Data Management Plan

Name of Contractor: University of Wyoming
Name of the Project: Connected Autonomous Traffic Signal Control Algorithms and Fleet Vehicles
Project Duration: 24 months   Start Date: 12/31/2018   End Date: 12/31/2020
DMP Version: 1
Date Amended, if any: N/A
Name of all authors, and ORCID number for each:

Milan Zlatkovic, ORCID: 0000-0002-6777-230X
Mohamed Ahmed, ORCID: 0000-0002-1921-0724
Zorica Cvijovic, ORCID: 0000-0003-3558-7967

WYDOT Project Number: RS03219

- **Name of all peer reviewed publications, which have been generated using data from this project to include:**
  N/A
- **Any Digital Object Identifier (DOI), assigned to any peer reviewed publication or data generated by this project:**
  N/A
- **URLs for all peer reviewed publications which have been generated using data from this project:**
  N/A
- **Dataset URL, if available:**
  N/A

What constitutes data will be determined by the Principle Investigator, Project Champion, and the Research Manager. In general, your plan should address final research data. This includes recorded factual material commonly accepted in the scientific community as necessary to validate research findings. Final research data do not include laboratory notebooks, partial datasets, preliminary analyses, drafts of scientific papers, plans for future research, peer review reports, communications with colleagues, or physical objects, such as gels or laboratory specimens. As part of your research, you may also generate unique data, which are data that cannot be readily replicated. Your DMP should also address unique data that may arise from your research.

WYDOT expects the timely release and sharing of data to be no later than the acceptance for publication of the main findings from the final dataset, unless the Principle Investigator will be embargoing the data. In such a case, the data cannot be embargoed for a period longer than 12

months. See Chapter 11 for information on retention and embargos.

1.   Introduction

The purpose of this research project is to:

> The purpose of the research project is to create field-ready, Connected-Autonomous Vehicles (CAV)-based traffic control programs that will improve operations and safety of trucks and fleet vehicles. The study will review the current protocols and recommend options applicable to Wyoming conditions. It will analyze, assess and develop traffic control algorithms that would use CAV technologies to improve operations and safety through signalized intersections. The focus will be on optimizing operations of freight and fleet (snowplows and striping) vehicles through signalized intersections. The analyzed strategies will include queue warning, speed harmonization and freight priority.

2.   Definitions

a.   Code or scripts include code used in the collection, manipulation, processing, analysis or visualization of data, but may also include software developed for other purposes.

b.   Copyright is a set of legal rights extended to copyright owners that govern such activities as reproducing, distributing, adapting, or exhibiting original works fixed in tangible forms.

c.   Data means the recorded factual material commonly accepted in the scientific community as necessary to validate research findings, but not any of the following: preliminary analyses, drafts of scientific papers, plans for future research, peer reviews, communications with colleagues. Recorded material excludes physical objects (e.g. laboratory samples). Research data also does not include trade secrets, commercial information, materials necessary to be held confidential; and personnel and medical information and similar information the disclosure of which would constitute a clearly unwarranted invasion of personal privacy.

d.   Data Archive is a site where machine-readable materials are stored, preserved or possibly redistributed to individuals interested in the materials.

e.   Data Management Plan is a document that specifies your plans for managing your data and files for a research project.

f.   Dataset means collection of data.

g.   Metadata refers to structured data about data that helps define administrative, technical, or structural characteristics of the digital content.

3.   Data Types and Storage

Rev. 6-4-2019

The types of data and/or datasets generated and/or used in this project include …

The data for this project will include geometrical characteristics of the six selected interchanges on I-80 (in Evanston, Rock Springs, Rawlins, Laramie and Cheyenne), such as horizontal and vertical alignments, cross section elements, roadside elements, ramp characteristics, auxiliary lane dimensions; operational characteristics (traffic volumes, traffic composition, speeds); and signal timing data.

The geometry data will be collected from the WYDOT's Roadway database, GIS maps, Google Earth and field visits. These data will be summarized in spreadsheet forms and also visualized in map forms (GIS and KML). They will be used to create traffic and driving simulation models. Geometry data is quite stable over time with minimal changes.

Operational data will be collected through available WYDOT databases and field observations. They will be summarized in spreadsheet and map forms, and will exist in the developed models. The dataset will include historical operational data, which will be used to assess the changes over time and perform projections for the future.

Signal timing data have been obtained from WYDOT in a form of Synchro models. These data will be replicated in the simulation models.

The research team will develop memos on data collection and reduction procedures, which will be included in the final report. Since the team will be using existing databases and field data collection, the datasets will be reproducible.

Spreadsheet software (such as Excel and Access), GIS tools and Google Earth (for KMLs) will be the software tools needed to create, process and visualize the data.

The data will be stored on UW team computers and storage UW's OneDrive, which have excellent protection. Backup data will also be stored to external devices (CDs and USBs) which will only be accessible to the research team.

