USDOT Tier-1 University Transportation Center (UTC) Transportation Cybersecurity Center for Advancing Research and Education (CYBER-CARE)

Data Management Plan

Grant number: 69A3552348332
Grant period: June 1, 2023 – May 31, 2029
Lead institution: University of Houston
Partner institutions: Embry-Riddle Aeronautical University, Rice University, Texas A&M University -Corpus Christi, University of Cincinnati, University of Hawai'i at Manoa
Effective date: Oct. 1, 2023
Contact person: Dr. Yongxin Liu, Embry - Riddle Aeronautical University, yongxin.liu@erau.edu and Dr. Yunpeng (Jack) Zhang, University of Houston, yzhan226@central.uh.edu
Creators: Dr. Yongxin Liu, Ph.D., Embry - Riddle Aeronautical University and Dr. Arlei Silva, Ph.D., Rice University

Table of Contents

Disclaimer

This Data Management Plan (DMP) has been written to comply with U.S. Department of Transportation funding requirements in the "Plan to Increase Public Access to the Results of Federally Funded Scientific Research" <<https://doi.org/10.21949/1520559>>. All Cyber-CARE principal investigators (PIs) funded by USDOT are expected to follow the guidance and rules laid out in this DMP. Detailed instructions will be provided upon the receipt of a research award, and the PI will be required to acknowledge compliance with the DMP requirements.

USDOT CYBER-CARE UTC

The Transportation Cybersecurity Center for Advanced Research and Education (CYBER-CARE) is a US Department of Transportation (USDOT) Tier-1 University Transportation Center (UTC) funded in 2023. CYBER-CARE primarily focuses on the USDOT statutory research priority area of "Reducing Transportation Cybersecurity Risks." CYBER-CARE aims to establish a fundamental knowledge basis and explore advanced theory to mitigate the impacts of large-scale cyberattacks on transportation infrastructure and connected and automated vehicle (CAV) systems.

CYBER-CARE's mission is to promote interdisciplinary research and education across both transportation and cybersecurity domains. The research projects at CYBER-CARE will develop conceptual frameworks, construct comprehensive datasets, explore novel analytical approaches, support the implementation of public policies and infrastructure investments, and build a high-quality industry workforce through education. All CYBER-CARE research projects can be organized into four thrusts: CAV cybersecurity, AI applications and transportation data security, advanced traffic management system (ATMS) cybersecurity, and next-generation transportation cybersecurity systems. CYBER-CARE will accelerate industry collaborations, foster new technologies, and provide professionals with the skills and opportunities needed to become successful leaders in their fields. Notably, as CYBER-CARE will prioritize engagement with underrepresented minorities, these communities stand to benefit from professional development training in transportation cybersecurity.

CARE Principles for Indigenous Data Governance: The current movement toward open data and open science does not fully engage with Indigenous Peoples rights and interests. Existing principles within the open data movement (e.g., FAIR: findable, accessible, interoperable, reusable) primarily focus on characteristics of data that will facilitate increased data sharing among entities while ignoring power differentials and historical contexts. The emphasis on greater data sharing alone creates a tension for Indigenous Peoples who are also asserting greater control over the application and use of Indigenous data and Indigenous Knowledge for collective benefit. This includes the right to create value from Indigenous data in ways that are grounded in Indigenous worldviews and realize opportunities within the knowledge economy. The CARE Principles for Indigenous Data Governance are people and purpose-oriented, reflecting the crucial role of data in advancing Indigenous innovation and self-determination. These principles complement the existing FAIR principles encouraging open and other data movements to consider both people and purpose in their advocacy and pursuits. CYBER-CARE will follow the full CARE Principles on https://www.gida-global.org/care.

Data Management Plan

The purpose of this Data Management Plan (DMP) is to assist Center-funded faculty, staff, students, and partners in producing research outputs in accordance with sponsor (U.S. Department of Transportation) and University Transportation Center (UTC) program requirements. In coordinating research, education, and workforce training, the Center aims to ensure that data of all types will be managed and organized for security, consistency, and public dissemination as required. This document details the general requirements for all activities funded by the Center, whether it's conducting research, educating students, training professionals, or providing outreach to connect industry, government, and academia.

CYBER-CARE researchers will follow the guidelines and policies in this Center Data Management Plan. For each individual research project or effort, researchers will also create a narrative Project DMP, based on the U.S. DOT guidelines at <https://ntl.bts.gov/ntl/public-access/creating-data-management-plans-extramural-research>. Researchers may reference or quote from the CYBER-CARE Center DMP as appropriate but should be sure to tailor their project DMPs to the unique qualities of their research and call out where their DMP differs from the guidelines set forth in this Center DMP. Individual project DMPs are meant to serve as living knowledge management tools and should be reviewed and updated regularly throughout the project's life cycle: at minimum, each time there is a significant change in the research project, the data collected, or in project personnel. Researchers wanting more guidance on the U.S. DOT Public Access Plan can find guidance at https://ntl.bts.gov/public-access

Data description.

