Intelligent Sensor Integration on Rural Multi-Modal System

Dataset and Contact Information

Please provide as much of the the following information as possible:

1. **Name of the project;**
2. **Grant number;**
3. **Name of the person submitting this DMP;**
4. **ORCiD of the person submitting this DMP (need an ORCID? Register here: https://orcid.org/);**
5. **Email and phone number of the person submitting this DMP;**
6. **Name of the organization for which the person submitting this DMP is working;**
7. **Email and phone number for the organization;**
8. **Link to organization or project website, if applicable; and,**
9. **Date the DMP was written.**

1. Name of Project: Intelligent Sensor Integration on Rural Multi-Modal System

2. Grant Number: SMARTFY22N1P1G41

3. Name of Submitter: Jim Marino, Deputy District Manager

4. ORCiD #: 0009-0004-4278-9465

5. Email: jmarino@tahoetransportation.org

6. Phone: - 775-557-4901

7. Organization: Tahoe Transportation District

8. Email: info@tahoetransportation.org

8. Phone: 775-589-5500

9. Website: https://www.tahoetransportation.org/

10. Date Written: 12/12/23

Data Description

Please provide as much information as possible:

1. **Provide a description of the data that you will be gathering in the course of your project or data from a third party that you will re-use, if any;**
    1. **If there will be no data collected or re-used from another source, state that this is case;**
2. **Address the expected nature, scope, and scale of the data that will be collected, as best as you can at this stage;**

3. **As best as you can, describe the characteristics of the data, their relationship to other data, and provide sufficient detail so that reviewers will understand any disclosure risks that may apply;**
    1. **If data might be sensitive, please describe how you will protect privacy and security, if you know that now;**
    2. **You may need to update your DMP later to add more detail;**
4. **Discuss the expected value of the data over the long-term.**

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.

Data Format and Metadata Standards Employed

Please provide as much information as you can:

1. **Describe the anticipated file formats of your data and related files;**
2. **To the maximum extent practicable, your DMP should address how you will use platform-independent and non-proprietary formats to ensure maximum utility of the data in the future;**
    1. **If you are unable to use platform-independent and non-proprietary formats, you should specify the standards and formats that will be used and the rationale for using those standards and formats.**
3. **Identify the metadata standards you will use to describe the data.**
    1. **At least one metadata file should be a DCAT-US v1.1 (https://resources.data.gov/resources/dcat-us/) .JSON file, the federal standard for data search and discovery.**

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.

- WMV (Windows Media Video): Video file formats for storing surveillance or traffic camera footage, essential for video analytics and understanding traffic patterns.

- GeoJSON or Shapefiles: Geospatial data formats used to represent geographic features, essential for mapping traffic incidents, pedestrian volumes, or collision locations. Some ESRI file formats can be converted to CSV or TIFF, and where possible we will. However, as different layers of information can

get lost during conversion, we intend to include the original ESRI file alongside any conversions in the final dataset. Additionally, there are open-source GIS tools that can open most ESRI file types.

- Notably, there is QGIS https://www.qgis.org/en/site/, which is an open-source software, which is free, to open the files created in ESRI tools, which are not free

- 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. If/Where proprietary formats are utilized there will be documentation addressing why and the software and version needed for a user to open and view the data and/or documentation files.

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.

The final data will have a DCAT-US v1.1 Json metada file, which is the federal standard for data search and discovery to be compliant with the USDOT Public Access Plan.

Access Policies

In general, data from DOT-funded projects must be made publicly accessible. Exceptions to this policy are: data that contain personally identifiable information (PII) that cannot be anonymized; confidential business information; or classified information. Protecting research participants and guarding against the disclosure of identities and/or confidential business information is an essential norm in scientific research. Your DMP should address these issues and outline the efforts you will take to provide informed consent statements to participants, the steps you will take the protect privacy and confidentiality prior to archiving your data, and any additional concerns. In general, in matters of human subject research, your DMP should describe how your informed consent forms will permit sharing with the research community and whether additional steps, such as an Institutional Review Board (IRB), may be used to protect privacy and confidentiality. Additionally, when working with, or conducting research that includes Indigenous populations or Tribal communities, researcher will adhere to the CARE Principles for Indigenous Data Governance https://www.gida-global.org/care and make an explicit statement to that effect in this portion of the DMP.

