

Center Data Management Plan for the National Center for Sustainable Transportation (NCST)

National University Transportation Center

Grant Obligations No. 69A3552344814 and No. 69A3552348319

Grant Period: June 1, 2023 – May 31, 2029

Lead: University of California, Davis

Partners: California State University, Long Beach; Georgia Institute of Technology; Texas Southern University; the University of California, Riverside; the University of Southern California; and the University of Vermont

Director: Dr. Susan Handy

Contact: Lauren Iacobucci, Assistant Director, Liacobucci@ucdavis.edu

Effective: June 1, 2023

The National Center for Sustainable Transportation (NCST) provides national leadership in advancing environmentally sustainable transportation through cutting-edge research, policy engagement, and educational programs. The NCST supports policy innovation to achieve equitable decarbonization through four kinds of research activities: (1) Build tools to support policy making, including datasets, lifecycle assessments, and predictive simulation models that support scenario planning. (2) Develop policy elements, such as evidence-based standards, model regulations, pricing strategies, funding strategies, and others. (3) Conduct evaluation studies that assess the outcomes of policies using techniques such as experimental designs and econometric modeling. And, (4) Undertake basic research that provides a solid foundation for each of these activities by improving our understanding of the transportation system.

The NCST complies with the requirements of the USDOT Public Access Plan (<https://ntl.bts.gov/public-access>), which establishes objectives to ensure public access to Publications and Digital Data Sets arising from DOT-managed research and development programs. To fulfill this requirement, the NCST establishes this Center-level Data Management Plan (DMP) which describes our strategy for processing and archiving digital datasets resulting from NCST research in a repository that enables and allows for public access and sharing.

All NCST-funded researchers are required to comply with this Center-level DMP by (1) acquiring an ORCID iD¹, (2) developing a project-level data management plan for each project before funding for the project is awarded, and (3) publicly archiving any final analysis datasets resulting from the funded project in compliance with the NCST's Center DMP. The NCST provides data management resources, guidance, and templates for its researchers on the center website, <https://ncst.ucdavis.edu/data-management-plan-info-guidance>. The NCST strongly encourages

¹ <https://orcid.org/>

researchers to use DMPTool.org² to develop their project-level DMP. The NCST has created a project-level DMP template on DMPTool.org that complies with the USDOT Public Access Plan.

At the completion of each project, in addition to archiving their final project data in a compliance publicly accessible data repository (notwithstanding any restrictions to archiving), PIs are required to document their project data by completing a “Data Summary” section in each final deliverable. The Data Summary section provides information on what kind of data was collected and/or used in the study, the format of the data, how the data can be accessed, and any restrictions on how the data can be reused and redistributed.

Data Description

Projects funded by the NCST may gather many different types of data throughout the course of the UTC grant, including, but not limited to the following.

Travel surveys: Online or on-paper surveys administered to samples of individuals or households to record travel behavior and choices, attitudes and perceptions, and demographic characteristics. Primary datasets sometimes include individual and/or household records with personally-identifiable information (PII), such as names, addresses, school and work details, and other details that could be used to identify an individual or household directly, or in combination with other datasets. Datasets can usually be shared after redaction of the PII.

Professional surveys: Online or on-paper surveys administered to random samples or purposive samples of individuals acting in their professional capacity as representatives of their organizations. Datasets typically include individual or organization records, with individual names, addresses, and other personal and business identifiers. Datasets can usually be shared after redaction of the PII.

Interviews: Structured, semi-structured, or open-ended interviews conducted by phone, video, or in-person with human subjects and/or key informants on topics relating to travel behavior, transportation policy, transportation practice, or similar topics. Datasets include transcriptions of interviews with name, addresses, and all other personal identifiers redacted. Datasets may include video and audio recordings. Datasets can usually be shared after redaction of the PII.

