Data Management Plan

Project Name Travel behavior on the Eastern Corridor in Madrid

Project Identifier SPIP2017-02160

Grant Title BES-2016-077150

Principal Investigator / Researcher Fernando Romero

Project Data Contact fernando.romero@upm.es

Description The main aim of this Project is to address the problem of recurrent congestion suffered by drivers at the accesses to the city of Madrid, Spain. Madrid Metropolitan Area is connected to Madrid City Centre through a radial network of six main transportation corridors. All these corridors have free highways (named A1 to A6), rail services, and suburban bus services. Furthermore, four of these corridors have an additional alternative, a toll highway (named R2 to R5) parallel to the existing free highways. Toll sections generally have excess capacity whereas drivers in toll-free sections experience congestion during peak hours. Based on a case study, this project seeks to determine to what extent it is possible to change the behaviour of passengers in the corridor to reduce the use of the free highway in favour of the other options (toll highway, bus, metro). To this end, a survey was carried out on different transport demand management measures among the dwellers of the Eastern corridor.

Institution McMaster University

Data Collection
What types of data will you collect, create, link to, acquire and/or record?

The research team has acquired data from a Computer-Aided Telephone Interview (CATI) by subcontracting a specialized company. Responses based on the questionnaire were gathered in tabular datasets.

We have also acquired spatial data from multiple open data repositories to include in our analysis including transport network, census section, average income, election results, demographics,

Based on the responses, we have created additional spatial data on the most suitable routes and modes of transport for each respondent with the help of Google Maps API.

Based on tabular and spatial data, different Tobit and multilevel regressions will be carried in order to model the travelers' preferences producing modeling data.

In order to develop the state of the art and literature, we collected text files such as articles and reports.

What file formats will your data be collected in? Will these formats allow for data re-use, sharing and long-term access to the data?

Our tabular data were provided by a company in SPSS format file and Excel file. To ensure long-term access to the data, a copy of the data will be made in CSV files.

Row spatial data are excel files with a georeferencing field and shapefiles (SHP).

Articles and reports are pdf files.

What conventions and procedures will you use to structure, name and version-control your files to help you and others better understand how your data are organized?

We created different folders to organize the different parts of the work: Literature, Models_R, Models_SPSS, Models_STATA, Layers, Maps. A folder whose name is Original contains a copy of the questionnaire and a copy of the original responses to the questionnaire.

We have made use of a reference manager software, Mendeley, to structure all the articles and reports we collected to elaborate the literature review. This also allows us to have access to the articles in any device through Mendeley's Website. They are classified by subject and we add tags to improve their identification. We bookmark those interesting ones. All the marks, tags, and highlights that we make in a device are synchronized in the remaining ones.

Regarding version controls, so far, the research team has used the track changes command in Word documents. Word documents that are exchanged between the different members of the research team are identified with the last modification date in their title and the initials of the research who has made the items.

We are considering the use of GitHub to track better the changes we make.

Documentation and Metadata
What documentation will be needed for the data to be read and interpreted correctly in the future?

A legend explaining the dataset fields or variables, their units, and their range was appended on a new sheet on the Excel file with the questionnaire data. Both original variables and new ones created by the research team are described.

It will also be needed for the data to be read the original questionnaire.

We have created an Excel file that summarises the variables considered in the different models conducted so far.

A description of the tasks carried out each day is recorded in a Word document called "agenda". Clarifications on the methodology used can also be found there.

How will you make sure that documentation is created or captured consistently throughout your project?

Whenever a variable is changed or a new variable is created, the legend is updated.

Every day of project work, the student researcher writes on the agenda document the tasks carried out. Modifications and tasks carried out by other research members are communicated to the student researcher by email. In order to improve the documentation of the project, we are considering to use GitHub to keep track of the changes we make.

If you are using a metadata standard and/or tools to document and describe your data, please list here.

So far, I am not using any metadata standard. The research team will discuss this aspect in the future.

Storage and Backup

What are the anticipated storage requirements for your project, in terms of storage space (in megabytes, gigabytes, terabytes, etc.) and the length of time you will be storing it?

This research does not entail a large storage capacity, nor does it require an extraordinary length of time for storage. it requires approximately 4 Gb of storage. When the project is finished, both inputs (original data) and outputs (results) will be stored on an external hard drive owned by the main researcher at the university office.

How and where will your data be stored and backed up during your research project?

Currently, we keep three copies of everything related to this project in two desktop and one laptop hard drives. One of the hard drives is backed up twice a week. Moreover, these data are stored at and synchronized with the institutional account of OneDrive of the student researcher.

How will the research team and other collaborators access, modify, and contribute data throughout the project?

So far, since the database is small, the research team transfers data by email. However, it is considered to use the institutional server provided by Universidad Politecnica de Madrid, UPMDrive, or a repository in GitHub.

Preservation

Where will you deposit your data for long-term preservation and access at the end of your research project?

It is not expected to deposit the data in a public repository.

Indicate how you will ensure your data is preservation ready. Consider preservation-friendly file formats, ensuring file integrity, anonymization and de-identification, inclusion of supporting documentation.

To ensure long-term access to the data, a copy of the tabular data will be made in CSV files.

Sharing and Reuse

What data will you be sharing and in what form? (e.g. raw, processed, analyzed, final).

It is not expected to deposit the data in a public repository. Processed and final data will be deposited in a private repository in GitHub owned by the main researcher. Since raw data is accessible from different open data web pages, it is not necessarily to share with the scientific community.

Have you considered what type of end-user license to include with your data?

Not yet. The research team will discuss this aspect close to the publication of the research.

What steps will be taken to help the research community know that your data exists?

It is expected to publish an article on the results of the project. Thanks to the paper, the scientific community will be able to know the existence of the database and could ask for access to the private repository.

Responsibilities and Resources

Identify who will be responsible for managing this project's data during and after the project and the major data management tasks for which they will be responsible.

The person responsible for managing the project's data during and after the project is the student researcher Fernando Romero.

How will responsibilities for managing data activities be handled if substantive changes happen in the personnel overseeing the project's data, including a change of Principal Investigator?

Within the Data Plan Management, it is considered that main researcher will be responsible for the data in case the current responsible leaves the project.

What resources will you require to implement your data management plan? What do you estimate the overall cost for data management to be?

In order to store the data, an institutional cloud server is provided free of charge. The institution has a public repository where publications can be uploaded free of charge.

To ensure long-term preservation, an external hard drive is required. A device with data from other projects will be used, therefore, there is no need to buy another one. The external hard drive will be owned by the main researcher and will be kept at university offices.

Ethics and Legal Compliance

If your research project includes sensitive data, how will you ensure that it is securely managed and accessible only to approved members of the project?

Since our questionnaire does not include questions that allow respondents to be identified, we do not have sensitive data.

If applicable, what strategies will you undertake to address secondary uses of sensitive data?

Not applicable

How will you manage legal, ethical, and intellectual property issues?

Not applicable

This document was generated by DMP Assistant (https://assistant.portagenetwork.ca)