Provide a description of the data that you will be gathering in the course of your project. You should address the nature, scope, and scale of the data that will be collected. 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. Discuss value of the data over the long-term. Please provide the name of all repositories where the data will be housed during the lifetime of the project.

Checklist
- What type of data will be produced?
- How will data be collected? In what formats?
- How will the data collection be documented?
- Will it be reproducible? What would happen if it got lost or became unusable later?

- How much data will it be, and at what growth rate? How often will it change?
- Are there tools or software needed to create/process/visualize the data?
- Will you use pre-existing data? From where?
- Storage and backup strategy?

4. Data Organization, Documentation, and Metadata

The plan for organizing, documenting, and using descriptive metadata to assure quality control and reproducibility of these data include …

> The majority of data collected, redacted and analyzed in this research will be in non-proprietary forms (spreadsheets, GIS and KML types). Some data will be duplicated and analyzed in proprietary software (PTV VISSIM and Synchro) for simulation and analysis. The simulation results will be presented in the same spreadsheet and map forms for easy access by wider audiences.
> The metadata scheme will contain all information about the data collection procedures, storage, file types and naming conventions. It will be in a text format with descriptions and explanations.
> The naming convention will be designed to be self-explanatory, so that the names of the directories and files correspond to the contents in the data file.
> The UW team will be guided by our experience and lessons learned about data organization from previous research projects.

Your DMP should describe the anticipated formats that your data and related files will use. To the maximum extent practicable, and in accordance with generally accepted practices in your field, your DMP should address how you will use platform-independent and non-proprietary formats to ensure maximum utility of the data in the future. 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.
NOTE: Attach the Metadata Schema, URL for data generated, and all peer reviewed publications from this project.

Checklist
- What standards will be used for documentation and metadata?
- Is there good project and data documentation format/standard?
- What directory and file naming convention will be used?
- What project and data identifiers will be assigned?
- Is there a community standard for metadata sharing/integration?

5. Data and/or Database Access and Intellectual Property

Rev. 6-4-2019

What access and ownership concerns are there…

This research project will not use nor create data which will contain confidential information. Nevertheless, access to the data will be limited to the UW team (PIs and the students who will be working on the project). It will be stored on the UW team computers and UW OneDrive, which are accessible only to the individual that owns the account, or to whom the permission has been given. Personal and high-security data will not be used in this research. The UW team and WYDOT TAC will maintain the intellectual rights to data analysis results. There will be no embargo periods for the data to uphold.

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. If necessary, describe any division of responsibilities for stewarding and protecting the data among Principal Investigators.

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. 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.

Checklist
- What steps will be taken to protect privacy, security, confidentiality, intellectual property or other rights?
- Does your data have any access concerns? Describe the process someone would take to access your data.
- Who controls it (e.g., PI, student, lab, University, funder) ?
- Any special privacy or security requirements (e.g., personal data, high-security data) ?
- Any embargo periods to uphold?

6. Data Sharing and Reuse

The data will be released for sharing in the following way …

The complete data sets collected in the study will only be shared with WYDOT. WYDOT can

then reuse the data for subsequent analysis or share with third parties for future studies. The UW team may also use the collected data for future research. There will not be a special data archive for this project.

Reduced data sets and analysis results will be included in journal publications to be presented to wider audiences. Research papers will be submitted to transportation-oriented journals throughout the research.

We do not foresee any special requirements for data sharing. We also do not expect to have any copyrights or legal requirements for the data. The software needed to work with the data will be spreadsheets and map tools, VISSIM simulation and Python programming language.

Describe who will hold the intellectual property rights for the data created by your project. Describe whether you will transfer those rights to a data archive, if appropriate. Identify whether any copyrights apply to the data, as might be the case when using copyrighted instruments. If you will be enforcing terms of use or a requirement for data citation through a license, indicate as much in your DMP. Describe any other legal requirements that might need to be addressed.

Checklist
- If you allow others to reuse your data, how will the data be discovered and, shared?
- Any sharing requirements (e.g., funder data sharing policy) ?
- Audience for reuse? Who will use it now? Who will use it later?
- When will I publish it and where?
- Tools/software needed to work with data?

7. Data Preservation and Archiving

The data will be preserved and archived in the following ways …

The complete data will be kept on UW researchers' computers and the UW OneDrive with protected access, and shared with WYDOT. Copies will also be stored on external storage devices in the UW team research labs. We expect to have copies of the datasets for about 10 years upon completion of the project. The formats will be in .xlsx type (spreadsheet), GIS, KML, VISSIM and Python.

The team will share reduced and analyzed data through journal and other types of publications, which will be available to wider audiences.

Describe how you intend to archive your data and why you have chosen that particular option. You may select from a variety of options including, but not limited to:

- Use of an institutional repository.
- Use of an archive or other community-accepted data storage facility.

