



## **Data Management Plan**

### **September 30, 2023**

**Lead Institution:**

North Carolina Agricultural & Technical State University  
Greensboro, North Carolina 27411

**Consortium Members:**

Clemson University, Clemson, South Carolina 29556  
Florida Atlantic University, Boca Raton, Florida 33431  
University of Alabama, Tuscaloosa, Alabama 35487  
University of Kentucky, Lexington, Kentucky 40506  
University of Tennessee, Knoxville, Tennessee 37996  
University of Georgia, Athens, Georgia 30602

**Type of UTC:** Regional Center Proposal  
**Federal Region:** Region 4

**Statutory Research Priority Area:**  
Improving Mobility of People and Goods

Effective Date: June 1, 2023

Grant Period: June 2023 – May 2029

Grant Number: 69A3552348304

**Purpose of Research**

The Center for Regional and Rural Connected Communities (CR<sup>2</sup>C<sup>2</sup>) research activities will focus on improving how people and goods are transported throughout our nation, with special emphasis on rural and underserved communities in Region 4. The primary research topics include planning, developing, and adopting connected and coordinated multimodal technological solutions to enable and improve equitable, efficient, and effective mobility services.

This document serves as the center-wide data management plan and all projects' Principle Investigators (PIs) are required to submit a detailed data management plan for their specific project that complies with this center-wide plan and the U.S. DOT Public Access Plan (<https://www.transportation.gov/mission/open/official-dot-public-access-plan-v11>) .

**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, XML, .csv, .json, or file formats that are effective in storing traffic data.
- 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; traffic simulation tools such as SUMO or VTD; or programming languages such as Java, Python, or C++.
- Experimental results will include sensor data such as LiDAR, camera, radar, and thermal data and performance data from the systems under test using industry standard formats such as shapefiles for GPS data.
- Presentation materials such as videos and slides in PowerPoint.
- Research articles in Word or PDF.

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 are anticipated.
- New traffic data will be uploaded during data collection phases of projects as needed.
- Source code of quantitative models and algorithms developed will be uploaded to the CR<sup>2</sup>C<sup>2</sup> website or GitHub. Model or algorithm updates will be attached to an upcoming quarterly report, and relevant source code will be updated correspondingly on the CR<sup>2</sup>C<sup>2</sup> website or GitHub.
- Sensor data such LiDAR, camera, and thermal data will be collected as needed.
- Presentation data and research articles will be generated as the research projects progress.

Potential users of the data

- Qualified researchers in the CR<sup>2</sup>C<sup>2</sup> research consortium; public/private transportation

## **Center for Regional and Rural Connected Communities (CR<sup>2</sup>C<sup>2</sup>)**

professionals; automotive industry, Original Equipment Manufacturers (OEMs), and transportation service providers; 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 CR<sup>2</sup>C<sup>2</sup> projects can be used by outside parties in support of transportation planning, engineering, technology development and adoption, 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 people and goods in rural and disadvantaged communities.
- The data will provide insight into the preferences of residents of rural and underserved communities, 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 and engineers with better analytic information which can be used to schedule their services more efficiently and effectively.
- Data collected and/or produced from CR<sup>2</sup>C<sup>2</sup> 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 contribute to Open Science in the transportation field and can be used in education initiatives of transportation-related disciplines, such as developing course projects and providing data analysis exercises or research opportunities for students.

### 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 as explained under the “Archiving and Preservation Plans” section of this document.
- 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

- The DMP is a living-document that should be reviewed and updated as necessary throughout a grant’s life-cycle.
- Adherence to this plan will be checked quarterly and at least sixty days prior to the end of each grant year by a CR<sup>2</sup>C<sup>2</sup> staff member. Adherence checks will include review of the content, number of datasets released, availability for each dataset in subsetting/preservation-friendly data formats; availability of public documentation; validity of data citation, and review of the project DMP.

### **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. Files stored as Microsoft Excel sheets will also be stored as csv files for accessibility reasons.
- Expert consultation will contain open-ended responses stored as text.
- Regarding circumstances of open access file formats vs. proprietary formats, where possible, files may

be converted to an open access format for public access, and both types released to the intended repository. If proprietary software is used, especially in the case that it cannot be transformed into an open access format, there will be information such as software version included so a user will know what they need to do in order to view the file.

