New England University Transportation Center (NEUTC) Data Management Plan

Lead Institution: University of Massachusetts Amherst
Partner Institutions: UConn, University of New Hampshire, UMaine, MIT, Norwich University, Bunker Hill CC, Holyoke CC
Effective Date: 11/13/2023
Grant Period: 6/1/2023 – 5/31/2029
Grant Number: 69A3552348301
Contact person: Cole Fitzpatrick, cfitzpat@umass.edu

The New England Regional UTC (NEUTC) is a diverse, multidisciplinary consortium committed to addressing the pressing issue of traffic safety. Our objective, in line with the Infrastructure Investment and Jobs Act (IIJA), is to drive transformative research, education, and technology transfer to address critical traffic safety needs in a time when roadway fatalities are distressingly high.

Our research and educational activities at NEUTC are guided by four principal safety themes, each addressing a critical challenge in transportation safety. These themes capture the various integral components of the transportation system, focusing on technology, infrastructure, vehicles, and users with a commitment to equity and public engagement. Our overarching theme is **promoting safety**, with the common underlying science being the study of behavioral, systemic, environmental, and mobility-driven factors on safety.

Data Description for NEUTC Funded Projects

All projects funded by NEUTC are required to comply with the USDOT Public Access Plan, which mandates the uploading of all data to a repository upon project completion. Researchers will be required to submit individual data management plans containing the following information:

1. **Name of the Data/Project:** Provide a clear title for the data collection project or the specific data set.
2. **Purpose of Research:** Briefly elucidate the aim behind the collection of the data.
3. **Data Nature and Scale:** Specify the kind of data you will generate (e.g., numerical data, image data, text sequences, video, audio, database, modeling data, source code, etc.).
4. **Data Creation Methods:** Illustrate how the data will be gathered or produced (e.g., simulation, observation, experimentation, software, physical collections, sensors, satellite, enforcement

activities, researcher-generated databases, tables, spreadsheets, instrument-generated digital data output such as images and video, etc.).

5. **Data Collection Period & Update Frequency:** Define the span of time during which data will be amassed and the regularity of data updates.
6. **Relationship with Existing Data:** If leveraging existing data sets, detail the connection between the new data and the extant data.
7. **Potential Data Users:** Enumerate who might benefit from or utilize this data.
8. **Long-Term Value of Data:** Discuss the enduring significance of the data not just for NEUTC and its initiatives, but also for the broader public.
9. **Public Access Restrictions:** If any constraints are imposed on public access to the data, provide the reasons and a solution to ensure public access, including anonymization of the dataset.
10. **Party Responsible for Data Management:** Indicate which entity or individual will be overseeing the data's management.
11. **Adherence Check:** Even though individual data management plans are not required, elaborate on the measures that will ensure data is handled in accordance with the guidelines set by NEUTC and USDOT Public Access Plan.

All UTC-funded projects must have a document labeled “Data Management Plan”, even if the DMP states that “no data will be collected during this project.”

Data Formats and Accessibility in NEUTC Funded Research

NEUTC research projects generate a variety of data, including but not limited to response times, travel behavior, vehicle operations, and crash data. This data is typically available in the following formats:

Open-Access Formats (Preferred for public access):

- Comma Separated Values (.csv)
- Portable Document Format (.pdf)
- Joint Photographic Experts Group (.jpg)

Proprietary Formats:

- MS Excel (.xls, .xml)
- Video files (.mpg, .avi, .mov, .wmv)
- MS Excel Macro-Enabled Workbook (.xml)

Researchers are required to report the formats used, specifying if they are open or proprietary. For proprietary data, a rationale must be provided to NEUTC. Any deviations from original data formats should be clearly documented in research reports.

Metadata Management: Each dataset will include metadata detailing context, content, and structure. Researchers may employ nonstandard schemas but must justify this choice in their final report. Metadata management is the responsibility of the principal investigators throughout the project lifecycle. All metadata will be stored in the NEUTC Harvard Dataverse repository.

Software and Tools: The project technical report must describe the tools or software necessary for data reading or viewing, prioritizing open-source or widely accessible options.

Quality Standards: Given the interdisciplinary nature of transportation engineering, data quality standards will vary. The principal investigator of each project is responsible for ensuring data accuracy and completeness, adhering to industry standards.

All principal investigators for individual projects will be required to:

- Final datasets are recommended to be in a non-proprietary data format, such as csv.
- If principal investigators are using proprietary data formats, they will be required to discuss their rationale.
- Include metadata describing the context, content, and structure of the final version of data shared with the public.
- Describe how they will document the alternative formats they are using and why.
- List what documentation they will be creating in order to make the data understandable by other researchers.
- Indicate what metadata schema they are using to describe the data. If the metadata schema is not one standard for their field, and discuss their rationale for using that scheme.
- Describe how the metadata is managed and stored.
- Indicate what tools or software are required to read or view the data.
- Describe their quality control measures.

