FWO DMP Template - Flemish Standard Data Management Plan

Version KU Leuven

Project supervisors (from application round 2018 onwards) and fellows (from application round 2020 onwards) will, upon being awarded their project or fellowship, be invited to develop their answers to the data management related questions into a DMP. The FWO expects a **completed DMP no later than 6 months after the official start date** of the project or fellowship. The DMP should not be submitted to FWO but to the research co-ordination office of the host institute; FWO may request the DMP in a random check.

At the end of the project, the **final version of the DMP** has to be added to the final report of the project; this should be submitted to FWO by the supervisor-spokesperson through FWO's e-portal. This DMP may of course have been updated since its first version. The DMP is an element in the final evaluation of the project by the relevant expert panel. Both the DMP submitted within the first 6 months after the start date and the final DMP may use this template.

The DMP template used by the Research Foundation Flanders (FWO) corresponds with the Flemish Standard Data Management Plan. This Flemish Standard DMP was developed by the Flemish Research Data Network (FRDN) Task Force DMP which comprises representatives of all Flemish funders and research institutions. This is a standardized DMP template based on the previous FWO template that contains the core requirements for data management planning. To increase understanding and facilitate completion of the DMP, a standardized **glossary** of definitions and abbreviations is available via the following link.

1. General Project Information	
Name Grant Holder & ORCID	Thomas Neyens (ORCID IDENTIFIER: 0000-0003-2364-7555)
Contributor name(s) (+ ORCID) & roles	Luc De Meester (ORCID IDENTIFIER: 0000-0001-5433-6843), Copromoter
Project number [1] & title	G0A3M24N - Trajectories of newly emerging evolving metacommunities – an interdisciplinary research project in empirical ecology and biostatistics
Funder(s) GrantID [2]	
Affiliation(s)	x KU Leuven ☐ Universiteit Antwerpen ☐ Universiteit Gent x Universiteit Hasselt ☐ Vrije Universiteit Brussel ☐ Other: ROR identifier KU Leuven: 05f950310

[1] "Project number" refers to the institutional project number. This question is optional. Applicants can only provide one project number.

[2] Funder(s) GrantID refers to the number of the DMP at the funder(s), here one can specify multiple GrantIDs if multiple funding sources were used.

Please provide a short project description	Recent efforts to unravel the assembly of species communities have resulted in limited amounts of explained variation, suggesting that important processes are overlooked and that statistical methods to detect these processes are lacking. This project capitalizes on a unique opportunity provided by a nature restoration project, where several clusters of new ponds are being created. From their creation onwards, all ponds will be subjected to detailed monitoring, including the quantification of abiotic conditions, community assembly through conventional and eDNA techniques, population assembly through genomics, and animal vector visits through camera trapping. In parallel, we will develop statistical methodology to comprehensively model evolving metacommunities and how alleles and species assemble through space and time. We will advance state-of-the-art Joint Species Distribution Models to simultaneously consider co-occurrence patterns in alleles and species, imperfect detection in metacommunity eDNA data, spatial pond connectivity through animal vectors, and the occurrence of legacy effects. We will carry out experiments to quantify priority effects impacting community and population assembly. The synergy between innovative ecology and statistics to understand complex natural mechanisms will yield invaluable insights into the dynamics of evolving metacommunities in nature and will further advance developments in the widely applied Joint Species Distribution Modelling framework.

2. 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 [3].

				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
		☐ Generate new data ☐ Reuse existing data	☐ Digital ☐ Physical	☐ Audiovisual ☐ Images ☐ Sound ☐ Numerical ☐ Textual ☐ Model ☐ Software ☐ Other:		☐ < 1 GB ☐ < 100 GB ☐ < 1 TB ☐ < 5 TB ☐ > 5 TB ☐ NA	
Field	empirical field data obtained from the experiments outlined in the project proposal, consisting of environmental data, land-use and other regional data, and data on community composition	New data	Digital	Numerical	csv	< 100 GB	

Annotations	field annotations	New data	Physical				200 pages
Genetics	metabarcoding and RADseq data, which consist of short, targeted and genome-wide, respectively, fragments of DNA base-pair information generated by Illumina sequencing	New data	Digital	Textual	csv	< 100 GB	
Camera	camera-trap images	New data	Digital	Images	jpg	< 5 TB	
Models	statistical model developments	New data	Digital	Models	R script	< 1 GB	

GUIDANCE:
The data description forms the basis of your entire DMP, so make sure it is detailed and complete. It includes digital and physical data and encompasses the whole spectrum ranging from raw data to processed and analysed data including analysis scripts and code. Physical data are all materials that need proper management because they are valuable, difficult to replace and/or ethical issues are associated. Materials that are not considered data in an RDM context include your own manuscripts, theses and presentations; documentation is an integral part of your datasets and should described under documentation/metadata.
RDM Guidance on data

[3] Add rows for each dataset you want to describe.

