Dense spatiotemporal modelling of freshwater snails that transmit schistosomiasis – a citizen-science approach

A Data Management Plan created using DMPonline.be

Creator: Noelia Del Carmen Valderrama Bhraunxs

Affiliation: KU Leuven (KUL)

Funder: Fonds voor Wetenschappelijk Onderzoek - Research Foundation Flanders (FWO)

Template: FWO DMP (Flemish Standard DMP)

Principal Investigator: Anton Van Rompaey, n.n. n.n.

Data Manager: Noelia Del Carmen Valderrama Bhraunxs

Project Administrator: Noelia Del Carmen Valderrama Bhraunxs

Grant number / **URL:** 11L3223N

ID: 198251

Start date: 01-11-2022

End date: 31-10-2026

Project abstract:

Vector-borne diseases affect more than 3.9 billion people around the world, with the vast majority living in sub-Saharan Africa. Among these diseases, schistosomiasis is only preceded by malaria in terms of prevalence and morbidity. Schistosomiasis is caused by parasites of the genus *Schistosoma* that use freshwater snails as intermediate hosts (IH). Since IH snail distribution determines where snail-borne diseases occur, information on snail population dynamics at fine spatial and temporal scales –and a thorough understanding of how these dynamics are influenced by biotic and abiotic factors- are highly needed. In this proposal, we suggest using data obtained from the ongoing Action Towards Reducing Aquatic snail-borne Parasitic diseases (ATRAP) project which uses a citizen science (CS) approach to monitor the spatiotemporal patterns of IH snail populations. Firstly, we will focus on the CS data quality analysis and validation to ensure a solid foundation for further analysis. Secondly, we will model the spatiotemporal dynamics of the presence and abundance of the IH snail populations using geographical, abiotic and biotic data as predictor variables. Finally, we will study the effect of extreme events (floods and droughts) and long-term climate variations on the IH population dynamics using novel remote sensing products and climatic projections under various RCP scenarios.

Last modified: 13-04-2023

Dense spatiotemporal modelling of freshwater snails that transmit schistosomiasis – a citizen-science approach
FWO DMP (Flemish Standard DMP)

1. 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.

				Only for digital data	Only for digital data	Only for digital data	Only for physical data
Dataset Name	Description	New or reused	Digital or Physical	Digital Data Type	Digital Data format	Digital data volume (MB/GB/TB)	Physical volume
		Please choose from the following options: • Generate new data • Reuse existing data	*Please choose from the following options:* • Digital • Physical	*Please choose from the following options:* • Observational • Experimental • Compiled/aggregated data • Simulation data • Software • Other • NA	*Please choose from the following options:* • .por, .xml, .tab, .cvs,.pdf, .txt, .rtf, .dwg, .gml, … • NA	*Please choose from the following options:* • <100MB • <1GB • <100GB • <1TB • <5TB • <10TB • <50TB • >50TB • NA	
Citizen_science_form	Formulary with questions about snail occurrence and abundance	New data	Digital	Other	.pdf	<100MB	
Citizen_science_spreadsheet	Quantitative citizen science data including GPS coordinates of the site, the number of snails from the genera *Biomphalaria, Bulinus* and *Radix*, sampling date and hour.	New data	Digital	Observational	.xls	<100MB	
Citizen_science_multimedia	Pictures of the found snails.	New data	Digital	Observational	.jpg	<100GB	
Scripts	Programs to process the spreadsheet from the citizen scientists, the environmental variables and generate simulation models.	New data	Digital	Software	.R	<100MB	
Environmental variables	Precipitation from CHIRPS, temperature from GCOM and NDVI from sentinel-2 during the study period.	Reuse existing data	Digital	Compiled/aggregated data	.csv	<100MB	
Simulation maps	The output of the species distribution modelling.	New data	Digital	Simulation data	.shp	<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:

Environmental variables from google earth engine: https://developers.google.com/earth-engine/datasets

Are there any ethical issues concerning the creation and/or use of the data (e.g. experiments on humans or animals, dual use)? Describe these issues in the comment section. Please refer to specific datasets or data types when appropriate.

