

Data Management Plan

Name of Contractor: University of Wyoming

Name of the Project: Automated Real-Time Weather Detection System using Artificial Intelligence

Project Duration: 2 years **Start Date:** NTP **End Date:** 2 years after NTP

DMP Version:

Date Amended, if any:

Name of all authors, and ORCID number for each:

Mohamed M. Ahmed, Ph.D., P.E. - orcid.org/0000-0002-1921-0724

Md Nasim Khan, Ph. D. candidate - orcid.org/0000-0001-5996-091X

WYDOT Project Number:

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

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:

Development of an Automated Real-Time Weather and Road Surface Condition Detection System using Artificial Intelligence in Wyoming

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

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

The video data will be collected from existing WYDOT webcams and from in-vehicle cameras mounted on snowplows from the roadways in the state of Wyoming. The collected data will be used for training and validation of the weather and surface condition detection models. The video data will include different interstate freeways, state highways, and urban arterials. Two separate databases will be created for identifying weather conditions and road surface conditions,

respectively. The data will be collected for at least one year to ensure enough sample size for each weather condition. The Data Acquisition System (DAS) will be developed in a way to ensure effective data collection. It is worth noting that a 512GB SD memory card can store close to 60 days of high-resolution video data before it ran out of memory. Since this research aims at detecting high-resolution trajectory-level snowy weather events (i.e., light-, medium-, heavy-snow, and blizzard), for each selected roadway classification, the research team will first identify the intensity of snow storms, and date and time of each snowy weather event; then, acquire video data recorded by WYDOT tablets.

The data will be archived in digital format with proper encryption. Documentation of data collection will be included in the final report. The data will be only accessible to the PI and designated graduate students who received their IRB and CITI training to protect human subject information.

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 ...

Data will be organized in separated folder on the main computers and encrypted. Only users with the right credentials will be able to access and decrypt the data.

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

What access and ownership concerns are there...

All data will be housed at secured computers that only accessible by the co-PIs and graduate students working on this project. UW implements a rigorous security measure. No identifiable personal information will be acquired/collected.

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. 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 University of Wyoming will hold intellectual property rights for the data. The PI and graduate students, from the civil engineering department, who are responsible for the study are currently using different datasets. Data will be used under supervision of the PI. Data will not be accessed or used by other parties other than the University of Wyoming.

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 ...

Collected data will be retained for 5 years after the end date of the project.

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.
- 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
Title¹	Automated Real-Time Weather Detection System using Artificial Intelligence
Creator/contact point	Mohamed Ahmed, Ph.D., P.E. (orcid.org/0000-0002-1921-0724) E-mail: mahmed@uwyo.edu
Publication Date(s)	
Description/Abstract	<p>The rapid evolution of Information Technologies (IT) presents opportunities of using Machine Vision and Artificial Intelligence to provide imaging-based automatic detection and analysis of real-time road weather conditions. Machine Vision is an integration of a series of technologies, software and hardware products; it is the science of getting computers to automatically detect patterns in data, and then use the uncovered patterns to predict future data, or to perform other kinds of decision making under uncertainty. The overall Machine Vision process includes planning the details of the requirements and project and then creating a solution. During run-time, the process starts with imaging, followed by automated analysis of the image and extraction of the required information. Given the advantages of Machine Vision technology, such as real-time processing of road surface level weather conditions, accuracy of weather detection, cost-effectiveness, etc., it has been extensively used in various fields of engineering for image classification, pattern recognition, and text categorization.</p> <p>With consideration of the limitations of the existing WYDOT weather detection systems, and in view of the opportunity of the emerging automatic video-image processing technologies, this research aims at developing an affordable weather detection system, which will use video images collected by the WYDOT snowplows and roadside webcams. It is worth mentioning that WYDOT operates and maintains too many roadside webcams. It is not feasible for WYDOT TMC operators to review and process them timely. The weather information will be processed based on Machine Vision techniques. Eventually, the product of this research will assist WYDOT staff with providing road users with more accurate and reliable road surface weather conditions, which will result in safer travel decisions and more conservative driving behavior to mitigate the negative impacts of snowy weather on traffic safety.</p>
Subject and Keywords	Weather Detection, Webcams, Snowplows, Artificial Intelligence, Image processing, Surface condition, Traffic

Elements	Example of what is expected for each element
	Safety
Identifier² and/or source	
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	9/2019
Related Documents	
Coverage	Wyoming
Language	English
Publisher/Distributor	FHWA and Wyoming Department of Transportation
Funding agency	FHWA and Wyoming Department of Transportation
Access Restrictions	Data asset is not available to members of the public
Intellectual Property and Other Rights	The University of Wyoming will hold the intellectual property rights of the data.
License	No License; data will not be shared.
Code and software needs	Python, R, MATLAB, C, C++, Java
Format	PDF, CSV, Video files, JPG, PNG
Choice of Repository	All the data will be stored on secured UW computers and hard drives in keyed graduate student offices. The data will be encrypted while at rest.

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/>.