# Plan Overview

*A Data Management Plan created using DMPonline*

**Title:** Trajectories of tropical marine communities under high temperature and eutrophic conditions

**Creator:**Stephanie Martinez

**Principal Investigator:** Lisa Becking, Stephanie Martinez, Mainah Folkers

**Data Manager:** Stephanie Martinez, Mainah Folkers

**Project Administrator:** Lisa Becking, Stephanie Martinez

**Affiliation:** Wageningen University and Research (Netherlands)

**Funder:** Netherlands Organisation for Scientific Research (NWO)

**Template:** Data Management Plan | Wageningen University and Research

**ORCID iD:** 0000-0002-4584-3912

**ORCID iD:** 0000-0001-7696-2984

**Project abstract:**

Tropical coral reefs are going through fundamental changes due to increase seawater temperature and nutrient load. Most studies have focused on the coral-to-algae dominated shift. It has been observed that other relevant groups such as sponges and cyanobacteria can also dominate the system. However, the mechanisms and conditions under which these groups will dominate are poorly understood. The overall aim of this project is to understand how community dominance shifts through a temperature and nutrient gradient and identify tipping points and processes leading to those changes. Using the unique environments of marine lakes -islands of seawater- and reefs of Indonesia, we will look into the current and past trajectories of benthic organisms. The current and past community structure, and environmental settings will be analyzed using a multifaceted approach: photoquadrats, 3D models, species metabarcoding, environmental DNA, and sediment cores.

**ID:** 109433

**Start date:** 01-05-2022

**End date:** 31-03-2026

**Last modified:** 29-11-2022

**Grant number / URL:** 4167016700

**Copyright information:**

# Trajectories of tropical marine communities under high temperature and eutrophic conditions

## A. Describe the research project

### 1. Describe the organisational context of your research project.

| | |
|---|---|
| Name researcher | Stephanie J. Martinez Hernandez |
| DMP version (or date last modified) | 21-11-22 |
| Chair group/Business unit | Aquaculture and Fisheries Group |
| Graduate school (WU only) | WIMEK |
| Supervisor/(co-)promotor(s) (WU only) | Lisa Becking |
| Start date of project | 01-04-2022 |
| End date of project | 31-03-2022 |
| Project number | 4167016700 |
| Funding body | NWO - VIDI |

### 2. Give a short description of your research project.

| | |
|---|---|
| Title | Trajectories of tropical marine communities under high temperature and eutrophic conditions |
| Summary | Tropical coral reefs are going through fundamental changes due to increased seawater temperature and nutrient load. Most studies have focused on the coral-to-algae dominated shift. It has been observed that other relevant groups such as sponges and cyanobacteria can also dominate the system. However, the mechanisms and conditions under which these groups will dominate are poorly understood. The overall aim of this project is to understand how community dominance shifts through a temperature and nutrient gradient and identify tipping points and processes leading to those changes. Using the unique environments of marine lakes -islands of seawater- and reefs of Indonesia, we will look into the current and past trajectories of benthic organisms. The current and past community structure, and environmental settings will be analyzed using a multifaceted approach: photoquadrats, 3D models, species metabarcoding, environmental DNA, and sediment cores. |

### 3. List the individual(s) responsible for the following data management tasks.

| | |
|---|---|
| Data collection | Stephanie Martinez (PhD candidate)<br>Mainah Folkers (PhD candidate)<br>Lisa Becking (Promotor)<br>Adriana Alzate Vallejo (Postdoc)<br>Ludi Aji (PhD candidate) |
| Data quality | Stephanie Martinez<br>Mainah Folkers |
| Storage and backup | Stephanie Martinez<br>Mainah Folkers<br>Adriana Alzate Vallejo |
| Data archiving/publishing | Stephanie Martinez |
| Data stewardship/support | Samara Hutting (data steward of AFI-WUR group) |

**4. Name of data management support staff consulted during the preparation of this plan and date of consultation.**

Dr Irene Verhagen
WUR Library - Data Management Support
data@wur.nl
2022-11-26

## B. Describe the data to be collected, software used, file formats and data size

**5. Will you re-use existing data for this project?**

- Yes. Please specify below which data (e.g. DOI/url) and the terms of use (e.g. licence).

