## Data Management Plan

A critical aspect of all scientific research is managing generated data and organizing and archiving new and historical data to enhance access and utility. Data collection represents an investment of resources and provides information for current and future scientists, engineers, and policymakers. 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 1) tractor and trailer video and motion sensor data; 2) vehicle safety inspections and registration data; 3) real-time telematics of selected vehicles; 4) emission data; and 5) human subject interview and smart sensor/camera user study results (IRB pending). 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 by the state agencies from which we have acquired them.

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

Several file formats will be used for this project. RAW, AVCHD, ProRes, and DNxHD are video formats for storing videos with little or no compression to keep all scientific information. Excel is a standard way to organize and store some of the engineering data collected. Similarly, data will be used in MATLAB®, R, Python, and Unity, among others.

All model outputs from this research will be managed as all other project metadata. Here, by metadata, we refer to the contextual data attributes collected throughout the study. Financial and Engineering data will come from literature and publicly available sources. Careful attention to file naming conventions and meta-data tagging for databases is critical to ensure the 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 data, we will protect privacy, confidentiality, and security. Engineering data from publicly available sources will be available to the public through a project website. All data will be made available immediately after publication of the work. If needed, the data will be organized into manageable files to facilitate broad use. Metadata in the form of explanatory files with an overview of the research project and a data dictionary will also be provided for all data.

### 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 the project will be made available to the research community in accordance with the relevant DOT Policies and Guidelines. Reusing, redistribution, and derivative work will be facilitated and encouraged where available and permissible, with attribution and fair use.

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

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 hosting and providing long-term archiving and disseminating research products through a system called *Kilthub*, which we may use on this project. Otherwise, we will use *github*.