Data Management Plan

Constantine Samaras, Carnegie Mellon University

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. 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 conform to the federal policies on data management and the dissemination and sharing of research results. Electronic files of raw data and analysis files will be stored on a secure server at Carnegie Mellon University's campus as well as securely shared on the project electronic platform for scientific collaboration. This collaboratory will be used for data sharing, discussion, collaborative writing, project management, etc. The collaboratory will be built on the well known, commercially-available, password protected, and remotely backed up Box software, which the PI uses for interdisciplinary research projects at various scales.

Different types of data will be managed in the course of the project. This includes external data used as model inputs, model descriptions, and modeling results. External data comprise data collected from empirical tests in proposed research, as well as different agency sources, peer-reviewed sources, and trade literature. Meta-data include computational source codes of MATLAB, Python, R, Excel, and other models, model descriptions, parameters and assumptions. Model results data include tables of data and model outputs.

Data derived from this project shall be retained for at least 3 years on a Carnegie Mellon server, and interested researchers will have access to the data through coordination with the PI. If and when the data is removed from the project website, researchers can contact the PI to access the data. If needed to facilitate use, the data will be organized into multiple manageable files. Explanatory files with an overview of the research project and a data dictionary will also be provided. Data associated with peer-reviewed journal papers will either be included in the journal paper or supplementary information, or available from the PI.

Data formats will include formatted Excel and .CSV spreadsheets, Python, R, and MATLAB code, for the project as well as resulting tables and graphs. Other generated formats will be Word documents, journal articles as well as oral and poster presentations. Meta-data that are not published will be released after request (e-mail exchange). None of the data requires protection due to confidentiality, intellectual property or export controls. Digital data will be stored on personal computers and on the CMU Box server as a backup. The investigator will produce materials for education and outreach activities associated with automated vehicle sustainability. These are all available from the PI upon email request.

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

Where not in violation of agreements made with third parties in order to collect data from their systems, the data related to project will be made available to the research community. Where available and permissible, reuse, redistribution, and derivative work will be facilitated and encouraged, with attribution and fair use.