Video observations: Video recordings of the operation of transportation facilities, such as streets, roundabouts, bike trails, and transit stations, which may include the movement of people and vehicles through these facilities. Datasets may include video files, as well as manual or automated coding of the video. Datasets can usually be shared after redaction of all PII.

Vehicle/person location and activity data: Time-stamped vehicle trajectories or person location collected by mobile device or GNSS receiver data loggers. Other vehicle parameters such as energy use and emissions or pollutant levels may also be collected through instrumentation of test vehicles. Data may also be produced through lab experiments or simulation models including input data and output data. Datasets can usually be shared after redaction of all PII.

² <https://dmptool.org/>

Traffic volume and speed data: These datasets are generally collected using traditional instruments such as loop detectors, infrared sensors, magnetometers, and video or more recent techniques, including cell phone tracking and Bluetooth signals. In some cases, data may be purchased or licensed from third parties; analytical results summarizing the purchased or licensed data will be made available and the researchers will provide all information needed by a third party to license the same data and replicate the analyses.

Demographic data: Data on individual and household characteristics as well as vehicle ownership and other travel-related characteristics can be acquired from state agencies and/or commercial vendors. Exceptions to data sharing requirements will be granted when access to such datasets is restricted and special permission is required for access. Researchers will provide all information needed by a third party to license the same data and replicate the analyses.

Air quality data: Datasets collected using instruments such as air quality sensors may be proprietary or open data, depending upon the source. Primary data collected by the research team will be openly shared, unless prohibited by law, regulation, or contract. Data licensed from third party providers will be summarized. Researchers will provide all information needed by a third party to license the same data and replicate the analyses.

Industry data: Data on business operations acquired from state agencies and/or commercial vendors. Exceptions to data sharing requirements will be granted when access to such datasets is restricted by law, regulation, or contract and special permission is required for access. Researchers will provide all information needed by a third party to license the same data from the industry source, provided that the industry source commercially licenses such data.

Modeling and simulation data: Data not containing personally-identifiable information may also be produced through lab experiments or simulation models including input data and output data. Exceptions to data sharing requirements will be granted when access to such datasets is restricted and special permission is required for access.

Existing datasets: New datasets created as a result of calculating new variables with data from pre-existing datasets, such as travel surveys. Exceptions to data sharing requirements will be granted when access to such datasets is restricted and special permission is required for access.

Data Format and Metadata Standards

Data from NCST projects will be stored in platform-independent and non-proprietary formats whenever possible, such as txt, CSV, mp3, mp4, etc. If project researchers are unable to use platform-independent and non-proprietary formats, they will specify in their project DMP(s) the standards and formats that will be used and the rationale for using those standards and

formats. NCST projects will generally follow the format of the NREL TSDC server³ and/or the FHWA Research Data Exchange.

The metadata scheme NCST will use, at minimum, the Project Open Data Metadata Schema.⁴

Policies for Access and Sharing

In general, final analysis data from NCST research projects funded wholly or in part by USDOT will be made publicly accessible, depending on the type of data.

Open Data

Datasets in the public domain are considered to be open data. Demographic data, air quality data, industry data, modeling and simulation data, and existing data often qualify as open data, but they may be proprietary (see below). Open datasets can be freely shared.

Protected Data

A dataset containing personally-identifiable information (PII) is a protected dataset. These datasets must be collected, managed, and shared under the provisions of human subject protocols that have been approved by an Institutional Review Board (IRB). PII is protected under federal law and regulation, and primary data containing PII cannot be shared.

Primary datasets collected through travel surveys, professional surveys, and interviews often include individual and/or household records with personally-identifiable information (PII), such as names, addresses, school and work details, and other personal details. Datasets can generally be shared after redacting the PII, including information that would enable the identification of participants by coupling the primary dataset with secondary datasets in the public domain. Redacted datasets can only be shared if they follow procedures defined in a human subject protocol and approved by an IRB. Any video and audio recordings that include personally-identifiable information (faces, names, etc.) are exempt from data sharing requirements due to IRB protocol requirements.

