CCST Data Management Plan

Lead Institution:
Johns Hopkins University, Baltimore, MD, 21218

Member Institutions:
Diné College*, Tsaile, AZ 86556
Massachusetts Institute of Technology, Cambridge, MA 02139
Morgan State University*, Baltimore, MD 21251
University of Texas at Austin, Austin, TX 78712
University of Utah, Salt Lake City, UT 84112
Minority Serving Institution

Type of UTC:
Tier 1 Center

The Statutory Research Priority Area:
Preserving the Environment

Contacts:
Shima Hamidi, PhD; CCST Director, shamidi2@jhu.edu
Ebrahim Azimi, PhD; CCST Associate Director, eazimi2@jhu.edu

Date Effective: June 1st, 2023

Grant Period: June 1st, 2023 – May 30th, 2028

Center for Climate-Smart Transportation UTC

The Center for Climate-Smart Transportation (CCST)'s proposed research, education, leadership and technology transfer programs and activities are inspired by the urgent call for an evidence-based research agenda that goes beyond scientific merits, focuses on solutions and is practice-ready and would result in changes in transportation policy and practice, making climate change the center of transportation decisions as emphasized in the *USDOT Strategic Plan Goals* and the *USDOT Climate Action Plan: Revitalizing Efforts to Bolster Adaptation & Increase Resilience*. CCST contributes to this vision in the following focus areas:

1. Promoting Climate Culture in All Levels of Transportation Decisions
2. Community-Centered Solutions to Environmental Justice
3. Accelerate the Mass Market Adoption of EVs & Alternative Fuels
4. Reduce VMT & GHG via Modal Shift and Changes in Travel Behavior
5. Smart Cities & Innovative Adaptation and Mitigation Technologies

Center-Level Data Management Plan (DMP)

Following USDOT guidelines, CCST UTC requires a DMP for every project in the Center and will perform the following four-stage process to ensure that each CCST funded project complies with the DMP requirements proposed by US DOT:

Proposal stage: Projects' Principal Investigators (PIs) submit a DMP form (as part of the submission requirements) conforming to the DMP requirements in the proposal. The DMP should include:

i. Data Descriptions;
ii. Data Formats and reasons for necessary proprietary formats, if applicable;
iii. Contextual Documentation, such as data dictionaries defining the variables; README.txt files giving rational for the project and explaining methodologies; code books defining how data was processed;
iv. Quality Control Measures;
v. If applicable, explanations on why certain datasets cannot be shared. Individual DMPs will be reviewed and should be approved by CCST's director and/or associate director before the start date of projects.

Award Stage: PIs receive DMP compliance instructions as part of the award letter, and PIs must acknowledge the award letter and commit to all requirements in the letter by signature.

Post-award Stage: CCST requires PIs to update project-level DMP information (described above) and share with the center any changes to DMP during the research period. CCST views DMPs as living documents that change over the lifetime of a research project. DMP updates serve vital knowledge management functions to help on-board new research staff, to help keep the CCST UTC director apprised of project changes, and to give the DOT UTC Program confidence that funded research is being well-managed. Project information should be reviewed

at least quarterly and updated at any point there are changes to the project as awarded, or when project staff change.

Project Completion Stage: PIs submit the final dataset (that could be used to replicate project's analysis and results) to CCST. CCST will conduct an internal review to ensure compliance with DOT requirements. The final version of the datasets should be archived in Zenodo (see section 5 in page 7) no later than 60 days from project completion.

Center-Level Archiving:

During the active phase of projects, CCST website is used to showcase the projects and progress reports. After the end of Center's life, CCST will use Johns Hopkins University Data Repository (JHUDS) to store all center-level data and information including final report, datasets, semi-annual and annual reports and a copy of CCST products.

Long-term archiving of data will be managed by Johns Hopkins Data Services (JHUDS) using the Johns Hopkins Research Data Repository. The Repository (https://archive.data.jhu.edu/), which runs on the Dataverse software, is among the list of DOT-conformant data repositories (https://ntl.bts.gov/ntl/public-access/data-repositories-conformant-dot-public-access-plan). The Repository provides public access to data through an established repository platform supported by storage and preservation practices that follow the Open Archival Information System reference model. The Repository runs on Deposited data is given standard data citations and persistent identifiers (DOIs). JHUDS provides system administration and consultative support for researchers preparing data for deposit. Data will be archived under a memorandum of understanding renewed every 5 years with the PI's consent. CCST opt to use JHUDS due to its responsiveness, excellent support, and customization capabilities.

