

F. Data Management Plan

Data generated as part of this project, including instrumentation code, modeling code, training data, model descriptions, etc. will be managed using existing University of Pennsylvania infrastructure for administering and maintaining digital research (e.g. *Penn Computing and Education Technology Services*, with automatic nightly back up of source-code and documentation repositories), at no additional cost to the project. Some of the resources and data management practices are already in place, and being utilized by the PI in his research group. We will adopt and extend these practices for short-term data collection, retention and management.

Expected Data

An outline of the data expected to be generated by the project is as follows:

1. *Instrumentation code in C and Matlab*: One of the main code artifacts generated by this project will be instrumentation routines for AV system modeling, testing and verification. This code will be used to collect data both from experimental runs to generate performance models. The code implementing this instrumentation will be maintained using version control (SVN). Milestone versions of the code will be released to the public as open source software under the GNUv3 public license.
2. *Modeling algorithms and code*: A second major code artifact of our project will be the modeling algorithms and toolchains used to build our AV scenario models. We will publish the mathematical formulations of these techniques for archival by the publisher, and will maintain implementations of these techniques (*i.e.*, code) in version control. Concurrent with the publication of these techniques, we will release code implementing them using open source licenses.
3. *Training data*: No confidential data will be recorded or stored as part of this project. All recorded Autonomous Vehicle (AV) signal data provided by Wabtec, MIT, Clemson, UVA and AutoWare is anonymized at the source before it is shared with the PI. No Institutional Review Board (IRB) approvals are required on the part of the PI.

Period of Data Retention

During the project, the data associated with individual tasks will be stored by the PI using his laboratory's local servers and other resources. At regular intervals (quarterly), all the project related data will be copied and archived using the University of Pennsylvania's digital repository.

Data Formats, Short-term storage and dissemination

All the data described above and its accompanying documentation will be incrementally made accessible to researchers and the general public as mentioned before. All the computer codes will be implemented and made accessible to the scientific community in the form live web-tools. Selected source code, associated input/output files and documentation will also be released as the project and this computational modeling technology matures.

Long-term Data Storage and Preservation of access

All data on University of Pennsylvania storage systems will be maintained and housed on their servers for the duration of the life of 5 years. The PI will coordinate efforts for collection, preservation, and centralization of access to the project in the long term.