

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: January 31, 2022

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

WYDOT Project Number: RS03220

- **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:

Perform seismic testing at 9 sites throughout Wyoming and attempt to correlate seismic data with pile driving records, rock strength, and depth of transitional material.

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 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.

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 ...

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.

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...

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.

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 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.

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 ...

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.

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	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.
Identifier² 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, <i>regardless of whether it has been made available</i> . 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/>.