Video observations and vehicle/person location and activity data sometimes also include PII. High-resolution data that contain personally-identifiable information (e.g., home and school locations) are exempt from data sharing requirements due to IRB protocol requirements. Datasets summarizing the high-resolution data (such as trip distances or speeds on roadway links) can often be made available as long as they are part of an approved IRB protocol. Datasets can generally be shared after redacting the PII, including information that would enable the identification of participants by coupling the primary dataset with secondary datasets in the public domain. Redacted datasets can only be shared after data processing methods have been defined in a human subject protocol and approved by an IRB.

³ https://www.nrel.gov/transportation/secure_transportation_data.html

⁴ <https://resources.data.gov/resources/dcat-us/>

All projects involving human subjects will abide by the requirements of the Institutional Review Board of the institutions of the Principal Investigators (PIs) of the projects. These projects must have an IRB-approved protocol for ensuring informed consent of participants and protecting privacy and confidentiality. When working with or conducting research that includes Indigenous populations or Tribal communities, PIs will adhere to the CARE Principles for Indigenous Data Governance.⁵ Data will be shared only after redaction of all individual identifiers, including names, residential addresses, exact geo-coordinates of residences or workplaces, and email addresses. If researchers are not able to deidentify the data in a manner that protects privacy and confidentiality while maintaining the utility of the dataset, they will be required to describe in the final report the necessary restrictions on access and use.

The PIs of the project will be responsible for acquiring IRB approval and adhering to IRB and other data sharing requirements. The PIs must report IRB approvals and other data sharing requirements in their project proposals and progress reports.

Proprietary Data

Proprietary data are licensed from third-party providers and may only be distributed in accordance with data licensing contract language. These data are usually available for license by other interested parties from the same vendor/source. These data are generally not openly shared; instead, the research teams share the license information so that third-parties can license the same data for their research. Data from public sources for which access is restricted and special permission is required are subject to the same sharing policies as proprietary data from industry sources.

Projects using proprietary data from commercial or public sources will abide by all conditions and requirements imposed on the use of the data. If the source organization, license agreement, or terms of sale prohibit the public sharing of the data, the project will be granted an exception from data sharing requirements.

Policies for Re-use, Redistribution, Derivatives

Intellectual property rights will generally be held by the PIs of the projects and/or their home institutions. Rights will not be transferred to a data archive. Copyrights may apply to the presentation or display of data from some projects, such as those using copyrighted instruments or proprietary data sources. PIs will identify whether any copyrights apply to the data, as might be the case when using copyrighted instruments. Projects using proprietary data from commercial or public sources will abide by all conditions and requirements imposed on the use of the data.

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.

⁵ <https://www.gida-global.org/care>

Plans for Archiving and Preservation

Dryad⁶, the data repository of the University of California, will be the primary repository for the NCST. Dryad meets the requirements of the USDOT Public Access Plan for repositories.⁷ Dryad is an open and curated data publishing platform for researchers to share and publish their data. The California Digital Library and Dryad co-developed this service which is designed to be a simple data publication tool for researchers to meet funder and publisher mandates. Each dataset goes through a curation process to check for findability, accessibility, interoperability, and reusability.

PIs may opt for other compliant repositories that conform with the USDOT Public Access Plan; other repositories must meet federal requirements and 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.

All PIs will be required to provide sufficient metadata to ensure the discoverability of the data. Future funding for the PI for other NCST projects is contingent on successful implementation of the project-level data management plan.

Change Log

2023-08-25: Original draft.

2023-08-28: Revisions based on feedback.

2023-09-01: Revisions based on feedback.

2023-09-12: Revisions, finalized, ready for distribution.

2023-09-30: Revisions based on feedback, finalized, ready for distribution.

⁶ <http://datadryad.org/stash>

⁷ <https://ntl.bts.gov/ntl/public-access/guidelines-evaluating-repositories>