

Research Data Management Plan

[Project Title]

Draft v0.x

[date]

photo/graphic
placeholder



Technical Report Documentation Page

1. Report No. FHWA/MT-yy-xxx/xxxx-xxx	2. Government Accession No. (Delete and insert information here or leave blank)	3. Recipient's Catalog No. (Delete and insert information here or leave blank)	
4. Title and Subtitle Preliminary Data Management Plan Template [Insert Project Title]		5. Report Date (Delete and insert information here or leave blank)	
		6. Performing Organization Code (Delete and insert information here or leave blank)	
7. Author(s) (Delete and insert information here or leave blank)		8. Performing Organization Report No. (Delete and insert information here or leave blank)	
9. Performing Organization Name and Address (Delete and insert information here or leave blank)		10. Work Unit No. (TRAIS) (Delete and insert information here or leave blank)	
		11. Contract or Grant No. (Delete and insert information here or leave blank)	
12. Sponsoring Agency Name and Address Research Programs Montana Department of Transportation (SPR) 2701 Prospect Avenue PO Box 201001 Helena MT 59620-1001		13. Type of Report and Period Covered Draft Report – Project Proposal	
		14. Sponsoring Agency Code 5401	
15. Supplementary Notes (Delete and insert information here or leave blank)			
16. Abstract (Delete and insert information here or leave blank)			
17. Keywords (Delete and insert information here or leave blank)		18. Distribution Statement (Delete and insert information here or leave blank)	
19. Security Class. (of this report) Unclassified	20. Security Class. (of this page) Unclassified	21. No. of Pages (Delete and insert information here or leave blank)	22. Price (Delete and insert information here or leave blank)

Table of Contents

1.0 Narrative	4
1.1 Background and Purpose.....	4
1.2 Goals and Objectives	4
1.3 Definitions.....	4
2.0 Roles and Responsibilities.....	5
3.0 Data Quality Objectives	6
3.1 Data Quality Indicators	6
3.1.1 Bias and Precision	6
3.1.2 Accuracy	6
3.1.3 Completeness	6
3.1.4 Representativeness	6
4.0 Data Description and Documentation	6
4.1 Data type, size, and formats	7
4.2 Data Collection and Processing	7
4.3 Data Standards	7
4.4 Metadata and Metadata Schema	7
4.5 Data Dictionary	8
4.6 File & Directory Naming Convention	8
5.0 Data Security.....	8
5.1 Data Protection Requirements	8
5.2 Data Storage and Backup.....	8
5.3 Secondary data (use restrictions)	8
6.0 Data Storage and Access.....	8
6.1 Persistent identifier (PID) / ORCID ID	8
6.3 Data Access Considerations.....	9
6.4 Data Repository	9
6.5 Data Retention.....	9
6.6 Data Archiving.....	9

List of Tables

Table 1. Definitions	4
Table 2. Data Owner and Steward	5

INSTRUCTIONS:

Each section has a leading question, statement, or guidance to provide context for information requested/required. Any such guidance is in red font and is to be deleted after filing in content, including this INSTRUCTIONS section, and the Table of Contents must be refreshed to reset page numbers **prior to submitting this DMP.**

Data Management Plans (DMPs) provide guidance and controls covering the lifecycle of data collected, curated, used, shared, and stored for research projects funding by the Montana Department of Transportation (MDT) Research Program. This DMP will document the project’s data quality objectives, data collection method(s), data processing and curation steps, quality control and determination of “fitness” for use, accessibility, security, data retention, and archival detail.

1.0 Narrative

1.1 Background and Purpose

Provide a description of the project’s background and purpose. Include a description of the data that you will be gathering during your project and how the project will use that data. Address the types, origins, and sizes of the data that will be collected or generated, as well as the anticipated file formats and standards that will apply to your data.

Consider the following questions when completing this section:

- What types of data (tabular, image, source code, database, text, etc.) will be collected or used?
- What file formats and file extensions are being used? Are they platform-independent (i.e., can be used by different kinds of computers)? Are they non-proprietary (i.e., do not require special software)? If not, provide a justification to explain why those formats are being used.
- How will the data be collected or obtained? What equipment, software, or other resources will be used?
- What limits, parameters, and/or sample sizes are planned for the data?
- How much data is expected to be collected or generated? What will be its rate of growth? How many files will there be and how much storage is needed?
- If using secondary data, describe its use and reuse terms or licensing.

1.2 Goals and Objectives

1.3 Definitions

Add any definitions that are specific to the research project.

Table 1. Definitions

Accuracy	A measure of agreement among repeated measurements of the same property under identical, or substantially similar conditions.
Bias	Represents systematic error (i.e., persistent distortion of a measurement process that causes errors in a particular direction).

