Plan Overview

A Data Management Plan created using DMPTool

Title: Qian_DDTA

Creator: Sean Qian - **ORCID:** [0000-0001-8716-8989](0000-0001-8716-8989)

Affiliation: Carnegie Mellon University (CMU)

Principal Investigator: Sean Qian

Data Manager: Sean Qian

Funder: United States Department of Transportation (DOT) (transportation.gov)

Funding opportunity number: 7/1/17-6/30/18

Template: U.S. Department of Transportation Public Access Guidance v1

Last modified: 10-08-2017

Copyright information:

Qian_DDTA

The main objective of this project is to develop theories, algorithms and system to utilize large-scale data for holistic decision making of multi-modal transportation systems. We propose to develop an integrative data platform and decision making systems for public agencies. The core is a centralized data engine supported by a web application to manage and analyze data across the Pittsburgh region. The centralized data engine, which essentially sets protocols for data exchange from various sources and translate them to information we need, is necessary to accommodate the needs of data fusion and analytics. The data will be stored and managed in distributed servers across CMU and those agencies who own the data. The data engine offers organization, visualization and analytics of a wide array of mobility data. Unlike the traditional single computer stand-alone software or tools for data preparation and system design, the data engine relies on browser-based human-computer interaction. The web application visualizing data and recommending decisions serves the front end of the data engine.

Where not in violation of agreements made with third parties in order to collect data from their systems, all the data related to project will be made available to the research community through the same web application designed for this project.

While integrating data extracted from various entities, such as Pennsylvania Department of Transportation, Port Authority of Allegheny County, etc., the raw data will first be converted to mySQL database file for cleansing, fusion and processing. The data in mySQL are further imported into OpenStreetMap for visualization. The PI and research assistants will code algorithms using ANSI C++ that is compatible for any platform. The output data will mainly be temporal-spatial flow and system performance metrics. They can be provided in the standard .cvs format or GIS formats to any interested party via the web application.

Throughout the duration of the proposed work, the PI will in a timely manner communicate any significant findings with the scientific community in accordance with USDOT policy through journal publications, national and international conference presentations, and seminars. The reported results will be made available to the research community, where possible and permitted and upon request.

When sharing the data, the PI commits to protect privacy, confidentiality, and security. If the research and discoveries found during the project might be secured with intellectual property, the PI will work with our respective Technological Transfer Offices to protect potential proprietary data. In addition, the PI will not post to any publicly available site any raw data that are not permitted to share, especially under data usage agreement with private sector.

Data derived from this project shall be retained for at least one year. The tool developed in this project will be open source and shared along with research results to research community through the web designed for this project. The data in this project does not contain private or confidential information.

Any used data will reside on PCs and workstations belonging to the PI's university. All data will be periodically and systematically backed up either onto multiple external hard drives, or a centralized backup cloud through the university, to ensure full data recovery in the event of equipment failure.