As part of compliance with the CYBER-CARE Data Management Plan (DMP), PIs will write data descriptions for funded projects. These should distinguish between newly collected data and data being re-used from other projects as well as the actual observations and generated data to be submitted to a data repository. The Data Description section should indicate who on the research team is responsible for managing and securing project data, as well as which team members are assigned which data access levels. As general guidance, CYBER-CARE PIs may also consider addressing the following questions in the related section of project specific DMP:

- Name the data, data collection project, or data producing program.
- Describe the purpose of the research.
- Describe the data that will be generated in terms of nature and scale (e.g., numerical data, image data, text sequences, video, audio, database, modeling data, source code, etc.).
- Describe methods for creating the data (e.g., simulated; observed; experimental; software; physical collections; sensors; satellite; enforcement activities; researcher-generated databases, tables, and/or spreadsheets; instrument generated digital data output such as images and video; etc).
- Discuss the period of time data will be collected and frequency of update.
- If using existing data, describe the relationship between the data you are collecting and existing data.
- List potential users of the data.
- Discuss the potential value of the data have over the long-term for not only the project team's institution, but also for the public.

- If the project team request permission not to make data publicly accessible, explain rationale for lack of public access.
- Indicate the party responsible for managing the data.
- Describe how the project team will check for adherence to this data management plan.

Data format and metadata standards

The CYBER-CARE researchers should use the metadata scheme DCAT-US (https://resources.data.gov/resources/dcat-us/) which is a government standard for metadata concerning datasets. Data from CYBER-CARE projects recommend non-proprietary formats, such as txt, csv, mp3, dat, JPEG, etc. For other data generated during research, a README text documents should be attached to illustrate the information fields and encoding paradigms for better reuse. Reports, publications, and presentations will be produced in .pdf format. If necessary for accessing and reusing the data, any programming code developed by a project will be archived alongside the data. As general guidance, CYBER-CARE PIs may also consider addressing the following questions in the related section of project specific DMP:

- List in what format(s) the data will be collected. Indicate if they are open or proprietary.
- If proprietary data formats are used, discuss the rationale for using those standards and formats.
- Describe how versions of data be signified and/or controlled.
- If the file format(s) is (are) not standard to the specific field, describe and document the alternative way of using.
- List what documentation will be created in order to make the data understandable by other researchers.
- Indicate what metadata schema are applied to describe the data. If the metadata schema is not one standard for your field, discuss your rationale for using that scheme.
- Describe how will the metadata be managed and stored.
- Indicate what tools or software is required to read or view the data and version.
- Describe data quality control measures.

Access Policies

Every project that includes human subjects will comply with the Institutional Review Board (IRB) requirements of the Principal Investigators' institutions. It is mandatory for these projects to possess a protocol approved by the IRB, which ensures the informed consent of participants and safeguards their privacy and confidentiality. Individual identifiers such as names, residential addresses, geo-coordinates of residences, and email addresses will be redacted before any data sharing occurs. The Principal Investigators are accountable for obtaining IRB approval and complying with IRB and other data sharing stipulations. Reporting of IRB approvals and compliance with data sharing requirements must be included in the project proposals and progress reports.

If CYBER-CARE researchers encounter any sensitive information during the scenario modeling, we will comply with the following rules: **Identity Protection:** Any data that could potentially disclose identities of individuals or organizations will be anonymized or pseudonymized to safeguard privacy. **Confidential Information Removal:** The project will respect confidential business information and national security concerns by avoiding disclosure of sensitive data. **Data Access Restriction:** Access to the data will be granted based on a controlled and restricted access policy, ensuring that only authorized researchers can access and use the data for research purposes. **Public Use Assessment:** Limited public use files may be generated from the data, but no sensitive and identifiable information will get propagated or disclosed. As general guidance, CYBER-CARE PIs may also consider addressing the following questions in the related section of project specific DMP:

- Describe what data will be publicly shared, how data files will be shared, and how others will access them.
- Indicate whether the data contain private or confidential information. If so:
  - Discuss how will the project team guard against disclosure of identities and/or confidential business information.
  - List what processes project team will follow to provide informed consent to participants.
  - State the party responsible for protecting the data.
- Describe what, if any, privacy, ethical, or confidentiality concerns are raised due to data sharing.
  - If applicable, describe how deidentification will be performed before sharing:
  - Identify what restrictions on access and use will be placed on the data.
  - Discuss additional steps, if any, will be used to protect privacy and confidentiality.

Projects that utilize proprietary data, whether from commercial or public entities, will adhere to all conditions and stipulations related to the data's usage. In instances where the source organization restricts public data sharing, the project will receive an exemption from data sharing obligations.

Policies for Re-use, Redistribution, Derivatives

The USDOT also 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.

Intellectual property rights of the projects will be retained by the Principal Investigators and/or their respective institutions. Nonetheless, once data is transferred to the archive and enters the public domain, any applicable copyrights will be duly identified. Projects that employ proprietary data, whether derived from commercial or public entities, will adhere to every condition and requirement stipulated for the data's usage.

