



Preliminary Data Management Plan Instructions



Preliminary Data Management Plan Instructions

Produced by
U.S. Department of Transportation
Intelligent Transportation Systems (ITS) Joint Program Office

Notice

This document is disseminated under the sponsorship of the Department of Transportation in the interest of information exchange. The United States Government assumes no liability for its contents or use thereof.

The U.S. Government is not endorsing any manufacturers, products, or services cited herein and any trade name that may appear in the work has been included only because it is essential to the contents of the work.

Contents

- Introduction 4**
- 1 Project Overview 4**
- 2 Data Overview 4**
- 3 Data Stewardship 5**
 - 3.1 Data Owner and Steward5
 - 3.2 Access Level.....5
 - 3.2.1 Can all of the data from this project be shared with the public, or is controlled-access required for at least some of the data?6
 - 3.2.2 Datasets Requiring Controlled-Access6
 - 3.2.3 Informed Consent6
 - 3.3 Re-Use, Redistribution, and Derivative Products Policies6
 - 3.4 Data Storage, Archiving and Preservation7

Preliminary Data Management Plan Instructions

Introduction

This document provides guidelines for creating a preliminary Data Management Plan (DMP) to make it easier for projects to prepare their preliminary DMP using a consistent format. These guidelines are based off of U.S. DOT's National Transportation Library's [Guidance for Creating Data Management Plans for Extramural Research](#), along with additional clarification on U.S. DOT's expectations for DMPs. These instructions inform applicants of elements the U.S. DOT would like to see in the preliminary DMP but do not serve as an approved government form or template. Applicants should use their best judgement in determining what information to include and whether or not additional information should be incorporated into the preliminary DMP.

A preliminary DMP is a document submitted as part of an application that describes how applicants plan to handle digital datasets over the course of a research project pursuant to federal and Departmental requirements. A preliminary DMP must be provided as part of an application for any Intelligent Transportation Systems (ITS) Joint Program Office (JPO)-funded project or project adhering to the Data Access Center of Excellence best practices, and should adequately identify: 1) the data to be collected, 2) how the data will further the goals of the U.S. DOT, 3) how the data will be made accessible, and 4) how the data will be stored.

The United States Department of Transportation (U.S. DOT) understands elements in the preliminary DMP may be updated as the research project develops. However, the U.S. DOT seeks a clear explanation of the research project's data and the data's alignment with the U.S. DOT's research goals. After funds are awarded, this preliminary DMP shall be updated regularly to capture pertinent changes to a project's data management procedures and/or data.

1 Project Overview

Please provide summary information about the research project and its goals, as well as how the project's data helps the U.S. DOT achieve its goals.

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

- **Project Title:** Please provide the title of the proposed project.
- **Project Goals and Objectives:** Briefly describe the goals and objectives of this project, as well as how the data collected in this project will help the U.S. DOT achieve its goals and objectives.
- **Project Description:** Briefly describe the research project, including project design and execution.

2 Data Overview

Please provide a description of the nature, scope, and scale of the data that will be collected and/or produced. To the extent possible, please describe the datasets that are anticipated to be collected throughout the research project and any details for these datasets.

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, video, audio, database, modeling data, source code, 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; sensors; satellite; enforcement activities; researcher-generated

Preliminary Data Management Plan Instructions

databases, tables, and/or spreadsheets; instrument generated digital data output such as images and video; etc.), 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 U.S. DOT (e.g. .csv, .txt, json, newline json, etc.). If not yet known, explain to the best of your ability.
- **Metadata:** Indicate what metadata schema you will use to describe the data and/or dataset in various contexts. If not yet known, explain to the best of your ability. U.S. DOT expects metadata in several contexts, including:
 - At the field level, describing the elements in the data and/or dataset
 - At the asset level, describing the contents of the data or dataset generally, in accordance with the [Project Open Data Metadata Schema](#) or another appropriate standard
 - At any other level that may be appropriate for the data and/or dataset

3 Data Stewardship

Please provide details around data stewardship. Data stewardship involves proper management of data throughout the data lifecycle. This includes, but is not limited to, maintaining data quality and safeguarding the data.

3.1 Data Owner and Steward

Given the partnership-driven nature of the U.S. DOT, many determinations around project data, including its ownership, will be made during negotiations between project partners. Applicants should indicate in their preliminary DMP the status of negotiations at the time of application.

Please define the data owner and data steward for the anticipated data and/or datasets that will be produced as part of the research project. It is acceptable for multiple datasets to have the same data owner and/or data steward.

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

- **Dataset Title:** Please provide the data and/or datasets title(s) that will be assigned the designated data owner and/or data steward. You may break out data and/or datasets that have separate data owners and/or data stewards.
- **Data Owner:** List who the data owner of the data or dataset is. 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 the data owner may be the U.S. DOT for U.S. DOT-funded research projects.
- **Data Steward:** List who the data steward for the data or dataset is. 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. The data steward may be the recipient of U.S.DOT-funded research projects or one of the recipient's partner organizations.

3.2 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. Please detail which data may require controlled-access, reasons for controlling access to data, and plans for safeguarding these data.

Preliminary Data Management Plan Instructions

Controlled-access is defined as restricting access to certain groups of persons due to data containing personally identifiable information (PII), information that threatens privacy of an individual or group, information that threatens confidentiality of a person or group, and/or contains confidential business information (CBI).¹

3.2.1 Can all of the data from this project be shared with the public, or is controlled-access required for at least some of the data?

