Center for Efficient Mobility
Data Management Plan

OCTOBER 2026

Center for Efficient Mobility

A USDOT University Transportation Center

Texas A&M Transportation Institute
Johns Hopkins Bloomberg School of Public Health
Georgia Tech College of Engineering
UTEP
UCR College of Engineering – Center for Environmental Research and Technology
Morehouse School of Medicine
NDSU North Dakota State University

Overview

This document describes the Center for Efficient Mobility (CEM)—formerly Center for Advancing Research in Transportation Emissions, Energy, and Health (CARTEEH)—policy on data handling, storage, and sharing. CEM is a Tier 1 University Transportation Center (UTC) that aims to promote interdisciplinary research in the area of transportation and health, including facilitating collaboration between health and transportation researchers by integrating data sources from the two research fields. The CEM center was originally set up under the Fixing America's Surface Transportation (FAST Act) in 2016 and has been renewed for funding under the Bipartisan Infrastructure Legislation (BIL) UTC Competition in 2023. The integration of transportation and health data presents novel challenges for effective and safe data handling and sharing. For this reason, CEM's Data Management Plan (DMP) will be a "living document" that will be updated as necessary over the life of the Center. One of the initial projects undertaken by CEM was the establishment of a data hub (https://carteehdata.org/) to facilitate the sharing and access of data in the interdisciplinary area of health and transportation. This data hub has since been moved to Zenodo and contains datasets from previous projects developed under the scope of CARTEEH. CEM will continue to enhance the data hub and ensure it remains a repository conformant with the U.S. Department of Transportation (DOT) Public Access Plan.

CEM's DMP will ensure that data collected during research activities involving the center complies with DOT policy and provides the maximum value for future research. The approved DMP will be available to collaborators via the CEM website and data hub.

DATA DESCRIPTION

CEM focuses on improving the mobility of people and goods through innovation, advanced technology, and data-driven decision-making. The Center pursues a mobility-centric, safety-driven, and technology-forward research portfolio aligned with current USDOT priorities of safety, efficiency, and economic impact. CEM's specific focus areas include data integration for mobility insights and digital infrastructure, advanced modeling and transportation optimization, the application of artificial intelligence to improve efficiency and safety, evaluation and commercialization of emerging technologies through an innovation accelerator, advanced air mobility and unmanned aerial systems (UAS), rural and underserved community mobility, freight and goods movement, and safety for automated vehicles and vulnerable road users.

Data hosted by the center includes both existing datasets collected under its prior scope (CARTEEH) and new data collected under its current focus on efficient mobility. Existing transportation and health-related datasets developed under the center's prior scope are retained in and remain accessible through the data hub. New data collected going forward reflects CEM's mobility-centric focus and includes traffic and mobility operations data, connected and automated vehicle data, advanced air mobility and UAS data, infrastructure

and freight data, and the sensing and telemetry sources described below. CEM will handle data through the CEM data hub. The goal of the data hub is to support the integration of data that advances efficient mobility, digital infrastructure, and technology deployment. As CEM continues to update and maintain the data hub, standard operating procedures (SOPs) for data collection, storage, and sharing within the center will be developed to ensure that all data is handled in line with DOT and other funding agencies' policies. These SOPs will be made available to collaborators and potential users of the data. Potential users of this data include students, academic faculty, system designers and developers, and other individuals interested in conducting transportation and mobility-related research.

Members of the CEM consortium will have a variety of existing datasets which will be leveraged for center research. The data hub will organize this existing data, along with existing publicly available data. Research projects funded by the center will also collect original data as part of their research activities. Original data will be collected in such a way that maximizes synergies with the existing data, avoids duplication, and leverages the knowledge gained from past efforts. CEM researchers will collect data using a range of sensing and data-capture methods, including unmanned aerial systems (UAS) and drone-based imagery, LiDAR and InSAR remote sensing, machine vision and roadside camera systems, connected and automated vehicle (CAV) sensors, GPS and positioning systems, real-time vehicle activity and fleet-composition monitors, weather data feeds, and direct observation. CEM research projects are expected to produce a variety of data types ranging in scale from compact tabular datasets to large-scale geographic and real-time streaming data describing vehicle activity across regional transportation networks. Data types will include numerical and tabular data, spatial and geospatial data, aerial and LiDAR point-cloud data, images and video, sensor and telemetry streams, and inputs for AI and large multimodal models. Data collected through funded projects will often require quality assurance processing. Other projects will involve summaries or integrations of existing complex data sets. In both cases, the computer code and models used for processing will be preserved alongside data outputs.

When datasets involve human subjects or contain sensitive information, specific data handling policies will be developed to ensure compliance with Institutional Review Board (IRB) requirements and applicable data-handling policies. For example, elements of data that contain personally identifying information (PII) will be removed before making datasets publicly available. In other cases, principal investigators (PI) may be required to document and publicize the characteristics of a data set without making it explicitly available through the data hub. This practice will allow collaborating researchers to understand the contents of a data set but will ensure the owner of the data can make case-by-case decisions on how it can be used for future research.

