Plan Overview

A Data Management Plan created using DMPTool

Title: Intelligent Sensor Integration on Rural Multi-Modal System

Creator: Jim Marino - ORCID: 0009-0004-4278-9465

Affiliation: United States Department of Transportation (DOT) (transportation.gov)

Data Manager: Kira Richardson, Josh Schmid

Project Administrator: Jim Marino, Tara Styer

Funder: United States Department of Transportation (DOT) (transportation.gov)

Funding opportunity number: DOT-SMART-FY23-01

Grant: SMARTFY22N1P1G41

Template: Digital Curation Centre

Project abstract:

The Tahoe Transportation District (TTD) and regional partners are seeking to enhance transportation safety and mobility along key corridors in the Lake Tahoe Basin. The roadway network entering, traveling within, and leaving the Tahoe Basin lacks the infrastructure required to acquire real-time and historical traffic and congestion data. This pilot project aims to support TTD and the regional partners in collecting accurate count data at each of the seven entry/exit points of the Tahoe Basin and along the Truckee/US80/SR267/SR89 roadways.

The project is designed to plan, prototype, test, and evaluate a limited deployment of a data collection sensor infrastructure to gather transportation and traveler-related information. The goal is to integrate this information into a single cloud-based open source or interface for reporting and
management. This information will be utilized by TTD, TRPA, partners, commuters, and travelers within the Tahoe Basin and adjoining areas to provide an integrated infrastructure for collecting vehicle data. This data will be incorporated into a database for various stakeholders.

Furthermore, the project will establish the framework for long-term data collection across the region and integrate multiple transportation data sources for efficient use by partner agencies. It will propel the region toward real-time parking availability for motorists and improve the ease of transit use, walking, and bicycling. Other long-term uses include sharing information about weather hazards, closures, construction, or crashes.

Start date: 09-15-0003

End date: 03-15-2025

Last modified: 01-04-2024

Copyright information:

The above plan creator(s) have agreed that others may use as much of the text of this plan as they would like in their own plans, and customize it as necessary. You do not need to credit the creator(s) as the source of the language used, but using any of the plan's text does not imply that the creator(s) endorse, or have any relationship to, your project or proposal.
Data Collection

What data will you collect or create?

The data collection process will be comprehensive and involve a range of critical metrics encompassing multi-modal traffic and safety data. The data will include items such as, but not be limited to, traffic signal status, traffic-volume data, O-D data, traveler profile data, traffic delays, local weather information, road closures due to snow, avalanches, landslides, flooding, road construction events, street events, vehicular or pedestrian accidents, transit movement and delays, ride-sharing locations and availability, micro-transit locations and availability, and parking locations and availability.

There are additional data sources from other agencies, such as vehicle counts and crash data from NDOT and Caltrans, in addition to other third-party sources like location-based data or GPS (connected vehicle) data. TTD, Placer County, RTC of Washoe County, and other transit providers can contribute transit ridership and vanpool use data. Bicycle and pedestrian count information will be integrated as well.

The collected data will support providing a shared understanding of extensive visitation and address peak demand, including road user travel volumes and patterns. It aims to evaluate road user safety, particularly for vulnerable road users such as bicyclists and pedestrians, and to support Vehicle Miles Traveled (VMT) and Greenhouse Gas (GHG) goals. Additionally, it aims to optimize available parking and transit use, aid in communication with the traveling public, ultimately aiming to protect Lake Tahoe’s famed clarity.

The data collection effort aims to capture a comprehensive view of multi-modal interactions within the transportation system. The scale will likely involve real-time and historical data (as available) across the entry and exit points of Lake Tahoe to gauge the effectiveness of real-time multi-modal traffic data collected by smart sensors. This will be facilitated through a single cloud-based open source or interface for pertinent transportation and traveler-related information. This approach intends to offer a broad scope of insights into the functionality and safety enhancements across different locations, traffic scenarios, and conditions.

Some example data characteristics that will be collected are below.

- Multi-modal corridor volume: Data indicating different modes of traffic flow at specific locations or times.
- Traffic count: The number of modes that pass a certain location.
Accident data: Data concerning crashes (vehicular or vulnerable road users).

Transit information

The data collected will be interconnected, forming a comprehensive map of multi-modal traffic interactions and safety incidents tracked over the course of the pilot period. There are inherent disclosure risks, especially concerning individual privacy and identifiable patterns. Personal details linked to incidents could pose privacy concerns, necessitating anonymization techniques to mitigate such risks before any dissemination or analysis.

The proposed project will not collect any personal identifiable information such as license plates and facial recognition.

The collected data will be archived and stored for future analysis and insights by many potential users such as traffic engineers, safety engineers, road operators, etc. The data collected will drive continual improvements, inform policy and decision-making, provide predictive insights to help mitigate safety risks or traffic congestion based on historical trends, foster innovation, potentially reduce commute times, provide information transparency for users, and ultimately enhance transportation safety and mobility along key corridors in the Lake Tahoe Basin.