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.	We will use an existing database, SAFRED, that provides data from earlier related studies. The SAFRED dataset is a compilation of multiple datasets that have been published separately. There is no DOI for the combined dataset, but most data have been submitted to GBIF and have a DOI; below the links: Manscape, https://www.gbif.org/dataset/312bb844-4980-4b1a-a64b-e79f9ac083a4 Pondscape, https://www.gbif.org/es/dataset/a621b3ba-8415-41f2-a4af-7ec9511ae868 Midden-Limburg, https://www.gbif.org/dataset/3236cdc4-2f4d-4bd6-b53f-54fd8d8a0aa8 Tommelen, https://data.freshwaterbiodiversity.eu/ipt/resource?r=tommelen
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.	☐ Yes, human subject data; provide SMEC or EC approval number: ☐ Yes, animal data; provide ECD reference number: ☐ Yes, dual use; provide approval number: ☒ No Additional information:
Will you process personal data[4]? 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).	☐ Yes (provide PRET G-number or EC S-number below) ☒ No Additional information:
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.	☐ Yes ☒ No If yes, please comment:

[4] See Glossary Flemish Standard Data Management Plan

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 to what data they relate and what restrictions are in place.	☐ Yes ☒ 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 to what data they relate and which restrictions will be asserted.	☐ Yes ☒ No If yes, please explain:

3. 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). *RDM guidance on documentation and metadata*.	A README.txt will be provided for each dataset

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. *REPOSITORIES COULD ASK TO DELIVER METADATA IN A CERTAIN FORMAT, WITH SPECIFIED ONTOLOGIES AND VOCABULARIES, I.E. STANDARD LISTS WITH UNIQUE IDENTIFIERS.*	☒ Yes ☐ No We will use metadata standards that are accepted by GBIF (https://www.gbif.org/standards), more specifically EML (and if more appropriate, Darwin Core); we may deviate from this if specific repositories require other standards.

4. Data Storage & Back-up during the Research Project	
Where will the data be stored? *Consult the interactive KU Leuven storage guide to find the most suitable storage solution for your data.*	☐ Shared network drive (J-drive) ☐ Personal network drive (I-drive) ☐ OneDrive (KU Leuven) ☐ Sharepoint online ☐ Sharepoint on-premis ☐ Large Volume Storage ☐ Digital Vault ☒ Other: During the research, we use the 'ManGO - Active Data Management Platform' offered by the 'High Performance Computing – Research Data Management' facilities within KU Leuven to store field and genomic data. This platform is designed to securely store data, and it facilitates data sharing between KU Leuven and (possible) partners. Camera-trap data will be stored in the Agouti format, with summary integration in SAFRED.

How will the data be backed up? *WHAT STORAGE AND BACKUP PROCEDURES WILL BE IN PLACE TO PREVENT DATA LOSS?*	☒ Standard back-up provided by KU Leuven ICTS for my storage solution ☒ Personal back-ups I make (specify) For extra security, data are also stored on external hard disks. ☐ Other (specify)
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 ☐ No If no, please specify:
How will you ensure that the data are securely stored and not accessed or modified by unauthorized persons? *CLEARLY DESCRIBE THE MEASURES (IN TERMS OF PHYSICAL SECURITY, NETWORK SECURITY, AND SECURITY OF COMPUTER SYSTEMS AND FILES) THAT WILL BE TAKEN TO ENSURE THAT STORED AND TRANSFERRED DATA ARE SAFE.* *Guidance on security for research data*	ManGO and Agouti are designed to securely store data. Access to KU Leuven data-sets is only granted to the researchers directly involved in the research. External hard disks are stored in a room with limited access; they are used as extra security only.
What are the expected costs for data storage and backup during the research project? How will these costs be covered?	Costs are estimated to be limited as the databases are not exceedingly large. They will be paid on the consumables budget of the project

