CENTER FOR ADVANCED TRANSPORTATION MOBILITY

Data Management Plan
February 28, 2017

Lead Institution
North Carolina Agricultural &Technical State University
Greensboro, North Carolina 27411
(Minority Institution – HBCU)

Consortium Members:

Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061

Embry-Riddle Aeronautical University, Daytona Beach, Florida 32114

University of the District of Columbia, Washington, D.C. 20008
(Minority Institution – HBCU)

Type of UTC:
Tier 1 Center Proposal

FAST Act Research Priority Area:
Improving Mobility of People and Goods

Purpose of Research

The Center for Advanced Transportation Mobility (CATM) research activities will focus on improving how people and goods are transported throughout our nation, with special emphasis on disadvantaged populations. The primary research topics include last mile connectivity, mobility assistance, and emergency management optimization.

Data Description

Types of data and methods for creating the data

- Focus groups, surveys, and interviews will yield demographic data, self-reported subjective responses, and audio data. Focus group data will be collected with a researcher guiding discussion among multiple participants. Survey data will be collected from participants using either paper or electronic surveys. Video and parametric data will be collected using commercially available as well as customized data recorders.
- Expert consultation will contain open-ended responses stored as text.
- Traffic data will be retrieved from available transportation data bases and written as text files.
- Human behavior data will be collected through literature review, historical data, and observation.
- Computer simulations will result in quantitative models and algorithms that will be implemented using software tools such as Matlab and AnyLogic, or programming languages, such as Java and FORTRAN.

Period of time data will be collected and frequency of update

- Data will be collected as needed throughout the duration of each project or until enough data has been collected to achieve the desired statistical power.
- Focus groups, survey administration, interviews, and expert consultations will typically take place over a 6-month period of time. Once the data collection phases of each project are complete, no further updates for the data collected within those phases is anticipated.
- New traffic data will be uploaded every 6 months during data collection phases.
- Source code of quantitative models and algorithms developed will be uploaded to the CATM website. Model or algorithm updates will be attached to an upcoming quarterly report, and relevant source code will be updated correspondingly on the CATM website.

Potential users of the data

- Qualified researchers in the CATM research consortium; public/private transportation; emergency/disaster planning; and emergency/disaster response researchers and agencies; as well as members of the greater research community are expected to utilize the data collected.

Potential long-term value of the data

- Data collected and/or produced by CATM projects can be used by outside parties in support of transportation planning, emergency/disaster planning, and policy making activities.
- User feedback obtained through the projects will be valuable in making further changes to the platforms developed to benefit the mobility of vulnerable road users.
- The data will provide insight into the preferences of vulnerable road users, the ability of new technologies to address their needs, and how technologies can improve their mobility.
- Data on user needs could serve to benefit a wide variety of research projects.
- The data is expected to provide transportation planners with better analytic information which can be used to schedule their services more efficiently and effectively.
- Data collected and/or produced from these projects can support researchers within the consortium to continue to expand their transportation-related research and provide the institutions with more collaborative research opportunities in transportation.

- The data obtained from the projects can be used in engineering education, such as developing course projects, and providing students data analysis exercises or research opportunities.

Party responsible for managing the data

- The PI and co-PIs will be responsible for managing the data for their individual projects until the project is completed. Upon completion, copies of the final datasets will be submitted to the Center for archiving on NC A&T State University's institutional repository.
- The confidentiality of all participant data will be protected and all data will be stored on secure servers for analysis.

How adherence to the data management plan will be checked

- Adherence to this plan will be checked monthly and at least sixty days prior to the end of each grant year by a CATM staff member. Adherence checks will include review of the content, number of datasets released, availability for each dataset in subsettable/preservation friendly data formats; availability of public documentation; and validity of data citation.

