Mathematical modelling of human-honeyguide cooperation and communication - Student Outline DMP

1. General guidelines

PURPOSE OF THIS TEMPLATE
The purpose of the Outline DMP is to indicate your initial plans for how your data will be collected, shared and stored, and to give you a chance to think about these data-focused aspects of the research process. As your begin doing your research, your data process may change, and it is perfectly acceptable to change your data management plan to accommodate the changes in your research process.
Indicate below that you understand the purpose of completing this Outline DMP template.

- I understand the Outline DMP template is a projection of my anticipated data management planning requirements and should be updated as my project develops.

2. Authors and supervisors

PROJECT NAME
Replicate the title of your project, dissertation or thesis exactly as it appears in your proposal document.

Mathematical modelling of human-honeyguide cooperation and communication

PERSONAL DETAILS
Indicate the name(s) and student number(s) of the student(s) who will be involved in this project, dissertation or thesis.

Emma-Beth Peters
PTREMM009

SUPERVISOR(S) DETAILS
Indicate who will supervise this project, dissertation or thesis. If you do not yet have a supervisor, leave this section blank.

Claire Spottiswoode
spottiswoode@cantab.net

3. Data Collection/Generation

ORIGINAL DATA
Indicate whether you will collect or produce original data for your study. If yes, briefly describe the type of data and how you plan to manage it.
If you are unsure at this time, indicate what you think you are most likely to collect. If you are not intending to gather or collect your own data, declare that here.

- I do not intend to collect original data

DATA RE-USE
Indicate if you intend to re-use existing data, either from online searches or from datasets provided by your supervisor, lab, or funder.
If you are not intending to re-use existing data, declare that here. Also note any restrictions that apply to the re-use of data.

- I intend to reuse existing data in my study (described below).

I am reusing data from a previous project undertaken by my supervisor. She has given me permission to reuse the data and all her relevant papers will be cited. This is shared on David Lloyd-Jones' GitHub Repository guiding_to_non_bees and is licensed under a GNU General Public License v3.0.

DATA SENSITIVITY & SECURITY

Indicate whether your research data may contain sensitive, personal, disclosive, or otherwise at-risk information.

If yes, briefly describe the type of sensitivity involved and the steps you will take to secure and control access to your data.
If you are unsure at this stage, indicate what you think is most likely. If your data is not sensitive, declare that here and state how you will still ensure responsible storage.

- My data is not sensitive or at-risk.

4. Data Storage

DATA SIZE ESTIMATE

Indicate the estimated size of your completed dataset, and indicate whether or not you will need to access additional data storage facilities. If such storage is not provided by your unit or department, you may need to factor in the cost of purchasing additional storage space.

- 20GB or less

I will not need to access additional data storage facilities.

DATA BACKUPS

Indicate how you plan to ensure your data is secure and retrievable in case of errors or hardware failure. Describe what procedures you will put in place to back-up copies of your data and where they will be stored.

- I intend to backup my data using a service provided by UCT (UCT GoogleDrive, UCT OneDrive, Netstorage, ZivaHub etc.).

The dataset is saved on my laptop, in my UCT OneDrive account and is also available on David John Lloyd-Jones' Github account, and is also backed up by my supervisor. The raw dataset, code and cleaned data will also all be linked to my GitHub profile and managed with version control.

5. Data Sharing/Publication

DATA SHARING

According to UCT's Research Data Management policy, research data should be made open by default, with provisions for making it closed in specific cases (such as ethical considerations or potential commercialisation).
Indicate whether you plan to publish your research data.
• If yes: say where you will publish it and what licence (e.g., Creative Commons) you will use.
• If no: explain why, and refer to any ethical issues, commercial or patent intentions, or data-use agreements that prevent publication.

- I intend to share my data (details below).

My project is using data from a previous study and is under a GNU General Public License v3.0 from GitHub https://github.com/dlloyd-jones/guiding_to_non_bees. This allows it go under: Commercial use, Modification, Distribution, Patent use, and Private use. All contributors will be correctly attributed if this project gets published.

DATA DESCRIPTION

What information will you include to help others understand and use your dataset?
(For example: a short description of your study, explanations of variables, survey questions, or keywords. These are known as metadata.)

As all the data I will be using has already been used in projects by my supervisor, all metadata is correct and freely available as attached to the necessary publications. The metadata and all README files are also available on David Lloyd-Jones' GitHub account: https://github.com/dlloyd-jones/guiding_to_non_bees. This repository will be forked onto my own GitHub account. Coding explanations will occur within my commented R code and README files kept up to date on GitHub. This documentation will ensure that future researchers can understand the data, that the owners of the data are acknowledged, and that my use of the data is recorded and explained. As I construct the mathematical model that is the main body of my research, I will include accessible information about the inputs, variables, parameters and functions.

6. Budget

BUDGET

Indicate any costs specifically relating to the management and curation of your data, such as purchasing additional storage space, digitisation of physical media, data storage or curation charges, and data audits. Most student research will be able to make use of free options provided by UCT and will not have to budget for data costs.

