

## DMP title

**Project Name** Built Environment Influences on Bicycling Behaviour in Hamilton, Ontario

**Principal Investigator / Researcher** Elise Desjardins

**Description** I am conducting an explanatory sequential mixed methods study to understand and describe how the built environment influences bicycling behaviour and route choice in Hamilton, Ontario. Most of the current literature that has looked at the attributes of the built environment that influence bicycling for transport is from European countries with a long tradition of such types of travel, and only recently more evidence from North America and Canada in particular has emerged. For this reason, there is a need to collect local evidence to understand whether the built environment attributes identified in the literature from other countries are relevant or different here in Canada. In the first phase, I will be doing secondary data analysis of bicycle trips in Hamilton collected through the Transportation Tomorrow Survey. I will develop a Poisson regression model to estimate bicycle travel. Based on the model analysis, I will conduct bikeability audits in areas across the city that have under- and over-predicted bicycling shares. In the second and final phase, I will conduct semi-structured interviews with local bicyclists to gather their perspectives about how the built environment influences their route choice. I will use the results of the bikeability audits to structure the interview.

**Institution** Portage

## Data Collection

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

I will acquire numeric data (trip counts) from the University of Toronto's Data Management Group. The *Transportation Tomorrow Survey* (TTS) is a collection of data information about urban travel for commuting purposes in Southern Ontario, Canada. The survey has been carried out every five years since 1986 by the Data Management Group. I will be acquiring data from Hamilton respondents in the 2016 survey. The survey collects three categories of travel information: i) household data; ii) demographics; and iii) trip data. The final category is the most useful for our purposes as it provides data on the number of trips by mode of transport, as well as origins and destinations for trips made by those modes. In addition, I will acquire spatial data from a variety of open data sources to visualize built environment attributes in Hamilton, including bike lanes, land use mix, and transit.

I will estimate a Poisson regression model that combines the numeric data (from TTS) and the spatial data (from various geographic sources), which will produce modeling data. Based on the model, I will conduct bikeability audits which will produce images and tabular data. These data will be used to describe the local built environment.

Finally, I will also be collecting qualitative data from semi-structured interviews with local bicyclists. The interview guide will be developed using the images and tabular data from the bikeability audits. The transcripts from these interviews will be coded and analyzed to create thematic data.

**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?**

Secondary data (i.e., from *TTS* and geographic sources) that is acquired or collected from other sources will be in the following formats: CSV, SHP, SHX, and DBF. This data is readily available to the public online, so the formats do not require specialized software or hardware. Other individuals would be able to access this data directly from the sources.

Qualitative data will be collected in MP3 formats (for audio recordings) and TXT for transcripts. Thematic analysis of interviews will be done in NVivo and in WORD. Although NVivo is a specialized software, the transcripts and coding memos will be available in open file formats for future re-use or sharing.

**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?**

All quantitative data files are created in R and are stored in a repository on GitHub, a version control system, to systematically record changes as they are made by either me or the thesis supervisor. These files are named with summary of content, while GitHub keeps track of the date and who made changes through commits. All secondary, numeric data (e.g., raw files) acquired for the project will also be stored in GitHub and are named with date of export, summary of content, source of the data (e.g., *TTS*, *StatsCan*). Files are stored in specific folders on GitHub to keep track of which files are used for each analysis. Folders are numbered based on order of research process, and are named with a summary of content.

All qualitative data files will be stored on a laptop hard drive and will be named with date, summary of content, author initials, and version number.

**Documentation and Metadata**

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

All documentation for quantitative data is done in R notebooks and these are stored in GitHub. In each notebook, the researchers describe the purpose of the file, data-level descriptions (where data was acquired from, what data was acquired), research methodology, variable definitions, steps taken for the data analysis, and also record their learnings or thoughts. The researchers record what was accomplished in the notebook and what will be done in the next file.

All documentation for qualitative data will be done in NVivo and in WORD. The student investigator will record analytic memos to document assumptions, positionality, and data saturation. The student investigator will also produce a guide in WORD to document coding and thematic analysis. This guide will be stored and available for future use.

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

GitHub documents who has worked on the project most recently for each version and what was changed through *commits*. A *Commit* needs a summary and a description of the changes. Sometimes the changes are small and the message is not very detailed, but for major changes, the researchers write well-documented messages to describe what was done during the update to a new version. The researchers will document consistently through Commit messages. The student investigator will be primarily responsible for ensuring accuracy and completeness of the documentation for quantitative data. When possible, the student investigator will review all changes made each week and ensure that documentation is complete.

