# DMP_Danielle_Beaulne_Thesis

**Project Name** THE IMPORTANCE OF GEOSPATIAL INPUTS IN ASSESSING FINE-SCALE LANDSCAPE GENETIC PATTERNS OF A TEMPERATE TREE FROG

**Project Identifier** 002

**Principal Investigator** / **Researcher** Danielle Beaulne

**Project Data Contact** Adriana Lopez Villalobos

**Description** The importance of geospatial inputs in assessing fine-scale landscape genetic patterns of a temperate tree frog. To address issues of landscape genetic patterns in fine-scale natural heterogenous environment.

**Institution** Portage

## Data Collection

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

A total of 112 individuals collected among 22 breeding aggregations across a fine-scale landscape (~150 km2). One the breeding aggregations were separated by 50 km. There is genetic data and location (GPS). There are three types of data.

1) Genetic data, which is stored in FastQ

2) Ecological data of individuals (tabular)

3) LiDAR data, optical data (spatial)

4) images of study sites

### 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?

The data was stored in csv files (ecological data), .las (Landscape measurements, LiDAR), .tif (optical), and FastQ data (genetic).

### 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?

A descriptor of the file, making sure there are no spaces or characters that might cause errors. For LiDAR data after edited, include a record of the contents in the filename. Keep a record of all files in a separate file.

## Documentation and Metadata

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

A document stating and explaining each field, since the LiDAR data does not have headers for each field. A record of the lab work that was performed to extract and sequence the genetic samples. This is partially outlined in the completed thesis.

### How will you make sure that documentation is created or captured consistently

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

This is difficult to achieve since details of data acquisition are often modified by lessons learned from previous projects, and it is a continually evolving process. When possible, data which had been collected with sufficient similarity can be amalgamated into a spreadsheet which reconciles differences in data collection parameters, modeling decisions or filtering decisions.

However, initial data collection campaigns were informed by advice from fellow grad students, mentors, and my supervisor to ensure that the initial data collection was compliant with all of the minimum requirements.

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

During collection, I used my own custom descriptions to best document the data. This documentation will be refined when I deposit the data and the metadata might be translated into a specific standard.

## 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 genetic data is stored in the computer cluster (QUEENS UNIVERSITY CAC), and is GB's in size, I believe

The geospatial is stored in a lab computer in Geology Dept. Queen's, and is GB's in size. This data has also been backed up on multiple external harddrives and other personal computers.

The ecological data is in a personal computer.

Data analysis and figures are currently stored on a personal computer and a lab computer, and some can be found in the completed thesis.

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

Data is stored between a personal laptop, a lab computer, the CAC, and multiple USB keys and external harddrives. The external harddrives are stored in a seperate location from the lab computer.

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

Currently collaborations have been limited to within my labs groups both within the department of biology and the department of geological sciences and geological engineering. There have also been few collaborators, with the main contributors consisting of myself and my two supervisors. As such, a lot of the file sharing has occured via email, and sensitive data has been transferred over USB key or external

harddrive.

## Preservation

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

Ecological data and lab data collected and recorded for the sampling and extraction of genetic information will be stored on the lab NAS, and may be backed up in other locations in the future.

Genetic sequencing data will be stored on the CAC.

### 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.

Reader-friendly file names. Ecological data saved in .xlsx or .csv. Genetic data stored on the fluster in .fastq. Geospatial data saved in non-proprietary formats, such as .tif or .asc where appropriate.

## Sharing and Reuse

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

Raw LIDAR data are co-owned by two faculty at Queen's, and so I do not have authorization to share or disseminate these data. Final data will be shared through publication, and potentially through other means such as the QUBS data repository or other databases upon consultation with supervisors.

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

To be discussed with co-authors.

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

Data will be shared through publication. Some biological data may be shared through databases such as the QUBS repository and other library databases upon consultation with co-authors and supervisors.

## 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.

Danielle Beaulne

### 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?

Prior to graduating and leaving as a grad student, my supervisors will have access to an organized version of all of my data. I will need to discuss with my supervisors what would happen to the data in the event that they both leave.

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

Currently, most of our data is stored on lab computers, a NAS, and in a lab server. The cost would be those costs incurred by implementing those systems.

## 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 co-authors have to agree to share data and make it available in the repositories mentioned above.

**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?**

That is something that needs to be addressed when writing the publications. Requires a meeting with co-authors.

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