Data Archive	A copy of data, a database, or documents preserved in a secondary, lower cost storage location, for infrequent historical reference and/or recovery.
Data Lifecycle	A conceptualization of how data is created and used which attempts to define a “birth-to-death” value chain for data, including acquisition, storage and maintenance, use, movement to archive, and destruction.
Data Retention	The length of time that data is stored or archived before purging (destructing).
Data Repository	A system that provides access to research data and has controls and processes to ensure authenticity and access on a continuing basis.
Precision	Represents random error (i.e., error among repeated measures of the same property under identical conditions, but not systematically in the same direction or of the same magnitude).
Quality Control (QC)	The process of verifying that collected data are complete, correct, and meet expected outcomes. Quality control focuses on defect identification and provides an assessment of collected data against data quality objectives.
Representativeness	The degree to which data accurately and precisely represent a characteristic of a population, parameter variations at a sampling point, a process condition, or an environmental condition.
Secondary Data	Data produced or procured by someone else. Secondary data may be freely shared, purchased, or obtained through a data use agreement. It is important to know what, if any, limitations secondary data has as it may restrict or prevent you from sharing derivative data.

2.0 Roles and Responsibilities

Data Owner: entity that has the decision authority on the data’s use, sharing, and disposition.

Data Steward: entity that is responsible for ensuring all aspects of data management are implemented (e.g., quality assurance and quality control actions, monitoring, and back-up) and that the data is protected per any security requirements.

Data custody refers to who has ownership and/or stewardship custody of the data at any point in the data lifecycle. Add Data Owner and Steward information to Table 2.

Lifecycle stages include Collect/Create/Acquire; Store/Maintain; Use; Archive.

Each data set collected, acquired, or created during this research project has an assigned Data Owner and Data Steward (Table 2).

Table 2. Data Owner and Steward

Dataset Name/Title	Data Owner	Data Steward	Lifecycle Stage ¹
<i>Deterioration Curves for Bridge Elements</i>	<i>Montana DOT (MDT)</i>	<i>Western Transportation Institute (WTI, MSU)</i>	<i>Collect/Create/Acquire</i>
XX	XX	XX	XX
XXX	XXX	XXX	XXX

1. **Lifecycle stages:** Collect/Create/Acquire; Store/Maintain; Use; Archive

3.0 Data Quality Objectives

Data quality objectives (DQOs) are qualitative and quantitative statements derived from a systematic process that defines the data and data rigor (i.e., quality) required to support project objectives and decision-making. The DQO process provides the problem statement (include in Section 1.1 above), decisions or estimates needed, data required, analytical approach or decision rule(s) to draw conclusions.

3.1 Data Quality Indicators

Data Quality Indicators (DQIs) are control measures (i.e., acceptance criteria) for data to be collected, created, or acquired by the research project to ensure the data is fit for the intended analysis and/or research decision(s) within a defined risk tolerance.

3.1.1 Bias and Precision

Bias and precision (collectively known as accuracy) are two principal attributes, or characteristics, of data quality, especially in environmental studies. **Bias:** Represents systematic error (i.e., persistent distortion of a measurement process that causes errors in a particular direction). **Precision:** represents random error (i.e., error among repeated measures of the same property under identical conditions, but not systematically in the same direction or of the same magnitude. To the extent that the data to be collected in this research are subject to measurement bias and precision error describe what those are and how they will be controlled for.

3.1.2 Accuracy

Accuracy is a measure of agreement among repeated measurements of the same property under identical, or substantially similar conditions. Describe, as necessary for this research proposal, the accuracy measure necessary to support the research in the collected/generated data required for this research.

3.1.3 Completeness

Completeness is a measure of the amount of valid data/information needed to be obtained for the intended analytical approach/tools. Describe the level or degree of completeness in the collected/generated data required for this research.

3.1.4 Representativeness

Representativeness addresses the extent to which measurements reflect the sampling unit from which they were taken, as well as the degree to which samples represent the target population. Describe, as necessary for this research proposal, how representativeness is addressed by specifying the number and location of samples within the study design.

4.0 Data Description and Documentation

Indicate how you will organize and document the data. **Data documentation** should describe the data to ensure that future data users, including the original creator, can fully contextualize and understand it.

Your documentation should include descriptive **metadata** as well as any other information necessary to understand and reuse the data (e.g., processing steps, values, external libraries, registries, software dependencies, etc.). As data organization affects data processing and analysis you should plan how folders and files are organized as well as the information within data files. Whenever possible you should use existing **metadata schemas** to ensure maximal interoperability with other data.

Consider the following questions when answering this section:

- How will the data be organized?
- Will any unique identifiers be used to identify files, folders, samples, etc.?
- Which metadata standards and schemas will be used (if applicable)?
- What details (metadata) are necessary for others to use your data?
- How will metadata be generated (automatically and/or manually)?
- Are there existing **standards** or formats for your metadata? (e.g., accepted domain standard, widespread usage, software-generated formats, etc.)
- What data dictionaries/taxonomies/ontologies will be used for this research data, if applicable?
- Are there any external tools, libraries, or software that will be required to use the data (e.g., software, instruments)?