Materials produced during the project will be distributed in line with the policies of the University/Participating institutions and USDOT. As general guidance, CYBER-CARE PIs may also consider addressing the following questions in the related section of project specific DMP:

- Name who has the right to manage the data.
- Indicate who holds the intellectual property rights to the data.
- List any copyrights to the data. If so, indicate who owns them.
- Discuss any rights be transferred to a data archive.
- Describe how the data will be licensed for reuse, redistribution, and derivative products. A list of Data Repositories Conformant with the DOT Public Access Plan can be found at< https://ntl.bts.gov/ntl/public-access/data-repositories-conformant-dot-public-access-plan>.

After the primary findings from the final research data set have been accepted for publication, the research data substantiating, supporting, and validating these findings will be made accessible within 60 days on CYBER-CARE center website.

Plans for Archiving and Preservation

CYBER-CARE recommends preserving useful data using dual-backup strategy, local backup, and web archiving. For local backup, CYBER-CARE recommends PIs to backup their workstations using local or institutional storage devices bi-weekly. CYBER-CARE requires PIs in each site create their local long-term data storage and backup plan compatible with center DMP.

For web archiving, CYBER-CARE recommends PIs to upload and publish all data and code to Data Repositories Conformant with the DOT Public Access Plan <https://ntl.bts.gov/ntl/public-access/data-repositories-conformant-dot-public-access-plan>. Digital objects that are being archived must be described with a minimum number of metadata that ensures its discoverability. Whatever archive option CYBER-CARE PI choose, that archive must support the capture and provision of the US Federal Government Project Open Data Metadata Schema <https://resources.data.gov/resources/dcat-us/>. In addition, the archive you choose must support the creation and maintenance of persistent identifiers (e.g., DOIs, handles, etc.) and must provide for maintenance of those identifiers throughout the preservation lifecycle of the data.

As general guidance, CYBER-CARE PIs may also consider addressing the following questions in the related section of project specific DMP:

- Discuss how you intend to archive your data and where (include URL).
- Indicate the approximate time period between data collection and submission to the archive.
- Identify where data will be stored prior to being sent to an archive. You should also:
- Describe how back-up, disaster recovery, off-site data storage, and other redundant storage strategies will be used to ensure the data's security and integrity.
- Describe how data will be protected from accidental or malicious modification or deletion prior to receipt by the archive.
- Discuss your chosen data archive's policies and practices for back-up, disaster recovery, off-site data storage, and other redundant storage strategies to ensure the data's security and integrity for the long-term.
- Indicate how long the chosen archive will retain the data.
- Indicate if the chosen archive employs, or allows for the recording of, persistent identifiers linked to the data.
- Discuss how the chosen data repository meets the criteria outlined on the Guidelines for Evaluating Repositories for Conformance with the DOT Public Access Plan page < https://ntl.bts.gov/ntl/public-access/data-repositories-conformant-dot-public-access-plan>.

Processes to Ensure Project-Level Adherence

To guarantee that each project funded by the US DOT adheres to the DMP requirements set forth by the US DOT, CYBER-CARE UTC will undertake four stages:

Proposal Stage:
PIs commit to conforming the DMP as stated in the proposal.

Award Stage:
PIs are provided with DMP compliance instructions in the award letter. Acknowledgement of the award letter and commitment to all stipulated requirements are affirmed by the PIs through signature.

During Research Project:
Given that research projects and teams undergo changes, it is essential to update the project-level DMP information as necessary. DMPs are dynamic documents, subject to alterations throughout the research project's duration. Regular updates to the DMPs facilitate efficient knowledge management, aid in integrating new research staff, keep the CYBER-CARE UTC director informed of project alterations, and assure the DOT UTC Program of the project's proper management. A review of project information should be conducted at least quarterly, with updates made whenever there are modifications to the awarded project or changes in project staff. PIs are required to submit data to the CYBER-CARE UTC director. CYBER-CARE UTC will undertake an internal review to verify compliance before posting the index of data to center website within a 60-day timeframe.

Project Completion Stage:
Upon completion, PIs submit new datasets created as part of a project funded by CYBER-CARE and links to other third-party public datasets to the CYBER-CARE UTC director. After completion an internal review is carried out by CYBER-CARE UTC to verify compliance before posting the data and index to CYBER-CARE center website within 60 days.

Change Log:

2023-10-06:
- Original document created.

2023-10-31:
- Updated Title Page to include more details and contact information.
- Updated CYBER-CARE UTC overview to included compliance with Indigenous Data Governance
- Updated Data format and metadata standards.
- Updated Plans for archiving and preservation, DOIs or PIDs are mandatorily required for publicly available data and code.
- Merged Policies for Sensitive Information Protection into Access Policies to maintain compliance with Creating Data Management Plans for Extramural Research by DOT.
- Provided a list of Data Repositories Conformant with the DOT Public Access Plan
- Provided general guidelines for CYBER-CARE PIs in project specific DMP as applicable.