

Title

Smart and equitable parks: quantifying returns on investments based on probabilistic mobility-dependent correlates of park usage using cyber-physical system technologies

Data management plan:

Types of data:

The research team will obtain, process, and analyze several types of data including: 1) GIS map data for understanding community network topologies and transportation properties; 2) traffic use data (for both public transportation and walkable/bikable pathways) measured using *Urbano* nodes (configured with passive infrared sensors), as well as traffic data, under different circumstances; 3) environmental data measured using *Urbano* nodes (configured with NO₂, SO₂, O₃, PM_{2.5}, and noise intensity sensors) at parks to inform nodal weights within the network. The team will partner with the City of Pittsburgh and the Pittsburgh Parks Conservancy to implement the proposed framework within a Pittsburgh community and collect all in situ measurements from a distributed network of *Urbano* nodes. Due to the investment of time and resources needed to collect data, it will be of the highest priority to ensure data quality, protect collected data, clearly communicate results derived from the data, and maintain the data in archival storage (all discussed subsequently). The data will be stored and managed in a cloud-based server and backed up on a local server at CMU. Looking at the data flow between *Urbano* nodes and the cloud, the Amazon Web Services (AWS) IoT Core will be used to connect *Urbano* nodes with the cloud, AWS Lambda and AWS S3 will be used to manage computing and storage, respectively, and Plotly will be used to visualize data on a website managed by the research team and accessible to stakeholders.

Data and metadata standards:

Metadata will include computational models in MATLAB and source code written in C, spatio-temporal measured data at each node location, model descriptions, weight and linkage parameters, and assumptions. The model output will be a system performance metric, where the system performance is evaluated using a data-driven probabilistic spatio-temporal network reliability analysis. The output of the network reliability analysis is the system-level probability of failure with respect to a given failure limit state (defined by an acceptable accessibility threshold), where degradation or failure of transportation linkages and the reduction of nodal weights increases the network's probability of failure. The aggregated data used to assess the system performance metrics are fully anonymous with no personally identifiable information and can be provided in a standard .csv format to any interested party.

Ethics and legal compliance:

Any ethical issues will be reported to the PI's university and subject to the university's Ethics policy.

Policies for access and sharing:

The PI will communicate all intellectual contributions with the scientific community through journal publications, conference presentations/papers, and seminars. When sharing data, the PI will protect privacy, confidentiality, and security. The research team will discuss making data publicly available with the deployment partners (i.e., City of Pittsburgh and Pittsburgh Parks Conservancy) prior to sharing information. Should the research findings lead to intellectual property, the PI will work with CMU's Technological Transfer Office to protect proprietary data. The PI will not publicly post any raw data that the PI does not have permission to share.

Policies for re-use, redistribution:

The aggregated data in this project are fully anonymous and do not contain any private or confidential information. The research results belong to the PI's university. When not in violation of agreements made with third parties in order to collect data from their systems, the project data will be made available to stakeholders and the research community.

Plans for archiving and preservation:

All data will be stored and managed in a cloud-based server and backed up on a local server and external hard drive at CMU. All data will be regularly backed up to ensure full data recovery in the event of failure.