Hyperlocal Monitoring of Traffic-Related Air Pollution to Assess Near-Term Impacts of Sustainable Transportation Interventions

A Data Management Plan created using DMPTool

Creators: Cesunica Ivey, Peng Hao

Affiliation: University of California, Riverside (UCR)

Funder: United States Department of Transportation (DOT)

Template: National Center for Sustainable Transportation - Project Data Management Plan

ORCID iD: 0000-0002-4740-2627

Project abstract:

Traffic and air pollution are two of the South Coast Air Basin's most difficult challenges for environmental sustainability. This challenge also exists at local levels, such as in the City of Riverside, where two major highways service the area (CA 60/I-215 and CA 91) and background air pollution is high in the afternoons due to pollution transport and photochemistry. The South Coast Air Quality Management District predicts continued increases in vehicle miles of travel (VMT) in the Basin, while secondary ozone levels are also beginning to increase after decades long reductions. Heavy-duty trips are also increasing due to increasing goods movement activity in inland Southern California. Therefore, it can be conjectured that traffic-related air pollution will continue to be a challenge for the City of Riverside, whose corridors service a high volume of logistics activity. This project proposes a low-cost, measurement-based approach for assessing the impacts of sustainable traffic interventions on local air pollution. Traffic-related air pollutions, NO_2 and $PM_{2.5}$, will be measured along an urban corridor while simultaneously implementing eco-driving strategies as vehicles pass through signalized intersections. The chosen testbed location is the City of Riverside Innovation Corridor, a six-mile roadway that services downtown Riverside, University of California, Riverside, and several businesses and community organizations. Evaluation of traffic and air quality feedbacks in the testbed will provide insight into the effectiveness of wider-scale implementation of smart transportation technologies in the City of Riverside for the improvement local air quality.

Last modified: 09-16-2020

Hyperlocal Monitoring of Traffic-Related Air Pollution to Assess Near-Term Impacts of Sustainable Transportation Interventions

Data Description

Provide a description of the data that you will be gathering in the course of your project. This could include, but is not limited to, new data you are collecting, or existing datasets (whether from a prior project or an external source). Refer to the NCST Center-wide Data Management Plan for additional examples.

Address the following, as applicable:

1. **Name the data, data collection project, or data producing program, if applicable.**
2. **Describe the data your project will generate in terms of nature and scale (e.g., numerical data, image data, text sequences, video, audio, database, modeling data, source code, etc.).**
3. **Describe methods for creating the data (e.g., simulated; observed; experimental; software; physical collections; sensors; satellite; enforcement activities; research-generated databases, tables, and/or spreadsheets; instrument-generated digital data output, such as images and video; etc.).**
4. **Discuss the period of time data will be collected and the frequency of any updates, if applicable.**
5. **If the project uses existing data, describe the relationship between the data you are collecting and the previously collected data.**
6. **List potential users of the data.**
7. **Discuss the potential value that the data will have over the long-term for the NCST and for the public.**
8. **If you request permission not to make data publicly accessible, explain the rationale for lack of public access. Provide sufficient detail so that reviewers will understand any disclosure risks that may apply.**
9. **Indicated who will be responsible for managing the data at the project level.**

Air Quality Data

Ambient air pollutant concentrations will be measured at 6 intersections using a solar-powered, cellular-enabled monitoring unit. Specifically, the resulting data will be 15-minute measurements of PM2.5 and NO2 concentrations, as well as temperature and relative humidity. Sensor information can be found at the vendor website (Clarity Movement, https://www.clarity.io/). Data will be collected at intersections beginning in October 2020 through the life of the funded project (June 2021). Clarity measurements are accessed and downloaded from a password-protected, cloud-based online dashboard. Later in the project, measurement efforts may be extended to include micrometeorology (wind speed and direction) and carbon monoxide measurements. Our data will be made available to other CE-CERT researchers and engineers at the City of Riverside who may be interested in the data. Roadside measurements, while made with low-cost instrumentation, may be of interest to City managers nationwide, especially those who have interest in reducing air pollution related to urban traffic. Dr. Ivey will manage all air pollution data collection and curation.