Please provide as much information as possible:

1. **Describe any sensitive data that may be collected or used;**
2. **Describe how you will protect PII or other sensitive data, including IRB review, application of CARE Principles guidelines, or other ethical norms and practices;**

1. **If you will not be able to deidentify the data in a manner that protects privacy and confidentiality while maintaining the utility of the dataset, you should describe the necessary restrictions on access and use;**

3. **Describe any access restrictions that may apply to your data;**
4. **If necessary, describe any division of responsibilities for stewarding and protecting the data among Principal Investigators or other project staff.**

To safeguard sensitive data and ensure privacy and security the following best practices will be utilized: anonymization techniques, control data access, secure storage, and data record minimization.

Additional mitigations will be added later as we solidify the privacy and security data strategy for this project.

This project will not collect or store personally identifiable information. No licenses 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 be public in compliance with USDOT Public Access Plan. 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.

Re-use, Redistribution, and Derivatives Products Policies

Recipients are reminded:

1. **Data, as a collection of facts, cannot be copyrighted under US copyright law;**
2. **Projects carried out under a US DOT SMART Grants is federally funded; therefore, as stated in grant language:**
    1. **Recipients must comply with the US DOT Public Access Plan, meaning, among other requirements, project data must be shared with the public, either by the researchers or by US DOT;**
    2. **That by accepting US DOT funding through this grant, recipients have granted to US DOT a comprehensive non-exclusive, paid-up, royalty-free copyright license for all project outputs (publications, datasets, software, code, etc.). This includes all rights under copyright, including, but not limited to the rights to copy, distribute, prepare derivative works, and the right to display and/or perform a work in public; and,**
    3. **In accordance with Chapter 18 of Title 35 of the United States Code, also known as the Bayh-Dole Act, where grant recipients elect to retain title to any invention developed under this grant, US DOT retains a statutory nonexclusive, nontransferrable, irrevocable, paid-up license to practice or have practiced for or on behalf of the United States any such invention throughout the world.**

Please provide as much information as possible:

1. **Describe who will hold the intellectual property rights for the data created or used during the project;**
2. **Describe whether you will transfer those rights to a data archive, if appropriate;**
3. **Identify whether any licenses apply to the data;**
    1. **If you will be enforcing terms of use or a requirement for data citation through a license, indicate as much in your DMP;**
4. **Describe any other legal requirements that might need to be addressed.**

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

By receiving USDOT funding, this projects has granted the USDOT a comprehensive non-exclusive, paid-up, royalty-free copyright license for all project outputs, including but not limited to publications, datasets, software, code, and other derivate products. This includes all rights under copyright, including but not limited to the rights to copy, distribute, prepare derivative works, and the right to display and/or perform a work in public. In accordance with Chapter 18 of Title 35 of the United States Code, also known as the Bayh-Dole Act, if this project elects to retain title to any invention developed under this grant, USDOT retains a statutory nonexclusive, nontransferrable, irrevocable, paid-up license to practice or have practiced for or on behalf of the United States any such invention throughout the world.

In compliance with the USDOT Public Access Plan, this project will transfer data rights to a public data repository, ensuring the data and project outputs are accessible to the public and other stakeholders.

Archiving and Preservation Plan

Please provide as much information as possible:

1. **State where you intend to archive your data and why you have chosen that particular option;**
2. **Provide a link to the repository;**
3. **You must describe the dataset that is being archived with a minimum amount of metadata that ensures its discoverability;**
    1. **Whatever archive option you choose, that archive should support the capture and provision of the US Federal Government DCAT-US Metadata Schema https://resources.data.gov/resources/dcat-us/**
4. **In addition, the archive you choose should support the creation and maintenance of persistent identifiers (e.g., DOIs, handles, etc.) and must provide for maintenance of those identifiers throughout the preservation lifecycle of the data;**
5. **Your plan should address how your archiving and preservation choices meet these requirements.**

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.

The final dataset will be deposited in the Repository & Open Science Access Portal (ROSA P) (https://rosap.ntl.bts.gov/) at https://doi.org/10.21949/1530332 and the record metadata aligned with the DCAT-US Metadata Schema.