be actively assessed from day-to-day, because conditions change from day-to-day.

Ideally, the meta-data (such as position information) should be unambiguous so that there is no doubt about the area of pavement being monitored. Geospatial coordinates on a map of the road network would be helpful.

8. What simulation software was used?

Aimsun was used in this study.

9. Traditional data comes from loops per lanes, but third-party data does not. How do you compare these two or how do you mix them together? Averaging?

The details are in the report. The short answer is that we use the flows and densities from the loops. This information is then projected upstream and downstream along the freeway. The third-party data is used to obtain vehicular travel times. Once speed and flow are known over the same length of roadway, one can then calculate delay.

10. Any observations on how 3rd party data is useful for estimating traffic volume measures (e.g., VMT); what penetration rates are you seeing with third party data?

In general, penetration rates are proprietary and not published in detail. There are equity challenges concerning the use of third-party data for flows. Whose flows are you measuring? Is the entire population represented fairly in these flows or is it biased?

11. Do third parties provide on-ramp and off-ramp traffic data?

Yes. Data is also provided on arterial streets.

12. Can We please know distribution of Hybrid in terms of percentage of third party and point detection data? let's say is it 5/5 or 4/6 etc.?

For third party data, the assumption was that 5% of vehicles were tracked. The assumption for loops was that they were working properly and measuring 100% of vehicles that pass over them.

## Research Connection: Hybrid Data Implementation

### Questions & Answers

13. I like your focus on point data quality over quantity. Does a hybrid data approach open the door to optimizing the Caltrans census network in terms of the number and location of point sensors needed to provide the quality calibration data for a hybrid approach?

This is a very interesting question. Point data is well suited for obtaining flows. It does make sense to optimize the deployment of point-sensors to obtain overall flow data needed for Caltrans' business decisions. The hybrid approach enables speed data to be captured through other means. This could open the door to a new integrated approach to collect enough data for both MPR type reporting and Census type reporting. This is a great area for future research.

14. Can you explain how we will be performing VMT analysis using this new methodology?

Details are in the report. However, the simple answer is that the hybrid calculation includes a process to project flow data upstream and downstream and to calculate measures (including VMT) over a more detailed domain of analysis. It did not make much of a difference for VMT, although it did slightly improve the accuracy of the calculation.

15. In this study, Is Probe data and Location Based Service (LBS) data both were used or only LBS?

In this study, speed data was synthesized using simulation to mimic what would be available from GPS-based travel times that are provided by third parties. This is different than what is sometimes called "cell tower" based data where only the general location of the device is known, but not its GPS coordinates.

16. Some of the third-party companies included Caltrans PeMS data as a part of their traffic analysis algorithm, would you separate out those different data sets to avoid "double-counting" the PeMS data?

This is a very interesting question. The proposed method was intended to utilize data from fixed, point sensors and from mobile sources. If the third-party is already using PeMS-based flows, then they could use that information to provide VHT as well as delay. It is difficult to separate out that data once it is already mixed into a product.

17. What do you suggest for improvements that can be implemented for VDS systems?

The main improvements would be (1) to prioritize having periodic and active assessments of data quality, (2) upgrading from a linear reference system to a more general (map-based) method for specifying location information about infrastructure assets.

## Research Connection: Hybrid Data Implementation

### Questions & Answers

18. Please speak more to the issue of integrating the base map data between Caltrans and third-party providers

The main challenge here is that to use both third party data and existing Caltrans data together, one needs to be specific about which section of pavement the data cover. One does not want to mix up data for pavement section A with data for pavement section B. To do this at scale requires that the meta-data (data about the data) is available, usable, and correct.

The available meta-data in PeMS, for example, uses a linear reference system that provides an approximate position of the sensor on the pavement. At freeway interchanges and connectors that approximate position information is difficult to use. More details are contained in the final report.

19. When NPMRDS is available free, why should we look for third party private vendors and spend money?

NPMRDS provides historical data, but not real-time data. If historical data are sufficient for your needs, then you would do well to use them.

20. So, you are saying that you can get the vehicle classification data. Can you provide Electric Vehicles (EV) percentage along the road/corridor?

I am not aware of third parties providing detailed vehicle classification data at the time of this study.

21. FHWA for census requires 95% data collection accuracy

No comment

22. Any potential cross connection with GTFS - Real-time data feeds for transit currently or in future plans by any third-party aggregator you know.

I am not aware of this.

23. What technologies are the 3rd parties using? is it mainly Bluetooth? Thanks

All the commercial vendors of "big-data" roadway traffic utilization and performance data depend heavily on smartphone applications, in-vehicle OEM navigational devices, and data from connected vehicles. To obtain these data, the vendors have business agreements with multiple cell phone manufacturers, carriers, and/or smartphone app providers.

24. How hybrid Data can be used if Freight Management operations?

No comment

## Research Connection: Hybrid Data Implementation

### Questions & Answers

25. District 5 is using more and more Side Fired Radar for VDS instead of loop detection. Has there been any comparison whether the data from Side Radar Data gives better quality than traditional loops?

No comment

26. Third part data are mostly proprietary and like a black box we have no idea how they process their data.

No comment

27. Are you familiar with Replica modeling tool? If so, how does it compare with Streetlight data?

I am not familiar with the Replica modeling tool.

28. How will this research help our operations of the State Highway System be more reliable?

Ultimately, better data can be used to help improve business decisions. This research provides a path forward for using new data resources.