**Data description**

This research aims at understanding the impact of ride sourcing services to the urban mobility, which leads to fundamental knowledge for next-generation traffic management and infrastructure design. The result includes the prediction of travel time, travel delay, vehicle-mile-traveled and emissions for both recurrent vehicles and ride-sourcing vehicles, at each road segment and intersection by time of day. The core is a centralized data/simulation engine to manage and analyze data collected from multiple data sources. The data will be stored and managed in distributed servers across CMU. The data/simulation engine offers organization, visualization and analytics of a wide array of mobility data.

**Data format and metadata standards**

While integrating data extracted from various entities, such as traffic counts data, traffic speed data, vehicle trajectories data 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 PIs and research assistants will code algorithms using ANSI C++ that is compatible for any platform. The output data will mainly be system performance metrics. The aggregated data, without any personally identifiable information, can be provided in the standard .cvs format or GIS formats to any interested party via the web application.

**Policies for access and sharing**

Throughout the duration of the proposed work, the PIs 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 PIs commits to protect privacy, confidentiality, and security. If the research and discoveries found during the project might be secured with intellectual property, the PIs will work with our respective Technological Transfer Offices to protect potential proprietary data. In addition, the PIs 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.

**Policies for re-use, redistribution, derivatives**

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.

**Plans for archiving and preservation**

Any used data will reside on PCs and workstations belonging to the PIs’ 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.