University Transportation Center for Underground Transportation Infrastructure (UTC-UTI)
DATA MANAGEMENT PLAN (DMP)

Creator(s): Marte Gutierrez

Affiliation: Colorado School of Mines

Consortium Members: California State University Los Angeles and Lehigh University

Last modified: February 27, 2017

1. DATA DESCRIPTION

Data types and sources include all data generated or used through the course of the research conducted and supported by UTC-UTI, and which are considered digital research data necessary to test, support and validate the research findings. These data will include collated and synthesized data sets from literature search and surveys, field monitoring, laboratory experiments, as well as relevant and important model and simulation input data and results. To ensure quality and integrity, data collection will be performed according to domain best practices.

The UTC-UTI will require each individual researcher to submit detailed data descriptions for their individual research projects per this plan as outlined in the guidance.

1) Name the data, data collection project, or 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., numerical data, image data, text sequences, video, audio, database, modeling data, source code, etc.).
4) Describe methods 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 images and video; etc).
5) Discuss the period of time data will be collected and frequency of update.
6) If using existing data, describe the relationship between the data you are collecting and existing data.
7) List potential users of the data.
8) Discuss the potential value of the data have over the long-term for not only your institution, but also for the public.
9) If you request permission not to make data publicly accessible, explain rationale for lack of public access.
10) Indicate the party responsible for managing the data.
11) Describe how you will check for adherence to this data management plan.

2. DATA FORMAT AND METADATA STANDARDS

Data from literature survey and review will be in stored as Excel Workbook and Spreadsheets.

Modeling/simulation input data will be stored in ASCII or RTF (Rich Text Format) batch file format directly readable by the modeling/simulation code. Modeling/simulation results will be stored in ASCII or RTF files as directly generated by the modeling/simulation code.

Reports, papers, graphs, figures, dissertations and other publications will be stored as Adobe PDF files.

Paper records including laboratory notebooks, raw data, hand drawings and sketches, and tabulations of data and results will be scanned and converted to electronic form and archived with other electronic data collected as Adobe PDF files.

1) All Researchers will be required to have final datasets that are not propietary in the standard data format of the field such as csv, txt, etc.
2) If Researchers are using proprietary data formats, they will be required to discuss their rationale.
3) Researchers be required to describe the data process log to clarify the final version of data shared to the public.
4) Researchers will be required to describe how they will document the alternative formats they are using and why.
5) Researchers will list what documentation they will be creating in order to make the data understandable by other researchers.
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, and discuss their rationale for using that scheme.
7) Researchers will have to describe how the metadata will be managed and stored.
8) Researchers will indicate what tools or software is/are required to read or view the data.
9) Researchers will describe their quality control measures.

3. POLICIES FOR ACCESS AND SHARING

Researchers will be required to address any access restrictions in the project DMP they submit to UTC-UTI.

For project DMPs, Researchers will address issues and outline the efforts they will take to provide informed consent statements to participants, the steps they will take the protect privacy and confidentiality prior to archiving their data, and any additional concerns (e.g., embargo periods for your data). If necessary, they will describe any division of responsibilities for stewarding and protecting the data among other project staff.

If Researchers will not be able to de-identify the data in a manner that protects privacy and confidentiality while maintaining the utility of the dataset, Researchers will describe the necessary restrictions on access and use.

If an individual research project includes human subject research, Researchers will be required to go through Colorado School Mines Institutional Review Board (IRB) or their home institutions IRB, if they have one. Researchers will be required to address the following:

1) Describe what data will be 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

a. Discuss how you will guard against disclosure of identities and/or confidential business information
b. State the party responsible for protecting the data.
c. List what processes you will follow to provide informed consent to participants.

3) Describe what, if any, privacy, ethical, or confidentiality concerns are raised due to data sharing.
4) If applicable, describe how you will de-identify your data before sharing. If not:
   a. Identify what restrictions on access and use you will place on the data.
   b. Discuss additional steps, if any, you will use to protect privacy and confidentiality.

4. POLICIES FOR RE-USE, REDISTRIBUTION, DERIVATIVES

Colorado School of Mines or the home institution of the Researchers holds the IP for data created by the project.

Researchers will be required to describe if they are transferring rights to the data archive, if they do they do not describe this, the home institution maintains the rights.

Researchers will be required to cite the data source and license under which they used the data in their project DMPs.

In general, Researchers will address the following in their project DMPs:

1) Name who has the right to manage the data.
2) Indicate who holds the intellectual property rights to the data.
3) List any copyrights to the data. If so, indicate who owns them.
4) Discuss any rights be transferred to a data archive.
5) Describe how your data will be licensed for reuse, redistribution, and derivative products.

5. PLANS FOR ARCHIVING AND PRESERVATION

Plans for archiving will support the capture and provision of the US Federal Government Project Open Data Metadata Schema. In addition, the archive will support the creation and maintenance of persistent identifiers (e.g., DOIs, handles, etc.) and will provide for maintenance of those identifiers throughout the preservation lifecycle of the data.

1) UTC-UTI will archive all publications and data on CERN's Zenodo, https://zenodo.org/communities/utc-uti, which is conformant with US DOT guidelines as described at https://ntl.bts.gov/publicaccess/evaluatingrepositories.html.
2) When a project submits a final report, the Researchers will have 60 days to archive their data on Zenodo.
3) Researchers will maintain and back-up data until it is uploaded to Zenodo.
4) Zenodo's procedures and policies for back-up, data recovery, retention, security and integrity are outlined in https://zenodo.org/policies.
5) Zenodo provides how back-up, disaster recovery, off-site data storage, and other redundant storage strategies will be used to ensure the data's security and integrity.
6) Zenodo will retain data for the lifetime of the repository. This is currently the lifetime of the host laboratory CERN, which currently has an experimental program defined for the next 20 years at least.
7) Each data upload in Zenodo gets a Digital Object Identifier (DOI) to make them easily and uniquely citable.

8) PIs shall copy and paste the language in this section into their project DMPs, unless they are using a different archive.