**Data description**

Four types of data will be obtained, processed and analyzed in this project: 1) GIS map data for understanding the network topology and work zone locations/time data; 2) traffic crash data for analyzing the impact of safety; 3) other data sources for categorizing the safety risks, such as INRIX time-varying speed, weather conditions; and 4) Census data and points of interest at the block level for the State of Maryland. The data will be stored and managed in distributed servers across CMU. The data 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 Census data, safety data, traffic counts data, traffic speed data, work zone data, Waze 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 PI 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 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.

**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 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.