Standards Used

- Transcripts will be created from the audio files and stored as text, and survey responses and moderator notes will also be stored as text and/or spreadsheet files.
- Expert consultation will contain open-ended responses stored as text.
- Parametric data (i.e., time series of vehicle kinematics and/or human behavioral data) and traffic data will be converted from the formats produced by the data collection systems into either text or .csv files that can be read using Excel.
- Human behavior data analysis results will be stored as .sav and .spo file formats. *Statistical Package for the Social Sciences* (SPSS) software is required to view these files.
- Source code will be stored in the formats of the computer language or software used to implement the source code. Software used to develop the source code is required to view it.
- VTTI video and parametric data will be collected in VTTI's previously developed proprietary format to minimize development costs on this project. The data will be exported into standard formats (e.g., CSV, spreadsheet, or relational database formats) for viewing outside VTTI.
- Simulation results and documents for models and algorithms will be preserved in computer data files as word-processed documents, spreadsheets, or presentations using either Microsoft Office formats or portable document format (aka "PDF"). *Microsoft Office* or *Adobe Acrobat Reader* will be required to view the files.
- Image data from simulations will be preserved in image files (e.g., .png and .jpg). In cases where audio recording is used, for example, stakeholder interviews, data will be preserved in audio files (e.g., .mp3). *Windows Photo Viewer* and *Windows Media Player* can be used to view the files.
- The codebook containing the data definitions, variable definitions, and any necessary metadata required to interpret the data analyses and parametric data will be created using plain text and stored along with the data files.

Access Policies

- IRB protocols will be submitted to each institution involved in projects utilizing human subjects. To ensure that all personnel involved in collecting the data understand participants' rights, personnel will be required to complete and pass a Basic Course for Humans Subjects Research.
- Informed consent forms will be required from human subjects involved in CATM research projects. These forms will inform participants that their data will be coded and/or aggregated prior to dissemination such that no personally identifiable information will be shared outside of the research team.

- Only the PIs for the individual projects will have long-term access to or ownership of the sensitive raw data associated with their project, which will be saved in password protected files on secure servers.
- De-identified data associated with each project's final report will be shared.
- CATM is expected to grant public access to the final versions of project data.
- The raw data from CATM projects will not be publically available.

Re-Use, Redistribution, and Derivative Products Policies
- Shared data will not include any personally-identifiable information, confidential business information, or classified information.
- De-identified data will be shared with the final report for each project
- The final report will be posted on the CATM website.
- In addition to the final report, journal articles written during the funding period or as a direct or immediate result of analyses for the report will be archived on NCA&T's institutional repository alongside the data (see Archiving section below).
- The de-identified data will be organized by project and made accessible through NCA&T's institutional repository.
- Outside parties interested in the raw data will need to contact the project's PI(s) to request it. Such requests will be considered on a case-by-case basis.
- Publication copyrights will be held by the USDOT and reproductions must be authorized.
- No copyright protected data is expected to be collected; however, any applicable copyright licenses that should arise will follow the terms of the license.
- There are no known reasons that the sharing, re-use, and redistribution of that data resulting from CATM projects might be prohibited.

Archiving and Preservation Plans
- Final data, associated definitions and documentation, relevant reports and any other project output will be archived in North Carolina A&T State University's institutional repository. All project output will be hosted indefinitely with a minimum 10-year commitment after the project ends, plus migration support in case of need for transition to a future archiving system. The URL for the institutional repository is http://cdm15116.contentdm.oclc.org/cdm/. The repository platform, CONTENTdm Digital Collection Management Software, is managed by the F.D. Bluford Library and hosted by Online Computer Library Center (OCLC). OCLC supports long-term data management and ensures the security and integrity of all hosted data. OCLC has an ISO-certified data center that is monitored 24/7 by system operators, security guards, and cameras. A certified team reviews processes for applications, systems, and procedures on a continuous basis. Multiple copies of data in the CONTENTdm preservation archive are geographically distributed in separate, secure facilities that are protected from both minor and major outages and disasters. All files stored in CONTENTdm are backed up to tape, cloned to a secure tape archive, and transported offsite.
- Data, associated definitions and documentation, relevant reports and any other project output deposited in the repository will be openly available and discoverable using the Dublin Core metadata schema. All files will be assigned a stable, persistent URL for long-term access.
- ORCIDs (unique researcher IDs) for all project investigators and contributors will be reported for CATM projects. This ID will be associated with all resulting publications, datasets, and other project outputs throughout its existence.