Identifying Factors to Improve Bicycle Lane Safety in Pittsburgh, PA

Data Collection

What data will you collect or create?

We will collect Pittsburgh crash data, survey data, and image data.

How will the data be collected or created?

Data will be collected using standards for protecting privacy. Folders and files with be structured hierarchical based on data acquired. Data quality will be periodically reviewed by at least 2 people.

Documentation and Metadata

What documentation and metadata will accompany the data?

Data will be in the form of images (PNG or JPG format) and ASCII text files. An internal secure wiki page will be set up to document the data formats, data locations, and data versioning.

Ethics and Legal Compliance

How will you manage any ethical issues?

Survey data will be collected anonymously. Imaging data is likely to capture the faces of people in outdoor public areas and vehicle license plates. We will protect privacy by blurring faces and license plates. We will submit an application for Not Humans Research to the IRB.

How will you manage copyright and Intellectual Property Rights (IP/IPR) issues?

The data and algorithms developed will be published and made available for research purposes.

Storage and Backup

How will the data be stored and backed up during the research?

Data will be stored on a secure server in a key controlled laboratory on CMU's Oakland campus. Access to the server will be restricted to only people on the project team. Data is stored on RAID hard drives. Data will be routinely backed up to a CMU server on the cloud. In the case of hard drive failure on the server, data can be rebuilt by replacing the damaged hard drive and rebuilding the RAID or copying data from the backup server in the cloud.

How will you manage access and security?

Data access will only be given to members of the project team.

Selection and Preservation

Which data are of long-term value and should be retained, shared, and/or preserved?

Any data used to compute final results will be kept for long term storage. These data will also be made available for research purposes.

What is the long-term preservation plan for the dataset?

Data will remain on a secure server for at least one year after the project end date. At such time, the data will be moved to an external hard drive for archival purposes.

Data Sharing

How will you share the data?

De-identified data will be shared with members of the research community. Data will be available one year after project start date.

Are any restrictions on data sharing required?

Data is restricted for research purposes only. Not commercial purposes. We will work with the technology transfer office to draft a data sharing license.

Responsibilities and Resources

Who will be responsible for data management?

Robert Tamburo (PI) will be responsible for implementing, reviewing, and revising the DMP.

What resources will you require to deliver your plan?

No additional expertise, hardware, or software are needed to deliver the DMP

Planned Research Outputs

Dataset - "Pittsburgh Bicycle Lane Dataset"

Survey responses, crash data statistics, and anonymized images captured of bike lanes and the road.

Planned research output details

Title	Type	Anticipated release date	Initial access level	Intended repository(ies)	Anticipated file size	License	Metadata standard(s)	May contain sensitive data?	May contain PII?
Pittsburgh Bicycle Lane Dataset	Dataset	2025-07-29	Restricted	None specified	1 TB	Creative Commons Attribution Non Commercial Share Alike 4.0 International	None specified	No	No