

NATIONAL INSTITUTE FOR TRANSPORTATION AND COMMUNITIES (NITC)

DATA MANAGEMENT PLAN

Drafted: 02/28/2017

1 Overview

This document outlines the process that will be used to ensure the secure archiving and the rapid dissemination of all data collected by projects that are funded through the 2016 UTC National Center grant.

2 Data Products

The focus of our NITC's program is to improve the mobility of people and goods to build strong communities. Our theme is motivated by several important trends, highlighted in the USDOT's Beyond Traffic 2045 report. Urban populations are increasing, yet so is income inequality. As the second-largest household expense, transportation contributes to inequality. Demographic shifts, along with efforts to address climate change and other policy challenges, are contributing to demand for multimodal solutions. Rapid advances in technology and shared mobility have the potential to address these challenges, yet pose new ones related to implementation and equity. With constrained resources, we need better data and tools to optimize mobility.

To carry out its theme, NITC will support research in several different but interrelated research areas that focus on increasing access to opportunities, improving multi-modal planning and shared use of infrastructure, advancing innovation and smart cities, and developing data, models, and tools. Considering the diverse nature of this research, we expect to see a diverse set of research methodologies and data. For example, methods applied in these projects may include the collection of primary data in the field, web-based data collections such as surveys, secondary document review, computer simulations, and the development of new software and other technological tools. We fund projects that use a rigorous investigatory framework and, as a result, we expect that many funded projects will also use more than one method to generate data. We therefore anticipate that projects will collect qualitative and/or quantitative data that will include, for example, image data, numerical data, and source code (Table 1).

Table 1: Types of Data Expected to be Shared under DOT's Open Access Requirement. #

| Data type | Data source (examples) | Data format |
|----------------|--|---|
| Video | Video recording of automobile traffic | .mov, .mpeg, .mp4 |
| Image | Map depicting location of study sites | .tiff, .jpeg, .png |
| Numerical | Count data from automobile, pedestrian, or bike traffic, responses on surveys, measurements from air-quality surveys | .csv or .ascii |
| Text sequences | Participant responses on surveys | .pdf/a, .ascii |
| Location data | location data for buses | avl, gps |
| Source code | Development of an online data portal | js/html/css python/php/java/nodejs/sql |

Note: #List is based on past and currently funded NITC projects. Future projects may generate additional data types not included here.

3 Data Access Policy

To ensure that all projects will comply with the parameters of DOT's Open Access Plan we developed a DMP guide and template for Principal Investigators (PIs) that outline essential data parameters and relevant policies. PIs are required to submit DMPs for every funded project. **Individual project DMPs will address and adhere to the following requirements.**

Data generated by funded projects will be made available for open access, with one exception. For some projects, data vendors may legally prohibit PIs from sharing data on an open platform. In this case, PIs will be asked to provide the details of the data source and specific data set parameters. For all other data, we will set the expectation that data should be available without undue delay.

PIs may request a proprietary period for pending or planned publications.

Most data will be archived in an open source and platform-independent format (see Table 1). We will only allow the submission of non-proprietary formats unless the PI of the project can justify the need for submitting data in proprietary format. For example, in some cases it may be necessary to retain datasets in a tool-specific format for use with specific applications (e.g., VISITRO, VISSIM, PTV Viswalk, Matlab).

In their DMP, PIs will need to outline their data quality assurance process and identify the individual(s) responsible for administering its steps. They will also be required to outline the expected data product and individual(s) in charge of archiving and submitting the final data product to NITC's research administrator once the project is completed. A data product has to be accompanied by metadata that includes adequate detail to ensure that the data product, all of its components, and its context are fully described.

To streamline the data archive across projects, PIs will be required to adhere to specific naming conventions for data products. PIs will also be responsible for citing the source(s) of the data appropriately in cases where data are downloaded from publically available regional, state, or national data repositories (e.g., US Census Bureau, EPA, US National Climate Data, Office of the Treasurer-Tax Collector).

3.1 Re-Use and Re-Distribution of Data

In some rare cases, PIs will want to license the re-use or distribution of data. PIs will have the option of setting conditions on the re-use of their materials by affixing a permission and copyright license statement, including a Creative Commons License to their work.

3.2 Handling of Sensitive Information

Projects collecting sensitive information will be required to comply with university policies governing the collection and handling of sensitive information (e.g., personal identifiers), consistent with Institutional Review Board (IRB) procedures and in accordance with relevant federal and state privacy laws and regulations. If sensitive information is collected, PIs will be responsible for de-identifying data in a manner that protects privacy and confidentiality while maintaining the utility of the dataset prior to submitting data product for archiving to NITC.

4 Data Archive and Archiving Process

Data of all NITC funded projects will be archived using PDX Scholar (<http://pdxscholar.library.pdx.edu>), a document and data archive that is curated, administered, and maintained by the Portland State University (PSU) Library or equivalent at PI's home university.

PDX Scholar meets all criteria necessary to be considered in conformance with the DOT Public Access plan. For example, this archive is specifically designed to collect, preserve, and make accessible scholarly output of Portland State Faculty, students, and affiliates, retains documents and data in perpetuity, and

assigns datasets a DOI. Back-ups are currently run off-site and will be moved to Amazon S3 cloud storage in the future.

PIs will be required to submit the data product with the final version of their project report. If there are changes to the data product relative to the initial proposal in the DMP, PIs have to justify these differences at the time of submitting the data product. The Research Administrator will review all components of the data product using the following criteria:

- Adherence to project's DMP
- Naming convention of files included with the data product
- Completeness of the metadata and, if included, data dictionary

Data product will be forwarded to the PSU library and uploaded to PDX Scholar within 30 days from submitting data files.

In addition to a DOI, each data product will also be assigned a persistent web address that will be entered into NITC's project and program management system. This web address will be listed with the link to the project's final report on NITC's website. As a result, interested parties will be able to access data products through PDX Scholar or the projects' website (<http://nitc.trec.pdx.edu/research>).