

DATA MANAGEMENT PLAN AND METADATA SCHEMA

Data Management Plan

Name of Contractor: University of Wyoming

Name of the Project: Assessment of Motorcycle Safety in Wyoming: Fatal and Severe Crashes, Contributing Factors and Potential Countermeasures

Project Duration: 15 months Start Date: March 2022 End Date: June 2023

DMP Version: 3

Date Amended, if any: N/A

Name and ORCID (Open Research and Contributor Identifier) Number for each author:

Milan Zlatkovic, ORCID: 0000-0002-6777-230X

Isaac Baah, ORCID: 0000-0003-1596-5633

Zorica Cvijovic, ORCID 0000-0003-3558-7967

WYDOT Project Number: RS04222

Introduction

What constitutes data will be determined by the Principal Investigator (PI), Project Champion, and the Research Manager. In general, your plan should address the data in the final research project.

The following forms of data/datasets should be reviewed when determining what data should be archived and listed in this Data Management Plan (DMP):

- a) Primary data used in the production of the report: Raw, verified data that has been obtained directly from a source. It can be captured through experiments, surveys, interviews, focus groups, or other direct interactions with individuals in the field. Does not include analysis data.
- b) Unpublished datasets: Materials and methods; clear description of the variables presented; supported by unpublished reports; and any other relevant material.
- c) Secondary Data: Pre-existing data not gathered or collected by the authors. Usually collected by another organizations or source.
- d) Metadata: Set of data that describes and gives information about the dataset – cataloging information.
- e) Dataset description document: Describes all variables in the dataset and the measurement units used.
- f) Codebook: A list of variable names, variable labels, and label values. Should specify the data position of each variable, describe the contents of each variable, and identify the range of possible codes and the meanings of the codes.
- g) Questionnaires: An unused copy of the questionnaire.
- h) Handbooks, guides, and manuals derived from research.

Determination of what counts as data and what should be archived will depend on the Principal Investigator's knowledge of the data and what he/she believes is valuable. As part of the

research project, your DMP should address unique data that may arise from your research.

Data that does not need to be archived or saved includes preliminary analyses of a project, drafts, plans for future research, peer reviews, interoffice communications, emails, letters, or other forms of correspondence. The Principal Investigator and the Project Champion will have the opportunity to discuss what data and/or other digital material should be excluded prior to finalizing the project.

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, and there will be no embargo period approved for this project without prior approval from the Research Center.

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, including information that falls under the HIPAA and PII confidentiality impact levels, and similar information the disclosure of which would constitute a clearly unwarranted invasion of personal privacy. Any information that falls under this definition shall not be considered open source and shall not be publicly available, Data Archive is a site where machine-readable materials are stored, preserved or possibly redistributed to individuals interested in the materials.
- d) Data Management Plan is a document that specifies your plans for managing your data and files for a research project.
- e) Dataset means collection of data.
- f) Metadata refers to structured data about data that helps define administrative, technical, or structural characteristics of the digital content.

I. For peer reviewed publications provide the following:

- A. Name of all peer reviewed publications that have been generated using data from this project:
Multinomial Logistic Regression Modeling of Motorcycle Crash Severities and Contributing Factors in Wyoming
- B. Any Digital Object Identifier (DOI) assigned to any peer reviewed publication or data generated by this project:
DOI: 10.31075/PIS.68.02.01
- C. All persistent uniform resource locators (URLs) for all peer reviewed publications

that have been generated using data from this project:

<https://doi.org/10.31075/PIS.68.02.01>

D. Dataset URLs, if available: N/A

II. The purpose of this research project is to:

The purpose of this research project is to assess the characteristics of motorcycle safety in Wyoming, with the focus on fatal and severe injury crashes, and provide a set of recommendations with a potential to reduce the frequency and severity of motorcycle-related crashes. The study reviews the current literature and practices in motorcycle safety, collects crash data for Wyoming, analyzes the data and develops statistical and machine learning models, provides recommendations for improvement in infrastructure, traffic control and signage, and performs a survey of motorcycle experts.

III. Data Types and Storage

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

The data for this project include crash data on the crash, vehicle and person levels from Wyoming for a time period of 12 years, obtained through the WYDOT CARE system, with additional sources including the Fatality Analysis Reporting System (FARS), National Motorcycle Institute Fatality Reporting System, Crash Investigation Sampling System, as applicable. The collected crash data do not include identifying information.

The roadway geometry data were collected from the WYDOT's Roadway database, GIS maps, Google Earth and field visits. These data were summarized in spreadsheet forms and also visualized in map forms (KML), as needed. Geometry data are quite stable over time with minimal changes.

Excel spreadsheet, GIS tools and Google Earth (for KMLs), statistical software RStudio, and Python programming language were the software tools used to create, process, analyze and visualize the data.

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

NOTE: Provide a description of the data that you will gather in the course of your project. 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. 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?

IV. Data Organization, Documentation, and Metadata

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

The majority of data collected, redacted and analyzed in this research are in non-proprietary forms (spreadsheets, GIS and KML types, images). Some data are duplicated and analyzed in specific software (RStudio and Python, which is available at no cost) for statistical analysis and machine learning applications. The results are presented in the same formats for easy access by wider audiences.