- No

Will you process personal data? If so, briefly describe the kind of personal data you will use in the comment section. Please refer to specific datasets or data types when appropriate.

- Yes

The personal data is the ID of each citizen scientist, in total my project works with 25 citizens but they are pseudonymised with an ID number (1-25). The purpose of this action was to be able to account for individual variability when doing statistical analysis. Nevertheless, this is in any moment use to identify a particular person.

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/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

2. 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).

Each one of the datasets will be described in detail in README.txt files. In the case of the first data set, it will be a description of the goal and the context of the formulary. For the spreadsheet, a detailed description of each column of the array will be included. For the multimedia files, a .csv file indicating the id_unique of each picture (an id which links the snail pictures with each observation in the spreadsheet) will be included. The scripts will be commented on properly for future use. The environmental variables will have a README.txt indicating resolution, temporal extension and specific source. Finally, the simulation maps will be also explained in detail with a legend and README.txt file indicating the statistical method and the predictor variables.

Will a metadata standard be used to make it easier to find and reuse the data? If so, please specify (where appropriate per dataset or data type) which metadata standard will be used. If not, please specify (where appropriate per dataset or data type) which metadata will be created to make the data easier to find and reuse.

- Yes

An id_unique .csv file to identify each individual sampling report (rows in the spreadsheet dataset).

3. Data storage & back-up during the research project

Where will the data be stored?

The data will be stored at KU Leuven One Drive which is a cloud to safely store, manage and share files. This free cloud supports up to 2 TB. In the case of the codes, these will be stored at GitLab.kuleuven.be which is the research code management and collaboration service of KU Leuven.

How will the data be backed up?

Backup is secured daily on the central servers of the university. All data will be stored electronically on the personal KU Leuven One Drive.

Is there currently sufficient storage & backup capacity during the project? If yes, specify concisely.
If no or insufficient storage or backup capacities are available, then explain how this will be taken care of.

- Yes

The KU Leuven One Drive is up to 2TB which will be sufficient for this project.

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

The data does not contain personal information, therefore, there are no concerns about confidentiality. Nevertheless, the data should not be modified, this is why the edition permit will not be granted but only visualisation.

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.

4. Data preservation after the end of the research project

Which data will be retained for at least five 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 the datasets will be stored for at least 10 years following the recommendations of the RDM of KU Leuven.

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

At the KU Leuven Research Data Repository RDR. According to the RDM policy of KU Leuven, data uploaded to the repository will remain available for a minimum of 10 years.

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

According to the RDM policy of KU Leuven, every researcher can store 50 GB per year for free, that storage capacity should be enough for the completion of this project. Nevertheless, if needed, the bench fee could be used to pay for extra storage capacity.

5. Data sharing and reuse

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

- Yes, in an Open Access repository

The data which will be available are the Citizen_science_form, Citizen_science_spreadsheet, citizen_science_multimedia, scripts and simulation maps.

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

Not 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 in the comment section per dataset or data type where appropriate.

- No

Where will the data be made available? If already known, please provide a repository per dataset or data type.

At the KU Leuven Research Data Repository RDR for all the datasets.

When will the data be made available?

After the publication of research results in a peer-reviewed journal.

Which data usage licenses are you going to provide? If none, please explain why.

The license for the data sets will be Attribution-NonCommercial 4.0 International (CC BY-NC 4.0) which allows to freely share and adapt when giving appropriate credit and indicating if changes were made. Also, the material should not be used for commercial purposes. In the case of the codes, a permissive license will be granted: the MIT license.

Do you intend to add a PID/DOI/accession number to your dataset(s)? If already available, you have the option to provide it in the comment section.

- Yes

Depending on the data repository and the type of data that would be made available, a unique identifier will be added to the data set.

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

According to the size of the files, there are no expected costs for data sharing. Nonetheless, if needed, the bench fee could be used to pay for it.

6. Responsibilities

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

Noelia del Carmen Valderrama Bhraunxs

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

Noelia del Carmen Valderrama Bhraunxs

Who will manage data preservation and sharing?

Tine Huyse and Anton Van Rompaey

Who will update and implement this DMP?

Noelia del Carmen Valderrama Bhraunxs