Project-level Data Management Plan

This document outlines how CCST researchers manage data collection and storage during and after a research project is completed. CCST requires individual researchers (PIs) to write a 2-to-3-page narrative project-level data management plan (DMP) and submit it to CCST along with other requirements of CCST proposal submission. DMP for each proposal will be reviewed for consistency and should be approved before the project's start date. DMP for individual projects consists of three key components including 1) *Data Description, 2) Data Format and Metadata Standards, 3) Policies for Access and Sharing, 4) Policies for Re-use, Redistribution, Derivatives* and *5) Archiving and Preservation Plans.*

PIs will also submit an updated version of DMP and final dataset(s) as part of deliverables (along with final report, policy brief and other project requirements) for each project. The CCST's associate director will review the project-level DMP and data submissions to ensure compatibility with the USDOT Public Access Policy and follow the center's data management plan.

1) Data Description

PIs will use the following framework (and items) to describe the data, users, data collection and processing, etc. in their DMPs:

1. Name the data, the data collection project, or the data-producing program.
2. Describe the purpose of the research.
3. Describe the data that will be generated, in terms of nature and scale (e.g., field measurements, laboratory measurements, modeling, numerical data, image data, text sequences, audio data, video data, database, modeling, archival/textual data, interview, survey, field observation, etc.).
4. Describe methods and procedures for creating the data (e.g., simulated, observed, experimental, software, physical collections, sensors, satellite, enforcement activities, researcher-generated databases, tables, and/or spreadsheets, instrument generated digital data output such as video and images).
5. Discuss the time window over which the data will be collected, and the frequency of update.
6. If using existing data, describe the relationship between the data you are collecting and previously collected data.
7. List potential users of the data.
8. Discuss the potential value the data have over the long-term, both for your project or institution, and for CCST as well as the public.
9. If you request permission not to make data publicly accessible, explain the rationale for lack of public access.
10. Identify the party responsible for managing the data at the project level.

2) Data Format and Metadata Standards

Data gathered from transportation-related research varies and includes, but is not limited to, the following: travel times, vehicle miles traveled, transit ridership, land-use related data, census demographic data, infrastructure financial data, infrastructure sensor locations and information, traveler behavior data, interview data, survey results, government agency documents and minutes, driver behavior data, aerial scans, spatial data and trip generation information.

The data is typically found in the formats listed below:

- MS Excel (.xls)
- MS Powerpoint (.ppt)
- MS Word (.doc)
- Video files (.xlm, .csv, .mpg, .avi, .mov, wmv)
- MS Excel Macro (.xml)
- Comma Separated Values (.csv)
- Portable Document Format (.pdf)
- Joint Photographic Experts Group (.jpg)
- Geographic Information System (.mxd, .lyr, .gdp .shp, .dbf)

PIs will use the following framework (and items) to describe data formats, metadata and

documentations in their DMP report:

1. List the formats of data that were collected and indicate whether they are open access or proprietary.
2. CCST requires that data provided should not be proprietary in nature. If a PI is anticipating using proprietary data formats, the researcher must provide the rationale for doing so, and indicate software that will be able to read the data.
3. Describe how different versions of the data will be identified and/or controlled.
4. If file formats are not standard to transportation, researchers should document the alternative they are using.
5. Provide documentation to make the data understandable by other researchers, including listing the tools used to generate the data.
6. Researchers will indicate what metadata schema they are using to describe the data. If the metadata schema is not one standard for their field, researchers should discuss their rationale for using that approach.
7. Researchers will describe how the metadata were managed and stored during the collection process.
8. Researchers will indicate the tools or software are required to read or view the data.
9. Researchers will describe their quality control measures implemented in their project to ensure its accuracy.

3) Policies for Access and Sharing

Following USDOT guidelines, CCST is committed to the principle that data collected using public money should be available to public and other researchers.

The PIs of each CCST funded project are responsible for the project-specific data management and security during the experimental process. In their final project deliverables, researchers must address the following:

1. Describe what data will be publicly 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:
   o Discuss how will you guard against disclosure of identities and/or confidential business information.
   o List what processes you will follow to provide informed consent to participants.
   o State the party responsible for protecting the data.
3. Describe what, if any, privacy, ethical, or confidentiality concerns are raised due to data sharing.
4. If applicable, describe how you will deidentify your data before sharing. If not:
   o Identify what restrictions on access and use you will place on the data.
   o Discuss additional steps, if any, you will use to protect privacy and confidentiality.

