

Center for Advancing Research in Transportation Emissions, Energy and Health

Data Management Plan

This document describes the Center for Advancing Research in Transportation Emissions, Energy and Health (CAR-TEEH) policy on data handling, storage, and sharing. CAR-TEEH is a Tier 1 University Transportation Center (UTC) that has an explicit goal to encourage collaboration between health and transportation researchers by integrating data sources from the two research fields. The integration of transportation and health data presents novel challenges for effective and safe data handling and sharing. For this reason, CAR-TEEH'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 CAR-TEEH is the establishment of a data hub to facilitate the sharing and access of data in the interdisciplinary area of health and transportation. It is anticipated that this data hub initiative will inform subsequent updates to the DMP. CAR-TEEH will develop the data hub into a repository conformant with the U.S. Department of Transportation (DOT) Public Access Plan.

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

Data Description

The Center for Advancing Research on Transportation Emissions, Energy, and Health (CAR-TEEH) is a Tier 1 University Transportation Center (UTC) focused on addressing emissions in the context of public health. CAR-TEEH's specific focus areas include policy and decision making, emissions and energy estimation, alternative technologies, exposure assessment, public health impacts, and data integration. A major focus of CAR-TEEH is collaboration between researchers at different universities, operating in different research fields. CAR-TEEH's goal is to use this diverse expertise to address research gaps and unify the fields of transportation and public health research and policy making.

Data collected within the center include transportation and health related data. Transportation data includes transportation networks, traffic activity (speed and volume on corridors and transportation networks), emissions data (volume and concentration of pollutants), and demographic data (sources and destinations of travelers). Examples of health-related data include demographic/epidemiological, and human exposure data, human activity data, biomarkers and genetic information. CAR-TEEH will handle data through a custom designed data hub, the development of which is funded by the center. The goal of the data hub is to allow center collaborators to explore the potential for transportation and health data integration. A significant component of the data hub project will be to develop standard operating procedures (SOP) for data collection, storage and sharing within the center that



will 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 health related research.

Members of the CAR-TEEH 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. CAR-TEEH researchers will record new data using methods utilizing sensors such as personal exposure monitors, geographic positioning systems, portable emission monitors (PEMS), questionnaires, and direct observation. CAR-TEEH research projects are expected to produce a variety of types of data ranging in scale from compact tabular datasets to large-scale geographic data describing vehicle activity data across regional transportation networks. Data types will include numerical and tabular data, spatial data, images, video, and questionnaires. 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.

Given the importance of health data to CAR-TEEH, specific data handling policies will be developed to ensure Institutional Review Board (IRB) and specific health related data handling policies are adhered to. For example, elements of data which contain potentially identifying information (PII) will be removed prior to 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.

CAR-TEEH 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 CAR-TEEH will work with the CAR-TEEH Program Manager to manage all data 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 CAR-TEEH data hub. The CAR-TEEH Program Manager will ensure research projects are compliant with the CAR-TEEH DMP, and will store project-specific DMPs in CAR-TEEH administrative records. In addition, CAR-TEEH research teams will have access to SOPs and this DMP through the CAR-TEEH data hub and website.



Data Formats and Metadata standards

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

The data generated by CAR-TEEH research projects will be uploaded and archived into the CAR-TEEH 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 CAR-TEEH research teams. After datasets are uploaded to the data hub, curators will verify compliance of each dataset with this CAR-TEEH DMP. Prior to 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.

CAR-TEEH 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. CAR-TEEH 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 CAR-TEEH 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 prior to 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 which raises any concerns regarding privacy, ethical, or confidentiality will not be made available to the public.

Re-Use, Redistribution, and Derivative Products Policies

After data is uploaded to the CAR-TEEH data hub, data management rights will be transferred to the curators of the CAR-TEEH data hub. CAR-TEEH 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 CAR-TEEH 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 prior to the CAR-TEEH 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 CAR-TEEH project at a minimum.

Prior to archiving data on the CAR-TEEH 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. CAR-TEEH research teams will follow the usual processes employed by each consortium member institution for these purposes. Upon archiving of the data, the CAR-TEEH data hub will handle these processes.

