

Introduction

This document provides guidelines for creating a Data Management Plan (DMP). These guidelines are based on the U.S. Department of Transportation's (DOT) [Guidance for Creating Data Management Plans for Extramural Research](#), along with additional clarification on the Federal Transit Administration's (FTA) expectations for DMPs. These instructions inform applicants of elements that FTA would like to see in the DMP but do not serve as an approved government form or template. Applicants should use their best judgment in determining what information to include and whether additional information should be incorporated into the DMP.

A DMP is a document that describes how applicants plan to handle digital datasets, software, or code generated over the course of a research project pursuant to federal and Departmental requirements. A DMP must be provided as a condition of receiving FTA funds under the Section 5312 Research Program and should adequately identify 1) the data to be collected, 2) how the data will further the goals of the specific project and the research program, 3) how the data will be made accessible, and 4) how the data will be stored.

FTA understands that elements in the DMP may need to be updated as the research project develops to capture pertinent changes to a project's data management procedures and/or data.

A DMP should include all of the following elements and information.

1. Project Overview

Please provide summary information about the research project and its goals as well as how the project's data helps FTA achieve the overall goals of the FTA Research program. Suggested elements for this section include, but are not limited to the following:

- **Project Title:** Please provide the title of the project.
- **Project Goals and Objectives:** Briefly describe the goals and objectives of this project and how the data collected help the DOT achieve its goals and objectives.
- **Project Description:** Briefly describe the research project, including project design and execution.
- **Project Performance Measurements:** Briefly describe any performance measurements the project will use and, if appropriate, how the data collected support these measurements.

2. Overview of Digital Datasets Software or Source Code

Please describe the data, software, and/or source code that will be collected and/or produced. Suggested elements for this section include, but are not limited to:

- **Data Title:** Provide the anticipated data and/or dataset title(s) for the research project.
- **Description:** Briefly describe the purpose of the data and/or datasets that are anticipated to be collected during the research project.
- **Type/Scale:** Specify the type and scale of the data that will be generated (e.g., numerical data, image data, text sequences, etc.) for the research project's data and/or datasets.
- **Collection Method:** Describe the methods for collecting and/or acquiring the data (e.g., simulated, observed, experimental, software, physical collections) and specify by dataset if appropriate.

- **Data File Format(s):** Provide the data file format(s) the data and/or datasets are anticipated to be in when they are made accessible to the DOT (e.g., .csv, .txt, json, newline json, etc.). If not yet known, explain to the best of your ability.
- **Metadata:** Indicate what data definitions, explanations, and other information you will provide to describe the data and/or dataset in various contexts. If not yet known, explain to the best of your ability.
- **Data Owner:** List the data owner of the data or dataset. The data owner is the person or organization that has the authority, ability, and responsibility to access, create, modify, store, use, share, and protect the data. Data owners have the right to delegate these privileges and responsibilities to other parties. Please note that the data owner may be FTA for FTA-funded research projects.
- **Data Steward:** List the data steward for the data or dataset. The data steward is, at the direction of the data owner, the person or organization that is delegated the privileges and responsibilities to manage, control, and maintain the quality of a data asset throughout the data lifecycle. The data steward may also apply appropriate protections, restrictions, and other safeguards depending on the nature of the data, subject to the direction of the data owner.

3. Access Level

Providing appropriate data access to safeguard data is a key aspect of data stewardship. In accordance with the [OPEN Government Data Act](#), datasets should be made publicly accessible unless specific concerns require the data to have controlled access.

If you plan to collect data that requires controlled access, please describe the reasons for controlling access to data and plans for safeguarding these data. Controlled access is defined as restricting access to certain groups of persons due to data containing personally identifiable information (PII), information that threatens the privacy of an individual or group, information that threatens the confidentiality of a person or group, and/or contains confidential business information (CBI).

Suggested elements for this section include, but are not limited to the following:

- **Access Level:** Can all the data from this project be shared with the public, or is controlled access required for at least some of the data? Please detail whether the data are:
 - All Public Access: There are no anticipated data requiring controlled access.
 - Some/All Controlled-Access: There are anticipated data in one or more datasets requiring controlled access (e.g., PII or CBI).

If your project plans to include controlled-access data, please provide the following information (if not, please skip to Step 4).

- **Dataset Title:** Provide the anticipated data/dataset(s) that will be at the specified access level.
- **Reasons for Controlled Access:** For any data requiring controlled access, detail what in the data is anticipated to require controlled access and why controlled access may be required for those elements. Provide details on all elements of the data and/or datasets that are anticipated to require controlled access.
- **Safeguarding Methods and Processes:** Describe the planned methods and processes that will be taken to safeguard the data requiring controlled access, such as redacting data before providing public access to the dataset. Elaborate on the specific methods—for example, blurring license plates in videos, geofencing sensor

data, removing persistent IDs, etc. Any privacy, ethical, or confidentiality concerns raised due to data-sharing should be described.

