National Center for Infrastructure Transformation (NCIT)
Type of UTC: National
Center Data Management Plan

Contact: Judy A. Perkins, Ph.D., PE, juperkins@pvamu.edu
Date Effective: 1 June 2023
Grant Period: 1 June 2023 – 31 May 2029
Grant Numbers: 69A3552344813 and 69A3552348318
Name of UTC Lead: Prairie View A&M University
List of Partner Institutions: Arizona State University, Blinn College District,
Michigan State University, Rutgers University, Texas A&M Transportation
Institute, and Texas A&M University

The National Center for Infrastructure Transportation (NCIT) will work collaboratively with its university partners and stakeholders to provide meaningful contributions in infrastructure design, maintenance and construction methods, and infrastructure planning policy across all transportation modes. NCIT's research will focus on three pillars, namely, infrastructure durability and resilience, technology, and policy.

NCIT will tackle research, education, and outreach activities to support these pillars. The Center requires each research project to have a Data Management Plan (DMP) that can reference the NCIT DMP to reduce copy/paste of repeated language, that states the unique and specific aspects of the research project for each section described in NCIT's DMP, and that the research project DMPs are living knowledge management tools that should be reviewed and updated regularly and each time there is a significant change in the research project, the data collected, or in project personnel. The updated DMPs shall be presented to USDOT for review.

Moreover, the Center has developed a strategy for managing and archiving digital data sets, which is described in this DMP. The DMP outlines NCIT's approach for handling, storing, and sharing data. To facilitate the sharing and accessibility of interdisciplinary transportation infrastructure data, NCIT has established a dedicated data repository. This endeavor aligns seamlessly with the U.S. Department of Transportation (USDOT) Public Access Plan, ensuring adherence to pertinent guidelines and regulations.

Data Description

NCIT activities will yield several types of data related to research, leadership, educational activities, workforce development, and technology transfer. Much of the data will be accessible from the NCIT website (ncit@pvamu.edu) by the public at large. Nonetheless, very large datasets or sensitive data may be only accessible to authorized personnel via secured (password protected) sites or may not be made accessible to the public at all under some circumstances.

Data collected or generated from **Research** activities may include, but are not limited to:
- Manual collection of laboratory or field data
- Automated/Software generated collection of laboratory or field data
- Algorithms and software (original and commercial source code)
- Performance metrics of NCIT research, such as: number of peer-reviewed publications and their citations, conferences, and presentations; media coverage of NCIT and NCIT projects; number of implemented strategies, products, patent applications, or techniques

Research data will be overseen under the specific management plans for each associated research project. The NCIT research results will also be incorporated into peer-reviewed publications, and presented at professional or academic conferences, and workshops. Data of this type will also be shared with USDOT OST-R in the bi-annual and annual performance metrics reports. Students will be instructed on how to preserve proper logbooks and data management during their research experience. They are required to document all their work and submit their records plus all electronic files/data used to their project manager. Considering the multi-institutional nature of this Center, each consortium member university is responsible for data storage and archiving appropriate with the established data management procedures and file specifications of that institution during the life of the individual projects under the NCIT grant. Data management and enforcement of data management expectations among Center members will be coordinated through this general DMP, email correspondence, and secondary Data Security Plans, if necessary. PIs of each NCIT research project are responsible for collecting, storing, retaining, and sharing the data from the research, education, or outreach activities on his or her project during the period of performance for individual projects. At the completion of each research project, the PI will be required to upload and store all the processed data electronically, on a server managed by PVAMU. The information technology (IT) staff at PVAMU will facilitate the storage and archiving of data at the direction of the Center Director and the requirements set forth within the scope of work of NCIT. Each Associate Director of NCIT is responsible for managing the project PIs at his or her site in ensuring the adherence to this DMP, i.e., all the necessary data are collected, stored, retained, and accessible. Should any of the project PIs leave the institution, the Associate Director will ensure that the practices outlined in this DMP are adhered to, such as providing copies of the data to the remaining PI or Center Director along with documents that describe what data have been collected, how the data are collected, where and in what format the data are stored, and how the data can be accessed.

Data collected or generated from **Education and Workforce Development** activities may include, but are not limited to:
- Educational materials to support student research and course modules in NCIT related topics, K-12 outreach, educational software, recordings of seminars and other professional events, etc.

- Continuing education, lecture notes, and other curriculum or training materials
- Performance metrics of NCIT education, such as: number of attendees in K-12 programs; transportation related courses offered by faculty affiliated with NCIT, number of undergraduate and graduate students participating in NCIT activities; number of transportation related degree and non-degree programs.
- Performance metrics of NCIT diversity programs, such as: number of female, minority, or disabled students in NCIT research
- Performance metrics of NCIT workforce development, such as: number of attendees at training sessions, webinars, career-building activities, or online courses; hours of technical assistance offered to agencies through project meetings and training workshops.

Data collected or generated from **Technology Transfer and Collaboration (T2C)** activities may include, but are not limited to:
- Technical documentation (journal papers, conference papers, reports, etc.) written by NCIT researchers, students, and collaborators.
- Instructional videos (seminars/training sessions, webinars, featured speakers, and other events)
- Social media posts, image data, animations or videos that broaden the impacts of NCIT activities.
- Quarterly e-newsletters
- Annual progress reports
- Performance metrics of NCIT T2, such as: number of inventions, patent applications, technologies etc. resulting from NCIT research activities; presentations given at professional and academic meetings; peer-reviewed publications; number of attendees at NCIT T2 events; number of stakeholders engaged in NCIT activities; number of partnerships formed; number of adoptions of NCIT research outputs; benefits to industry or society from NCIT activities.

