TRANSPORTATION INFRASTRUCTURE DURABILITY CENTER

DATA MANAGEMENT PLAN

A Data Management Plan created using DMPTool

Creator: Amanda Collamore, TIDC Communications Education, Outreach, & Reporting Manager

Affiliation: University of Maine System (UMS)

Template: Department of Transportation (DOT)

Last modified: 11-30-2018

TRANSPORTATION INFRASTRUCTURE DURABILITY CENTER

DATA MANAGEMENT PLAN

DATA DESCRIPTION

Each research project will generate data unique to the research being conducted. Generally, projects will generate quantitative data from field performance and laboratory tests. Data generated will include, but is not limited to, test samples, physical collections, experimental measures, computer and model simulations, photos, videos, and PI notes. Data gathered during the course of this research will be converted into standard, readily accessible formats including, but not limited to, pdf documents, plain text files, jpeg image files, and mpeg videos.

Under this Data Management Plan (DMP), each project is expected to summarize what data is anticipated to be generated, the plan for data storage and archival, and how data will be made readily available after the project is complete.

TIDC's DMP is written with the University of Maine's (TIDC's lead University) policies as the primary reference source. All member Universities have similar policies, standards, and plans in place. To assist with creating DMPs, The University of Maine (https://acg.umaine.edu/resources/data-management/data-management-plans/), the University of Connecticut (https://lib.uconn.edu/services/research-data/consulting/), the University of Massachusetts Lowell (http://libguides.uml.edu/c.php?g=560720&p=3858362), the University of Rhode Island (https://uri.libguides.com/datamanagement/plan), and the University of Vermont (http://researchguides.uvm.edu/datamanagement/DMP) utilize the DMPTool (a service of the University of California Curation Center of the California Digital Library). Western New England University has an internal data management plan, with all of their data stored on department and college-wide servers, as well as the research group's servers.

DATA FORMAT AND METADATA STANDARDS

Data will be organized into individual experiments containing question or rationale behind performing a particular experiment, the method used to perform the experiment, the results obtained, and the meaning of the results. This will be stored as a hard copy (laboratory notebook) in the office of the PI, as well as a digital copy of the PI's office computer and backed up on an external hard drive. The raw data will also be stored digitally in these locations and organized in folders referencing the data generator, the project title, a short description of the experiment, and a reference to the digital hard copy laboratory notebook where the originally recorded information can be found.

The project will create documentation detailing the sources, coding, and editing of all data, in sufficient detail to enable another researcher to replicate them from original sources; and descriptive metadata for each study including a title, author, abstract, descriptive keywords, and file descriptions. The project will include bibliographic information for any publication by the

project based on that data. For scientific data, we will deploy netCDF, a data format that is self-describing and machine-independent. For non-scientific sound (in WAV), video, or images separate from the documentation deposited as JPEG 2000 files (with lossless compression) or uncompressed TIFF files. Documentation will be deposited in PDF/a, or plain-text formats, to ensure long-term accessibility, such as a readme file in each directory to describes detailed metadata information, such as data formats, access methods, usage, authors, notes, and revision history.

Immediately after collection, quantitative data will be converted to Stata, SPSS, R, GraphML, and/or CSV formats. These formats are fully supported by the Maine Dataverse Network (MDVN), which performs archival format migration; metadata extraction; and validity checks. Deposit in these formats will also enable on-line analysis; variable-level search; data extraction and re-formatting; and other enhanced access capabilities.

The MDVN repository system's "templating" feature will be used for consistency of information across studies. The MDVN system automatically generates persistent identifiers, and Universal Numeric Fingerprints (UNF) for studies; extracts and indexes variable descriptions, missing-value codes and labels; creates variable-level summary statistics; and facilitates open distribution of metadata with a variety of standard formats (DDI v 2.0, Dublin Core, and MARC)and protocols (OAI-PMH and Z39.50).