Policies for Access and Sharing

The principal investigator is responsible for how the data is managed and secured during the experimental process. Once the project is completed, the data will be publicly available via the NEUTC repository in the Harvard Dataverse. NEUTC researchers are required to upload their data within 60 days of their project end date. Because some transportation-related research requires the use of human subjects, permission from the Institutional Review Board (IRB) where the research originated will be obtained prior to publishing onto the public sites for data sharing.

Principal investigators will be required to address any access restrictions in the data management section of proposals they submit to NEUTC. In this section principal investigators will address issues and outline the efforts they will take to provide informed consent statements to participants, the steps they will take the protect privacy and confidentiality prior to archiving their data, and any additional concerns (e.g., embargo periods for their data). If necessary, they will describe any division of responsibilities for stewarding and protecting the data among other project staff. If the data contains personally identifiable information (PII), the principal investigators will be required to anatomize the dataset and if they are not able to de-identify the data in a manner that protects privacy and confidentiality while maintaining the utility of the dataset, faculty will describe the necessary restrictions on access and use. If an individual research project includes human subject research, researchers will be required to go through University of Massachusetts IRB or their home institution's IRB, if they have one.

Principal investigators will be required to address the following:

1. Describe what data will be shared, how data files will be shared, and how others will access them.

2. Indicate whether the data contain private or confidential information. If so
    a. Discuss how you will guard against disclosure of identities and/or confidential business
    a. information
    b. List what processes you will follow to provide informed consent to participants.
    c. State the party responsible for protecting the data.
2. Describe what, if any, privacy, ethical, or confidentiality concerns are raised due to data sharing.
3. If applicable, describe how you will de-identify your data before sharing. If not:
    a. Identify what restrictions on access and use you will place on the data.
    b. Discuss additional steps, if any, you will use to protect privacy and confidentiality.

Policies for Re-use, Redistribution, and Derivatives in NEUTC Funded Projects

Intellectual Property and Public Access:
The intellectual property rights for all data generated in NEUTC-funded projects are held by the principal investigator's home institution. This data is governed by the General Provisions of Grants for 2016 University Transportation Centers, particularly Item #16 on Patents and Copyrights (pages 11-15). The principal investigator (PI) retains ownership of all publicly provided data.

Licensing and Rights Transfer:
In their final report, PIs must detail any rights to be transferred to the data archive. They must also specify how the data will be licensed for reuse, redistribution, and the creation of derivative works. When proprietary data is used, PIs are required to cite the data source and the license under which they utilized the data in their project data management plans.

Data Management Plan Requirements:
Principal investigators should address the following in their individual data management plans.

1. Identification of the party responsible for data management.
2. Clarification of who holds the intellectual property rights to the data.
3. Listing of any copyrights on the data and the ownership of these rights.
4. Discussion of any rights to be transferred to a data archive.
5. Description of the licensing terms for data reuse, redistribution, and derivatives.

Adhering to FAIR Principles:
In line with the FAIR principles, NEUTC is committed to ensuring that the data is:

- **Findable:** Easy to locate for both humans and computers, with well-maintained metadata and standard identification methods.
- **Accessible:** Stored in a way that data and its metadata are understandable and retrievable for authorized users.
- **Interoperable:** Compatible with other datasets and tools for analysis, enabling integration with other data.
- **Reusable:** Well-documented and richly described so that it can be reused effectively and ethically in different contexts.

By adhering to these principles, NEUTC aims to maximize the impact and utility of the data generated from its funded projects.

Plans for Archiving and Preservation in NEUTC-Funded Projects

Data Management and Archiving:
Principal investigators (PIs) or their delegates are responsible for managing data before, during, and after collection, ensuring compliance with their Institutional Review Board's standards. All data from NEUTC-funded projects will be archived in Harvard's Dataverse. PIs have 60 days post-project to archive their data and are responsible for maintaining it until upload. They must also detail measures to protect the data from accidental or malicious modification or deletion before it is archived.

Harvard Dataverse Repository and Backup:
Harvard University Information Technology (HUIT), Harvard Library, and the Institute for Quantitative Social Science (IQSS) jointly manage the Dataverse repository. This includes maintaining a full backup of all data and directories, ensuring data integrity. The repository's digital archiving policy aligns with Harvard's mission to preserve archival collections and maintain data quality using best digital archival practices. Dataverse backs up all application/system files and databases nightly, storing them off-site for 45 days. Research data files are replicated every four hours to a secondary off-site storage array. Additionally, data is integrated into the DRS Storage Infrastructure for long-term tape storage at the Harvard Depository. The Dataverse preservation policy is available at Harvard Dataverse Preservation Policy.

Post-Archiving Data Management:
After archiving, data management shifts to the Harvard Dataverse team. They ensure continued accessibility and integrity of the datasets. Once published, datasets receive a DOI persistent identifier from the California Digital Library's (CDL) EZID service (DataCite member), solidifying long-term access. Published datasets cannot be unpublished and can only be deaccessioned under exceptional circumstances, such as legal mandates for data destruction.

Change Log for Data Management Plan:
11/6/2023 – Draft DMP submitted to USDOT
11/6/2023 – USDOT provided comments on numerous sections
11/13/2023 – Revised DMP submitted to USDOT
11/16/2023 – DMP approved by USDOT