4) Policies for Re-use, Redistribution, Derivatives

The PI of each CCST research project will provide the following information along with the final report:

1. List the names of those who have the rights to manage the data. Data manager roles are assigned by project PI.
2. Indicate who holds the intellectual property rights to the data.
3. List copyrights to the data, if any. If there are copyrights, indicate who owns them.
4. Discuss any rights to be transferred to the data archive. The rights will remain with the data owner(s) unless otherwise noted.
5. Describe how your data will be licensed for reuse, redistribution and derivative products.
   a. Data will use the most appropriate license such as Creative Commons for redistribution and derivative products, where applicable. CCST suggests the PIs to use Creative Commons Zero (CC0) and CC-BY licenses that encourage reuse with the least restrictions.
   b. PIs will be required to cite the data source and license under which they used the data in their project DMPs.

Johns Hopkins University and PI's home institution hold the IP for data created by their CCST-funded projects unless it is specified otherwise by PIs.

All intellectual property rights to the processed data, reports and products that result from projects funded by CCST will be shared between the PI and their institution, and the funding agency (CCST and DOT). Intellectual property rights to raw data will be shared between the PI and CCST. All project reports will be publicly available on CCST's website. Processed data will be made publicly available as specified in the Archiving and Preservation section.

PIs and other authors funded by the CCST may copyright books, publications or other materials developed from CCST/DOT funding, but DOT reserves a royalty-free, nonexclusive and irrevocable license to reproduce, publish, or otherwise use the work for public purpose. PIs can retain the entire right title and interest for each innovation, but DOT must have a nonexclusive, irrevocable, paid-up license to practice the invention throughout the world.

5) Archiving and Preservation Plans

PIs for CCST-funded projects will use Zenodo (https://zenodo.org) for data storing. While JHUDS is our preferred data repository, its service is limited to JHU affiliates. Zenodo is a reputable community-accepted data storage facility and provides access for CCST researchers and PIs from all partner institutions. Zenodo is powered by the invenio open source digital library framework (http://invenio-software.org/) and supported by CERN (http://home.cern/), OpenAIRE (https://www.openaire.edu/), and EU Framework Programme for Research and Innovation (https://ec.europa.eu/programmes/horizon2020/).

CCST has chosen the Zenodo repository for storing data related to CCST funded projects. Zenodo is managed by the CERN Data Center, which has significant experience managing Big

Data as open access. Zenodo does not take ownership of the data. Data at Zenodo is backed up nightly. Other advantages of Zenodo that makes it compliant to the USDOT data repository requirements are:

1. Data in Zenodo is guaranteed for at least 20 years as articulated via Zenodo's Frequently Asked Questions, https://zenodo.org/faq.
2. Datasets in Zenodo are given unique Digital Object Identifiers (DOIs) by DataCite.
3. Zenodo conforms with the National Transportation Library's Guidelines for Evaluating Repositories for Conformance with DOT Public Access Plan, as listed at https://ntl.bts.gov/publicaccess/repositories.html.
4. Researchers can sign into Zenodo with their ORCID (https://orcid.org/) or GitHub (https://github.com/) credentials.
5. Data stored in Zenodo is part of CERN's disk storage service EOS (see http://information-technology.web.cern.ch/services/eo-service).
6. Zenodo is partially an Open Archive Information System (OAIS) model for data archiving (ISO14721 – http://www.iso.org/iso/catalogue_detail.htm?csnumber=57284). Zenodo is working on a Data Seal of Approval compliance (See http://www.datasealofapproval.org/en/).
7. Zenodo provides a wide ranges of user licenses to specify data accessibility. CCST suggests the PIs to use Creative Commons Zero (CC0) and CC-BY licenses that encourage reuse with the least restrictions.

The DMP compliance instructions included in award letter for individual CCST-funded project cover the following information/instructions for data archiving. PIs must acknowledge and commit to all these requirements in the letter by signature.

1. The data will be shared with CCST and will be uploaded to Zenodo before the research project's DRAFT FINAL REPORT is delivered to CCST.
2. The PIs on each CCST funded project should ensure that data to be archived temporarily at their home institution or it is stored securely on a designated device (computer, external hard drive, etc.).
3. The PIs on each CCST funded project should ensure that data collected will be backed up prior to being archived. The scope of work for each project should describe how the PIs intends to prevent loss of data prior to archiving.

Change log:

2023-09-08 0: Original draft

2023-09-19 1: Revised with DOT comments