

## ATMA for Work Zone Safety

### 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. ATMA for Work Zone Safety
2. SMARTFY22N1P1G16
3. Heather Pickering-Hilgers
4. 0009-0002-9449-1463
5. heather.pickeringhilgers@state.co.us 303-808-1696
6. Colorado Department of Transportation
7. 303-757-9011
8. <https://www.codot.gov/>
9. 10/30/2023

### 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.

We intend to collect and archive operational log messages from the ATMA systems deployed in Colorado, Wisconsin, Minnesota, and Oklahoma. The log files are automatically generated by the

ATMA's onboard control computer and record detailed information about system behavior, vehicle dynamics, and navigation status during deployment.

Each log file captures time-stamped records that include message identifiers, GPS-derived location (latitude, longitude, and altitude), navigation states, desired and actual headings, velocity, current and desired gap, number of GPS satellites, if the GPS is valid, cross track errors, current brake command, current steering command and the following vehicle state. Data is gathered for both the lead and follow vehicles. The filename follows the naming convention yyyy-MM-dd\_hh-mm-ss\_log.csv.

The logs will allow the team to systematically evaluate the ATMA performance across diverse operating environments and traffic conditions. The data will support quantitative analyses of localization accuracy, control stability, and overall system reliability, providing statistically meaningful insights into autonomous maintenance vehicle behavior.

All data will be anonymized and quality-checked prior to submission to the U.S. DOT's ROSA P repository for long-term preservation and public access. The dataset is approximately 1.4 GB in total and contains roughly 3,800 CSV files organized by state and date of collection.

## 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.**

We will collect ATMA operational log files in comma-separated value (CSV) format. These log files are automatically generated by the ATMA's onboard computer and periodically exported to a secure working computer for processing and analysis. The project team will export the data from the vehicle to a working computer weekly and use Python program to analyze the performance. Processed data products and analysis results are organized by deployment state and date to maintain traceability. We will also be creating an ATMA Deployment Tool Kit which will summarize any findings. This will be created in Google Docs and saved in PDF.

## 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 to 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.

The project team will collect the data and only the relevant project team members will have access to that data. All data will be archived on each agency's protected and encrypted servers. No sensitive or PII Data is expected to be collected but if any is generated it will be secured in accordance with the data policies in the state that the data is being gathered. Because of this, the project team isn't anticipating any concerns regarding privacy. Any public accessible data, will adhere to the data sharing goals as specified within the Notice of Funding Opportunity (NOFO), CDOT intends to provide regular project updates. The project summary data will be compiled by CDOT staff involved in the project and updated regularly.

## **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.

The state or university that creates the data will own the intellectual property rights to that data. No licenses will apply to that data.

## 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 project stakeholders include Colorado's Chief Data Office (CDO) whose expertise in data management will be leveraged. The CDO will also support development of the data dictionary, defining essential metadata, and both the creation and management of the persistent identifiers. This will support the US Federal Government Project Open Data Metadata Schema. Every attempt will be made to ensure that the data management process adheres to the USDOT data policies.

All data will be held with each state or university project partner during the project and stored and archived according to the state or university policies. The project will also use the USDOT repository ROSA P (<https://rosap.ntl.bts.gov/>) to archive any final project data.