- 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.
- Simulation results for models and algorithms will be preserved in computer data files as word-process 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.
- Research articles and presentation materials and documents will be preserved as PDF for accessibility, with optional Word or PowerPoint associated files.
- Developed algorithms and versioning will be archived using GitHub.

#### **Access Policies**

- In general, data from research projects will be made publicly accessible as explained under the *Archiving and Preservation Plans* Section. Exceptions to this policy are data that contain personally identifiable information, confidential business information, or classified information as well as intellectual property (IP).
- 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 Human Subjects Research.
- Informed consent forms will be required from human subjects involved in CR<sup>2</sup>C<sup>2</sup> 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. In the case of the departure or replacement of a PI, a designee – by one of the Associate Directors of the Center – will take over the security and accessibility of original raw data and passwords.
- De-identified data associated with each project's final report will be shared.
- CR<sup>2</sup>C<sup>2</sup> is expected to grant public access to the final versions of project data.
- The raw data from CR<sup>2</sup>C<sup>2</sup> projects will not be publicly available.

**Re-Use, Redistribution, and Derivative Products Policies**

- NCA&T State University or the home institution of the CR<sup>2</sup>C<sup>2</sup> researchers will hold the IP and copyright for data and other materials created by the project. The USDOT hereby reserves a royalty-free, nonexclusive and irrevocable license to reproduce, publish, or otherwise use and to authorize others to use the work for government purposes. CR<sup>2</sup>C<sup>2</sup> researchers will be required to release the data in an open license for reuse, redistribution and derivative products which will be based upon the open licenses and provided by the data archive.
- 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 CR<sup>2</sup>C<sup>2</sup> website, in ROSA-P (<https://rosap.ntl.bts.gov>), and in our institutional repository (<https://digital.library.ncat.edu/cr2c2/>).
- 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 for which the sharing, re-use, and redistribution of data resulting from CR<sup>2</sup>C<sup>2</sup> projects might be prohibited.
- The author or NCA&T State University or the home institution of the CR<sup>2</sup>C<sup>2</sup> researchers may copyright any books, publications, or other copyrightable materials developed under CR<sup>2</sup>C<sup>2</sup>, though USDOT has a royalty-free, nonexclusive and irrevocable license to reproduce, publish, or otherwise use and to authorize others to use the work for government purposes.

**Archiving and Preservation Plans**

- Final, cleaned, and verified research data, associated metadata, definitions and documentation, relevant reports and any other project output will be archived as forward compatible formats (e.g., .csv, .txt or .xml files) in ROSA-P (<https://rosap.ntl.bts.gov>), North Carolina A&T State University's institutional repository (<https://digital.library.ncat.edu>), GitHub (<https://github.com/>), and/or any of data repositories conformant with the DOT Public ACCESS Plan (<https://doi.org/10.21949/1520566>).
- In case of using Aggie Digital Collections, the repository platform, Digital Commons, is managed by the F.D. Bluford Library and hosted by bePress. bePress supports long-term data management and ensures the security and integrity of all hosted data. bePress' production servers are housed in a highly-secure facility with multiple connections and backup generators or in redundant Amazon Web Services availability zones. Files uploaded to Digital Commons are stored in a redundant storage cluster as well as backed up to Amazon Glacier that specializes in data archiving and backup. bePress' comprehensive approach ensures the preservation and availability of data for future use. Data, associated definitions and documentation, relevant reports and any other project output

## **Center for Regional and Rural Connected Communities (CR<sup>2</sup>C<sup>2</sup>)**

deposited in the repository will be openly available and discoverable using the Dublin Core metadata schema in addition to the DCAT-US schema (<https://resources.data.gov/resources/dcat-us/>) as required government standard for metadata concerning datasets. . All files will be assigned a stable, persistent URL for long-term access, as well as a persistent identifier (i.e., DOI, handle, etc.). Further details about bePress' preservation policies are found here [https://www.bepress.com/reference\\_guide\\_dc/safeguarding-content-digital-commons/](https://www.bepress.com/reference_guide_dc/safeguarding-content-digital-commons/).

- 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 <https://digital.library.ncat.edu/cr2c2/>
- ORCIDs (unique researcher IDs) for all project investigators and contributors will be reported for CR<sup>2</sup>C<sup>2</sup> projects. This ID will be associated with all resulting publications, datasets, and other project outputs throughout its existence.

### **DMP Change Log**

- 2023-09-30: Original document created
- 2024-02-01: Updated to include USDOT comments