Plan Overview

A Data Management Plan created using DMPTool

DMP ID: https://doi.org/10.48321/D1F56ED1AC

Title: WMATA -- Transit Data Standards and Analysis Infrastructure Framework

Creator: Michael Eichler - **ORCID:** 0000-0002-6802-4872

Affiliation: United States Department of Transportation (DOT) (transportation.gov)

Funder: United States Department of Transportation (DOT) (transportation.gov)

Funding opportunity number: DOT-SMART-FY23-01

Grant: https://www.transportation.gov/grants/SMART

Template: SMART Grants Stage 1 Data Management Plan (DMP)

Project abstract:

The project seeks to address two major challenges involving Transit Data, where Transit Data is defined as data generated by sensors and automated transit operational systems (e.g., automatic passenger counters, automatic fare collection, automatic vehicle location). First, the SMART grant will support the continued development, refinement, and prototyping of Transit Data standards (formats and frameworks) and second, the development of open-source, interchangeable software modules at WMATA. The project will result in published standards and guidance made available to all transit agencies, academic researchers, and private companies, as well as reusable software modules that can be implemented or customized by other agencies, researchers, and/or private industry.

Start date: 07-01-2024

End date: 12-31-2025

Last modified: 10-10-2024

WMATA -- Transit Data Standards and Analysis Infrastructure Framework

Project Contact Information

Please provide as much of the the following information as possible:

1. **Name of the project;**
2. **Grant number;**
3. **Name of the person submitting this DMP;**
4. **ORCiD of the person submitting this DMP (need an ORCID? Register here: https://orcid.org/);**
5. **Email and phone number of the person submitting this DMP;**
6. **Name of the organization for which the person submitting this DMP is working;**
7. **Email and phone number for the organization;**
8. **Link to organization or project website, if applicable; and,**
9. **Date the DMP was written.**

Name of the project; WMATA -- Transit Data Standards and Analysis Infrastructure Framework

Grant number; 69A3552441015

Name of the person submitting this DMP; ORCiD, Email and Phone

- Michael Eichler
- 0000-0002-6802-4872
- meichler@wmata.com
- 202934-0796

Name, phone, URL, project website of the organization for which the person submitting this DMP is working;

- Washington Metropolitan Area Transit Authority
- 202-673-7000
- http://wmata.com
- https://github.com/TIDES-transit/TIDES
-

Date the DMP was written.

July 30, 2024

Data Description

Please provide as much information as possible:

1. **Provide a description of the data that you will be gathering in the course of your project or data from a third party that you will re-use, if any;**
    1. **If there will be no data collected or re-used from another source, state that this is case;**
        1. **If you answered "No data" above, then you are finished and may skip the most of the steps and sections below, after you:**
            1. **Save your DMP as it exists;**
            2. **Submit it to your Grant Manager or the NTL staff for review.**
2. **Address the expected nature, scope, and scale of the data that will be collected, as best as you can at this stage;**
3. **As best as you can, 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;**
    1. **If data might be sensitive, please describe how you will protect privacy and security, if you know that now;**
    2. **You may need to update your DMP later to add more detail;**
4. **Discuss the expected value of the data over the long-term.**

This project will not gather/create original datasets nor re-use any third party data. The project will create the following:

1. Refined and updated TIDES data standards, including data spec documentation
2. Framework for writing open source software to read/write/refine data stored in standard formats

3. Source code to read/write/refine data stored in standard formats

The work product will be in text files that represent the standards, framework, and data manipulation logic.

The TIDES data standard references the general transit feed specification (GTFS), and data tables implementing TIDES will reference archived GTFS datasets.

This project's outputs will help create a marketplace for open-source software that allows transit agencies to more easily use the data generated by software/hardware systems that count passengers, track vehicle locations, and collect fares.

Data Format and Metadata Standards Employed

Please provide as much information as you can:

1. **Describe the anticipated file formats of your data and related files;**
2. **To the maximum extent practicable, your DMP should address how you will use platform-independent and non-proprietary formats to ensure maximum utility of the data in the future;**
    1. **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.**
3. **Identify the metadata standards you will use to describe the data.**
    1. **At least one metadata file should be a DCAT-US v1.1 (https://resources.data.gov/resources/dcat-us/) .JSON file, the federal standard for data search and discovery.**

Anticipated file formats:

- TIDES specification & documentation: JavaScript Object Notation (JSON), markdown (MD)
- Framework: text and graphics (PDF)
- Source code: structured query language (SQL)

These files will be stored in clear, open text readable by any text editor.

The final data package will include a DCAT-US version 1.1 (https://resources.data.gov/resources/dcat-us/) .JSON metadata file, which is the federal standard for datasets and required for compliance with the USDOT Public Access Plan.

Access Policies

In general, data from DOT-funded projects must be made publicly accessible. Exceptions to this policy are: data that contain personally identifiable information (PII) that cannot be anonymized; confidential business information; or classified information. 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 the protect privacy and confidentiality prior to archiving your data, and any additional concerns. 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. Additionally, when working with, or conducting research that includes Indigenous populations or Tribal communities, researcher will adhere to the CARE Principles for Indigenous Data Governance https://www.gida-global.org/care and make an explicit statement to that effect in this portion of the DMP.