5. 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...). *Guidance on data preservation*	☒ All data will be preserved for 10 years according to KU Leuven RDM policy ☐ All data will be preserved for 25 years according to CTC recommendations for clinical trials with medicinal products for human use and for clinical experiments on humans ☐ Certain data cannot be kept for 10 years (explain)
Where will these data be archived (stored and curated for the long-term)? *Dedicated data repositories are often the best place to preserve your data. Data not suitable for preservation in a repository can be stored using a KU Leuven storage solution, consult the interactive KU Leuven storage guide.*	☒ KU Leuven RDR ☐ Large Volume Storage (longterm for large volumes) ☐ Shared network drive (J-drive) ☐ Other (specifiy):
What are the expected costs for data preservation during the expected retention period? How will these costs be covered?	Costs are estimated to be limited as the databases are not exceedingly large. They will be paid on the PIs budgets (team generating most data: research group Luc De Meester)

6. 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. *NOTE THAT 'AVAILABLE' DOES NOT NECESSARILY MEAN THAT THE DATA SET BECOMES OPENLY AVAILABLE, CONDITIONS FOR ACCESS AND USE MAY APPLY. AVAILABILITY IN THIS QUESTION THUS ENTAILS BOTH OPEN & RESTRICTED ACCESS. FOR MORE INFORMATION: HTTPS://WIKI.SURFNET.NL/DISPLAY/STANDARDS/INFO-EU-REPO/#INFOEUREPO-ACCESSRIGHTS*	☒ Yes, as open data: Models ☐ Yes, as embargoed data (temporary restriction) ☒ Yes, as restricted data (upon approval, or institutional access only): Filed, Annotations, Camera, Genetics ☐ No (closed access) ☐ Other, please specify:
If access is restricted, please specify who will be able to access the data and under what conditions.	Individuals affiliated with Freshwater Ecology, Evolution & Conservation (Research group Luc De Meester) L-BioStat (Research Group Thomas Neyens) can access the restricted data upon approval by Luc De Meester and Thomas Neyens
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.	☒ Yes, privacy aspects ☐ Yes, intellectual property rights ☐ Yes, ethical aspects ☐ Yes, aspects of dual use ☐ Yes, other ☐ No If yes, please specify: The data that we gather relate to specific ponds that are owned by private persons / organisations. We will therefore not disclose precise coordinates of the ponds, but use spatial specifications that provide information on the general location but are not precise enough to identify the owner.

Where will the data be made available? If already known, please provide a repository per dataset or data type.	☐ KU Leuven RDR ☒ Other data repository (specify): Models (newly developed statistical models) will be made available on GitHub ☐ Other (specify)
When will the data be made available?	☒ Upon publication of research results ☐ Specific date (specify) ☐ Other (specify)
Which data usage licenses are you going to provide? If none, please explain why. *A DATA USAGE LICENSE INDICATES WHETHER THE DATA CAN BE REUSED OR NOT AND UNDER WHAT CONDITIONS. IF NO LICENCE IS GRANTED, THE DATA ARE IN A GREY ZONE AND CANNOT BE LEGALLY REUSED. DO NOTE THAT YOU MAY ONLY RELEASE DATA UNDER A LICENCE CHOSEN BY YOURSELF IF IT DOES NOT ALREADY FALL UNDER ANOTHER LICENCE THAT MIGHT PROHIBIT THAT.* *Check the RDR guidance on licences for data and software sources code or consult the License selector tool to help you choose.*	☐ CC-BY 4.0 (data) ☒ Data Transfer Agreement (restricted data) ☐ MIT licence (code) ☐ GNU GPL-3.0 (code) ☐ Other (specify)
Do you intend to add a PID/DOI/accession number to your dataset(s)? If already available, please provide it here. *INDICATE WHETHER YOU INTEND TO ADD A PERSISTENT AND UNIQUE IDENTIFIER IN ORDER TO IDENTIFY AND RETRIEVE THE DATA.*	☐ Yes, a PID will be added upon deposit in a data repository ☐ My dataset already has a PID ☒ No

What are the expected costs for data sharing? How will these costs be covered?	Most data repositories are free of charge. The limited costs that may arise will be paid on the consumables budget of the project.

7. Responsibilities	
Who will manage data documentation and metadata during the research project?	Luc De Meester & Thomas Neyens
Who will manage data storage and backup during the research project?	Luc De Meester & Thomas Neyens
Who will manage data preservation and sharing?	Luc De Meester & Thomas Neyens
Who will update and implement this DMP?	Thomas Neyens