
Plan Overview

A Data Management Plan created using DMPTool

Title: GUERRA_The built environment and pedestrian safety in the Philadelphia region

Creator: Erick Guerra

Affiliation: University of Pennsylvania (upenn.edu)

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

Funding opportunity number: 7/1/2017 - 1/31/2019

Template: U.S. Department of Transportation Public Access Guidance v1

Last modified: 09-26-2017

Copyright information:

The above plan creator(s) have agreed that others may use as much of the text of this plan as they would like in their own plans, and customize it as necessary. You do not need to credit the creator(s) as the source of the language used, but using any of the plan's text does not imply that the creator(s) endorse, or have any relationship to, your project or proposal

GUERRA_The built environment and pedestrian safety in the Philadelphia region

Data description

Describe the data that will be gathered in the course of the research project, including whether the data should be preserved for long-term access.

This project relies on the following data sources:

1. Geographic point and data files of reported traffic collisions, injuries, and fatalities for Philadelphia and for surrounding PA counties from 2010 to 2013. The State of Pennsylvania maintains and provided the data through the Delaware Valley Regional Planning Commission. A public data interface is available here: <https://www.dot6.state.pa.us/>
2. Publicly available US Tiger line and State road segments. Includes data on road classifications and daily traffic flow on limited sample of roads.
3. Publicly available American Community Survey Census tract data and shape files.
4. Publicly available Zip Business Patterns data.
5. Publicly available LEHD Origin-Destination Employment Statistics

Data are being processed in two different ways with a spatial coverage of the 5 PA counties in the Philadelphia MSA:

1. Aggregated to the Census tract level.
2. Associated with the collision point files and matched to street lines and intersections.

Data format and metadata standards

Describe the standards and machine-readable formats that will be used in the course of the research project.

Data will be formatted along multiple spatial dimensions as described in the data description. Metadata is provided by the public agencies that produced and provided the data. We have generated a data dictionary for our primary final databases and maintain all original databases along with the R scripts used to process the data.

Policies for access and sharing

Discuss the access policies that will apply to the data, so as to protect against the disclosure of identities, confidential business information, national security information, etc. and whether public use files may be generated from the data.

Data are all publicly available in some format. The research team has primarily shared data using Box and Dropbox. We maintain all current data and coding scripts on a shared, password-protected Box folder.

For some crash data, we have also provided an interactive online data tool: <http://crashphilly.erickguerra.net/>

This website includes older crash data that we are not using in the research project due to inconsistencies in geospatial reporting before 2010.

Policies for re-use, redistribution, derivatives

Discuss the policies for re-use, re-distribution and derivative projects.

The researchers working on the project will use the data for several projects. We will provide publicly accessible versions of final datasets to go with any peer-reviewed publications.

Plans for archiving and preservation

Outline the plans for archiving and preservation, specifying where research data will be deposited, and specify that data will be deposited at the time of initial publication of any related peer-reviewed journal article.

An archive of all data, scripts, and final datasets will be kept on Box, as well as Dropbox.