NCIT will develop a customized data repository with the objective of facilitating exploration into transportation infrastructure data by center collaborators. The data repository includes a crucial component focused on establishing standard operating procedures (SOPs) for data collection, storage, and sharing within the center and to the public. These SOPs are designed to ensure compliance with policies set forth by the DOT and other funding agencies. Collaborators and potential data users, including students, academic faculty, system designers, developers, and other individuals interested in transportation infrastructure-related research, will have access to these SOPs for the data repository.

Data Formats and Standards
Datasets published within the data repository will be assigned versions and will be associated with consistent metadata that documents various aspects of the data. This metadata may include but is not limited to information such as the collection location, date, and responsible parties, details about the experiments or observations described by the data, data resolution, quality assurance methods employed, and data ownership particulars. Additionally, NCIT's metadata format will align with the Data Catalog Vocabulary (DCAT) Schema v.1.1 guidelines (https://resources.data.gov/resources/dcat-us/).

To ensure consistent versioning, only data repository curators will have the authority to publish new versions of existing datasets. They will collaborate and coordinate with the NCIT research teams in this process. After datasets are uploaded to the data repository, the curators will verify compliance with the NCIT DMP. Prior to publication, each Principal Investigator (PI) will be required to confirm that the public dataset accurately represents their expectations and faithfully reflects the provided data.

Each research team within NCIT will collect data in various formats. Commercially accessible packages (EXCEL, MATLAB, ArcGIS, SPSS, SQL database, etc.) will be used to perform data analysis; these packages are industry standard, and their data formats are not expected to become outdated in the foreseeable future. Research data will be captured and kept in databases or flat file format from raw laboratory notes or as ASCII, TIFF, OIB files from experimental analysis. Plots will be captured using commercial mathematical packages, e.g., EXCEL (.csv or.xlsx) MATLAB, Mathematica, ANSYS, ArcGIS, SPSS, SQL database, ORIGIN, etc., and images will be saved in standard file formats (PDF, TIFF, JPEG, etc.). Software will be written in C, C++, MATLAB, EXCEL, or another readily available programming language or platform. Presentations, seminar materials, and instructional documentation will be created and saved in Microsoft PowerPoint, PDF, or Word formats. The data will be submitted to conferences and journals in WORD or PDF format manuscripts. Prior to being made publicly accessible, the data will be converted into open-source formats, unless a waiver is granted by the DOT. Additionally, both the native file from any commercially accessible software and the open access file will be released to all intended repositories.

Next, to prioritize the handling of infrastructure data within NCIT, dedicated data handling policies for each project will be established to ensure compliance with Institutional Review Board (IRB) guidelines and specifics related to data handling regulations, if necessary. For instance, data elements containing personal identifying information (PII) will be removed before datasets are made publicly accessible. Additionally, there may be instances where PIs are required to document and disclose the characteristics of a dataset without directly sharing it through the data repository. This approach enables collaborating researchers to gain an understanding of the dataset's contents while allowing the data owner to make individual decisions on its future research usage on a case-by-case basis.

Some NCIT research projects may receive partial funding from industry partners, leading to the generation or reanalysis of proprietary data. If proprietary data collection or analysis occurs, a waiver will be sought from the DOT to exempt the storage of such data from public access requirements. In the event the granting of a waiver from DOT does not occur, NCIT researchers will provide thorough documentation addressing why proprietary formats are used, the software, and the version needed for the end user to access the data as well as documentation files.

Data Access and Protecting Sensitive Data
All data collected within NCIT will be accessible through the data repository and the NCIT website, with access control managed through security and privilege levels for sensitive data. Data made publicly available will undergo a de-identification process to ensure the exclusion of private or confidential information. It is a necessary and mandatory step prior to the release of data to the public.

In cases where sensitive data is involved, only brief descriptions of the data may be made available, enabling the PI responsible for collecting the data to make individual decisions on data sharing on a case-by-case basis. Any data that raises concerns related to privacy, ethics, or confidentiality will not be made accessible to the public.

Data Sharing, Re-use, and Redistribution

Upon uploading the data to the NCIT data repository, the curators responsible for managing the repository will assume the data management rights. Consequently, the NCIT data will be made openly available for sharing. We anticipate that there will be no access restrictions to most data generated by NCIT. Exceptions may apply, as per PVAMU Patent and Copyright Policies and those of consortium member universities. Anyone requesting re-use, re-distribution, or production of results from the work will be required to acknowledge the original contribution of the NCIT project. The access to the experimental data obtained from NCIT-sponsored activities will have disclaimers and policies regarding the fair use of the data in other publications or products.

Data Preservation and Archiving

The design and implementation of the data repository will adhere to the criteria outlined in the Guidelines for Evaluating Repositories for Conformance with the DOT Public Access Plan (https://doi.org/10.21949/1503646).

Datasets will be prepared and submitted to the data repository for archiving prior to the end date of each NCIT research project, as specified in the project timeline. Upon publication, each dataset will be assigned a Digital Object Identifier (DOI).

Before archiving data on the NCIT data repository, research project data will be stored in the data management systems of the project's PI institution. Each consortium member institution employs backup, disaster recovery, off-site data storage, and other redundant strategies to safeguard against accidental or malicious modification or deletion of data. The NCIT research teams will follow the established processes utilized by their respective consortium member institutions for these purposes. Once the data is archived, the NCIT data repository will handle these processes accordingly.

Change Log:
2023-09-26 1254: Original Draft
2024-01-23 1429: Revised version submitted
2024-02-14 1823: Final version submitted