Plan Overview

A Data Management Plan created using DMPTool

Title: Expand Work Zone Data Exchange to Local Roads in Wisconsin Using Smart Work Zone ITS

Creator: Erin Schwark

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

Principal Investigator: Steven Parker, Alex Ariza, Craig Schanning, Erin Schwark

Data Manager: Steven Parker, Alex Ariza, Craig Schanning, Erin Schwark

Project Administrator: Steven Parker, Alex Ariza, Craig Schanning, Erin Schwark

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

Funding opportunity number: NOFO #20.941

Grant: https://www.transportation.gov/grants/smart/fy23-smart-stage-1-notice-funding-opportunity-nofo

Template: SMART Grants Stage 1 Data Management Plan (DMP)

Project abstract:

The project will expand the existing Wisconsin WZDx data model to incorporate local road work zone characteristics and prototype the capability on a select set of rural and urban projects in Wisconsin using smart work zone devices purchased through the project. The project will also evaluate the ability to scale the proposed solution to all WisDOT improvement projects statewide with local road work zone components. Lastly, this project will provide a model to incorporate other work zones along the local road network into the Wisconsin WZDx using smart
work zone technology.

Start date: 09-15-2023

End date: 03-15-2025

Last modified: 01-05-2024

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 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.

1. Work zone event data for each pilot work zone deployment will be collected by the WisDOT Smart Grant project team from five sources:

   • Planned lane closure event history from the Wisconsin Lane Closure System (WisLCS).
   • Reported lane closure event data from the Wisconsin Work Zone Data Exchange (WZDx).
   • Connected work zone device integration logs from the Advanced Traffic Management System (ATMS).
   • Ground truth data collected through pilot work zone site visits by the WisDOT Smart Grant project team.
   • Clarifying information from interviews with project stakeholders including work zone field staff, TMC staff, and project engineers.

This data will be analyzed by the WisDOT Smart Grant project team to evaluate the performance of the WZDx and ATMS connected device integration for each pilot deployment with respect to the technical performance measures described in the Evaluation Plan.

2. The expected nature, scope, and scale of data for each dataset is described below:

   • The WisLCS provides a comprehensive description of planned lane closures in Wisconsin including type of work, roadway and lane level impacts, schedule, location, geometry, restrictions, and detours. It currently serves as the primary source for all highway lane closures statewide in the Wisconsin 511 traveler information system and the Wisconsin WZDX. The WisLCS will be enhanced at the start of the project to include additional attributes for local road lane closures. Closure history is available for download through WisLCS user interfaces. The Wisconsin Traffic Operations and Safety (TOPS) Laboratory at the University of Wisconsin-Madison maintains the WisLCS on behalf of WisDOT.
   • The Wisconsin WZDx provides a standardized GeoJSON data feed containing up-date-information about active work zone events. The WZDx will be enhanced at the start of the project to incorporate field verified work zone status attributes through connected work zone device integration in the ATMS. It will also be enhanced to include additional attributes for local road lane closures. The UW TOPS Lab has an existing process that archives the WZDx GeoJSON data feed from the ATMS at 1-minute intervals. This archive is managed in a large
scale PostgreSQL database.

- The ATMS vendor (Arcadis) will provide access to the connected device integration logs for purposes of this project.
- The WisDOT Smart Grant project team will conduct site visits to the pilot work zone locations to collect ground truth sample observations including: GPS measurements, logged observations, and recorded video.
- Clarifying information from pilot project staff and stakeholders will be in the form of written documents.

3. Data collected for this study will consist primarily of publicly available work zone event data. Additional data from device integration logs, work zone site visits, and stakeholder interviews will provide clarifying information and/or additional technical details related to the work zone events. Data collected for this study is not expected to include personal identifying information (PII). For collaboration purposes, project results and data will be shared between the project team via a secure UW Box account.

4. Work zone event history from the WisLCS and WZDx is already archived at the TOPS Lab in collaboration with WisDOT due to its value for work zone related research, analysis, and performance management. Understanding the accuracy of these datasets is therefore a critical need. The additional project datasets generated by this project (ATMS logs, ground truth observations, and clarify interviews) will potentially serve as a future reference and template for evaluating Wisconsin work zone event data quality.

