# Data Management Plan:

# Cost-Effective Designs of Smart City Technologies for Vehicular Communications

# Data Description

The following types of data will be used and/or produced in the project:

* Data from U.S. DOT reports: effectiveness of vehicle-to-vehicle and vehicle-to-infrastructure applications for road safety;
* Data collected from a citywide deployment: mobility and communications data from Porto, Portugal;
* Technical and Cost Data: numerical parameters and data (other than the data above) used for the engineering and economic modeling.

In what follows, it is described (i) how each type of data above will be organized, documented, standardized, stored, protected, shared, and archived, (ii) who will take responsibility for carrying out the activities listed in (i), and when these activities will take place over the course of the project (and beyond).

## Data from U.S. DOT reports

One of the tasks proposed in this project is to examine outcomes of specific road safety applications. For this task we will use data publicly available in U.S. DOT reports including (Harding et al., 2014), which contain estimates of outcomes from safety applications for different conditions such as the type of application considered and vehicle speeds.

This data will be obtained primarily from reports that resulted from the U.S. DOT’s Safety Impact Methodology. The methodology combines driving simulation under several scenarios with crash data statistics. In addition, complementary data on safety outcomes for conditions other than those considered by the U.S. DOT may also be collected from other sources, which include (Yue, Abdel-Aty, Wu, & Wang, 2018).

## Data Collected from a Citywide Deployment

Another task proposed for this project is to simulate V2X communications under realistic conditions of vehicle usage and mobility. For this project, we will use proprietary data collected from a real vehicular network using Dedicated Short Range Communications (DSRC) and cellular technologies. The data refers to the state of the network as operating in March, 2015, in the city of Porto in Portugal (Future Cities Project, 2017). In that network, buses equipped with OBUs (communications onboard units) offer free Wi-Fi to passengers, and route data to one of 27 DSRC roadside units (RSUs) connected to the Internet. When a vehicle cannot connect to an RSU, data is sent over cellular. All city buses have their positions tracked with onboard GPS devices, as well as about 60% of the city’s taxis. We use a dataset with measurements of data transferred over DSRC and cellular, and GPS position data of 400+ buses and 400+ taxis. Porto data will be used in three ways. First, GPS positions will be used to determine the positions of the vehicles in the network simulation part of our engineering-economic method. Second, strength of the wireless signal received from RSUs is measured in the buses. Third, coordinates of Porto roads intersections are used for modeling the locations of RSUs hotspots in the simulation model.

The data from Porto buses and taxis that were used in this paper are summarized in Table I.

|  |
| --- |
| Table I. Data Used for the Analysis |
| Data Item | Number of Observations | Description |
| Data volume/ position/signal per 15-second per bus  | 400+ buses: 240106 data points | Per 15-second interval, per bus: GPS position, received signal strength from RSU or peer bus that the vehicle is connected to. Data were collected from October 2014 to March 2015 |
| RSU positions | 27 RSUs | Per RSU: GPS position and height as of March 2015 |
| Position per second per taxi  | 400+ taxis: 120106 data points | Per second, per taxi: time, GPS position, and an identifier of the vehicle. Data were collected in March 2012 |
| Porto roads | 5,000 intersections, 7,000 roads | GPS positions of urban intersections and length of roads |

## Technical and Cost Data

For this project we will develop engineering and economic models which will use data collected or estimated from a variety of sources. Those data will be used to:

* Define scenarios of operation of road safety applications using V2X communications technologies. These data may include road geometry and sizes, number of vehicles, vehicle distances, speeds, acceleration, etc.
* Set parameters for the network simulation, from antenna height and transmission power in the physical layer to the protocol used in the transport layer.
* Set parameters for the economic models, such as the average cost of a cellular tower, average cost of a roadside unit, average number or sectors in a cell, cost of communications links for different capacities in bits per second or GB per month.

# Data format and metadata standards

Of the data available in U.S. DOT reports and other sources, the portion relevant to this project will be collected and organized in tables that make traceable the source, dates, and conditions of which each data entry applies to. These data will be stored in text files, in space or comma-separated values tabular formats, or Excel spreadsheet files that can be read by MATLAB. (This is the tool planned to be used for the modeling proposed in the project description.)

