



Caltrans Division of Research,
Innovation and System Information

Research

Notes

Modal

MAY 2022

Project Title:
Advancing Adoption of General
Transit Specification Feed (GTFS)
Programs in California

Task Number: 3515

Start Date: September 1, 2020

Completion Date: August 31, 2022

Task Manager:
Bradley Mizuno,
Transportation Engineer, Electrical
bmizuno@dot.ca.gov

Advancing Adoption of General Transit Specification Feed (GTFS) Programs in California

This project develops and tests a method for assessing General Transit Specification Feed (GTFS) accuracy in relation to vehicle location information.

WHAT IS THE NEED?

The General Transit Specification Feed (GTFS) and its more recent extension to real-time data allow transit agencies to make their route and schedule information (as well as real-time updates) available to external trip planners. Research has found that the availability of this information, particularly via smartphone applications increases the legibility, predictability, and, consequently, use of transit services. Furthermore, the availability of this schedule and real-time information is a critical basis for any Mobility-as-a-Service offering.

California has an unusually large array of transit providers that often offer connecting and overlapping services. There is a need to provide the underlying trip data so that public and private entrepreneurs can make these data available to the public to maximize the use of the associated transit resources as well as promote integrated mobility services.

Finally, these data provide an important historical record of transit service provision and performance throughout the state. Caltrans has a long-term interest in archiving this material to manage the transportation system.

Through the California Integrated Travel Project (Cal-ITP), information about GTFS coverage, existing gaps, and barriers to adoption have been assessed through interviews with agencies and data inventorying. California's recently released minimum GTFS guidelines provide information and support for agencies implementing the standard. Additionally, Mobility Data, a



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

non-profit organization working with Cal-ITP, is developing a GTFS Validator to assess conformity with the GTFS specification itself as well as a grading scheme to assess accuracy of fixed assets such as route and station names. However, there's currently limited information or tools available to assess the quality and accuracy of the schedule and arrival prediction data.

WHAT ARE WE DOING?

This project will facilitate the adoption of GTFS static and GTFS real-time into the practices of transit agencies across the state as well as create a statewide clearinghouse and historical archive of GTFS information. All transit agencies receiving Federal Transit Agency or Caltrans support will be included.

This project will:

- Review relevant approaches to assessing GTFS accuracy
- Propose an appropriate assessment methodology
- Scan transit agencies to identify five test properties
- Apply methodology at test properties
- Evaluate methodology application
- Consider possible refinements
- Prepare guidance document

WHAT IS OUR GOAL?

The myriad benefits of GTFS for transit trip planning only occur if those data realistically represent transit service. This project develops and tests a method for assessing GTFS accuracy in relation to vehicle location information.

WHAT IS THE BENEFIT?

The proposed project carried out via this task order aligns with two Caltrans priorities:

- Modality – this research will leverage transit by measuring their performance.
- Innovation – this research is embracing the dedicated short-range communications data and other real-time data to help Caltrans solve major planning issues.

WHAT IS THE PROGRESS TO DATE?

This quarter saw the collection of five full days of General Transit Feed Specification Realtime (GTFS-RT) data from five different California transit agencies, the application of the proposed metrics to the full five days of the Modesto MAX data, and some refinement of the measures as well as some additional data cleaning protocols in light of using the 45 million rows of data from Modesto.