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.
1. Anticipated file formats:

- WisLCS closure records are managed in an Oracle 19c relational database. Records are also available for download through the WisLCS in comma separated value (CSV) format. All data elements in the WisLCS are based on non-proprietary and/or open standards. Lane closure geometry is stored in GeoJSON format.
- Wisconsin WZDx data is based on the WZDx 4.2 specification. Data is collected at 1-minute intervals and archived in a PostgreSQL relational database in GeoJSON format.
- ATMS connected work zone device integration service logs are semi-structured timestamped plain text files.
- Ground truth data collected through pilot work zone site visits will include GPS measurements, logged observations, and recorded video.
- Clarifying information from pilot project staff and stakeholders will be in the form of written documents.
- Project results will be in the format of written documents and derived datasets.

2. How will the project use platform-independent and non-proprietary formats to ensure maximum utility of the data in the future?

- All relational datasets are readily available in CSV or GeoJSON format. The technical analysis for this project will conducted primarily within PostgreSQL, a prevailing open source relational database system using standard SQL.
- Site visit field observations will be recorded in written documents or Excel files. Video recordings will be taken using GoPro devices and stored in MP4, a prevailing, non-proprietary format.

3. This project will use the Project Open Data Metadata Schema v1.1 (DCAT-US v1.1) to describe the datasets.

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](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.

1. This project will not collect personal identifying information (PII), with the possible exception of the following:
   1. Video recordings from work zone pilot site visits may capture images of the research team members. Those images will either be removed before sharing or the research team members will sign a waiver of their release.
   2. After-action stakeholder interviews may capture the names of work zone field staff and/or TMC control room staff. References to individuals will be removed before sharing.
   3. Internal configuration details in the ATMS device logs that may lead to a security concern will be removed before sharing.
2. All PII will be removed or formally waived before releasing project data for public access.
3. There are no access restrictions to the data.
4. The TOPS Lab will retain an archive of the Smart Grant project datasets for three years from the date of final payment. Retention and access to project records will be in accordance with requirements established under 49 Code of Federal Regulations 18.42.

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.

1. Intellectual property rights for data created or used:
   1. The Wisconsin WZDx data feed is owned by the Wisconsin Department of Transportation.
   2. The Wisconsin Lane Closure System database is owned by the University of Wisconsin-Madison.
   3. The Advanced Traffic Management System database is owned by Arcadis.
   4. All other data collected for this project will be owned by the respective research team members that performed the data collection.

2. All data created and used for this project will be transferred into a separate archive repository maintained by the TOPS Lab on behalf of the project. The Wisconsin Department of Transportation and the U.S. Department of Transportation (USDOT) shall have a non-
exclusive, perpetual right to unlimited use of any and all datasets in the repository.
3. There are no licenses that apply to the datasets.

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.

1. Project data will be archived by the TOPS Lab at the University of Wisconsin-Madison. This option has been chosen since the TOPS Lab maintains a mature data management platform (see bullet 4 below) that serves as the system of records for several large scale Wisconsin Department of Transportation datasets. The TOPS Lab will also conduct the data collection and analysis tasks for this project.
2. The project archive repository will be available from the following link: https://topslab.wisc.edu/research/.
3. Persistent identifiers are already incorporated in the WisLCS, WZDx, and ATMS for work zone events and connected devices. This concept will be extended to include all records in the project archive repository. The archive will also provide metadata per the DCAT-US v1.1 schema standard.
4. The TOPS Lab has a large-scale data management platform (WisTransPortal) that has been developed over the past 20 years through sponsorship of the Wisconsin Department of Transportation. The WisTransPortal infrastructure consists of a combination of virtual and on-premises servers and storage arrays managed by full-time TOPS Lab IT professional staff through the UW-Madison Division of Information Technology (DoIT) Campus Data Center. Technical data for this project will be managed in the TOPS Lab relational database cluster.
(Oracle and PostgreSQL) for retention, retrieval, analysis, and backups. All databases and servers are managed to meet or exceed campus network and data security requirements.