
CIMP242-BG

A Data Management Plan created using DMP Assistant

Creator: Megan Perra

Principal Investigator: Eliezer Gurarie

Contributor: Megan Perra

Affiliation: Other Organisation

Funder: Government of the Northwest Territories

Template: Portage Template for Open Science Workflows

Project abstract:

We propose to study disturbance on barren-ground caribou with a two-pronged acoustic monitoring approach. (1) We plan to add miniature sound recorders and accelerometers (measure movements) to collars already being deployed on barren-ground caribou. These devices will allow us to measure the behavioral impacts of industrial noise and insect harassment on barren-ground caribou, as well as detect adult and calf mortality events directly. (2) We will use stationary acoustic recording units (ARUs) to evaluate the climate variables that affect insect activity and directly measure the levels of human-made noise in the region. These tools will improve our understanding of the conditions under which we see the most insect harassment, the landscape features important for insect relief, and the cumulative impact of insect harassment and human activity of caribou behavior. We will also identify the industrial and human-made sounds that caribou encounter most, as well as the kinds of sounds which have the greatest impact on behavior. We are researching both Bluenose East (BNE) and Bathurst herds to evaluate differences in their range that may contribute to differences in their population trends. This will be the first research of its kind to directly evaluate caribou behavior at a fine-scale, allowing us to record real-time responses to disturbance, and to make inferences about body condition and calf survival. This research will assist northern decisions in the context of cumulative effects by modeling behaviors that influence adult female body condition and calf survival, which has been shown to be related to insect harassment (Johnson et al. 2022). Understanding the population-level effects of insect harassment and sound disturbance will support managers' ability to predict calf survival rates and population trends, thereby informing decisions about land use. This will contribute to monitoring and mitigating the effects of sound disturbance on caribou, particularly that which comes from mining operations. This project will be led by the Fate of the Caribou (FotC), a collaborative effort to study barren ground caribou across North America. The principal personnel for FotC include Megan Perra (PhD Student at SUNY-ESF), Eliezer Gurarie (PI at SUNY-ESF) and Anne Gunn (co-PI). We will also work closely with Government of the Northwest Territories Department of Environment and Natural Resources (GNWT-ENR), the Wek'èezhii Renewable Resources Board, and the North Slave Métis Alliance (see letters of support). We request \$299,995 from the Northwest Territories Cumulative Impact Monitoring Program.

Identifier: 11241

Start date: 01-06-2023

End date: 31-05-2026

Last modified: 04-05-2023

Grant number / URL: CIMP242-BG

Copyright information:

The above plan creator(s) have agreed that others may use as much of the text of this plan as they would like in their own plans, and customize it as necessary. You do not need to credit the creator(s) as the source of the language used, but using any of the plan's text does not imply that the creator(s) endorse, or have any relationship to, your project or proposal

Responsibilities and Resources

Who will be responsible for data management? Will the Principal Investigator (PI) hold all responsibility during and beyond the project, or will this be divided among a team or partner organizations?

Megan Perra will be responsible for data management

In the event that the PI leaves the project, who will replace them? Who will take temporary responsibility until a new PI takes over?

Megan Perra would take responsibility in such an event.

List all expected resources for data management required to complete your project. What hardware, software and human resources will you need? What is your estimated budget?

We require open source software (Program R, and Python) and will be using GitHub to host our code in a public repository. Labelled audio datasets will also be hosted on GitHub website as a browse-able 'library'. Raw audio and accelerometer data will be backed up on physical drives and hosted on a server at our institution at no cost to us. Total cost of physical drives is a one-time expense of \$400.

Data Collection

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

We will acquire audio data from stationary recorders and audilogger data (accelerometers and audio data) from audiloggers attached to caribou GPS collars that are deployed by ECC at GNWT.

Answer the following regarding file formats:

- A. What file formats do you expect to collect (e.g. .doc, .csv, .jpg, .mov)?
- B. Are these file formats easy to share with other researchers from different disciplines?
- C. In the event that one of your chosen file formats becomes obsolete (or is no longer supported) how will you ensure access to the research data?
- D. Does your data need to be copied to a new media or cloud platform, or converted to a different file format when you store or publish your datasets?

- 1. Audio data will be collected as .flacs or .wav files and stored as .flac; accelerometer data will be collected as a .csv and stored as .rda
- 2. Yes
- 3. Data will be physically backed up on drives
- 4. It needs to be stored on a local server and audio data will be compressed to a .flac to make it easy to transport and store

Answer the following regarding naming conventions:

- A. How will you structure, name and version-control your files to help someone outside your research team understand how your data are organized?
- B. Describe your ideal workflow for file sharing between research team members step-by-step.
- C. What tools or strategies will you use to document your workflow as it evolves during the course of the project?

- 1. Files will be structured based on date and unique equipment (recorder or audilogger) ID to ensure that SD cards can be properly identified if labels are somehow lost.
- 2. Files will be shared through a server that can be remotely accessed, code will be shared through github
- 3. We will use GitHub to document our workflow and produce R or Python markdowns to assist our collaborators in understanding said workflow

Documentation and Metadata

What support material and documentation (e.g. README) will your team members and future researchers need in order to navigate and reuse your data without ambiguity?