CEM research projects may receive partial funding from industry partners. These projects may generate proprietary data or reanalyze proprietary data. If this type of data collection/analysis occurs, the portion of the project under which data collection is completed will be funded entirely by the industry partner. In such cases, a waiver from storing public data will be sought from the DOT.

The PI of each research project funded through CEM will work to ensure appropriate data management throughout the life of the project. This will include all activities related to the preparation of the data for public access. After the data is archived in the repository described in this DMP (and verification that the data complies with minimum requirements), responsibility will be transferred to the curators of the CEM data hub. CEM Leadership will ensure research projects are compliant with the CEM DMP and will store project-specific DMPs in CEM administrative records. In addition, CEM research teams will have access to SOPs and this DMP through the CEM data hub and website.

DATA FORMATS AND METADATA STANDARDS

Data will be collected in a variety of formats by each CEM research team and will be converted into open-source formats before becoming publicly accessible unless a waiver is granted by DOT as described in the previous section.

The data generated by CEM research projects will be uploaded and archived into the CEM data hub that is currently under development. The data hub will be custom-developed for the center to archive and share data using persistent identifiers, data set version control and adherence to metadata standards. Access to the data hub will be granted through a web browser interface and username and password controlled. Datasets published within the data hub will be 'versioned' and will be associated with a consistent meta-data format that documents the characteristics of the data. This meta-data will include where, when, and by whom it was collected, the experiments or observations described by the data, the resolution of the data, quality assurance methods, and data ownership details. To ensure that versioning is handled consistently, only data repository curators will be able to publish new versions of existing datasets, in collaboration and coordination with the CEM research teams. After datasets are uploaded to the data hub, curators will verify the compliance of each dataset with this CEM DMP. Before publishing, each PI will be required to verify that the public dataset produced matches their expectations and is an accurate representation of what they provided.

CEM researchers will ensure that archived data is understandable and usable by other researchers through the creation of descriptive products, including metadata, a description of data collection method(s), links to project descriptions, scopes of work, publications, and a data dictionary. These products will be managed and stored alongside the data in the data hub. Standards for metadata and associated products will be developed and published within the data hub. CEM will adopt meta-data standards that bridge the

transportation and health fields. Where possible, the data hub will integrate software that helps researchers adhere to this DMP.

ACCESS AND SHARING POLICIES

All data collected within the CEM will be made accessible via the data hub. Access to data will be controlled through security/privilege levels. Data made available to the public will not contain private or confidential information. De-identification of data will be a necessary and required step before making data publicly available. Some sensitive data may be made available only through brief descriptions of the data, allowing the PI that collected the data to make case-by-case decisions on data sharing. Data that raises any concerns regarding privacy, ethics, or confidentiality will not be made available to the public.

RE-USE, REDISTRIBUTION, AND DERIVATIVE PRODUCTS POLICIES

After data is uploaded to the CEM data hub, data management rights will be transferred to the curators of the CEM data hub. CEM data will be made available for open sharing under the Creative Commons Zero (CC0) universal public domain dedication. Under CC0, data and derivative products will be available for reuse and redistribution without restriction.

ARCHIVING AND PRESERVATION PLANS

The CEM data hub will serve as the primary repository for all data. The data hub will be housed on RAID servers, with appropriate disaster recovery procedures that ensure the recoverability of the data and the data hub. Regular disaster recovery tests will be undertaken to ensure the adequacy of these procedures. The data hub will be designed and implemented to meet the criteria outlined in the Guidelines for Evaluating Repositories for Conformance with the DOT Public Access Plan.

Dataset preparation and submission to the data hub repository for archiving will occur before the CEM research project end date, as described in each research project timeline. Upon publishing, each dataset will be assigned a Digital Object Identifier using the EZID DOI minting service run by the California Digital Library. Any data will be preserved for the duration of the CEM project at a minimum.

Before archiving data on the CEM data hub, research project data will be stored on the data management systems of the project PI's institution. Back-up, disaster recovery, off-site data storage, and other redundant strategies are used by each consortium member institution, which protects data from accidental or malicious modification or deletion. CEM research teams will follow the usual processes employed by each consortium member institution for these purposes. Upon archiving the data, the CEM data hub will handle these processes.

TRAINING

To ensure CEM researchers, students, and other stakeholders are equipped with the knowledge and skills needed to effectively manage research data, a training module will be implemented about best practices for data collection, organization, storage, usage, sharing, and preservation. This training module will help users understand the benefits of open data sharing and how to prepare data for publication and raise awareness about the importance of data security and ethics in conducting research. All researchers, students, and stakeholders who have access to the data will need to go through the training which will help to maintain data integrity, improve collaboration, promote data sharing, and support compliance.

CHANGE LOG

09/27/2023 – Original draft (adapted from FAST Act Grant DMP)

10/01/2026 – Revised data management plan branding to reflect UTC name change and description of data types and formats changed to be representative of current center focus areas and priorities.