- **Informed Consent:** If your project contains human subject research, describe how informed consent forms will permit sharing with the research community and whether additional steps, such as an Institutional Review Board (IRB) review, are used to protect privacy and confidentiality.

4. Re-Use, Redistribution, and Derivative Products Policies

In general, data should be made available through an open license that is available at no cost to the public and with no restrictions on copying, publishing, distributing, transmitting, citing, or adapting. However, recipients have the option to impose some conditions on the re-use and sharing of their research results, such as requiring that others who use their results attribute the original creators or restricting others from modifying their work or using the results of their work for commercial purposes. A listing of licenses can be found on the Creative Commons [page](#).

Suggested elements for this section include, but are not limited to the following:

- **License Used:** List the license anticipated to be used for this dataset and provide a URL (or several) with more information on the license. Specify whether the license is an open license or a non-open license. An open license is a legal guarantee that a data asset is made available (A) at no cost to the public and (B) with no restrictions on copying, publishing, distributing, transmitting, citing, or adapting. A non-open license does not adhere to all open license criteria.
- **Reasons for a Non-Open License:** If a non-open license is anticipated to be used for this dataset, provide the reason for using it.

5. Data Storage, Archiving and Preservation

Storing and retaining the data is a key part of the data steward's responsibilities to manage, control, and maintain the quality of a data asset throughout the data lifecycle.

FTA recommends that recipients preserve their data by providing it to the [National Transportation Library Repository and Open Science Access Portal \(ROSA P\)](#), where it will be retained in perpetuity and be searchable based on a Digital Object Identifier (DOI), dataset title, and other descriptive metadata.

Alternately, your organization may choose to archive your data in its own institutional repository or a third-party repository or may choose self-dissemination. Institutional repository is an option for agencies that have their own data repository and require that data be housed in the organization's archive. In these cases, DOT would provide a link to the institutional repository in Research Hub and ROSA P. For potential third-party repositories, please see the [Data Repositories Conformant with the DOT Public Access Plan](#). Self-dissemination is not a preferred option, as it requires that a specific point of contact be available for the foreseeable future to provide the data to others upon request.

ROSA P Option

If you choose to archive the data via ROSA P, please include the following generic language in your Data Management Plan:

The dataset will be archived in the [National Transportation Library \(NTL\) Repository and Open Science Access Portal \(ROSA P\)](#). Prior to archiving, the data are stored on the secured BTS networks and drives, which are backed up nightly. The US DOT systems are secured from outside users and backed up daily.

Files in ROSA P are backed up on NTL drives at DOT daily; at the Centers for Disease Control, the repository managing facility, daily; and on Amazon Web Service Cloud servers in Virginia and Oregon daily.

The dataset will be retained in perpetuity.

NTL staff will mint persistent Digital Object Identifiers (DOIs) for each dataset stored in ROSA P. These DOIs will be associated with dataset documentation as soon as they become available for use.

The DOIs associated with this dataset include <https://doi.org/10.21949/> [NTL staff will assign after funding.]

[To request a DOI for each research dataset, please send an email to your FTA project manager and NTLDataCurator@dot.gov; Subject line: Dataset DOI Request. Include dataset title and your organization's name in the email body.]

The assigned DOI resolves to the repository landing page for the [insert title of project] dataset so users may locate associated metadata and supporting files.

ROSA P meets all the criteria outlined in the [Guidelines for Evaluating Repositories for Conformance with the DOT Public Access Plan](#).

Alternative to ROSA P

If your organization plans to archive the data in a repository other than ROSA P, your DMP should:

- Indicate the approximate time period between data collection and submission to the archive.
- Identify where data will be stored prior to being sent to an archive.
- Describe how back-up, disaster recovery, off-site data storage, and other redundant storage strategies and policies will be used to ensure the data's security and integrity, or provide the link to the repository's documentation.
- Describe how data will be protected from accidental or malicious modification or deletion prior to receipt by the archive, or provide the link to the repository's documentation.
- Indicate how long the chosen archive will retain the data.
- Indicate if the chosen archive employs, or allows for the recording of, persistent identifiers linked to the data, or provide the link to the repository's documentation on this topic.
- Discuss how your chosen data repository meets the criteria outlined in the [Guidelines for Evaluating Repositories for Conformance with the DOT Public Access Plan](#) or is on the list of [Data Repositories Conformant with the DOT Public Access Plan](#).