Wangari's Data Management Plan

A Data Management Plan created using DMPonline.be

Creator: Wangari Wambui

Affiliation: KU Leuven (KUL)

Template: KU Leuven BOF-IOF

Grant number / URL: ZB/23/010 REIS+BENC

ID: 207752

Start date: 01-09-2023

End date: 31-08-2027

Project abstract:

Vector-borne diseases are a major global public health burden, affecting more than 1 billion people worldwide. Malaria and schistosomiasis are the main causes of morbidity and mortality, especially children and pregnant women in developing countries. They are transmitted by vectors (mosquito and snail, respectively), whose life cycles include an aquatic stage, resulting therefore in overlapping distributions and thus coinfections. Therefore, the World Health Organization recommends integrated disease control strategies, i.e. simultaneous drug treatment complemented with locally adapted vector control, to combat these diseases sustainably. However, targeted vector control relies on a thorough knowledge of vector distribution. Because conventional malacological and entomological field surveys are expensive and time-consuming, this project wants to develop innovative parasite and vector diagnostic techniques to monitor infection directly in the field, based on environmental DNA and isothermal amplification.

Last modified: 11-06-2024

Wangari's Data Management Plan

Research Data Summary

List and describe all datasets or research materials that you plan to generate/collect or reuse during your research project. For each dataset or data type (observational, experimental etc.), provide a short name & description (sufficient for yourself to know what data it is about), indicate whether the data are newly generated/collected or reused, digital or physical, also indicate the type of the data (the kind of content), its technical format (file extension), and an estimate of the upper limit of the volume of the data.

Dataset name / ID	Description	New or reuse	Digital or Physical data	Data Type	File format	Data volume	Physical volume
		Indicate: ***N****(ew data) or* ***E****(xisting data)*	Indicate: **D**(igital) or **P**(hysical)	Indicate: **A**udiovisual **I**mages **S**ound **N**umerical **T**extual **M**odel **SO**ftware Other (specify)		Indicate: <1GB <100GB <1TB <5TB >5TB NA	
eDNA	Environmental DNA (eDNA) extracted from filter samples	N	Physical	Other (physical DNA)	Dried DNA in tubes	Physical space	
Genomic DNA	Genomic DNA extracted from trematode, snail and mosquito specimen	N	Physical	Other (physical DNA)	Dried DNA in tubes	Physical space	
Sequences	DNA sequences for trematodes, snails and mosquitoes	N:sequenced specimen; E:from GenBank	Digital	T	fastq.gz	<100MB	

If you reuse existing data, please specify the source, preferably by using a persistent identifier (e.g. DOI, Handle, URL etc.) per dataset or data type:

Parasite WormBase:https://parasite.wormbase.org/index.html
GenBank:https://www.ncbi.nlm.nih.gov/genbank/

Are there any ethical issues concerning the creation and/or use of the data (e.g. experiments on humans or animals, dual use)? If so, refer to specific datasets or data types when appropriate and provide the relevant ethical approval number.

- No

Will you process personal data? If so, please refer to specific datasets or data types when appropriate and provide the KU Leuven or UZ Leuven privacy register number (G or S number).

- No

Does your work have potential for commercial valorization (e.g. tech transfer, for example spin-offs, commercial exploitation, …)? If so, please comment per dataset or data type where appropriate.

- No

Do existing 3rd party agreements restrict exploitation or dissemination of the data you (re)use (e.g. Material or Data transfer agreements, Research collaboration agreements)? If so, please explain in the comment section to what data they relate and what restrictions are in place.

- No

Are there any other legal issues, such as intellectual property rights and ownership, to be managed related to the data you (re)use? If so, please explain in the comment section to what data they relate and which restrictions will be asserted.

- No

Documentation and Metadata

Clearly describe what approach will be followed to capture the accompanying information necessary to keep data understandable and usable, for yourself and others, now and in the future (e.g. in terms of documentation levels and types required, procedures used, Electronic Lab Notebooks, README.txt files, codebook.tsv etc. where this information is recorded).

