# TraCR Data Management Plan

**Date Effective: June 1, 2023**
**Grant Period: June 1, 2023 – May 31, 2029**
**Grant Number: 69A3552344812**

**UTC Title**: National Center for "Transportation Cybersecurity and Resiliency" or TraCR

**UTC Director**: Dr. Ronnie Chowdhury (Clemson University)

**Primary Contact:** Dr. Ronnie Chowdhury (Clemson University); E-mail: mac@clemson.edu; Address: 216 Lowry Hall, Clemson, SC 29634

**UTC Associate Directors**: Dr. M Hadi Amini (Florida International University), Dr. Alvaro Cardenas (the University of California at Santa Cruz), Dr. Gurcan Comert (Benedict College), Dr. Mansoureh Jeihani (Morgan State University), Dr. Steven Jones (the University of Alabama at Tuscaloosa), Dr. Judith Mwakalonge (South Carolina State University), Dr. Bhavani Thuraisingham (the University of Texas at Dallas) and Dr. Satish Ukkusuri (Purdue University).

The National Center for Transportation Cybersecurity and Resiliency (TraCR) focuses on building an ironclad defense for the nation's transportation systems against cyberattacks by pioneering advanced cybersecurity strategies and solutions to defend our infrastructure and society. TraCR's research operates at the cutting edge, addressing basic science and implementable technologies, procedures, and policies that will be developed, tested, disseminated, and deployed thanks to our technology transfer and policy-focused efforts. TraCR's consortium members are exploring the multifaceted vulnerabilities that remain unaddressed throughout the nation's transportation cyber-physical-social systems. TraCR continuously monitors the fast-moving world of transportation cybersecurity challenges and threats as they appear across different modes, geographies, and applications. To address shortcomings in these systems, TraCR's research thrusts include security and resiliency, user and data privacy, society and environment, and evolving quantum computing threats and opportunities. TraCR's research addresses basic science as well as the creation of cutting-edge software and hardware cybersecurity technologies.

## Types of Data

Data produced from the projects undertaken by TraCR is expected to include, but is not limited to, websites, publications, research posters, computer codes, and archival-quality data acquired and curated from various texts and other sources. Text documents, including technical papers, theses, posters, and software documentation, would include figures, tables, and graphs; computer codes would include source, test routines, data, building, installation instructions, and ancillary scripts.

For TraCR projects, datasets generated from experiments or obtained from existing sources may fall into, but not limited to, the following categories:

- ***Cybersecurity Event Data:*** These data include cybersecurity events and related information collected from different sources.
- ***Connected and Autonomous Vehicle/Sensor Data***: This category includes data generated from sensors and devices embedded in connected and autonomous vehicles. Connected and autonomous vehicles may include, but are not limited to, passenger vehicles, transit vehicles, and heavy-duty freight trucks.
- ***Transportation Infrastructure Data***: The TraCR team will conduct real-world as well as simulation experiments to collect data related to various transportation infrastructures, such as traffic signal controllers and vehicle-to-everything edge devices.
- ***Human Factor Related Data***: Human performance/behavior data collected from human subject studies may be used. Examples include human factor-related data collected from field experiments and driving simulator studies.
- ***Social Media Data:*** Data from social media applications may be used by the center.
- ***Road Network Data and Simulation Data:*** The cybersecurity algorithms will be tested using real-world transportation network data. Additionally, different computer-based simulation platforms will provide data that could be used for evaluation and validation of a few of TraCR's research projects.
- ***News and Weather Data:*** News and weather data may be used by the center.

**Data Formats and Standards**

Since various tools and techniques will be involved in generating and collecting large amounts of data, metadata (e.g., data headers from simulation and real-world testing data) will be made available in their original formats for reusability and originality. Detailed data collection methods and procedures will be documented corresponding to each metadata set available in at least one of the following open file formats, plain text (.TXT), PDF, HTML, MS Word, or LaTeX. When commercially available packages are used, the corresponding files, packages, or libraries will be based on industry standards that are not expected to be obsolete soon. For any proprietary formats, the corresponding software, package, libraries, versioning, etc., or related information will be provided so that a user will know how to access or view the file. The aggregated data will be in plain text format and may be compressed for space-saving purposes. Data dictionaries will be provided for all collected datasets, and standards used for data and metadata format and content (in absence of existing standards or standards deemed inadequate) will be documented with proposed solutions and remedies. Metadata like data dictionaries will be stored on both institutional data servers and the USDOT Research Data Exchange (RDE) platform following the DCAT-US Schema v1.1 (Project Open Data Metadata Schema).