POLICIES FOR ACCESS AND SHARING

All data will be available through request of the PI. Summarized results will be made publicly available through local presentations and reports as well as scholarly publications. All applicable University of Maine and research ethics protocols will be followed. Final, peer-reviewed manuscripts accepted for publication and other media-based deliverables will be uploaded to DigitalCommons@UMaine in electronic format for free, worldwide dissemination via open access. DigitalCommons@UMaine (http://digitalcommons.library.umaine.edu/), the University's institutional repository (IR) will provide long-term, stable delivery of final, resulting, peer-reviewed manuscripts accepted for publication in peer reviewed journals.

DigitalCommons@UMaine software is a product of bepress of Berkeley, CA. The service offers unlimited storage with offsite backups of files on Amazon Glacier as well as quarterly fullsite archives. Long-term accessibility to resulting, peer-reviewed content will be facilitated by bepress' (Open Access Scholarly Publishing Services) commitment to the provision of long-term stable URLs and the preservation of all objects uploaded to the repository in the original format. In addition, the service is committed to making PDF documents web-accessible on a permanent basis.

All TIDC research projects will be submitted to the Transportation Research Board's (TRB Research in Progress (RiP) database within one month of the project's selection. Final reports for all research projects will be published on TIDC's website (tidc-utc.org) in an easily accessible, searchable format. Final Research documents will also be distributed to the following entities, as required by the Grant Deliverables and Reporting Requirements, (1) the Transportation Library, (2) the Volpe National Transportation Systems Center, (3) the Federal Highway Administration Research Library, and (4) the National Technical Information Service within two months of the completion of the research. In addition to sending the final research documents to the required

entities, TIDC will change the status of the project from "active" the
"completed" in the RiP database. TIDC will also submit railroad related research findings to the
Federal Railroad Administration eLibrary (https://www.fra.dot.gov/eLib/Find).

POLICIES FOR RE-USE, REDISTRIBUTION, DERIVATIVES

The PI and/or the University of Maine will retain the right to decide when, where, and to who the
data may be re-used or redistributed. This includes both to companies who may be interested in
the outcome of the experiments for product development and improvement or private individuals
engaged in research or technology transfer.

PLANS FOR ARCHIVING AND PRESERVATION

During the execution of this project, all primary data will be placed on an external hard drive and
stored in the office of the PI. In addition, the Advanced Computing Group (ACG) at the
University of Maine will be used to provide cloud computing and storage capabilities for this
project. As such, all data produced will be stored on multiple redundant disk arrays housed in a
secure, climate-controlled data center with backup power. Operation and maintenance of the
machines as well as backups of all data is provided by the ACG in order to ensure that all data
are continuously available. Both an internal firewall and an external firewall are in place to
ensure the security of the data.
The archival and preservation of all data will be managed by ACG@UMaine. The
ACG@UMaine primary data repository is the Maine Dataverse Network (MDVN). The MDVN
is hosted by the ACG@UMaine on a secured shared storage array acquired under the funding
support of a previous NSF MRI grant. The storage array is fast and reliable, and it has a total of
120 TB of raw storage interconnected with multiple 4 Gb/s Fiber Channel paths and offsite
backups.
The ACG@UMaine commits to good archival practice including 24x7 machine operation
support versioning and deaccession compliant, regular off-site backup, and regular content
migration to ensure that data are available for access consistently and kept secure. All data will
be stored, during and post project execution, within campuses network infrastructure, which
employ firewalls and secure authentication and authorization methods for login and access. Data
deposited into the MDVN will be retained indefinitely, never destroyed (unless required by law),
backed up on a daily basis and replicated across multiple locations for long-term access. Access
to data in the long-term will be provided in accordance with the permissions, terms of use,
policies and procedures established by the PIs at the time of archival.
The ACG@UMaine services include personnel training and support for all technical aspects of
the data produced during the course of this project necessary to maintain security, dissemination,
andpreservation. Should the archiving entity be unable to perform, transfer agreements with the
University of Maine System are in place to easily migrate to another entity within the university
system.