Vehicle Data

In this project, onboard GPS and On-board diagnostics (OBD) data loggers will be equipped on our study's connected vehicles to collect real-time vehicle activity data and vehicle operation data. The vehicle activity and operation data previously collected by CE-CERT will also be applied in model validation and data analysis in this project. The data from infrastructure is a key input for all vehicle-to-infrastructure (V2I) based CAV applications. For example, the eco-approach and departure system uses signal phase and timing (SPaT) information to guide vehicles passing through the intersection in an energy efficient manner. In the field experiment, the in-vehicle data collection system will receive real time data via Dedicated Short Range Communication (DSRC) or 4G communications. Geographic data, such as intersection location and road grade will also be transmitted to vehicles in the MAP message. All the transportation data will be collected before and after sensor deployment for comparison. They will be used in model development, validation, and evaluation. Dr. Hao will manage all vehicle data collection and curation.

Data Format and Metadata Standards

Your DMP should describe the anticipated formats that your data and related files will use. To the maximum extent practicable, and in accordance with generally accepted practices in your field, your DMP should address how you

will use platform-independent and non-proprietary formats to ensure maximum utility of the data in the future. If you are unable to use platform-independent and non-proprietary formats, you should specify the standards and formats that will be used and the rationale for using those standards and formats.

Address the following, as applicable:

1. **List in what type of format(s) the data will be collected, and indicate if they are open or proprietary.**
2. **If you are using proprietary data formats, discuss your rationale for using those standards and formats.**
3. **Describe how versions of the data will be signified and/or controlled.**
4. **If the file format(s) you are using is(are) not standard to transportation, describe how you will document the alternative you are using.**
5. **List what documentation you will be creating in order to make the data understandable by other researchers.**
6. **Indicate the type of metadata schema you are using to describe the data. If the metadata schema is not one that is standard for your field, discuss your rationale for using that schema.**
7. **Describe how the metadata will be managed and stored during the collection process.**
8. **Indicate what tools or software is required to read or view the data.**
9. **Describe the quality control measures you will implement in your project to ensure its accuracy, etc.**

Air Quality Data

Air quality measurements will be downloaded from the Clarity dashboard and stored in CSV files, which can be read by many programs. Data version will be controlled by including the download date in the file name and by not overwriting existing files. Air quality data are provided with measurement units and qualifiers included in the headers of the CSV file. All air quality data will be intermittently re-downloaded to ensure that retrieval algorithms are not changed by the vendor.

Vehicle Data

Vehicle activity, operation and signal data will be collected in .csv or .txt format. Below is a list of the main elements in the dataset:

- Vehicle dynamic information: time, location, speed, acceleration rate, steering angle
- Vehicle operation data: pedal positions, operating mode, connectivity/automation capabilities, engine speed.
- Signal data: downstream intersection id, signal phase and count down information

The raw vehicle data are scrubbed to protect privacy, filtered for quality control, and then put into a relational database for easy access and update.

Policies for Access and Sharing

Protecting research participants and guarding against the disclosure of identities and/or confidential business information is an essential norm in scientific research. Your DMP should address these issues and outline the efforts you will take to provide informed consent statements to participants, the steps you will take to protect privacy and confidentiality prior to archiving your data, and any additional concerns (e.g., embargo periods for your data). If necessary, describe any division of responsibilities for stewarding and protecting the data among Principal Investigators or other project staff.

If you will not be able to de-identify the data in a manner that protects privacy and confidentiality while maintaining the utility of the dataset, you should describe the necessary restrictions on access and use. In general, in matters of human subject research, your DMP should describe how your informed consent forms will permit sharing with the research community and whether additional steps, such as an Institutional Review Board (IRB), may be used to protect privacy and confidentiality.