**Roles and Responsibilities**

The Center Director and the Associate Directors for the partner institutions will be responsible for making their data publicly available for access. The Center Director will be responsible for the lead institution, Clemson University. TraCR's Senior Program Manager, Megha Patel (E-mail: megha@clemson.edu), will oversee the implementation of this data management plan.

TraCR researchers are responsible for submitting a project Data Management Plan with research funding proposals. For guidance, see US DOT's "Creating Data Management Plans" page at https://doi.org/10.21949/1520571 .

**Policies for Access and Sharing and Provisions for Appropriate Protection/Privacy**

All actions of data access and sharing in TraCR's research will conform to Institutional Review Board (IRB) requirements at each participating institution. All institutions of TraCR will obtain IRB approval before utilizing any human subject data. All private identifiable information regarding human subjects will be strictly removed from the data for publication and data-sharing purposes to protect personal privacy; the participants will be de-identified such that the data cannot be traced back to the participants. All use of the data will conform to the purposes documented in the original informed consent. Further, when working with, or conducting research that includes Indigenous populations or Tribal communities, TraCR researchers will adhere to the CARE Principles for Indigenous Data Governance https://www.gida-global.org/care.

TraCR researchers associated with each project are responsible for articulating the measures to be taken to ensure the confidentiality and privacy of the data. These researchers are also responsible for mentioning any concerns, such as an embargo period for publishing the data.

All data necessary to reproduce research activities will be made available to the community. If the data sizes are small, they will be hosted directly via TraCR's website and other public resources such as the USDOT Repository & Open Science Access Portal (ROSA P https://rosap.ntl.bts.gov/ ). If the data are large, metadata and samples will be hosted on the above-mentioned avenues, while the full dataset will be provided via an appropriate repository. The availability of raw data will be determined on a case-by-case basis. All raw data generated from public sources (e.g., public traffic infrastructure, weather stations, news) and federally funded resources (e.g., testbeds) will be made available. Raw data provided to TraCR via research agreements with private entities will not be made available. However, data collection and query procedures will be provided so that other institutions with access to these resources can reproduce the research results.

Access to, and sharing and distribution of code, data, and documentation will be via websites and publishing in conferences, journals, and monographs. The TraCR Director and the Associate Directors of the partner institutions intend to publish and present the research results at relevant workshops, conferences, etc. Students will be expected to participate in research and to present results. Theses and dissertations from TraCR projects will be maintained and made available by the library at each partner institution.

Each partner institution will be responsible for following the TraCR data management plan. Each partner institution will be responsible for storing, securing, backing up, and sharing data related to the research of its researchers. The original author of each document or code and the corresponding partner institution

will retain intellectual property ownership.

**Policies and Provisions for Re-Use, Re-Distribution**

Data from research conducted within TraCR will be significant for studies involving threat modeling, detection and mitigation, laboratory and field experiments, and fundamental and applied studies. The data will be made accessible following the protocols of the underlying data source. In many cases, aggregate datasets will be made available to the research community. Scholars are welcome to use the available data for their independent research in relevant areas with citations to the originator of the data sets.

For each TraCR project, if researchers utilize any third-party data, they are responsible for citing the appropriate source with license information, if necessary. The researchers are expected to address the following in their project data management plan:

- Contact information of the personnel responsible for managing the data.
- Contact information for the personnel who will hold intellectual property rights to the data.
- Copyright information for third-party data.
- Information related to licensing for re-use and re-distribution, and derivative products.

**Plans for Archiving and Preservation of Access and Data Integrity**

For each project, the principal investigator will be responsible for storing, securing, and backing up the data related to the corresponding project. The generated data, project reports, and other research products (e.g., published papers, software, original proposals, quarterly project reports, and manuals) will be stored on data servers located in the involved partner institutions' campus facilities. Persistent identifiers such as digital object identifiers (DOIs) will be included where applicable. The TraCR team will make any non-sensitive data publicly accessible or accessible under approval via multiple resources, such as data repositories of the partner institutions, open data hub or storage. Data related to ongoing analysis or sensitive data will not be shared except in special circumstances, such as research collaborations. The lead university (Clemson University) maintains a dedicated Information Technology (IT) department with the capability to provide regular data backup services. This will be utilized for preserving the data's long-term security and reliability. The information on Clemson University's and other partner institutions' data servers will be available via the RDE. Metadata like data dictionaries will be stored on both institutional data servers and the RDE platform following the DCAT-US Schema v1.1 (Project Open Data Metadata Schema).

Change Log:

***October 10, 2023***

- Added information for the primary contact.
- Added information related to TraCR and its research focus.
- Added information related to TraCR's Senior Program Manager.

***October 17, 2023***

- Added information related to open file and proprietary formats.

- Added information related to metadata file format.
- Added information related to data repositories.
- Added information related to persistent identifiers.