Mineta Consortium for Transportation Mobility: Data Management Plan

Data description

The Mineta Consortium for Transportation Mobility (MCTM) will conduct research, education, workforce development, and technology transfer activities to improve mobility of people and goods and make our nation's transportation system safe, efficient, accessible, and convenient for all. The Consortium will conduct research to achieve the following goal and objectives:

Goal: Lead the nation in research that identifies safe, reliable transportation solutions that increase mobility of people and goods and strengthen the nation's economy

- *Objective 1:* Leverage new technologies and innovative processes to achieve a seamless, multimodal surface transportation system that integrates with other "smart city" investments.

- *Objective 2:* Create a safer, more reliable, and more resilient surface transportation system that improves equity through increased access to jobs, housing, services, and other opportunities.

- *Objective 3:* Reduce the impact of transportation on climate change by identifying feasible alternative modes and fuels and effective ways to reduce vehicle miles traveled.

- *Objective 4:* Extend surface transportation access to people of all abilities and socioeconomic levels, connecting people to where they live, work, and play.

- *Objective 5:* Optimize passenger and freight movements to improve mobility of people and goods through development of more accurate data models and advanced application of analytical tools.

Data developed as part of MCTM's research activities will include the following:

- Surveys of individuals, agencies and/or organizations;

- Geographic data;

- Data developed as a result of laboratory/fields experiments focused on research topics identified above. These data may include, for example: photos, videos, spreadsheets of experiment/laboratory results, etc.

Data will be collected throughout the duration of the five-year grant. The long-term value of the data will be to provide transportation decision-makers, the general public, and others interested in transportation research the ability to replicate research findings or utilize the data (with attribution) in subsequent studies.

Researchers funded through this grant will be required to indicate compliance with the data management plan prior to approval of any individual research project. Researchers funded through the grant will be expected to manage the data according to the approved IRB protocol prior to submission to SJSU ScholarWorks (scholarworks.sjsu.edu). SJSU ScholarWorks currently publishes reports for the Mineta Transportation Institute (http://scholarworks.sjsu.edu/mti_publications/). Once publications have been uploaded to the site, the accompanying data sets will be sent to ScholarWorks staff (scholarworks@sjsu.edu) or Emily Chan, Scholarly Communications Librarian (emily.chan@sjsu.edu) for repository inclusion. Management of the data will be in accordance with the use agreement for the project. Each record/report in SJSU ScholarWorks will specify these data use details.

Data format and metadata standards

Data from MCTM research activities will be primarily stored in standard, accessible data formats including, but not limited to: .csv, .txt, .rtf, .doc, .xls, etc. Geographic data will be stored in common formats including, for example, geodatabases, raster files, etc. Image files will be stored in common formats including, for example, .tif, .jpg, etc. While every effort will be made to archive data in open formats, proprietary formats will be utilized when those are the best way to present the data to retain the information or when the proprietary format (e.g. .doc) is commonplace.

Data will be anonymized and stripped of personally identifiable information in accordance with San José State's University policies, specifically the Policy for Protection of Human Research Subjects (http://www.sjsu.edu/senate/docs/S08-7.pdf). SJSU's Institutional Review Board has a federal-wide assurance through the U.S. Department of Health and Human services and complies with all federal regulations for research involving human subjects.

Metadata for each dataset will be provided following the standards set forth in the Project Open Data Metadata Schema v1.1 (https://project-open-data.cio.gov/v1.1/schema). Criteria included will focus on authorship, subject, scope, as well as spatial or temporal extent of the data.

Metadata about the collected project data will be represented in a either a JSON file or as a README.TXT file. The JSON or README.TXT file would accompany the dataset to provide maximum context.

Metadata standards, schema, and identifiers would be indicated in the citation record in ScholarWorks.

Data will be stored in a publicly accessible archive (San José State University's ScholarWorks repository: http://scholarworks.sjsu.edu/) and users will be able to download the data to their own computers for use. Direct viewing of the data within the repository will not be available.

All funded research will undergo peer review conducted by two academics and one practitioner.

Policies for access and sharing

All MCTM research that involves human subjects will be required to obtain approval from San José State University's Institutional Review Board (or equivalent entity from partner institutions). The University's Institutional Review Board ensures that the research is in compliance with university policy and federal regulations established to ensure the safety of research participants and the ethical and responsible conduct of investigators. Research that involves human subjects and collects data with any personally identifiable information must document how individual privacy is protected. Researchers are expected to use protocols such as pseudonyms or coding systems to maintain confidentiality. Researchers must specify in their IRB application how data is collected and kept safe (e.g. encrypted files, password-protected computer, locked cabinet, etc.), who has access to data, and a retention plan for the data. In addition, researchers will be expected to document how that data will protect privacy and confidentiality while maintaining the use of the dataset for final distribution.

Policies for re-use, redistribution, derivatives

Intellectual property rights for the data will comply with San José State policy (http://www.sjsu.edu/senate/docs/F98-3.pdf). As such, ownership of copyrightable works are owned by the authors (e.g. SJSU faculty, staff, or students). Data produced under the auspices of this grant and subject to the requirement to post the data to a publicly accessible repository will utilize a CC-BY-NC license. This Creative Commons license requires attribution to the original author and allows the distribution, remix, re-use, and derivative works, as long as it is not for commercial purposes. The CC-BY-NC license will be noted in the record of the data posted to the online repository along with contact information for the corresponding author who can field inquiries regarding the data and its ownership.

Plans for archiving and preservation

MCTM data will be archived through San José State University's ScholarWorks repository. ScholarWorks is a data repository conformant with the U.S. Department of Transportation's Public Access Plan.

Prior to submission to ScholarWorks, all data will be stored by individual principal investigators in a manner compliant with the approved IRB (for human subjects

research). For submission to ScholarWorks, the data will be described and identified according to this data management plan.

SJSU Library contracts with bepress (https://www.bepress.com/) to host SJSU ScholarWorks. bepress uses cloud-based technology to provide the Digital Commons platform (https://www.bepress.com/products/digital-commons/), which is arguably the world's leading hosted repository platform. bepress highlights these key aspects to their preservation and security plan (https://www.bepress.com/reference_guide_dc/safeguarding-content-digital-commons/):

- All of the production servers are maintained at a high availability colocation facility with multiple backbone connections and backup generators or in redundant Amazon Web Services availability zones. The facility is secure and requires physical tokens (badges) and biometric identification for access.
- bepress maintains failover web, database, and storage servers to continue to serve content in case of failures.
- The databases have real-time redundancy that runs continuously, and bepress takes full nightly backups of the entire database. The nightly backups are stored away from the colocation facility in a separate physical location.
- The storage server for the production files has full hardware redundancy. brpress creates hourly and daily snapshots and offsite backups of the production filesystems. In this case, 'offsite' means offsite from the co-lo facility. Production files are mirrored every half hour, databases every 2 hours. A second copy of offsite backups is sent to Amazon S3. (Amazon S3 data is redundantly stored across multiple Amazon facilities and multiple devices in each Amazon facility.)
- All of the uploaded files are stored in triplicate in the redundant storage cluster, as well as backed up offsite to a third-party cloud service, Amazon Glacier, that specializes in data archiving and backup. Glacier performs regular, systematic data integrity checks and is built to be automatically self-healing.