

Data Management Plan

Name of Contractor: University of Wyoming

Name of the Project: Enhancing Crash Data Reporting to Highway Safety Partners in Wyoming by Utilizing Big Data Analysis and Survey Techniques

Project Duration: 24 Months **Start Date:** January, 2020

End Date: January, 2022

DMP Version: Version 1

Date Amended, if any:

Name of all authors, and ORCID number for each:

Ahmed Farid (0000-0002-5101-3237)

Sahima Nazneen (0000-0001-7424-3089)

Khaled Ksaibati (0000-0002-9241-1792)

Mohammad Mahdi Rezapour (0000-0002-722X)

WYDOT Project Number:

- **Name of all peer reviewed publications, which have been generated using data from this project to include:**
- Rezapour, M., and Ksaibati, K. (2021). The Latent Class Multinomial Logit Model for Modeling Front-Seat Passenger Seatbelt Choice, Considering Seatbelt Status of Driver. *Future Transportation*. 1(3), 559-569; <https://doi.org/10.3390/futuretransp1030029>
- Rezapour, M., and Ksaibati, K. (2021). Modeling crashes involving children, finite mixture cumulative link mixed model. *International Journal of Injury Control and Safety Promotion*. Volume 28, 2021 - Issue 4. <https://doi.org/10.1080/17457300.2021.1964088>
- Rezapour, M., and Ksaibati, K. (2021). Hamiltonian Monte Carlo with Random Effect for Analyzing Cyclist Crash Severity. *Signals*. 2(3), 527-539; <https://doi.org/10.3390/signals2030032>
- **Any Digital Object Identifier (DOI), assigned to any peer reviewed publication or data generated by this project:**
- <https://doi.org/10.3390/futuretransp1030029>
- <https://doi.org/10.1080/17457300.2021.1964088>
- <https://doi.org/10.3390/signals2030032>
- **URLs for all peer reviewed publications which have been generated using data from**

this project:

- <https://doi.org/10.3390/futuretransp1030029>
- <https://doi.org/10.1080/17457300.2021.1964088>
- <https://doi.org/10.3390/signals2030032>

- **Dataset URL, if available:**

- Nothing to report.

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:

To assess the data needs of the Wyoming Department of Transportation’s safety partners, identify gaps in crash reporting and recommend appropriate guidelines to present traffic safety data.

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 ...

Two types of data will be produced for this project.

1. Data from questionnaires sent to WYDOT's safety partners.
2. The second dataset will be obtained from WYDOT's electronic crash reporting system and will be in a csv format.

The data documentation will have information describing the data collection procedure, data processing, what is contained within the dataset, time of collection, persons involved in the collection and data quality assurance processes. The data will be reproducible. Questionnaires will be sent out again for the data collection if the original data collected becomes unusable. Also, the crash data may be extracted again from WYDOT's electronic crash reporting system if the data is lost or becomes unusable.

The combined size of the data is expected to be a maximum of a few gigabytes. The data is not expected to change over the project period. Microsoft excel will be used to process and visualize the data. Standard software programs including R and SAS will be utilized for statistical analysis. Pre-existing crash data will be used for the study. The data will be obtained from the Wyoming electronic crash reporting system. Pre-processed data will be stored on the WYDOT data warehouse. This will create a useful backup for the collected data that will be readily accessible.

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 data formats will be in Microsoft excel which can be read easily by different software packages. The data will be prepared and stored in accordance with the SQL standard. A data directory will be generated that will serve as a link to the master database. This directory will be based on a standard query structure. The project number will be adopted as the format for the file naming convention. Changes to files will be tracked by appending a sequence of numbers or letters to the project number. Data will be shared by providing access to the principal investigator and other students working on the project.

Data validation will be done by applying identifiable rules that will ensure that the data is accurate. Data that does not meet minimum levels of acceptable quality will be flagged and removed from the database.

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...

No intellectual property issues are expected to exist with the data to be collected for this project. Crash data from the Wyoming electronic crash reporting system will be owned by WYDOT who will restrict personal and business details from the database. WYDOT will be responsible for allowing access to the data for processing and modification, and for access restrictions. Access to the data will involve seeking authorization from WYDOT who may then allow access to a part or the whole dataset. There will be no embargo periods 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 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 identify 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 ...

Property rights for the data created will be held by WYDOT. The data will be collected and stored in a manner that allows for the efficient sharing of the data with other interested parties. The data will be submitted to WYDOT after the project and may be shared through the WYDOT Data Warehouse. Interested parties will seek permission for access to the data from WYDOT. The data will be available for access and modification using standard data and statistical software such as Microsoft Excel and SAS.

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 ...

After the completion of the project, the data may be archived by WYDOT in the National Transportation Library repository as recommended by the US Department of Transportation public access plan.

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	Enhancing Crash Data Reporting to Highway Safety Partners in Wyoming by Utilizing Big Data Analysis and Survey Techniques
Creator/contact point	Khaled Ksaibati (0000-0002-9241-1792), khaled@uwyo.edu
Publication Date(s)	March, 2022
Description/Abstract	Road safety is a crucial topic of transportation engineering. The Wyoming Department of Transportation (WYDOT) collects such data from police crash reports and roadway inventories. WYDOT also provides those data to its partner groups in the form of data records or summary statistics documented in periodical reports. The groups include the Wyoming Seat Belt Coalition, the Wyoming Highway Patrol, the Wyoming Association of Sheriffs and Chiefs of Police, the Wyoming Transportation Safety Coalition, the Governor’s Council on Impaired Driving, Wyoming’s counties, the Wyoming Bicycle and Pedestrian System Task Force and motorcycle groups. In this research, surveys were prepared, distributed to and collected from those groups asking about the quality of the data they receive from WYDOT, particularly when it comes to data provision frequencies and unreported data that would be beneficial to those groups. In addition, big data analyses were conducted to evaluate human factors influencing crash occurrences and data provision frequencies. This research’s efforts culminated in lists of recommendations to WYDOT regarding the provision of higher quality data at appropriate frequencies to its partners.
Subject and Keywords	Crash Data, Road Safety Groups, Surveys, Big Data Analytics
Identifier² and/or source	<ul style="list-style-type: none"> • https://doi.org/10.3390/futuretransp1030029 • https://doi.org/10.1080/17457300.2021.1964088 • https://doi.org/10.3390/signals2030032
Collection and Related Documents	Nothing to report

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.

Edition	Nothing to report
Related Documents	Nothing to report
Coverage	The State of Wyoming
Language	English
Publisher/Distributor	Wyoming Department of Transportation
Funding agency	Wyoming Department of Transportation
Access Restrictions	Restricted public (Data asset is available under certain use restrictions)
Intellectual Property and Other Rights	Some or all data, models, or code that support the findings of this study are available from the corresponding authors upon reasonable request.
License	The license status with which the dataset or API has been published.
Code and software needs	R software
Format	The machine-readable file format.
Choice of Repository	Nothing to report

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/>.