MICROMICS - MICROCLIMATE- AND GENOMICS-INFORMED DISTRIBUTION MODELING TO IMPROVE PREDICTIONS OF SPECIES' RANGE DYNAMICS AND EXTINCTION RISK UNDER ENVIRONMENTAL CHANGE

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	
Microclimate maps	High-resolution microclimate temperature maps of the study areas	N	D	N	.TIF	<100GB	
ForestTemp	Sub-canopy microclimate temperatures of European forests	E	D	N	.TIF	<1TB	
Thermal response curves	Response curves for all forest specialist species	N	D	N	.csv	<1GB	
Suitability and climate change exposure maps	For all forest specialist species	N	D	N	.TIF	<1TB	
Adaptability and dispersal maps	For both forest specialist species.	N	D	N	.TIF	<100GB	
Sequence datasets	For both forest specialist species.	N	D	T	.fasta	<1TB	
Genetic variant datasets	For both forest specialist species.	N	D	N	.vcf	<100GB	
Pooled allele frequency datasets	For both forest specialist species.	N	D	N	.txt	<100GB	

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:

ForestTemp: https://doi.org/10.6084/m9.figshare.14618235.v2
WorldClim: https://www.worldclim.org/

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).

Per work package, output data will be collected, including a text file with a clear description of what the data represent and how they were generated. A text file with a description of the type and format of all input data and scripts used to generate the output data will be kept inside the same folder. The name of the folder will specify the applied algorithm and the date the results were generated (A .txt file explaining the naming will be maintained).

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.

- No

For data collected on the field and analysed in the lab, a text file will be made containing the date that the data was collected, the coordinates and/or study area, the people that collected the data, a short explanation of each variable and its units, the file format, and user rights and acknowledgements.
For scripts, a text file will be made containing the date the script was made, the developer of the script, a short explanation of the target of the script, necessary software, necessary input data and explanation of the output, reference to data to which this script was applied, and user rights and acknowledgements.
For the generated results, a text file will be made containing the date that the results were obtained, the coordinates and/or study area, the author of the data, a short description of the
results, a reference to the script used to generate the results (if available), the file format, and
the user rights and acknowledgements.
For the genetic data obtained by sequencing companies, a pdf report will be provided to detail on the sequencing and bio-informatics procedure.

Data Storage & Back-up during the Research Project

Where will the data be stored?

- OneDrive (KU Leuven)
- Other (specify below)
- Large Volume Storage

The time-stamped master copy of the data will be kept on the Teams of the Micromics project. Copies can be made and kept on personal OneDrive accounts.
The raw read files obtained by sequencing companies will be deposited on GenBank.
The genetic variant and allele frequency files will be stored on Onedrive, with a copy on a large volume storage device.

How will the data be backed up?

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

The data storage is backed up via Teams and OneDrive (see above)

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

Yes, personal OneDrive accounts have 2 TB and the Teams 5 TB of storage.

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

The data are on secure servers, with centralized security controls and multi-layered encryption.

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

They are covered by the university/department

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

At the end of the project, all data will be stored on the archive K-drive of the Division Forest, Nature and Landscape together with the necessary metadata

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

- Other (specify below)

K-Drive (Archive)

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

The costs will be covered by the project PI and amount max to 1000 euros.

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
- Yes, as restricted data (upon approval, or institutional access only)
- Yes, as embargoed data (temporary restriction)

Data, metadata and code will be deposited to online open-access public repositories such as Dryad, Zenodo and Github. Biodiversity can, in addition to the deposition as described above, also be uploaded to biodiversity repositories such as GBIF (http://gbif.org/) and trait data repositories such as TRY (https://www.try-db.org/). Sequence read data deposited on GenBank will be embargoed for the duration of the project.

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

The files uploaded to figshare, Zenodo and Github will be public. Data stored on the university's central servers will be accessible after permission of the promoters.

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

Where will the data be made available?

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

- Other data repository (specify below)

See answer above

When will the data be made available?

- Upon publication of research results
- Specific date (specify below)
- Other (specify below)

Data and code underlying publications in each of the WPs will be deposited at the time of publication as much as possible, such that they are fully open and can be accessed and re-used from the time of publication onwards. Only in the case that data will be (partly) re-used in subsequent publications (e.g. some of the data such as the temperature data per plot) or due to other restrictions (e.g. journal policies), a reasonable delay/embargo in access to the data to publish our findings first and/or to secure intellectual property protection might be applied. All data and code will be openly accessible at the latest at the end of the project (if restrictions apply, with a maximum of 6 months embargo beyond that date).

Which data usage licenses are you going to provide?

Created using DMPonline.be. Last modified 20 January 2023

4 of 5

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 related to data sharing, apart from data handling and data transfer.

Responsibilities

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

The PhD researchers and promotors will be responsible for data documentation & metadata.

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

The PhD researchers and promotors will be responsible for data data storage. Data back up happens automatically.

Who will manage data preservation and sharing?

The PhD researchers and the promotors will be responsible for compiling a folder with all data and corresponding metadata that needs to be preserved. Our teams' data storage team will be
responsible for storing the data thereafter, with the supervision of the promoters

Who will update and implement this DMP?

The PIs bear the overall responsibility for updating & implementing this DMP.