They will need to pull the GitHub repository associated with each dataset, and R and Python markdowns.

How will you undertake documentation of data collection, processing and analysis, within your workflow to create consistent support material? Who will be responsible for this task?

Data collection methods will be documented by Megan Perra in formal documents (research permits and animal handling permits), and processing and analysis will be documented by Eliezer Gurarie and Megan Perra within GitHub, where our workflows are documented.

Do you plan to use a metadata standard? What specific schema might you use?

We plan to use the Ecological Metadata Language (EML) as our metadata specification.

How will you make sure that a) your primary data collection methods are documented with transparency and b) your secondary data sources (i.e., data you did not collect yourself) — are easily identified and cited?

We will include this information in the README files of our GitHub repositories, which precedes all the information housed there.

Storage and Backup

List your anticipated storage needs (e.g., hard drives, cloud storage, shared drives). List how long you intend to use each type and what capacities you may require.

We will back up all data on physical hard drives for the entirety of the project. Data will additionally be housed on local servers at our institution that can be remotely accessed.

What is your anticipated backup and storage schedule? How often will you save your data, in what formats, and where?

We will backup and store data on hard drives as soon as it is retrieved, and then upload it to our local servers (in .flac or .rda formats).

Keeping ethics protocol review requirements in mind, what is your intended storage timeframe for each type of data (raw, processed, clean, final) within your team? Will you also store software code or metadata?

We will store all raw data in perpetuity within Eliezer Gurarie's lab, and house all processed data on our local servers, and all code and metadata in our GitHub repositories.

Sharing, Reuse, and Preservation

How will your data (both raw and cleaned) be made accessible beyond the scope of the project and by researchers outside your team?

Labeled audio datasets will be made publicly available through code hosted on github that allows others to download it from our server. Labeled audio from audiologgers will be publicly available on a GitHub hosted webpage that allows users to download or browse annotated audio clips. All processed audio data will be made publicly available from our servers 3 years after the end of our project funding. It may additionally be donated to the archives of Environmental Climate Change Canada or hosted Wildtrax as a publicly available dataset.

Is digital preservation a component of your project and do you need to plan for long-term archiving and preservation?

It is, and we are planning to archive our data long term through collaborators at ECCC or Wildtrax.com.

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

All audio data (raw, processed, analyzed and final) will be made publicly available through aforementioned sources. Accelerometer data will be available on a case-by-case basis.

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

All audio data will be made available under a Creative Commons license. Sound and behavior labeling models will be under an MIT license.

What tools and strategies will you take to promote your research? How will you let the research community and the public know that your data exists and is ready to be reused?

We will publish in peer reviewed journals and through peer reviewed IEEE conferences.

Ethics and Legal Compliance

Are there institutional, governmental or legal policies that you need to comply with in regards to your data standards?

Question not answered.

Will you encounter protected or personally-identifiable information in your research? If so, how will you make sure it stays secure and is accessed by approved team members only?

No protected or personally identifiable information will be collected.

Before publishing or otherwise sharing a dataset are you required to obscure identifiable data (name, gender, date of birth, etc), in accordance with your jurisdiction's laws, or your ethics protocol?

Are there any time restrictions for when data can be publicly accessible?

Question not answered.

Planned Research Outputs

Audiovisual - "Labeled Audio Data from Stationary Recorders"

Labeled and annotated stationary ARU data

(MIT license)

Audiovisual - "Labeled audiologger audio data"

Labeled and annotated audiologger audio data

(MIT license)

Data paper - "Peer Reviewed Papers"

Peer reviewed journal articles using audiologger and stationary ARU data, 3 total

Software - "Sound Labeling Models"

Sound labeling models for all datasets

(MIT Licenses)

Software - "Behavior Labeling Model"

Behavior labeling model using accelerometer data

Dataset - "Audio Data "

All raw data from stationary audio recorders

Dataset - "Audio Data from Audiologgers"

Raw audio data from audiologgers

Planned research output details

Title	Type	Anticipated release date	Initial access level	Intended repository(ies)	Anticipated file size	License	Metadata standard(s)	May contain sensitive data?	May contain PII?
Labeled Audio Data from Stationary Recorders	Audiovisual	2026-05-01	Open	None specified	10 GB	None specified	None specified	No	No
Labeled audiologger audio data	Audiovisual	2025-05-01	Open	None specified	500 GB	None specified	None specified	No	No
Peer Reviewed Papers	Data paper	2027-05-01	Open	None specified		Creative Commons Attribution 4.0 International	None specified	No	No
Sound Labeling Models	Software	2025-05-05	Open	None specified		None specified	None specified	No	No
Behavior Labeling Model	Software	2024-10-04	Open	None specified		None specified	None specified	No	No
Audio Data	Dataset	2029-05-01	Embargoed	None specified	20 TB	Creative Commons Attribution 4.0 International	None specified	No	No
Audio Data from Audiologgers	Dataset	2029-05-01	Embargoed	None specified	10 GB	Creative Commons Attribution 4.0 International	None specified	No	No