Address the following, as applicable:

1. **List the roles that your project's data creation team members will have in data management, including any limitations on team member access due to the presence of personal or confidential information.**
2. **Describe what data will be shared, how data files will be shared, and how others will access them.**
    - **The data to be shared should, at a minimum, be the data required to reproduce your final results, subject to those restrictions imposed by data quality and the need to protect national/homeland security, individual privacy, and confidentiality.**
3. **Indicate whether the data contain private or confidential information. If so:**
    - **Discuss how you will guard against disclosure of identities and/or confidential business information.**
    - **Describe the processes you will follow to provide informed consent to participants.**
    - **State the party responsible for protecting the data.**
4. **Describe if there are any privacy, ethical, or confidentiality concerns due to the sharing of data.**
5. **If applicable, describe how you will de-identify your data before sharing. If this is not applicable to your**

project, then:

- **Identify what restrictions on access and use you will place on the data.**
- **Discuss additional steps you will use to protect privacy and confidentiality.**

Vehicle data are scrubbed for privacy before distribution beyond the project team. Vehicle and air quality data will be shared using a dynamic cloud service, such as Dropbox or Google Drive.

Policies for Re-use, Redistribution, Derivatives

Describe who will hold the intellectual property rights for the data created by your project. Describe whether you will transfer those rights to a data archive, if appropriate. Identify whether any copyrights apply to the data, as might be the case when using copyrighted instruments. If you will be enforcing terms of use or a requirement for data citation through a license, indicate as much in your DMP. Describe any other legal requirements that might need to be addressed.

Address the following, as applicable:

1. **Name who has the right to manage the data.**
2. **Indicate who holds the intellectual property rights to the data.**
3. **List copyrights to the data, if any. If there are copyrights, indicate who owns them.**
4. **Discuss any rights to be transferred to a data archive.**
5. **Describe how your data will be licensed for re-use, redistribution, and derivative products.**

Drs. Ivey and Hao have the right to distribute all data collected under this project. University of California, Riverside holds the intellectual property rights to the data. If applicable, the project team will work with the UCR Research and Economic Development office to establish agreements for licensing of derivative data and products.

Plans for Archiving and Preservation

Describe how you intend to archive your data and why you have chosen that particular option. You must describe the dataset that is being archived with a minimum amount of metadata that ensures its discoverability.
Address the following, as applicable:

1. **The data must be archived before the research project's DRAFT FINAL REPORT is delivered to the NCST Program Manager. Discuss how you intend to archive your data and where if not on Dryad (include URL).**
2. **Indicate the approximate time period between data collection and submission to the archive.**
3. **The PI of each NCST-funded project should ensure that the data to be archived temporarily at their home institution is stored securely on a designated device (computer, external hard drive, etc.). Identify where data will be stored prior to being deposited to an archive.**
4. **The PI of each NCST-funded project should ensure that the data collected will be backed up prior to being archived. Describe how back-up, disaster recovery, off-site data storage, and other redundant storage strategies will be used to ensure the data's security and integrity.**
5. **Describe how data will be protected from accidental or malicious modification or deletion prior to receipt by the archive.**
6. **If you will not be using Dryad,**
    - **Discuss your chosen data archive's policies and practices for back-up, disaster recovery, off-site data storage, and other redundant storage strategies to ensure the data's security and integrity for the long-term.**
    - **Indicate how long your chosen archive will retain the data.**
    - **Indicate if the chosen archive employs, or allows for the recording of, persistent identifiers linked to the data.**
    - **Discuss how your chosen data repository meets the criteria outlined in the Guidelines for Evaluating Repositories with the DOT Public Access Plan.**

The Bourns College of Engineering (BCOE) at the University of California, Riverside, in partnership with the UCR Libraries and the California Digital Library, has established a new, custom data management system designed to make results of our research readily and reliably accessible. This system combines two services of the California Digital Library for the first time: an eScholarship series (https://escholarship.org/), a curated site where all BCOE investigators can store and disseminate their data, and EZID permanent identifiers for long-term identification and management of data resources. Therefore, the team will use the California Digital Library for long-term storage and dissemination of the data collected during this project.

All NCST researchers must have a unique ORCID (Open Researcher and Contributor ID) identification. ORCID.org provides a registry of persistent unique identifiers for researchers and scholars, and automates linkages to research objects such as publications, grants, and patents. Registration is free and takes about 5 minutes. If other researchers are added to a project after its initiation, they are also required to obtain an ORCID.

List all Principal Investigators, Co-PI(s), and key contributors, and their respective ORCIDs.

PI: Cesunica Ivey, 0000-0002-4740-2627
Co-PI: Peng Hao, 0000-0001-5864-7358