The samples of the dataset collected from the Porto deployment will be stored in text files, in space or comma-separated values tabular formats.

MATLAB code that reads those files for use into the engineering and economic models will serve as metadata description of each field of the files.

The other technical and cost data that will be used to define application scenarios, to set parameters in the network simulation and economic analysis do not have a specific machine-readable format, as these data will be described directly in the research papers, posters and presentations produced in the project.

# Policies for access and sharing

The data to be used to estimate effectiveness of vehicle-to-vehicle and vehicle-to-infrastructure applications for road safety are publicly available from U.S. DOT reports (Harding et al., 2014) and other sources such as those listed in (Yue et al., 2018) which are also publicly available. None of this data includes Personally Identifiable Information (PII).

The data from the Porto deployment were collected by Veniam Networks, a private company that operates the vehicular wireless network in a partnership with Porto’s bus operator. Samples of the dataset were copied by CMU researchers working in the project for research purposes only; we do not have permission to make this information publicly available. To protect privacy, CMU researchers were careful not to obtain any information that could uniquely identify a vehicle or an individual, so the CMU dataset includes no Personally Identifiable Information (PII).

All other technical and cost data that will be used to define application scenarios, to set parameters in the network simulation, to be used for the economic analysis, and all the results from the technical and costs analysis will be made public in research conferences and journals. If these outlets are insufficient, we will make essential data that was not included in published papers available on a CMU website under the control of the PI.

# Policies for re-use, redistribution, derivatives

The data to be used to estimate effectiveness of vehicle-to-vehicle and vehicle-to-infrastructure applications for road safety are publicly available from U.S. DOT reports and other sources.

The data from the Porto deployment are for use by the CMU researchers working in the project for research purposes only. That data will be made available to the same researchers for re-use in future research projects, but we do not have permission redistribute this data to third parties unless approved by Prof. Ana Aguiar of the University of Porto.

All other technical and cost data that will be used to define application scenarios, to set parameters in the network simulation, to be used for the economic analysis, and all the results from the technical and costs analysis will be made public in research conferences and journals. Therefore, these data and results will have their re-use and re-distribution encouraged.

# Plans for archiving and preservation

The data from U.S. DOT reports and other sources that will be collected to estimate application effectiveness will be stored in the Box cloud storage made available by CMU. The Box account is linked to the Andrew ID of one of the researchers working in the project.

The data for from the Porto deployment are stored in a server in Portugal, under responsibility of Prof. Ana Aguiar of the University of Porto. The text files in use for this project will be stored in the Box cloud storage mentioned above.

The files used to produce the peer-reviewed journal articles resulting from this project will be stored as above before those articles are publicly available. For archival proposes, by the end of the project the files will be transferred to a Box account controlled by the project’s PI.

The research papers that document the other technical and cost data that will be used to define application scenarios, to set parameters in the network simulation and economic analysis will be made available at the PI’s website <http://www.ece.cmu.edu/~peha/papers.html>, and on archival research databases such as IEEE’s IEEExplore (<https://ieeexplore.ieee.org>), SSRN (<https://ssrn.com>), or other databases including Elsevier, Wiley and the Transportation Research Board.

# References

Future Cities Project. (2017). Discussing the largest vehicular network in the world. Retrieved January 5, 2017, from https://futurecities.up.pt/site/discussing-the-largest-vehicular-network-in-the-world-2/

Harding, J., Powell, G., Yoon, R., Fikentscher, J., Doyle, C., Sade, D., … Wang, J. (2014). Vehicle-to-Vehicle Communications: Readiness of V2V Technology for Application. U.S. DOT NHTSA.

Yue, L., Abdel-Aty, M., Wu, Y., & Wang, L. (2018). Assessment of the safety benefits of vehicles’ advanced driver assistance, connectivity and low level automation systems. *Accident Analysis and Prevention*, *117*(April), 55–64. https://doi.org/10.1016/j.aap.2018.04.002