Southeastern Transportation Research, Innovation, Development & Education Center Data Management Plan

A Data management plan created using the DMPTool

Creator(s): Ines Aviles-Spadoni, Plato Smith, Lily Elefteriadou

Affiliation: University of Florida

Last modified: March 08, 2017

Southeastern Transportation Research, Innovation, Development & Education Center Data Management Plan

Data description

The STRIDE Center requires that PIs for all funded projects submit detailed data descriptions upon completion of the research. This requirement is included in the scope of work of each project.
This deliverable should include the following elements:

1. A list of all data obtained, along with the data producing program, if applicable.
2. Describe the purpose of the research.
3. Describe the data your project will generate 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 were collected and the frequency of any updates, if applicable.
6. If the project used existing data, describe the relationship between the data collected under the project and previously collected data.
7. List potential users of the data.
8. Discuss the potential value that the data generated by your project will have over the long-term for the STRIDE Center and for the public.
9. If you request permission not to make data publicly accessible, explain the rationale for lack of public access.
10. Indicate who will be responsible for managing the data at the project level.

The STRIDE Center's research coordinator will monitor all research project final reports to ensure they have addressed all of the above questions.

Data format and metadata standards

Data gathered from transportation-related research varies and includes, **but is not limited to** the following: speeds, volumes, travel times, vehicle miles traveled, transit ridership, incidents, crashes, signal timings, video logs, land-use related data, infrastructure sensor locations and data, traveler behavior data, driver behavior data, and trip generation information.
The data is typically found in the formats listed below:

- MS Excel (.xls)
- Video files (.xml, .csv, .mpg, .avi, .mov, .wmv)
- MS Excel Macro (.xml)

Comma Separated Values (.csv)

- Portable Document Format (.pdf)
- Joint Photographic Experts Group (.jpg)

PIs of each STRIDE research project will be required to include the following information in their final report:

1. List in what type of format(s) the data were collected, and indicate if they are open or proprietary.
2. Data should not be proprietary in nature. If a PI is anticipating using proprietary data, they must explain their rationale in doing so.
3. Describe how versions of the data will be signified and/or controlled.
4. If the file formats are not standard to transportation, document the alternative you are using.
5. Provide documentation to make the data understandable by other researchers.
6. Indicate the type of metadata schema you are using to describe the data. If the metadata schema is not one that is standard for your field, explain your rationale for using that specific scheme.
7. Describe how the metadata were managed and stored during the collection process.
8. Indicate what tools or software is required to read or view the data.
9. Describe the quality control measures you implemented in your project to ensure its accuracy, etc.

Policies for access and sharing

The PI of each STRIDE funded research project is responsible for how the project-specific data is managed and secured during the experimental process. In your final project deliverable provide the following:

1. List the roles the data creation team members on your project have had in data management, including any limitations on team member access due to the presence of personal or confidential information.
    - The Data Management Librarian at the University of Florida will assist STRIDE and its consortium members with the management of data throughout the data lifecycle (e.g. creation, description, accession, preservation) including but not limited to training, where appropriate and applicable.The data management librarian has a Ph.D. in Information Science and experience as department head for a digital library center at an academic research library.
2. State whether the data can be shared with the public.
    - Data will be accessible via an open access repository with open archive initiative protocol for metadata harvesting (OAI-PMH) capabilities.
3. Describe what data will be shared, how data files will be shared, and how others will access them.
    - 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. Does your data contain private or confidential information? If they do:

- Tell us how you will guard against disclosure of identities and/or confidential business information.

Private and restricted data will be de-identified before deposit in either a general and/or institutional repository, if applicable. ResearchVault (an FISMA rated UF computing resource) is a secure storage computing environment for encrypting, managing, working with restricted data available to STRIDE.

- Describe the processes you will follow to provide informed consent to participants.
  - Informed consent will provide at the point of data capture.
- State the party responsible for protecting the data.
  - The PI is responsible for managing the data.

1. Describe if there are any privacy, ethical, or confidentiality concerns due to the sharing of data.
   - Privacy, ethical, or confidentiality concerns of sharing data will adhere to current institutional, university, and funding agency guidelines and polices.
2. If applicable, describe how you will de-identify your data before sharing. If this is not applicable to your project, then:

- Identify what restrictions on access and use you will place on the data.
- Discuss additional steps that you will use to protect privacy and confidentiality.

Policies for re-use, redistribution, derivatives

The PI of each STRIDE 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.
   - The person responsible for managing the data or data manager is selected by the PI.
2. Tell us who holds the intellectual property rights to the data.
   - 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
   - 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.
   - The rights will remain with data owner(s) unless otherwise noted.
5. Describe how your data will be licensed for reuse, redistribution and derivative products.
   - Data will use the most appropriate license such as Creative Commons for redistribution and derivative products, where applicable.

Plans for archiving and preservation

1. The STRIDE Center 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.eu/), 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 STRIDE research coordinator.
2. The PI on each STRIDE-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 STRIDE-funded project should ensure that the 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 the data be backed up? How often, will the data be managed off-site and how will you prevent it from being lost, and the security measures that 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, the STRIDE Center has chosen the Zenodo repository for storing data related to STRIDE-funded projects. Zenodo is managed by the CERN Data Center, and it has a lot of experience with managing Big Data as open access. Zenodo does not take ownership of the data. Data at Zenodo is backed up each night.
6. Data in Zenodo is guaranteed for at least 20 years as articulated via Zenodo FAQ, https://zenodo.org/faq.
7. Datasets in Zenodo are given unique Digital Object Identifiers (DOIs) by DataCite (https://www.datacite.org/). DOIs enable data access, linking, and sharing between publication systems.
8. Zenodo conforms with the National Transportation Library's Guidelines for Evaluating Repositories for Conformance with the 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/eos-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). Zenodo is working on a Data Seal of Approval (See: http://www.datasealofapproval.org/en/) compliance.