## Data Management Plan

 A critical aspect of all scientific research is the management of generated data and the organization and archiving of new and historical data in ways that enhance access and utility. Data collections represent an investment of resources and provide information for current and future scientists, engineers and policy makers. Their storage and accessibility are critically important (to allow current researchers to make conclusions and future researchers to build on existing work. The data management strategy goals for this project are to 1) ensure data quality, 2) provide appropriate access/protection to the data, 3) communicate clearly regarding the nature of the data, and 4) maintain the data in archival storage.

### Types of Data, Samples, Physical Collections, Software, Curriculum Materials, and Other Materials to be produced in the Course of the Project

This project will involve data related to: registration of vehicles in Pennsylvania; vehicle-specific parameters (e.g., number of cylinders, wheelbase, speed) from decoded VINs; results of safety and emissions inspections. Electronic files of raw data and analysis files will be stored on a secure server on campus until publication (see below), partly because their dissemination is prohibited from the state agencies from which we have acquired them.

### The Standards to be Used for Data and Metadata Format and Content

There are several file formats that will be used for this study. Excel is a standard way to organize and store some of the engineering data collected for this study (and is one of the tools that has been used in proof of concept activity). Similarly, data will be used in MATLAB®, R, and Python, among others.

All model outputs from this research will be managed as all other project metadata. Here, by metadata, we refer to the data as those that are collected over the course of the study. Engineering data will come from the literature and publicly available sources. Careful attention to file naming conventions and meta-data tagging for databases is critical to ensure replicability of results. The PIs maintain and enforce naming conventions related to model simulations within their research groups.

### Policies for Access and Sharing Including Provisions for Appropriate Protection of Privacy, Confidentiality, Security, Intellectual Property, or Other Rights or Requirements

Throughout the duration of the proposed work, the PIs will, in a timely manner, broadly communicate findings in accordance with DOT policy through journal publications, conference presentations, and seminars. After publication, the underlying data will be made available to the research community, where possible and permitted. When storing, sharing, and posting any data, we will commit to protect privacy, confidentiality, and security.

Engineering data obtained from publicly available sources will be made available to the public through a project website. These data will be made available immediately after publication of the work. If needed to facilitate use, the data will be organized into multiple manageable files. For all data, metadata in the form of explanatory files with an overview of the research project and a data dictionary will also be provided.

### Policies and Provisions for Re-use, Re-distribution, and the Production of Derivatives

Where not in violation of agreements made with third parties, the data related to project will be made available to the research community in accordance with the relevant DOT Policies and Guidelines. Where available and permissible reuse, redistribution, and derivative work will be facilitated and encouraged, with attribution and fair use.

### Plans for Archiving Data, Samples, and Other Research Products, and for Preservation of Access to Them

Data derived from this project shall be retained for at least the minimum period specified in the DOT Data Sharing and Archiving Policies and Guidelines on a Carnegie Mellon server, and interested researchers will have access to the data through coordination with the PIs. CMU Libraries have committed to host and provide long term archiving and dissemination of research products through a system called *Kilthub*, which we intend to use on this project.