We will use environmental and benthic cover data from previous expeditions done by our research group in marine lakes and reefs of Indonesia. Molecular data and developed primers of marine lake taxa are also available from previous research within the group. They are allocated within the WUR's drives, and parts are also found in previous publications from our group. For example:

- https://doi-org.ezproxy.library.wur.nl/10.1111/mec.14556
- https://doi.org/10.1016/j.marpolbul.2019.110700
- https://doi-org.ezproxy.library.wur.nl/10.1111/jbi.13873

When using previous datasets we agree to comply with the terms of a CC BY-NC 4.0 license.

**6. Will new data be produced?**

- Yes

New data will be produced by means of field measurements, video and image footage, molecular work, writing, laboratory and field journal.

**7. When producing new data, describe the data you expect in terms of type, software and format.**

| Data type (e.g. numerical, video, etc.) | Software (e.g. Excel) | (Open) file format (e.g. csv) |
|---|---|---|
| Benthic survey cover percentage and volume (numerical) | Store: Excel<br>Analysis: R | .csv |
| Environmental data (numerical) | Store: Excel<br>Analysis: R | .csv |
| Molecular data | Store: Text file format, excel<br>Analysis: R, Phyton | .fasta .txt .csv |
| Videos | Watch: VLC<br>Cut and editing: ffmpeg, Adobe Premier | .MP4 |
| Images | Color correction: XnView, Adobe Photoshop<br>Analysis: CoralNet; ReefCloud | .png .jpeg .TIFF |
| 3D objects (models) | Agisoft PhotoScan | .obj |

**8. Estimate how much data storage you require in total.**

- 100-1000 GB

Most storage space will be occupied by images and video footage of underwater benthos.

## C. Storage of data and data documentation during research

**9. Where will the data and accompanying documentation/metadata (see section E.) be stored and backed up during the research project?**

- W:drive (WUR network drive)
- Other. Please specify location and back-up frequency below.
- OneDrive for Business (WUR cloud storage)
- Sharepoint/Teams (WUR collaborative platforms)

During fieldwork, data will be stored in temporal hard drives. After returning from the field, data will be transferred to W:drive and Naturalis Drive. Documents will be stored in Teams for easy access and collaboration with colleagues.

## D. Structuring your data and information

**10. Give a representation of the folder structure you intend to use, or the link.**

*In W:drive*
W:/PROJECTS/AFI_massive_storage/marine_time_machine/StephanieMartinez
*In local computer synced with WUR and Naturalis cloud drives*
C:/MyPC/Naturalis/PhD/Stephanie/

- data/
  - master_matrices/
  - environmental_data/
    - reefs/
      - location_code/
    - marine_lakes/
      - location_code/
  - benthic_cover/
    - raw/
      - reefs/
        - location_code/
          - images/
            - raw/
            - color_correction/
          - videos/
      - marine_lakes/
        - location_code/
          - images/
            - raw/
            - color_correction/
          - videos/
    - processed_data/
      - reefs/
        - location_code/
      - marine_lakes/
        - location_code/
  - models_3D/
    - reefs/
      - location_code/
        - image/
        - video/
    - marine_lakes/
      - location_code/
        - image/
        - video/
  - molecular/
    - sediments/
    - cyanobacteria/
    - benthic_community/
    - ancient_dna/
  - sediment_cores/
    - images/
    - environmental/
- scripts/
  - environmental/

    - community_ecology/
    - molecular_pipelines/
    - markdowns/
- documents/
    - lab_book/
    - protocols_methods/
        - environmental/
        - molecular/
        - sediment_cores/
    - proposal/
    - manuscripts/
        - chapter1/
            - archive/
        - chapter2/
            - archive/
        - chapter3/
            - archive/
        - chapter4/
            - archive/
    - presentations/
        - archive/