**How will the data be collected or created?**

The data will be collected via sensor-based infrastructure with AI software integration yet to be decided and subject to prototyping and evaluation for the remoteness and weather characteristics of the region.

**Documentation and Metadata**

**What documentation and metadata will accompany the data?**

Our intent is to use industry-standard data collection and storage formats to allow for interoperability with other applications. The anticipated file formats for the collected data and related files will likely be a combination of various formats, including but not limited to:

- **CSV (Comma-Separated Values):** Suitable for structured data such as collision incidents, violations, or pedestrian volumes, enabling easy manipulation and analysis using various software tools.

- **JSON (JavaScript Object Notation):** Useful for storing hierarchical data or metadata related to V2X technologies, traffic signal systems, or video analytics configurations.
XML (extensible Markup Language): A potential format for encoding diverse data structures, especially useful for describing video analytics parameters or system configurations.

MP4 or AVI (Audio Video Interleave): Video file formats for storing surveillance or traffic camera footage, essential for video analytics and understanding traffic patterns.

Database Files (SQL, SQLite, etc.): Organized databases might be utilized to store and manage complex relational data, facilitating efficient querying and retrieval of information.

GeoJSON or Shapefiles: Geospatial data formats used to represent geographic features, essential for mapping traffic incidents, pedestrian volumes, or collision locations.

PDF (Portable Document Format): A potential format for documentation, reports, or manuals related to the systems or technologies under study.

The choice of file formats will depend on the nature of the data, its structure, and the intended analyses or applications.

Utilizing platform-independent and non-proprietary formats ensures the accessibility and usability of data across different systems and software, enhancing its long-term utility. The project will adhere to industry standards for data collection and storage formats, prioritizing those widely supported and platform-independent, accessible through various tools and programming languages such as CSV or JSON. Avoiding vendor-specific formats, the project will mandate documentation for metadata in all datasets.

Additionally, we plan to ensure that the solutions employed incorporate a clear and simple API (application programming interface), facilitating seamless integration with other applications and solutions.

**Ethics and Legal Compliance**

**How will you manage any ethical issues?**

This project will not collect or store personally identifiable information. No license plate numbers or facial recognition will be collected. Also, video clips will be deleted from the system after the
traffic and safety analysis is completed.

Data access will also be restricted. Data will only be accessible by personnel linked to the project from TTD/TRPA and appropriate consultants and vendors on the project team on permission basis.

During the period of performance for Phase 1, the responsibility for stewarding and protecting the data will be established per user agreements with consultant and vendor. Data sharing agreements will be formally established between partner agencies.

**How will you manage copyright and Intellectual Property Rights (IP/IPR) issues?**

Data collected during the project will likely not be subject to IP rights. Data that is anticipated to be collected throughout the pilot period will be subject to public disclosure…. Data will be anonymized.

Collected data will not have any licenses restrictions. Specific vendor analytical tools (NOT DATA) that provide unique mining capabilities and insights will mostly likely be subject to IP terms and conditions

TTD is not anticipating enforcing any data terms through a license. This will be updated during project executions

**Storage and Backup**

**How will the data be stored and backed up during the research?**

The project team is in the process of scoping this. We will edit and modify this plan as we move further into the project.

Once the project team completes its plan for any data archiving activities, we will ensure compliance with the US Federal government DCAT-US Metadata Schema.

**How will you manage access and security?**

To be developed in the stage 1 prototyping, evaluation, and planning scope. This DMP will be updated to reflect access and security.

**Selection and Preservation**

**Which data are of long-term value and should be retained, shared, and/or preserved?**
To be developed in the stage 1 prototyping, evaluation, and planning scope. This DMP will be updated to reflect retained, shared, or preserved data.

**What is the long-term preservation plan for the dataset?**

To be developed in the stage 1 prototyping, evaluation, and planning scope. This DMP will be updated to provide long-term data preservation.

**Data Sharing**

**How will you share the data?**

To be developed in the stage 1 prototyping, evaluation, and planning scope. This DMP will be updated to provide information on data sharing.

**Are any restrictions on data sharing required?**

To be developed in the stage 1 prototyping, evaluation, and planning scope. This DMP will be updated to reflect any restrictions on data sharing.

**Responsibilities and Resources**

**Who will be responsible for data management?**

At this time staff from the Tahoe Regional Planning Agency (TRPA) shall be responsible for data management. TRPA is the MPO for the region.

We anticipate through the partnerships developed in this bi-state and multi jurisdictional effort that the data collected would be made available to all partners responsible for transportation planning and delivery within the Lake Tahoe Basin. We anticipate data sharing agreements with our partners.

**What resources will you require to deliver your plan?**

We have contracted with a third-party firm (Parametrix) specializing in the planning and delivery of ITS/ATMS data aggregation projects. We anticipate utilizing third-party cloud based providers to be determined through prototyping and evaluation within the stage 1 scope.

We will be updating this section as the evaluation of hardware and software applications progress.