- I do not anticipate any data costs as my data is less than 10GB, and I will be using a storage system provided by UCT (UCT GoogleDrive, UCT OneDrive, Netstorage, ZivaHub, etc.) to curate my data.

Mathematical modelling of human-honeyguide cooperation and communication - Student Full DMP

1. Project Details

PROJECT NAME - Replicate the title of your project, dissertation or thesis exactly as it appears in your proposal document.

Mathematical modelling of human-honeyguide cooperation and communication

PERSONAL DETAILS - Indicate the name(s) and student number(s) of the student(s) who will be involved in this project, dissertation or thesis.

Emma-Beth Peters
PTREMM009

SUPERVISOR(S) DETAILS - Indicate who will supervise this project, dissertation or thesis. If you do not yet have a supervisor, leave this section blank.

Claire Spottiswoode

2. Project Summary

RESEARCH SUMMARY
Briefly summarise your study. Include the study's objectives, design, and methods.

Honeyguides are wax-eating birds that cooperate with human honey-hunters to access the contents of bees' nests, and respond preferentially to culturally varying, specialised signals directed at them by honey-hunters, enhancing cooperation. Here we ask, first, whether social learning by honeyguides is required to explain the emergence of this specialised system of interspecies communication. We model interactions between honeyguides and honey-hunters in northern Mozambique, parametrised by the ecology of this specific empirical system. We find that social learning by honeyguides is essential to explain the emergence of specialised interspecific communication shown by this system. Specifically, social learning is required for a large population of honeyguides to learn the signals of a relatively small population of honey-hunters, which purely asocial learning cannot achieve. We find that specialised signals and specialised understanding of those signals reciprocally reinforce one another in a manner analogous to genetic coevolution between species. This (i) provides a proof of principle of a concept of cultural coevolution, in which cultural traits coevolve in interacting species, and (ii) shows that cultural transmission can enhance interspecies communication, including between humans and other animals.

3. Data Collection/Generation

ORIGINAL DATA
If you are collecting your own data, describe the data you are gathering for your study. Briefly describe the type, scope and amount of the data you are producing.

- I am not collecting my own data.

The data were collected by members of the African Honeyguides team. The portion I am using consists of foraging GPS movement data, records of honeyguide interactions and of data on 3,317 trees containing bees' nests, collected by 21 Yao honey-hunters from Mbamba village using a custom application loaded on ruggedized smartphones, between June 2017 and September 2022. This will be used to measure the attributes and behaviours of honeyguides contributing to successful mutualism with humans. The data is contained in a 14kb xls file. This is shared on David Lloyd-Jones' GitHub Repository guiding_to_non_bees and is licensed under a GNU General Public License v3.0.

DATA REUSE
If you are re-using data from third-party sources in your study, record pertinent details here such as the source of the data, the extent of the data, usage rights or restrictions pertaining to the data, and how it incorporates into your study.

- I am using existing data in my study.

I am reusing data from a previous project undertaken by my supervisor. She has given me permission to reuse the data and all her relevant papers will be cited. This is shared on David Lloyd-Jones' GitHub Repository guiding_to_non_bees and is licensed under a GNU General Public License v3.0.

4. Data Quality Control and Formats

QUALITY CONTROL
Describe what measures you are taking to ensure the data you collect are of high-quality.

The data were collected using an app developed with help from Andrew Rayner from HabitatInfo. This ensures that no incorrect inputs were recorded and that quality was maintained throughout. Furthermore, data was collected in field by honey-hunters who had training in using the app. Post-collection, the data was cleaned and assured by experienced researchers including a doctoral student and my supervisor.

FILE FORMATS

Indicate the formats in which your data will be collected and processed.
Clarify whether you will use specialised, proprietary software to produce and access your data and whether you will convert to open, accessible formats for long term access and preservation. In the case of physical objects (such as artworks or models) indicate whether these will be digitised or otherwise preserved for accessibility.

The data is contained in a 14kb xls file and will be processed using open source R packages.

5. Data Management and Documentation

STORAGE AND BACKUP
Describe how your data is being stored and backed-up. If you are using a data service provider, provide details on how long they will retain the data.

The dataset is saved on my laptop, in my UCT OneDrive account and is also available on David John Lloyd-Jones' Github account, and is also backed up by my supervisor. The raw dataset, code and cleaned data will also all be linked to my GitHub profile and managed with version control.

DATA MANAGEMENT
Describe how you organise and manage your data.
Specify folder structures and any file-naming conventions or community data standards you have adopted.

My data will be organised into folders within my main project folder on my Onedrive. All names will be made up of a combination of letters, numbers, and underscores which is readable by all the software I am planning on using, including R and Excel. The folder structure will be as follows: Project > (Data > (Data_Raw, Data_Clean)), (Code), (Outputs>(Graphs, Equations)). The code files will be edited on my laptop desktop so that they correctly uploaded to GitHub. This is inline with DataCite public data file structure and recommendations.

DATA DOCUMENTATION
Describe what supporting notes or files (documentation) you will provide. Documentation helps others interpret your dataset, with details such as variable definitions, coding explanations, and README files.

