



Iowa Department of Transportation

Research & Analytics Bureau

Guidance on Data Management Plan (DMP) Requirements for Federally Funded Research Projects

Version 03, 2022/04/01

Summary

Iowa DOT Research expects the timely release and sharing of research data. Data management plans support this goal by providing a space for you to document which data will be gathered, analyzed, reused, stored, shared, and preserved for your research project. The Iowa DOT's data management plan was made to complement and support the US DOT's Public Access Plan.

Scope

What constitutes the data covered by this plan will be determined by the Principal Investigator and the Iowa DOT Research Project Manager but in general, your plan should focus on:

- 1) Digital research data, including software and code products, that are of sufficient quality to validate and replicate research findings, regardless of whether they are used to support scholarly publications.
- 2) Data that is difficult to replicate or of long-term interest to research communities.

If the data supports a publication then it should be released no later than at the time of manuscript acceptance. However, if there is a valid reason to restrict access to supporting data for a longer period of time, an embargo of up to twelve (12) months is acceptable. Reasons to further limit access to data, such as not sharing, longer embargo periods or restricted access to data should be documented in the plan, justified, and approved by the Iowa DOT Research Project Manager.

Data that is made publicly available must be reported to the [U.S. DOT's Repository and Open Science Access Portal](#) (ROSA P). To report your data please visit <https://rosap.ntl.bts.gov/submitContent> and follow current protocols for submitting content in compliance with the USDOT Plan to Increase Public Access to the Results of Federally-Funded Scientific Research.

Format

Your plan should be as detailed as needed to show compliance with the [US DOT Public Access Plan](#) and Iowa DOT R&A requirements, while staying within the page limit of three (3) pages total. All data management plans (DMP) must include a one (1) page cover sheet and a 1-2 page narrative consisting of the following sections:

1. Data description
2. Data documentation
3. Data security
4. Data sharing
5. Data preservation
6. Roles and responsibilities

Cover sheet formatting

Each data management plan must be preceded by a cover sheet containing the documentation table found on page 5. Fill out the table as completely as possible before beginning the project and update it throughout the project as needed. For example, the plan should be updated at the end of the project, and after derivative publications are created and/or published.

Narrative

Below you will find descriptions and a set of questions that can be used to help complete your response to each section of the narrative portion. These questions are not exhaustive and you are encouraged to add more details when needed in order to adequately complete the section. If a section does not apply to your project, you should provide an explanation rather than skip it. Words and phrases in bold are defined in a glossary which can be found in the Appendix.

1. Data Description

Provide a description of the data that you will be gathering in the course of 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 answering this section:

1. What types of data (tabular, image, source code, database, text, etc.) will be used or collected?
2. 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, you should provide a justification to explain why the format(s) are being used.
3. How will the data be collected or obtained? What equipment, software, or other resources will be used?
4. What limits, parameters, and/or sample sizes are planned for the data?
5. 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?
6. If using **secondary data**, describe its use and reuse terms or licensing.

2. Data Documentation and Organization

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:

1. How will the data be organized?
2. Will any unique identifiers be used to identify files, folders, samples, etc.?
3. Which metadata standards and schemas will be used (if applicable)?
4. What details (metadata) are necessary for others to use your data?
5. How will metadata be generated (automatically and/or manually)?

6. Are there existing standards or formats for your metadata? (e.g. accepted domain standard, widespread usage, software-generated formats, etc.)
7. What data dictionaries/taxonomies/ontologies will you use for your data, if these are used within your field?
8. Are there any external tools, libraries, or software that will be required to use the data (e.g., software, instruments)?

3. Data Security

Describe how the data will be stored, backed up, and protected in order 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:

1. 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).
2. 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.
3. Describe how identified risks will be mitigated. If your data has no known risks you should include a statement to that effect.

4. Data Sharing

Iowa DOT encourages open data sharing - i.e. sharing data with the least possible restrictions. In this section identify which research data will be shared beyond the research team, who will have access, how and when it will be disseminated, and under what circumstances. Make sure to also address how you will guard against disclosure of confidential or sensitive information such as human-subjects data, personally identifiable information (PII), confidential business information, intellectual property, etc. when sharing data. *Provide a justification if your data cannot or should not be shared beyond the research team.*

If using a data repository to share data, include information in this section that addresses how the repository conforms with the U.S. [DOT guidelines for evaluating repositories](#) including how the repository makes data sets discoverable, which **persistent identifier(s)** are used, and if the repository is on the [list of repositories that conform with the DOT's Public Access Plan](#). *Provide a justification if you plan to share data without using a data repository.*

Consider the following questions when answering this section:

1. Which data will be shared and when will it be available?
2. Who will have access to the shared data?

3. Will data use agreements or other restrictions be used to limit who may access the data and how it can be reused? If so, what are the terms for access and use and how can someone obtain access to the files?
4. Will any additional steps need to be taken to prepare the data for sharing? (e.g. removal of identifying information, export to more common or non-proprietary formats, etc.)