Analytic memos and weekly progress of qualitative data analysis will be captured in NVivo and WORD by the student investigator. The student investigator will create reminders to ensure that documentation is consistently done and complete.

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

As of now, the researchers are not using a metadata standard. However, this will be discussed further and may change. The researchers may use packages in R that have templates for documentation or publication for specific journals.

**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 storage space required for this thesis research will be in megabytes. Data will be collected over 4 months (September - December 2019) so there is no need for large amounts of storage space. Any data collected or created for this project will be stored for a minimum of two years, which will be sufficient time for potential publications and additional analysis if desired. It may be stored for longer if deemed necessary by the thesis supervisor.

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

Quantitative data will be stored in three different locations: on a laptop hard drive, on an external hard drive, and on a version control system (GitHub). Qualitative data will only be stored in the former two locations. GitHub may not be the most intuitive or best place for qualitative data, so the researchers will have to discuss where to store de-identified transcripts of interviews. There may be a version control system useful for qualitative data. Data will be backed up daily on GitHub and laptop hard drive as changes are made, as well as backed up weekly on an external hard drive.

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

All researchers will be able to access, modify, and contribute data to the project through a private repository on GitHub. The repository was created by the thesis

supervisor and the student investigator has been granted permission to modify and contribute to files as well. The version control system keeps track of the changes.

## **Preservation**

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

Quantitative data will be stored on a private repository on GitHub for long-term preservation and access. Both researchers will also retain copies of all versions of the data (e.g., raw, transformed/processed, analyzed, and final) on their laptop hard drives and external hard drives.

De-identified qualitative data will be stored on the researchers' laptop hard drives and external hard drives. There may be similar environments to GitHub for qualitative data. The researchers will have to discuss where to preserve the transcripts. Transcripts of interviews will be kept instead of audio recordings.

### **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.**

Quantitative data created in this project will be stored and preserved primarily in comma-separated formats or in compressed JPG files. Data analysis will be documented and stored in R notebook. Files will be stored on GitHub, which is easy to access with permission from the repository owner. Supporting documentation will include the 2016 data guide from the *Transportation Tomorrow Survey*, as well as a complete list of all sources of secondary data.

De-identified qualitative data will be stored in text format, when possible, or in WORD format. Supporting documentation will include a guide of codes that were used for the thematic analysis.

## **Sharing and Reuse**

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

Some of the quantitative data, such as the numeric data from the Transportation Tomorrow Survey and the spatial data from various sources, is raw and can be obtained by others directly from the sources. The researchers will not need to share that data.

Processed and analyzed quantitative data will be shared in files stored on GitHub, and in publications of the research results. Final data will be shared in GitHub.

Qualitative data from the bikeability audits and the semi-structured interviews will be processed and analyzed before sharing so that it is de-identified.

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

The researchers have not considered an end-user license but will discuss closer to the publication of the research.

### **What steps will be taken to help the research community know that your data**

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

The data will be stored in a private repository on GitHub. The researchers will cite this repository in all publications.

## **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 student investigator will be primarily responsible for managing the project's data during the project. The thesis supervisor will manage the data after the project's completion as he will remain at the institution and the student investigator will have graduated. The thesis supervisor is also the primary owner of the repository on GitHub where the data will be stored for long-term preservation. The student investigator will retain access to the repository. No training is needed to prepare anyone else for these duties.

**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?**

It is unlikely that there will be a change in Principal Investigator. In the event that the thesis supervisor begins working at a different institution, he will retain the responsibility for overseeing the management of the data. The thesis supervisor will continue to own the private repository where the data is stored and the student investigator will keep access.

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

Only human resources will be required to implement the data management plan. There is no cost to storing the data in a repository on GitHub or on a hard drive back up.

## **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?**

All qualitative data will be de-identified before long-term preservation and public sharing. Any quotations will not be attributed to named participants. The student investigator and thesis supervisor will be the only people to have access to the identifiable research data. Identifiable data (e.g., audio recordings) will be stored securely on password protected laptop hard drives of each researcher.

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

The Letter of Information for the project contains details about long-term preservation of data and potential sharing in future. Consent from research participants will be

obtained. Any data that may be shared or used in the future will not contain any identifiable information.

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

Data will be owned by the student investigator and thesis supervisor. Ethical requirements for sharing or re-using data have been addressed through participant consent and clearly documented in the Research Ethics Boards application that was given clearance.

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