Please provide as much information as possible:

1. **Describe any sensitive data that may be collected or used;**
2. **Describe how you will protect PII or other sensitive data, including IRB review, application of CARE Principles guidelines, or other ethical norms and practices;**
    1. **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;**
3. **Describe any access restrictions that may apply to your data;**
4. **If necessary, describe any division of responsibilities for stewarding and protecting the data among Principal Investigators or other project staff.**

1. No sensitive data will be collected or used as part of this project.
2. The project will not create or re-use any PII or sensitive data.
3. No access restrictions will apply to the project outputs.
4. All data specifications, documentation, and source code generated through this project will be stewarded by MobilityData and hosted on the TIDES Transit github: https://github.com/TIDES-transit

Re-use, Redistribution, and Derivatives Products Policies

Recipients are reminded:

1. **Data, as a collection of facts, cannot be copyrighted under US copyright law;**
2. **Projects carried out under a US DOT SMART Grants is federally funded; therefore, as stated in grant language:**
    1. **Recipients must comply with the US DOT Public Access Plan, meaning, among other requirements, project data must be shared with the public, either by the researchers or by US DOT;**
    2. **That by accepting US DOT funding through this grant, recipients have granted to US DOT a comprehensive non-exclusive, paid-up, royalty-free copyright license for all project outputs (publications, datasets, software, code, etc.). This includes all rights under copyright, including, but not limited to the rights to copy, distribute, prepare derivative works, and the right to display and/or perform a work in public; and,**
    3. **In accordance with Chapter 18 of Title 35 of the United States Code, also known as the Bayh-Dole Act, where grant recipients elect to retain title to any invention developed under this grant, US DOT retains a statutory nonexclusive, nontransferrable, irrevocable, paid-up license to practice or have practiced for or on behalf of the United States any such invention throughout the world.**

Please provide as much information as possible:

1. **Describe who will hold the intellectual property rights for the data created or used during the project;**
2. **Describe whether you will transfer those rights to a data archive, if appropriate;**
3. **Identify whether any licenses apply to the data;**
    1. **If you will be enforcing terms of use or a requirement for data citation through a license, indicate as much in your DMP;**
4. **Describe any other legal requirements that might need to be addressed.**

All intellectual property rights for the project outputs will be held by MobilityData, the international non-profit that governs transit data standards.

The TIDES standard currently has an Apache-2.0 license. Through the transition of TIDES and the projects outputs to MobilityData, other licenses may be applied.

Archiving and Preservation Plan

Please provide as much information as possible:

1. **State where you intend to archive your data and why you have chosen that particular option;**
2. **Provide a link to the repository;**
3. **You must describe the dataset that is being archived with a minimum amount of metadata that ensures its discoverability;**
    1. **Whatever archive option you choose, that archive should support the capture and provision of the US Federal Government DCAT-US Metadata Schema https://resources.data.gov/resources/dcat-us/**
4. **In addition, the archive you choose should support the creation and maintenance of persistent identifiers (e.g., DOIs, handles, etc.) and must provide for maintenance of those identifiers throughout the preservation lifecycle of the data;**
5. **Your plan should address how your archiving and preservation choices meet these requirements.**

All project outputs will be archived on the TIDES gitub site: https://github.com/TIDES-transit

This platform will provide open access to the greater TIDES software development community to maintain and update the standards, specifications, documentation, frameworks, and source code as the needs evolve.

A DCAT-US metadata file will be uploaded to the GitHub repository alongside the research outputs in accordance with the USDOT Public Access Plan. Additionally, at project close-out, all project outputs will be submitted to Zenodo which will permit project tracking via a PID.

In the event that MobilityData is no longer equipped to govern and maintain the repository of project outputs, the project will be archived via zipfile upload to *ROSA P*.

Planned Research Outputs

Software - "Open Source Data Transformation Source Code"

source code for open-source data transformation scripts

Model representation - "TIDES Data Standard Specification"

JSON and MD specification of data standard.

Data paper - "TIDES Analysis Infrastructure Framework"

Document describing shared platform for developing software modules that operate on TIDES-formatted data.

Planned research output details

Title	Type	Anticipated release date	Initial access level	Intended repository(ies)	Anticipated file size	License	Metadata standard(s)	May contain sensitive data?	May contain PII?
Open Source Data Transformation Source Code	Software	2025-12-30	Open	GitHub		Custom Data Use Agreements/Terms of Use	None specified	No	No
TIDES Data Standard Specification	Model representation	2025-12-30	Open	GitHub		Custom Data Use Agreements/Terms of Use	None specified	No	No
TIDES Analysis Infrastructure Framework	Data paper	2025-12-29	Open	GitHub		Custom Data Use Agreements/Terms of Use	None specified	No	No