Does habitat availability decline towards the range edge of Rhinanthus minor?

A Data Management Plan created using DMP Assistant

Creator: Olivia Rahn

Principal Investigator: Dr. Anna Hargreaves

Data Manager: Olivia Rahn

Affiliation: McGill University

Funder: NSERC/Alberta Conservation Association/McGill University

Template: Portage Template

Project abstract:

Every species occupies a unique, limited geographic space known as the species' range, which has limits in both elevation and latitude. Understanding what limits a species range remains a central goal of ecology, particularly because we expect species to shift their ranges into cooler areas when climate change causes temperatures to rise. However, this relies on a species ability to access new, more suitable habitat when its native range becomes uninhabitable. Models often predict that patches of suitable habitat decrease in size and increase in distance from one another towards the edges of a species' range, eventually reaching a point where the species cannot disperse into the next habitat patch. If true, this would prevent expansion beyond the existing range edge, but this theory has seldom been tested in the real world. My study provides one of the first empirical tests of habitat availability using the elevational range of one model system, the annual plant *Rhinanthus minor*. Elevational ranges are ideal study systems because they encompass a wide range of habitat and environment in a small area. I will test trends in habitat availability across *R. minor's* range using seed transplant experiments and high-resolution images of the habitat in each transplant plot. My results will reveal whether the availability of suitable habitat plays a role in setting *R. minor*'s range limits. This will allow us to better understand what might limit species' abilities to shift their ranges in response to a changing climate.

Identifier: 9709

Start date: 01-09-2021

End date: 06-09-2023

Last modified: 06-10-2022

Grant number / URL: https://www.nserc-crsng.gc.ca/students-etudiants/pg-cs/cgsm-bescm_eng.asp ; https://www.ab-conservation.com/grants/grants-in-biodiversity/

Does habitat availability decline towards the range edge of Rhinanthus minor?

Data Collection

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

Data types:

I will collect demographic and lifetime fitness data from transplant experiments using the montane annual plant *Rhinanthus minor*, conducted in Kananaskis, Alberta. In this transplant experiment, each individual plant has a unique identifier based on its position along an elevational transect, plot number, and the location of the plant within the plot. These data will be stored as tabular data and will include the following measurements related to the fitness of each transplanted individual: seeds produced per individual, individual height, individual leaf nodes, number of reproductive structures, and number of leaves. Columns will also include the site name, position in the elevational transect, date data was collected, and plant ID. This will be stored as tabular data in *.csv files.

Each transplant plot (240 total) will also be associated with a set of multispectral images taken using a Micasense Altum multispectral camera. Each multispectral image is stored as 5 separate photos- one for each reflective band used (red, blue, red-edge, infrared) and one thermal photo. Each plot has 3 sets of multispectral images associated with it, with one set taken during seedling emergence, one taken while plants are flowering, and one taken once fruits are mature.

I have also collected temperature data using iButton temperature loggers set out in 1/2 of the transplant plots (120 loggers total). Each iButton is associated with a specific transplant plot. These data will be stored as tabular data in *.csv files.

This project will also utilize tabular data collected from demographic surveys of natural *R. minor* populations. These surveys have occurred over multiple years, and will include columns containing measurements of the number of seeds produced per individual, height, plant density in a 1m^2 area, and individual leaf nodes, site name and the date data was collected.

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?

Data formats:

Demographic data (from both natural surveys and transplant plots) will be hand written in field notebooks, and then copied into *.csv files. Photos of notebooks were also taken while in the field to provide a backup in case notebooks were lost or became illegible.

*.csv files will be stored in 3 locations (Olivia Rahn's (OJR) external hard drive, OJR's computer hard drive, and OJR's Onedrive account). The Onedrive folder will be shared with Dr. Hargreaves (ALH) and any other contributors or collaborators who join the project.

The multispectral photos collected will also be stored on OJR's external hard drive, computer hard drive, and Onedrive account. These are in the file format .tif, which requires geospatial software (e.g. ENVI, ArcGIS, QGIS) to open.

Tabular data will always be stored in *.csv format to ensure accessibility, and pictures will be stored in *.tiff format.

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?

*.csv files containing demographic data from natural plant populations will be named according to the year collected and the date the data was finished being entered, and will contain the string "rhinanthus_demographic_survey", such that the final format is "rhinanthus_demographic_survey_YEARCOLLECTED_YYYYMMDD".

*.csv files containing data from transplant experiments will also be named be named according to the year collected and the date the data was finished being entered, but will contain the string "rhinanthus_transplant_data" with the final format "rhinanthus_transplant_data_YEARCOLLECTED_YYYYMMDD"

Multispectral photos will be labelled according to the plot number, site number, and date taken ("PLOTNUUMBER_SITENUMBER_YYYYMMDD").

Data will be stored in an R project named "Rhinanthus_Habitat_Availability". All raw data will be stored in a subfolder called "raw_data" and any cleaning scripts and related analyses will be saved in a subfolder called "scripts". Cleaned .csv files will be saved into a subfolder called "cleaned_data".