As all the data I will be using has already been used in projects by my supervisor, all metadata is correct and freely available as attached to the necessary publications. The metadata and all README files are also available on David Lloyd-Jones' GitHub account: https://github.com/dlloyd-jones/guiding_to_non_bees. This repository will be forked onto my own GitHub account. Coding explanations will occur within my commented R code and README files kept up to date on GitHub. This documentation will ensure that future researchers can understand the data, that the owners of the data are acknowledged, and that my use of the data is recorded and explained. As I construct the mathematical model that is the main body of my research, I will include accessible information about the inputs, variables, parameters and functions.

METADATA STANDARDS
Explain what structured information (metadata) you will provide about your data. Metadata helps others understand the context in which the data was collected, including methods, instruments, and provenance.

Our dataset will include metadata following the DataCite standard. Each dataset will have structured information such as:

- Dataset_Title: guided_trips_with_GPS
- Creator_Name: David Lloyd-Jones and the African Honeyguide team
- Collection_Method:
  An icon-driven application (designed by David Lloyd-Jones, Andrew Rayner of HabitatInfo, and Claire Spottiswoode) contains a series of steps mimicking the natural progression of a honey-hunt with a range of possible outcomes (e.g., the icons for the following questions could occur in this sequence: “Were you guided by a honeyguide?” [yes], “What happened next? [found tree],“Did you harvest the honey?” [yes], “Did you cut the tree down?” [yes], “How much honey did you take?” [mass in kg]). Appendix to Chapter 2 of the 2025 paper To bees or Not to Bees: Greater Honeyguides Sometimes Guide Humans to Animals Other Than Bees, but Likely Not as Punishment includes a diagram with all possible decisions. The icon-driven app interface, also contained a few widely understood Kiswahili words as prompts, which proved to be highly compatible with low literacy rates. The app was loaded and run on handheld Android devices (Cedar CT5, Juniper Systems, Logan, Utah, USA) featuring accurate GPS capabilities.
- Instrument_Details: Android devices (Cedar CT5, Juniper Systems, Logan, Utah, USA)
- Date_Collected: between 2017 and 2022
- Location: Niassa Special Reserve, northern Mozambique. The study area is defined by the range of Yao honey-hunters’ foraging trips, radiating outwards from Mbamba village (12°12’0S, 38°01’0E)

By including these columns, our metadata will make the dataset easier to find, understand, and reuse.

6. Data Security and Confidentiality

SENSITIVE DATA
Indicate to what extent your data are sensitive.
Describe how you will control access to your data. Indicate whether you anticipate a need for encryption or password-controlled access, and if so, how you will enforce that access.

- My data is not sensitive or at-risk

The data I am using are publicly available under a GNU General Public License v3.0 from GitHub at https://github.com/dlloyd-jones/guiding_to_non_bees.

ETHICS AND PRIVACY
Describe, as per your Ethics Clearance form or other similar documentation, any ethical or privacy issues that your data are subject to (if any). Summarise the main risks to the confidentiality and security of information related to human participants, the level of risk, and how this risk will be managed. If your project did not require ethical clearance, you may ignore this section.

- I do not require ethical clearance for my research.

I am doing a theoretical study and will not collect any new data, so do not need ethical clearance. The data being used are from David-Lloyd Jone's 2025 thesis. He obtained ethical clearance for data collection. Data were collected under Administração Nacional das Áreas de Conservação (ANAC) Permit numbers 008/2015, 11/11/2016, 15/2019, 09/2020 and 08/2022, and University of Cape Town Human Ethics Committee permit numbers FSREC 22-2017, FSREC 012-2022 and Science Faculty Animal Ethics Committee approval numbers 2017/HG/CSpot and 2022/V14/CS.

7. Data Sharing and Open Access

DATA PUBLICATION

Do you intend to publish your research data?

If you are working with sensitive data or are re-using third-party data, are there any restrictions or conditions attached to your datasets that affect data sharing, reuse, or attribution? If so, please describe them.

If you are not sharing your data, provide the appropriate justification as per the UCT Research Data Management guidelines.

- I do intend to publish my data.

My project is using data from a previous study and is under a GNU General Public License v3.0 from GitHub https://github.com/dlloyd-jones/guiding_to_non_bees. This allows it go under: Commercial use, Modification, Distribution, Patent use, and Private use. All contributors will be correctly attributed if this project gets published.

DATA REPOSITORY

Where will the research data be deposited or made available at the end of the project?

The data repository is already available under a GNU General Public License v3.0 from GitHub https://github.com/dlloyd-jones/guiding_to_non_bees. It will also by linked to my personal GitHub profile where the aforementioned repository will be forked, and any new analyses, code and information will be shared. I will share any new versions of the data on ZivaHub with permission by my supervisor and the original data collectors.

DATA LICENCE

Indicate under which licence you intend to share your research data.

- CC BY

8. Relevant Institutional or Study Policies

Indicate the relevant departmental, unit, or institutional policies that influence your data management activities.

The following policies apply: the UCT Intellectual Property Policy; the UCT Open Access Policy; and the UCT Research Data Management Policy.