Rev. 6-4-2019

- Self-dissemination.

You must describe the dataset that is being archived with a minimum amount of metadata that ensures its discoverability. Whatever archive option you choose, that archive must support the capture and provision of the National Transportation Library metadata requirements. In addition, the archive you choose must support the creation and maintenance of persistent identifiers and must provide for maintenance of those identifiers throughout the preservation lifecycle of the data. Your plan should address how your archiving and preservation choices meet these requirements.

Checklist
- How will the data be archived for preservation and long-term access?
- How long should it be retained (e.g., 3-5 years, 10-20 years, permanently) ?
- What file formats? Are they long-lived?
- Are there data archives that my data is appropriate for (subject-based? Or institutional)?
- Who will maintain my data for the long-term?

NOTE:
Researchers evaluating data repositories as the option(s) for storing and preserving their data should examine evidence demonstrating that the repository:

a. Promotes an explicit mission of digital data archiving.

b. Ensures compliance with legal regulations, and maintains all applicable licenses covering data access and use, including, if applicable, mechanisms to protect privacy rights and maintain the confidentiality of respondents.

c. Has a documented plan for long-term preservation of its holdings.

d. Applies documented processes and procedures in managing data storage.

e. Performs archiving according to explicit workflows across the data life cycle.

f. Enables the users to discover and use the data, and refer to them in a persistent way through proper citation.

g. Enables reuse of data, ensuring appropriate formats and application of metadata.

h. Ensures the integrity and authenticity of the data.

i. Is adequately funded and staffed, and has a system of governance in place to support its mission.

j. Possesses a technical infrastructure that explicitly supports the tasks and functions described in internationally accepted archival standards like Open Archival Information System (OAIS).

NOTE: This DMP is created as a derivative from the DMP belonging to the University of Minnesota and can be found at https://www.lib.umn.edu/datamanagement/DMP

Metadata Schema

Elements	Example of what is expected for each element
Title[1]	Human-readable name of the asset. Should be in plain English and include sufficient detail to facilitate search and discovery. A name given to the publication or data element. All substitute or alternative titles must have a different Metadata Transmittal Schema.
Creator/contact point	An entity/person(s) primarily responsible for making the content of the resource. Contact person's name, ORCID number, and email for the asset.
Publication Date(s)	The date associated with the final report/dataset.
Description/Abstract	Human-readable description (e.g., an abstract) with sufficient detail to enable a user to quickly understand whether the asset is of interest. May include abstract, table of contents, reference to a graphical representation of content or a free text account of the content.
Subject and Keywords	The topic of the content of the resource. Tags (or keywords) help users discover your dataset; please include terms that would be used by technical and non-technical users.
Identifier[2] and/or source	A unique identifier for the dataset/publication. Examples: URI, URL, DOI, ISNB, ISSN.
Collection and Related Documents	If there is a secondary dataset, cite source. The collection of which the dataset is a subset should be listed. Include all identifiers and/or sources.
Edition	Most recent date on which the dataset was changed, updated or modified.
Related Documents	Related documents such as technical information about a dataset, developer documentation, etc.
Coverage	Spatial location, temporal period, jurisdiction.
Language	The language of the dataset/publication.

[1] *To include alternate title; conference title; and journal title, if they are different.*

[2] *To include record numbers; report numbers; NTIS number; TRIS Accession Number; OCLC Number; ISBN; ISSN; contract number; and DOI if available.*

Elements	Example of what is expected for each element
Publisher/Distributor	FHWA and Wyoming Department of Transportation List all other publishing companies that this publication has been sent to.
Funding agency	FHWA and Wyoming Department of Transportation
Access Restrictions	The degree to which this dataset could be made publicly available, *regardless of whether it has been made available*. Choices: public (Data asset is or could be made publicly available to all without restrictions), restricted public (Data asset is available under certain use restrictions), or non-public (Data asset is not available to members of the public).
Intellectual Property and Other Rights	This may include information regarding access or restrictions based on privacy, security, or other policies. This should also serve as an explanation for the selected "accessLevel" including instructions for how to access a restricted file, if applicable, or explanation for why a "non-public" or "restricted public" data asset is not "public," if applicable.
License	The license or non-license (i.e. Public Domain) status with which the dataset or API has been published.
Code and software needs	List all code specific information. Is there specific software needed to run the database or data.
Format	The machine-readable file format. May include media type or dimensions. Used to determine the software, hardware or other equipment needed to display or operate the resources.
Choice of Repository	If you have a preference, list the repository where you will archive your data/datasets.

NOTE: Each separate report, dataset, collection, existing collection, and software developed must have its own table. All fields in this Schema must be completed at the time of the final report.

NOTE: This Metadata Schema is created as a derivative from the Common Core required fields which can be found at https://project-open-data.cio.gov/schema/.

Rev. 6-4-2019