The naming convention of the data used in this research is self-explanatory, so that the names of the directories and files correspond to the contents in the data file.

The UW team was guided by our experience and lessons learned about data organization from previous research projects.

NOTE: Your DMP should describe the anticipated formats for data and related files. 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, 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?

V. Data and/or Database Access and Intellectual Property

What access and ownership concerns are there?

This research project does not use nor create data which contain confidential information. Nevertheless, access to the data is limited to the UW team (PIs and the students who worked on the project). The data are 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 were not used in this research. The UW team and WYDOT TAC maintain the intellectual rights to data analysis results. There are 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 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?

VI. 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 is not a special data archive for this project.

Reduced data sets and analysis results have been included in journal publications to be presented to wider audiences. Research papers have been submitted to transportation-oriented

journals.

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 are spreadsheets and map tools, RStudio statistical software 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?

VI. Data Preservation and Archiving

The data will be preserved and archived in the following way(s).

The complete data are kept on UW researchers' computers and the UW OneDrive with protected access, and shared with WYDOT. Copies are also 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 are in the .xlsx type (spreadsheet), GIS, KML, RStudio 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.
- 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).

VII. Generative Artificial Intelligence (AI) Tools

Authors who use AI tools in the writing of a manuscript, production of images or graphical elements of any report, or in the collection and analysis of data, must be transparent in disclosing in the Materials and Methods (or similar section) of the report how the AI tool was used. Authors are fully responsible for the content of their manuscript, even those parts produced by an AI tool, and are thus liable for any breach of publication ethics.

Specifically, you must set out the following:

- **Clearly indicate the use of language models in the manuscript**, including which model was used and for what purpose. Please use the methods or acknowledgements section, as appropriate.

- **Verify the accuracy, validity, and appropriateness of the content** and any citations generated by language models and correct any errors or inconsistencies.
- **Provide a list of sources used to generate content** and citations, including those generated by language models. Double-check citations to ensure they are accurate and properly referenced.
- **Be conscious of the potential for plagiarism** where the LLM may have reproduced substantial text from other sources. Check the original sources to be sure you are not plagiarizing someone else's work.
- **Acknowledge the limitations of language models in the manuscript**, including the potential for bias, errors, and gaps in knowledge.

Please note that AI bots, such as ChatGPT should not be listed as an author on your submission.

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	Motorcycle-related crash data in Wyoming
Creator/contact point	Milan Zlatkovic 0000-0002-6777-230X mzlatkov@uwyo.edu
Publication Date(s)	December 2022
Description/Abstract	The database contains twelve years of motorcycle-related crash data (2008 – 2019) obtained from WYDOT’s Crash Analysis Reporting Environment (CARE) database. The data are classified into crash, person and vehicle levels. No personal information is contained in the databases.
Subject and Keywords	Wyoming; Motorcycles; Crashes; CARE database
Identifier² and/or source	WY2301F.
Collection and Related Documents	N/A
Edition	December 2022
Related Documents	<i>2019 Highway Safety Crash Data Survey</i> , Highway Safety Program, Wyoming Department of Transportation
Coverage	State of Wyoming (rural and urban areas), 2008-2019
Language	English
Publisher/Distributor	FHWA, Wyoming Department of Transportation, University of Wyoming, Mountain Plains Consortium (MPC) University Transportation Center (UTC)
Funding agency	FHWA and Wyoming Department of Transportation
Access Restrictions	The complete dataset is not available to members of the public
Intellectual Property and Other Rights	Wyoming Department of Transportation and the University of Wyoming. The complete dataset is non-public.
License	N/A

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
Code and software needs	Microsoft Excel, Google Earth, RStudio, Python
Format	.xlsx, .csv, .kml, .rdata, .rmd, .py, .ipynb
Choice of Repository	Not shared through public repositories.

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 that can be found at <https://project-open-data.cio.gov/schema/>

Elements	Example of what is expected for each element
Title	Roadway and intersection geometry
Creator/contact point	Milan Zlatkovic 0000-0002-6777-230X mzlatkov@uwyo.edu
Publication Date(s)	December 2022
Description/Abstract	The database contains roadway geometry data for selected rural roads, urban streets and intersections in Wyoming, including horizontal and vertical alignment and cross sections
Subject and Keywords	Wyoming; Roadway Geometry; Horizontal Alignment; Vertical Alignment; Cross Sections
Identifier and/or source	WY2301F.
Collection and Related Documents	N/A
Edition	December 2022
Related Documents	N/A
Coverage	State of Wyoming (rural and urban areas)
Language	English
Publisher/Distributor	FHWA, Wyoming Department of Transportation, University of Wyoming, Mountain Plains Consortium (MPC) University Transportation Center (UTC)
Funding agency	FHWA and Wyoming Department of Transportation
Access Restrictions	The complete dataset is not available to members of the public
Intellectual Property and Other Rights	Wyoming Department of Transportation and the University of Wyoming. The complete dataset is non-public.
License	N/A
Code and software needs	Microsoft Excel, Google Earth
Format	.xlsx, .csv, .kml
Choice of Repository	Not shared through public repositories.