4.1 Data type, size, and formats

Provide details on the data type(s), size, and formats to be collected, created, or acquired (including file format). How much data is expected to be collected or generated? What will be its rate of growth? How many files will there be and how much storage is needed?

4.2 Data Collection and Processing

Provide details on how data will be collected or created to support the research objective and meet data quality objectives/indicators.

4.3 Data Standards

Standards: An agreed-upon way of doing things, with the purpose of establishing shared expectations. Using standards for data helps ensure consistency in the format, organization, structure, and meaning of data and metadata. Existing standards (such as those issued by international bodies like ISO) should be used where applicable. **Describe the data standards that will be applied** to data collected or generated by this research project.

4.4 Metadata and Metadata Schema

Metadata, commonly called "data about data," is information which describes data. Good metadata enables others to understand and reuse data that they themselves did not create. Metadata may include field or abbreviation meanings, units, information on how data was collected or transformed, and more. Metadata is an essential part of the documentation process as are in-line comments in computer code to explain functions and variables.

Metadata schemas are established frameworks for recording metadata. They enable easy sharing and reuse of data since all the data are recorded the same way. If your data does not fit an existing schema, then your team must agree upon the minimum required amounts of metadata needed and how this information will be documented and formatted.

4.5 Data Dictionary

What data dictionaries/taxonomies/ontologies will be used for this research data, if applicable.

4.6 File & Directory Naming Convention

5.0 Data Security

Describe how the data will be stored, backed up, and protected to prevent loss and unauthorized access. Topics to address include data privacy, confidentiality, integrity, security risks, as well as safeguards and mitigation measures (e.g., encryption, backups, disaster recovery, de-identification methods, error checking, off-site storage, access limitations, etc.).

Consider the following when answering this section:

- Describe provisions for maintaining the security and integrity of the research data (e.g., encryption and backups, how the data will be protected from accidental or malicious modification or deletion, including data recoverability, disaster recovery/contingency planning, and any off-site storage sites or technology).
- Describe any known privacy, confidentiality, and security risks or restrictions associated with the data such as institutional review board (IRB) requirements, licensing or contractual restrictions, personally identifiable information (PII), proprietary information, intellectual property concerns, etc. **Include both original and secondary data in this assessment.**
- Describe how identified risks will be mitigated. If your data has no known risks, you should include a statement to that effect.

5.1 Data Protection Requirements

5.2 Data Storage and Backup

5.3 Secondary data (use restrictions)

6.0 Data Storage and Access

6.1 Persistent identifier (PID) / ORCID ID

Persistent identifier (PID) is a long-lasting identifier that will continue to reference its assigned item (such as a document or dataset) regardless of its location on the internet. Unlike other identifiers, PIDs rely on a separate service that must be updated to maintain linkage and to redirect inquiries when the location of an item changes.

ORCID stands for Open Researcher and Contributor ID. You can create your own ORCID at <https://orcid.org/>. ORCID.org provides a registry of persistent unique identifiers for researchers and scholars and automating linkages to research objects such as publications, grants, and patents.

Registration is free and takes about 5 minutes. ORCID [FAQs](#) are provided by National Transportation Library.

6.3 Data Access Considerations

Long-term access is a goal to keep data discoverable and usable in the future. For Montana DOT projects, PIs are responsible for ensuring all shared final data, reports, tech transfer summaries, and other textual products are publicly accessible for a period of 10 years from the end of the contract period.

My institution's policy is that the data and all supporting materials from all research are owned and must remain with the institution if I leave. How does this policy affect what I can say about data management?

Data management by an institution is one avenue by which data preservation and access can be achieved. However, this DMP must address the institutional strategy for providing access to relevant data and supporting materials.

Should I consider contributing my research data to a data archive?

Maybe. Archives are organizations that collect and distribute data. They understand what is needed to prepare data for wider distribution and documentation for users. They provide stable, reliable, and cost-effective means for distributing data. They also provide protections for the dataset and technical assistance for requestors.

(source: <https://ntl.bts.gov/ntl/public-access/faqs>)

6.4 Data Repository

A system that provides access to research data and has controls and processes to ensure authenticity and access on a continuing basis.

6.5 Data Retention

Define the data retention schedule for data accessibility prior to archiving.

6.6 Data Archiving

Should I consider contributing my research data to a data archive?

Maybe. Archives are organizations that collect and distribute data. They understand what is needed to prepare data for wider distribution and documentation for users. They provide stable, reliable, and cost-effective means for distributing data. They also provide protections for the dataset and technical assistance for requestors.

(source: <https://ntl.bts.gov/ntl/public-access/faqs>)