5. Data Preservation

Articulate your plans for **data preservation**, i.e. how data will be preserved for **long-term access**. Concentrate on details such as which data will be preserved, who will have access to it, how it will be prepared, and where it will be stored. If using a data repository for preservation make sure to address its preservation capabilities such as retention commitments, certifications, access restrictions (if applicable), etc.

Consider the following questions when answering this section:

1. Which data will be preserved and for how long?
2. What systems (people, technology) will monitor and ensure data integrity?
3. How will you ensure that there is enough documentation accompanying the shared data that other professionals in your field will be able to use and make sense of it?
4. Are there any other steps that need to be completed such as obtaining prior permission or verification to use a specific data repository or preservation method?
5. Have the costs for data preservation, including time and fees, been addressed in the budget?

6. Roles and Responsibilities

Describe the division of major responsibilities for managing, stewarding, and protecting your data among Principal Investigators and other members of the project team. Include contingency plans for the departure of key personnel from the project and for actions and responsibilities that extend past the end of the research contract.

Consider the following questions when answering this section:

1. Who is responsible for checking for adherence to the DMP? Who will keep the DMP updated throughout the project?
2. Who will provide quality control, monitoring, back-up, and other forms of quality assurance for the data?
3. Who is responsible for maintaining and monitoring data chosen for preservation?
4. Who will report public data sets to the U.S. DOT's [Repository and Open Science Access Portal](#) (ROSA P)?

**Data Management Plan (DMP) for
Iowa DOT Research Projects**

Research Data Management Documentation Table

Name of Principal Investigator(s) or Contractor(s) and ORCID number		
Current Project Title, and all previous project titles		
Iowa DOT Project Manager		
Iowa DOT Project number		
Other contract or grant numbers		
Iowa DOT Research-assigned project Digital Object Identifier (DOI), or researcher acquired DOI		
TRB Research in Progress (RiP) Title, Accession Number, and URL		
Project Duration (projected)	Start Date:	End date:
Do the data management requirements of the US DOT “Plan to Increase Public Access to the Results of Federally-Funded Scientific Research” apply to this project	Yes or No; and if No, why not:	
Name(s) of Federal funder(s), Funding Program Name(s), Agency Code(s) and/or Contract/Grant numbers		
DMP Version		
Date DMP amended, if any		
Name and ORCID number of each author		
Persistent links or identifiers assigned to this project, datasets, reports, or peer reviewed publications generated by this project		
Name and URL of all peer reviewed publications which have been generated from this project		

This table is to be filled out as completely as possible before the beginning of the project, and updated as needed, including at the end of the project, and after, as derivative publications are created.

[Note: Guidance for using this template can be found at: <http://publications.iowa.gov/id/eprint/40404>]

Appendix

Glossary

Data (or, research data): The recorded factual material, software, or code as commonly accepted in the scientific community that is of sufficient quality to validate and replicate research findings, regardless of whether the data are used to support scholarly publications.

Research data do not include laboratory notebooks, preliminary analyses, completed case report forms, drafts of scientific papers, plans for future research, peer reviews, communications with colleagues, or physical objects, such as laboratory specimens.

Data archive: see **Data repository**

Data documentation: supplementary files that describe data with sufficient detail to ensure that future users, including the original creator, can fully contextualize, understand, and reuse it. Codebooks, data dictionaries, user guides, and readme files are all examples of data documentation. Metadata is an essential part of the documentation process as are in-line comments in computer code to explain functions and variables.

Data preservation: data preservation is the process of ensuring the usability of data beyond the lifetime of the research activity. Data preservation is a process that involves steps such as evaluating file formats and documentation and ensuring data integrity and recovery.

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

Metadata: is often called “data about data” in that it is structured information that describes a resource. For example, metadata describing a book in a library catalog includes information (data) about the book’s author, its title and date of publication. Metadata for research data is similar in that it should answer questions about where the data came from, how it was made or obtained, who or what created it, etc.

Metadata schema: a set of rules that define a structure for formatting metadata. Multiple data sets that use the same metadata schema can be combined which improves the utility and value of data.

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

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.

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.

Long-term access: A goal to keep data discoverable and usable in the future. For Iowa 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 ten (10) years from the end of the contract period.