Please check the box for the access level that best applies to the project's data:

- **All Public Access:** There is no anticipated data requiring controlled-access.
- **Some/All Controlled-Access:** There is anticipated data in one or more datasets requiring controlled-access (e.g. Personally Identifiable Information [PII] or Confidential Business Information [CBI]).

3.2.2 Datasets Requiring Controlled-Access

The information below is only required if the "Some/All Controlled-Access" box is selected in 3.2.1. Suggested elements for this section include, but are not limited to:

- **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, please detail what in the data is anticipated to require controlled-access and why controlled-access may be required for those elements. Please 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. Please 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.

3.2.3 Informed Consent

If your project contains human subject research, please describe how 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.²

3.3 Re-Use, Redistribution, and Derivative Products Policies

The U.S. DOT is interested in who holds the intellectual property rights for the research project data, whether these rights will transfer to a data archive, whether any copyrights apply to the data, and other aspects that apply to the rights to use the data. Please provide details on these elements in this section.

In accordance with the [OPEN Government Data Act](#), to the extent practicable data must 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 such asset. In this section, please provide the applicable licenses for each dataset and specify if an open license will be used. If the project is U.S. DOT-funded, an open license should be used for the

¹ <https://www.its.dot.gov/data/>

² <https://ntl.bts.gov/public-access/creating-data-management-plans-extramural-research>

Preliminary Data Management Plan Instructions

dataset. If a non-open license is anticipated to be used for the dataset, please include a justification for why a non-open license will be used.

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

- **Dataset Title:** Please provide the title(s) of the dataset(s) that are anticipated to use this license.
- **License Used:** List the license that is anticipated to be used for these data and/or datasets (e.g. Creative Commons – BY version X.0) 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” means 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 such asset.³ A non-open license does not adhere to these open license criteria.
- **Reason(s) for a Non-Open License:** If a non-open license is anticipated to be used for this dataset, provide the reason for using a non-open license.

3.4 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.

In this section, list all of the data storage systems that will be used to store the research project’s data, provide details on those data systems, and specify how long the data will be stored in each system. Please note that U.S. DOT-funded projects should provide access to the U.S. DOT for a minimum of five years after the award period of performance expires. The U.S. DOT understands that dates provided in this section are anticipated dates and will be updated after award.

A public access system provides full data access to the public. A controlled-access system restricts access to certain groups of persons due to data containing personally identifiable information (PII), information that threatens privacy of an individual or group, information that threatens confidentiality of a person or group, and/or contains confidential business information (CBI).⁴ An applicant system is managed by the applicant, a U.S. DOT system is managed by the U.S. DOT, and a third-party system is a system that is managed by a person, group or organization other than the U.S. DOT or the applicant.

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

- **Data Storage System Name:** Provide the name of the anticipated data storage system(s) the dataset(s) will be stored in. Provide a URL/URLs or other means of accessing the data storage system.
- **Data Storage System(s) Type:** For each data storage system, state whether the data storage system is:
 - “Applicant – Public System”
 - “Applicant – Controlled System”
 - If you are using an applicant-managed system, please state whether the system is currently operational, under development, or will be developed as part of this project.
 - “U.S. DOT-managed – Public System”
 - “U.S. DOT-managed – Controlled-Access System”
 - “Third-party – Public System”

³ <https://www.congress.gov/bills/115th-congress/house-bill/1770/text>

⁴ <https://www.its.dot.gov/data/public-access/>

Preliminary Data Management Plan Instructions

- "Third-party – Controlled-Access System"
- **Dataset Title(s):** Provide all data and/or datasets that are anticipated to be stored in each data storage system.
- **Initial Storage Date:** Provide the anticipated initial date that data will be available in each data storage system. If you are live-streaming data, this may be the initial ingest date. You can write this date as a period after award (e.g. three months after award, six months after award, etc.)
- **Frequency of Update:** Please provide how frequently the data will be updated in the data storage system once ingestion begins. Please note the U.S. DOT prefers obtaining the data as quickly as possible after collection, preferably in near real-time and/or through automated uploads. This is in accordance with the OPEN Government Data Act's requirement to maximize accessibility to federally-funded datasets.
- **Archiving and Preservation Period:** Provide the duration for which the data and/or dataset will be maintained in each data storage system.

3.4.1 Data Storage System Description

Provide a description of each data storage system the project is intending to use, including any URLs or other means that provide additional details and documentation of the system, details on data backup of the system, and details on access and controlled-access, if applicable. If using an applicant or third-party data storage system, please provide additional details that demonstrate the system adheres to the U.S. DOT's policies for data storage systems, including:

- [Guidelines for Evaluating Repositories for Conformance with the DOT Public Access Plan](#)
- For Controlled-Access Systems, describe how the system will enable privacy protection, controlled-access and collaboration at least equal to U.S. DOT's controlled-access system, the [U.S. DOT Secure Data Commons](#)

3.4.2 Cybersecurity Policies

State any cybersecurity policies that apply to the data storage system the project is intending to use or the entirety of this DMP. Be specific in which data storage system(s) these policies apply to.



Preliminary Data Management Plan Instructions

U.S. Department of Transportation
ITS Joint Program Office-HOIT
1200 New Jersey Avenue, SE
Washington, DC 20590

Toll-Free "Help Line" 866-367-7487
www.its.dot.gov

FHWA-JPO