

---

## **C-TEDD UTC DMP**

*The Data management plan created using the DMPTool*

Center for Transportation Equity, Decisions and Dollars (C-TEDD)

Contact(s): Dr. Shima Hamidi; Robert Edamala, Chief Information Security Officer & Director of Information Security, University of Texas at Arlington

Copyright information: The above plan creator(s) have agreed that others may use as much of the text of this plan as they would like in their own plans, and customize it as necessary. You do not need to credit the creators as the source of the language used, but using any of their plan's text does not imply that the creator(s) endorse, or have any relationship to, your project or proposal.

---

### **Data description**

The C-TEDD UTC will require each individual researcher to submit detailed data descriptions for their individual research projects, following the guidelines below:

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 period of time 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. Identify 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 C-TEDD 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.

Individual researchers are required to write a 2 to 3 page narrative project-level data management plan and submit it to the repositories as a deliverable along with their research proposal. The C-TEDD's program manager will review the project-level data management plans to ensure they meet the requirements of the USDOT Public Access Policy and are in compliance with the center's data management plan.

### **Types of Data, 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 data, 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 of each C-TEDD research project will be required to include the following information in their final report:

1. List the formats(s) of data that were collected, and indicate whether they are open or proprietary.
2. C-TEDD 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 list 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 what 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.

### **Policies for access and sharing**

Following USDOT guidelines, C-TEDD is committed to the principle that data collected using

public money should be available to the public, and to other researchers.

The PI of each C-TEDD funded project is responsible for how the project-specific data is managed and secured during the experimental process. In their final project deliverable, researchers must address the following:

1. List the roles played by each team member in data management during the project span, including any limitations on team member access due to the presence of personal or confidential information
  - a. The data management librarian at the University of Texas, Arlington will assist C-TEDD and answer any questions that arise from consortium members with data management during the data lifecycle (e.g. creation, description, accession, preservation), including but not limited to training, where appropriate and applicable.
2. State whether the data can be shared with the public.
  - a. Data will be accessible via an open access repository with open archive initiative protocol, or similar, for metadata harvesting capabilities (OAI-PMH).
3. Describe what data will be shared, how data files will be shared, and how others will access them.
  - a. Open data will be shared through a general data repository. Metadata may also be shared through the institutional repository. Raw datasets will be stored and accessed through discipline-specific repositories, where available.
4. Do your data contain private or confidential information? If they do:
  - a. Investigators will need to discuss how they will guard against disclosure of identities and/or confidential business information.
  - b. Each PI will be required to follow their institution's established information security program and commit to securing the data consistent with least-privilege principles and best practices following the PI's institutions data classification standard. The PI will consult the PI's regulatory services department to ensure compliance with Institutional Review Board (IRB) protocols for safeguarding identities and/or protecting national security information as well as ensuring compliance with export control requirements.
  - c. Investigators need to state the party responsible for protecting the data.
  - d. Investigators need to list what processes they will follow to provide informed consent to participants.
  - e. Private and restricted data will be de-identified before deposit in either a general and/or institutional repository, if applicable.
  - f. The PI is responsible for managing the data
  - g. Researchers need to describe any privacy, ethical, or confidentiality concerns due to sharing of data.
    - i. Privacy, ethical, or confidentiality concerns of sharing data will adhere to current institutional, university and funding agency guidelines and policies.
  - h. If applicable, researchers need to describe how they will de-identify their data before sharing. If this is not applicable to your project:
  - i. Identify what restrictions on access you will place on the data. The PI will be required to implement access control to restrict data access following a

- least privilege model.
- ii. Discuss any additional steps you will use to protect privacy and confidentiality.
- iii. The PI will be required to adhere to their institution's information security policy to protect the data.

### **Policies for re-use, redistribution, and derivatives**

The PI of each C-TEDD 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.
  - a. The person responsible for managing the data (i.e. the data manager) is selected by the PI.
2. Tell us who holds the intellectual property rights to the data.
  - a. The data creator, institutional and/or funder holds intellectual property, where articulated and/or stipulated.
3. List copyrights to the data, if any. If there are copyrights, indicate who owns them.
  - a. The data creator, institutional and/or funder holds intellectual property, where articulated and/or stipulated.
4. Discuss any rights to be transferred to the data archive.
  - a. 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.
  - b. Researchers will be required to cite the data source and license under which they used the data in their project DMPs.

The University of Texas, Arlington, or the home institution of the project researchers, holds the IP for data created by C-TEDD-funded projects. Researchers will be required to describe if they are transferring rights to the data archive at C-TEDD. If investigators do not specify this, their home institution maintains rights to their data.

All intellectual property rights to the processed data, reports and products that result from projects funded by C-TEDD will be shared between the PI and their institution, and the funding agency (C-TEDD and DOT). Intellectual property rights to raw data will be shared between the PI and C-TEDD. All project reports will be publically available on C-TEDD's website. Analyzed data will be made publically available upon written request.

PIs and other authors funded by the C-TEDD may copyright books, publications or other materials developed from C-TEDD/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.

## Plans for secure archiving and preservation

Under the current plan, C-TEDD will use the Zenodo repository for storing its data. 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/>).

Zenodo (<https://zenodo.org>)

1. The data will be uploaded to Zenodo before the research project's DRAFT FINAL REPORT is delivered to the C-TEDD program manager.
2. The PI on each C-TEDD funded project should ensure that data to be archived temporarily at their home institution is stored securely on a designated device (computer, external hard drive, etc.).
3. The PI on each C-TEDD 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 PI intends to prevent loss of data. (Will data be backed up? How often? Will the data be managed off site and how will you prevent it from being lost? What security measures will be implemented?)
4. PIs must also describe how data will be protected from accidental or malicious modification or deletion prior to receipt by the archive.
5. As indicated above, C-TEDD has chosen the Zenodo repository for storing data related to C-TEDD 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.
6. Data in Zenodo is guaranteed for at least 20 years as articulated via Zenodo's Frequently Asked Questions, <https://zenodo.org/faq>.
7. Datasets in Zenodo are given unique Digital Object Identifiers (DOIs) by DataCite.
8. 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>.
9. Researchers can sign into Zenodo with their ORCID (<https://orcid.org/>) or GitHub (<https://github.com/>) credentials.
10. Data stored in Zenodo is part of CERN's disk storage service EOS (see <http://information-technology.web.cern.ch/services/eo-service>).
11. Zenodo is partially an Open Archive Information System (OAIS) model for data archiving (ISO14721 – [http://www.iso.org/iso/catalogue\\_detail.htm?csnumber=57284](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/>).
12. Project bi-annual and final Reports will be publically available on C-TEDD's website. Analyzed data and data deposits following publication of peer-reviewed journal articles will be made available upon written request to C-TEDD. At the end of the center's life, copies of all processed data will be uploaded to Zenodo.