Link to database design is here: https://drive.google.com/file/d/1mwYTqgsEBcShCeH3f2rErmO3lBpr0i-n/view?usp=sharing

Documentation and Metadata

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

README files will be created for the R project "Rhinanthus_Habitat_Availability," as well as for each subfolder (e.g. "raw_data", "cleaned_data", "scripts"). README files will detail the contents of each folder, the purpose of any analyses conducted, and the workflow used to clean and organize the data. README files associated with the "raw_data" and "cleaned_data" folders will also list and explain the variable and column names used.

I will annotate all R code used for data cleaning and analyses so that it can be reproduced in the future. My email will be included on all publications and published Dryad data if any users of the code run into issues or have comments on the reproducibility of the code.

Each metadata document will include the date the data was accessed, date or range of dates in which data was collected, names of the contributors of the data, variable definitions, format and file type of associated documents, and explanations of the structure and purpose of code provided.

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

I will consistently update the README files associated with this project. I will ensure that variable names and naming conventions are consistent across README files and metadata documents. Records of experimental design, data collection, data entry and storage will be kept and included in relevant metadata documents.

When relevant, details of study design, data collection and analysis will be detailed in published manuscripts associated with this work. For all manuscripts associated with this work data and code will be made public, along with the R project associated with the data cleaning and analysis.

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

I will use the Ecological Metadata Language (EML), as it is the most appropriate metadata standard for the type of data I am collecting. EML is used to describe ecological data in a standardized manner.

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?

Data from both the transplant experiments and surveys of natural populations will both take up ~200 kb of space. These data will be stored on my computer, external hard drive and Onedrive account permanently, as I am be the primary collector of these data. Once the project has been published or finished, data will also be stored publicly on Github (R code, manuscript, and raw data files) and Dryad (raw and cleaned data files).
The multispectral photos collected will take up 200 GB of space and will be stored on an external hard drive, on my Onedrive account, and on a collaborators external hard drive in Squamish, BC. I anticipate that these data will be stored on these hard drives for 4 years, and will be available on GitHub once the project has been concluded.

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

Data will be stored on my computer hard drive, external hard drive, and Onedrive account. All multispectral photos collected for the project will be stored on the aforementioned hard drives, OneDrive account, and a collaborator's external hard drive. All code and data will also be pushed to my Github account and the Github page will be made public once the project has been completed.

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

Myself and ALH have a shared Onedrive folder where we upload all data. All collaborators will also be able to view and edit the associated Github project. We anticipate that all data for this project has been collected by OJR or ALH, but if additional data is added by collaborators this data will be vetted by OJR or ALH prior to inclusion in any databases.

Preservation

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

All data will be stored on both Dryad (metadata, raw and cleaned *.csv files). The R project associated with this work, along with the *.Rmd files associated with any manuscripts written as a result of this work will be stored on Github. If multiple manuscripts are written, as separate R project and Github repository will be associated with each manuscript.
Projects on Github will be private until associated manuscripts are published.

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.

All tabular data will be stored in *.csv format, images will be stored in *.tiff file formats, and text stored as *.txt files. Data used in this project is not sensitive and therefore does not require anonymization or de-identification.

Sharing and Reuse

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

I will share both raw and cleaned (processed) data. I will share the tabular demographic data from both natural population surveys and transplant experiments, all multispectral images, and temperature data stored in *.csv format.

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

I will use a Creative Commons License (CC BY 1.0). Data are not sensitive and therefore a CC license is sufficient. A CC BY license enables users to modify and redistribute data in any form, with proper credit given to the original collector of the data.

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

Data will be stored publicly on Github, and Github repositories associated with this project will be linked whenever associated manuscripts are shared (e.g. on the Hargreaves lab website, ResearchGate, academic Twitter accounts). Data will also be linked in a data availability statement included with each manuscript published.

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.

OJR will be primarily responsible for managing data associated with this project. They will store all data on the appropriate hard drives and online servers (external hard drives, computer hard drive, OneDrive account), and will be responsible for ensuring that appropriate metadata is created and stored with associated data and scripts. OJR will create Github repositories for all associated manuscripts, upload all data onto Dryad, and will act as the corresponding author for any manuscripts associated with this work.

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?

ALH will have shared access to all data, scripts and metadata associated with this work. In a scenario where OJR is no longer the primary collector of this data or the primary manager of the project, ALH will become the primary data manager. ALH and OJR agree that ALH will not sign off on OJR's thesis until all data associated with the thesis (including this project) has been archived properly.

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

We may require additional storage in order to properly store *.tiff files associated with multispectral images. If additional storage is necessary, we will use the Compute Canada Cluster, which costs $225/core/year. If we use this storage mechanism for 4 years, the total cost will be $900.
All other storage mechanisms (Github, Onedrive, Dryad) are free of charge.

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?

This project does not include sensitive data, and therefore it will be publicly accessible. However, Github pages associated with manuscripts resulting from this project will be private until manuscripts are published.

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

N/A.

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

Data are not sensitive and therefore our research team is not concerned about legal, ethical and property issues. Ownership of data will be shared amongst collaborators.