The DNA sequences datasets will be described in FASTA format. The identity of each sequence will be detailed in its own row above the sequence bases.
Each of the dried DNA sample will have an identification ID in form of a barcode that can be scanned to retrieve the sample from the DNA collection.

Will a metadata standard be used to make it easier to find and reuse the data?
If so, please specify which metadata standard will be used.

If not, please specify which metadata will be created to make the data easier to find and reuse.

- Yes

An excel file with each dried DNA information including the barcode ID and the physical location of the sample.
For the DNA sequences, an excel file containing the GenBank accession numbers

Data Storage & Back-up during the Research Project

Where will the data be stored?

- OneDrive (KU Leuven)

The data involved in this research will be stored on the personal KU Leuven One Drive, which is a cloud to safely store, manage and share data/files.

How will the data be backed up?

- Standard back-up provided by KU Leuven ICTS for my storage solution

The data will securely be backed up daily on the KU Leuven central servers.

Is there currently sufficient storage & backup capacity during the project?

If no or insufficient storage or backup capacities are available, explain how this will be taken care of.

- Yes

The storage space provided for KU Leuven One Drive is 2TB, which is sufficient for this research.

How will you ensure that the data are securely stored and not accessed or modified by unauthorized persons?

The data generated by this research will be treated as non-commercial, which means that it can be freely shared and adapted, provided that appropriate credit is given and any changes are indicated.
For the dried DNA samples, only authorized personnel have access to the DNA storage vault

What are the expected costs for data storage and backup during the research project? How will these costs be covered?

The One Drive platform is free and sufficient in terms of storage. Nonetheless, if needed, the bench fee could be used to pay for extra storage capacity.

Data Preservation after the end of the Research Project

Which data will be retained for 10 years (or longer, in agreement with other retention policies that are applicable) after the end of the project?

In case some data cannot be preserved, clearly state the reasons for this (e.g. legal or contractual restrictions, storage/budget issues, institutional policies...).

- All data will be preserved for 10 years according to KU Leuven RDM policy

Where will these data be archived (stored and curated for the long-term)?

- KU Leuven RDR
- Other (specify below)

The physical dried DNA will be archived at the Africa Museum

What are the expected costs for data preservation during the expected retention period? How will these costs be covered?

According to KU Leuven RDM policy, 50 GB is provided for each researcher yearly for free, and this storage capacity should be enough for the completion of this project. Nonetheless, if needed, the bench fee could be used to pay for extra storage capacity.

The physical space for storing dried DNA is provided by the Africa Museum without cost

Data Sharing and Reuse

Will the data (or part of the data) be made available for reuse after/during the project?
Please explain per dataset or data type which data will be made available.

- Yes, as open data

The results obtained from this research will be published in a peer-reviewed journal. Additionally, the data will be accessible in the KU Leuven RDR and can be used while giving appropriate credit and indicating whether changes are made

If access is restricted, please specify who will be able to access the data and under what conditions.

No applicable

Are there any factors that restrict or prevent the sharing of (some of) the data (e.g. as defined in an agreement with a 3rd party, legal restrictions)?

Please explain per dataset or data type where appropriate.

- No

No restrictions required for data sharing, as data in this research does not contain any personal or confidential information. The data uploaded on the KU Leuven RDR, will remain available for minimum of 10 years after the research is completed

Where will the data be made available?

If already known, please provide a repository per dataset or data type.

- KU Leuven RDR (Research Data Repository)

When will the data be made available?

- Upon publication of research results

Which data usage licenses are you going to provide?

If none, please explain why.

- CC-BY 4.0 (data)

Do you intend to add a persistent identifier (PID) to your dataset(s), e.g. a DOI or accession number? If already available, please provide it here.

- Yes, a PID will be added upon deposit in a data repository

What are the expected costs for data sharing? How will these costs be covered?

There are no expected costs for data sharing

Responsibilities

Who will manage data documentation and metadata during the research project?

Cecilia Wangari Wambui

Who will manage data storage and backup during the research project?

Cecilia Wangari Wambui

Who will manage data preservation and sharing?

Hugo F. Gante and Tine Huyse

Who will update and implement this DMP?

Cecilia Wangari Wambui