

Data Management Plan

Name of Contractor: Shawn Griffiths

Name of the Project: Characterization of Soil and Rock for Transportation Infrastructure Using Seismic Methods in Wyoming

Project Duration: Start Date: January 13, 2020 End Date: March 31, 2023

DMP Version: 1

Date Amended, if any:

Name of all authors, and ORCID number for each:

Shawn C. Griffiths, Ph.D., 0000-0003-0694-4769

Rishav Poudal - 0000-0002-0825-3307

WYDOT Project Number: RS03220

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- b) Unpublished datasets: Materials and methods; clear description of the variables presented; supported by unpublished reports; and any other relevant material.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 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.

I. For peer reviewed publications provide the following:

- A. Name of all peer reviewed publications that have been generated using data from this project: NA
- B. Any Digital Object Identifier (DOI) assigned to any peer reviewed publication or data generated by this project: NA

- C. All persistent uniform resource locators (URLs) for all peer reviewed publications that have been generated using data from this project: NA
- D. Dataset URLs, if available: NA

II. The purpose of this research project is to:

Perform seismic testing at nine sites throughout Wyoming and attempt to correlate seismic data with pile driving records, rock strength, and depth of transitional material.

III. Data Types and Storage

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

The recorded data will be miniSEED or SEG-2. Data collection will be documented using field notebooks and datasheets, which will be scanned and stored with the raw data. The raw collected data is not reproducible and the loss of this data will require field investigations to be repeated. The analyzed data is reproducible and can be produced from the raw data by performing new analyses. If the analyzed data is lost or corrupted new analyses can be performed. It is anticipated that many of the raw files will be converted into text or Matlab files for use in the analyses. It is also anticipated that Matlab scripts to process and manage the data will also be developed and stored. Figures will be developed to help visualize the results and other important data (shear wave velocity profiles, dynamic soil properties, ect...), these files will be stored as *.jpg or Matlab *.fig files.

The growth rate of the data depends on the depth and number of analyses performed. Which is unknown until the analyses are complete. However it is anticipated that all the data will occupy no more than 1-3 terabytes of storage space for this project. Long-term storage and backup of the data will be handled at the University of Wyoming using local storage devices. Access to the raw data will be available to engineers at WYDOT upon request. It is not anticipated that the pre-processed or post-processed data will be published or accessible to the broad research community. However, it is anticipated that research findings, figures and some tabulated data will be published in a comprehensive reports, articles, and journals.

Pre-existing data includes “as-built” bridge plans from 1958 which are stored and accessible by bridge engineers at WYDOT, these plans may be used to help plan future site investigations. Other data collected includes PDA, field borings, rock strengths and pile driving records. WYDOT stores all this data and will make it available to the PI during the project as needed.

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:

Raw data formats include SEG-2 and miniSEED files which are common files types for the collection of data in seismic testing. From these raw files text and data analyses files will be generated. The organization of these generated files will be determined as the files are generated and will be consistent throughout the project. Likely data will be organized based on the project, site, and test number. This structure of folders and subfolders will allow for future identification of the tests performed and data collected can be easily identified if needed. Documentation of how the files structure is made and used will be determined between the graduate assistant and the principal investigator. Data will be generated in formats including; txt, jpeg, bmp, matlab fig files. Geopsy, a free inversion and wave analyses software will be used to develop multiple files for analyses and includes many file types.

There is currently no one single community sharing standard for storing and sharing this type of data however, the SEG-2 and miniSEED file formats are standard and can be shared with others (WYDOT) upon request.

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.

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?

Because the raw data will not be shared with the broad research community and no business or participants are sharing confidential information there are no concerns about database access. However, password protection will be used on short- and long-term storage devices. Long term data storage will be controlled on-site and controlled by the principal investigator.

NOTE: 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 raw data and processed data will not be shared with others except under written consent from the Wyoming Department of Transportation, and those hired to review the data as outlined in the project proposal. Published data from this project will include figures, tables and descriptions of the collected data. The data is likely to be published after the project is complete and in a yet unidentified journal. It is also likely that conference publications may be used to disseminate these results.

NOTE: 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).

Long-term storage of the data will be accomplished using a non-portable storage device. This will reside at the University of Wyoming and will be a backup for the data that is collected in the field as well as the analyzed data. Because much of the data can be reproduced from analyses,

long term storage will be backed up to the cloud or a similar on-line tool for the SEG-2 and miniSEED data. This raw data represents a large expense and is the only data that will be stored on the cloud. All other data can be generated from this data and will not be stored in cloud based archive. The file structure and long-term storage will be maintained by the principal investigator. While the use of an institutional repository to back up all data generated from the project is acceptable, the long term cost associated with monthly/yearly subscriptions make the use of a non-portable storage device much more affordable for most of the generated data.

The dissemination of the research results will also include publications that include persistent identifiers that are maintained by the publishers. Often digital object identifiers are used to reference and find that information. Although a journal has not been identified for publication, it is anticipated that the publisher will maintain a persistent identifier.

NOTE: 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?

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	Characterization of Soil and Rock for Transportation Infrastructure Using Seismic Methods in Wyoming
Creator/contact point	Shawn C. Griffiths, Ph.D. Associate Instructional Professor Sgriff18@uwyo.edu 0000-0003-0694-4769
Publication Date(s)	3/24/23
Description/Abstract	WYDOT utilizes traditional soil borings, drive point analyses, rock coring methods, and lab investigations to identify and characterize the sub-surface soil. Previous experience with these traditional investigation techniques is good; however, current guidance from FHWA encourages the use on non-traditional investigation techniques. These can include a variety of subsurface methods and tools including seismic testing. This study included the data collection and analyses of compression wave and surface wave testing at nine sites throughout Wyoming. The compression wave testing yielded poor quality data and did not result in an ability to see many layers within the subsurface. However, the surface wave testing resulted in a measure of stiffness at each of the nine sites and was a very good predictor of the bedrock depth at each site, with eight of the nine sites in agreement between subsurface investigation and the pile driving records. Correlations between the shear wave velocity and shear strength, rock quality designation and percentage of rock recovered are very poor and yielded no useable correlation data. Lastly, the boundary between weathered and more competent bedrock could not be determined using any of the seismic methods.
Subject and Keywords	Seismic Testing, sub-surface investigation, shear wave velocity, bridge investigation.
Identifier² and/or source	N/A
Collection and Related Documents	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
Edition	N/A
Related Documents	N/A
Coverage	State of Wyoming
Language	English
Publisher/Distributor	FHWA and Wyoming Department of Transportation
Funding agency	FHWA and Wyoming Department of Transportation
Access Restrictions	restricted-public data set
Intellectual Property and Other Rights	Raw data may be accessed with written permission from Wyoming Department of Transportation and submitted to the author.
License	The license or non-license (i.e. Public Domain) status with which the dataset or API has been published.
Code and software needs	Raw data is miniseed and SEG-2, multiple softwares can use this data for analyses.
Format	Text files delimited with spaces or commas. Size and dimensions are site/project specific
Choice of Repository	Desktop portable hard drive and Dropbox

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/>