DMP title

Project Name Urban Form Research

Project Identifier Ferrari_McMaster

Principal Investigator / **Researcher** Tatiana Ferrari

Project Data Contact tatianak.ferrari@gmail.com

Description This research project seeks to investigate how the urban form, the design of the built environment can control and shape the individual's travel behavior. There is a lack of empirical work on the subject in developing countries, which has different spatial dynamics that could bring other issues to the debate.

Institution McMaster University

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

The data to be collected are:

1) Remote Sensing Data: images extracted from the Global Human Settlement Layer (GHSL) (https://ghsl.jrc.ec.europa.eu/)

2) Brazil Census Data, from IBGE (https://censo2010.ibge.gov.br/)

3) Annual report on social information - RAIS, from the Ministry of Labor and Employment (http://trabalho.gov.br/rais)

4) Street Map: GIS Data

5) Urban area boundaries: Shapefile of Metropolitan areas limit drawn up by IBGE (https://www.ibge.gov.br/geociencias/organizacao-do-territorio/estrutura-territorial.html)

The collected data will be used to create quantitative data (measures and indicators) about urban form and travel behavior. The indicators will be used as inputs into an iterative k-means classification. The result will be a cluster map with a 1km x 1km grid resolution and an excel file with the attribute means classification.

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?

File formats are: Shape files, .tif, .ods, .xls and ASCII

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?

The structure of the folders will be by content and separate by raw, processed and final material.

The raw data refers to different data collected from third sources. The folder containing this data should not be changed in any hypotheses. It will be separated by content name.

Any change should be saved in the processed folder. The data name will use a

convention of a summary of the content and a track of the version by a sequentially v01, v02, v03,.

The final outputs need to be separated and saved in the final data folder along with the metadata or description and the processing code.

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

Metadata will be generated to describe the processed data and will be stored alongside the final data.

Also, the methodology and script code can be consulted to interpret the results.

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

It will be a good practice to include comments in the script code and in the descript text as we go along. Usually, the softwares that we are going to use, like R and Latex, provide a means for adding comment lines together with the text or code. Thus, we are going to add comments frequently to explain the approach chosen, difficults and impedances that we may find, and any other relevant note.

Another procedure will be to create a simple logbook, where each entry corresponds to an steps in the research and need to be filled with the information's about: identification of the process, main goals, quick description, software used, main results, file saved in the end, major documents consulted in the process and other relevant comments.

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

1) LogBook: aims to document the research steps with general information

2) Metadata: description of the data used

3) Script: describe the code and changes made to process the data

4) Text: Final paper with the main goals, methodology, results, and conclusion.

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?

The data can be stored indefinitely.

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

To store the data, 2 computers (work and personal) and an online backup (GitHub, Dropbox) will be used.

At the end of the process, an external hard drive will be used to store all documentation and research data indefinitely.

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

The supervisors' professors will be the collaborators in the project. They will have fully access to contribute and modify the data during the project.

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

We will provide the data to the public by posting it online in: 1) GitHub - data repository, and 2) journal, once the article has been accepted for publication.

Also, an external hard drive will be used to store all documentation and research data indefinitely.

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.

The file formats of the final data are generated in friendly file formats since you are using open software packages. For long-term storage the formats are good and can be read by programs like QGis and R.

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

All kinds of data in this research can be shared: raw, processed, analyzed and final data. Since all the data used is open and free domain.
Because of storage capacity, we will provide guidance on how to acquire the raw data and we will share the analyzed and final data.

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

Question not answered.

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

The disclosure process will be done by an article published in a journal and the data will available in GitHub.

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 core team of the project will be responsible to handle data management during and after the project.

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?

The substantive changes in the data will be made by Tatiana Ferrari and Antonio Paez direct involved in this project. Given major changes, the data repository, the metadata, and the coding process can guide another researcher through 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?

The resources needed will be a computer, software and network communication. There are no data management costs associated with this research.

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?

Not applicable.

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

Question not answered.

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

The use of any content of this research should be cited. The use of a DOI number may be a good practice to ensure the intellectual property of the research.

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