**11. Describe the file naming conventions you intend to use.**

- **master_matrices:** .csv and .txt files combining datasets. The most updated version of the dataset will be here for quick access.
- **data:** contains folders with raw and processed (i.e., color correction) data. Subfolders will reflect different data sections (i.e., environmental, benthic cover). Each section will be separated into study ecosystems (reefs/ and marine_lakes/). Each location will be allocated in its own folder (location_code/):
    - Location code setup:
        - **RA2022_S_X**
            - RA2022 = Raja Ampat 2022 (Geographic area + sampling year)
            - S = ecosystem type
                - reef
                - lake
            - X = location identification
                - r1 = reef 1
                - wr01 = Wararapop 01
    - Sample identification setup:
        - RA2022_S_X_**A_N**
            - RA2022_S_X = previous explanation
            - A = type of analysis that the sample is collected for
                - nut = nutrients
                - cyan = cyanobacteria
                - comm = community
            - N = sample number (001,002,003...)
- **scripts:** .txt and markdown files with codes for data analysis
- **documents:** contains text (.docx .pdf) files for manuscripts, protocols, etc. For manuscripts, each version must be specified as _date_v**number**. If comments or reviews are incorporated, name of editor should be included in the document file name. Old versions will be moved to the archive/

folder of each section.
- **presentations:** contains versions of presentations (.pptx). A version number should also be added, and old versions will be moved to archive/
- **Readme:** .txt file in the main folder that explains the code format and folder structure of the project.

### 12. Describe the file versioning system you intend to use.

Each document, presentation, and data matrix version must include "filename_date_v**number**". The version number format should use leading zeroes. For example, v01, v02, v03, etc

## E. Data documentation and metadata

### 13. Describe what data documentation and metadata will accompany the data.

- Data collected during fieldwork will be annotated in physical forms with relevant information (date, location, sample type, depth, observations, etc). This data will also be passed onto a digital form and saved in the cloud drive. A photo of the physical copy will also be taken for safekeeping.
- **README files** will be added to each main folder describing the content and abbreviations of that section. The following will be added to the readme file:
  - Purpose of the files, description of sections, and how files relate to each other
  - Information on the folder structure
  - Files formats present
  - Explanations of all abbreviations within folders and files
  - Explanation of data used, including units of measurements
  - Software requirements
  - Data creator(s) and manager(s) + affiliation(s)
  - Data provenance (keeping records of changes in data)

### 14. Describe what data quality controls will be used.

- Standardized excel sheets will be used for data entry.
- Physical, electronic, and visual copies of each field sheet will serve as cross-validation.
- Proper replication of sampling will allow for exploring data variability and calibration.
- Synchronized (cloud and repositories) scripts and markdowns will standardize analysis processes and minimize typos.
- Collaborators and other people collecting data can follow the standardized excel sheet formats and analysis scripts.

## F. Working with sensitive data (personal data, ethics), ownership, sharing and access

### 15. Are there reasons (privacy, ethics, contractual agreement, commercial interests, public security, IP rights) to restrict access to the data or limit which data will be made publicly available?

- No

### 16. Will you process and/or store personal data during your research project?

- No

### 17. Is this project registered in SmartPIA?

- No

### 18. Are there other ethical issues that need to be taken into account?

- No

I am going to be working with invertebrate species, sediment samples, and fossil data. There will be no animal manipulation.

### 19. Who has ownership and controls access over the data?

Wageningen University & Research, and Naturalis Biodiversity Center

### 20. Will there be any intellectual property (IP) rights associated with the data?

- No

## G. Data archiving and publishing

**21. Do you have selection criteria, which allow you to determine which part of the data should be preserved once the project has ended?**

- No

**22. What data will be archived internally (e.g. WUR network drive) for a minimum of 10 years?**

- All (raw) data produced during the project will be archived internally.

**23. What data will be published and made available for re-use via a data repository?**

- All data produced during the project will be published in a data repository.

**24. When will the data be available for re-use, and for how long will the data be available?**

- Data available as soon as the article is published.

**25. Which data repository do you intend to use to make the data findable and accessible?**

- **Github:** Scripts, analysis, raw data
- **GenBank:** DNA sequences

**26. Which metadata standard will be used to describe the data during archiving / depositing in a data repository?**

Dublin Core

**27. Which licence/terms of use will be applied to the data?**

CC BY-NC

**28. If analysis software is generated in this project, describe your publishing strategy,**

**below.**

If software is generated, all relevant information will be shared for re-use via online repositories such as Github.

## H. Data management costs

### 29. What resources (in time and/or money) will be dedicated to data management and ensuring that data is reusable?

Time to curate, upload, and backup the data will be spent during the duration of the PhD project. Extra hard-drives for fieldwork data collections will be necessary. The costs will be covered by the project budget.

### 30. If there are additional costs related to preparing the data for reuse, how will these costs be covered?

The costs will be covered by the project budget