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	**Renske Hoevers - https://orcid.org/0000-0002-0425-0192**
Contributor name(s) (+ ORCID) & roles	**Gert Verstraeten (Promotor) - https://orcid.org/0000-0002-6529-7381**
Project number [1] & title	3E240442 - SedaDNA as a new method to reconstruct how past human impact has shaped today's river systems
Funder(s) GrantID [2]	12A6I25N
Affiliation(s)	☒ KU Leuven ☐ Universiteit Antwerpen ☐ Universiteit Gent ☐ Universiteit Hasselt ☐ Vrije Universiteit Brussel ☐ Other: ROR identifier KU Leuven: 05f950310
Please provide a short project description	Humans have shaped the present-day European river systems for so long and to such extent that their functioning and sensitivity cannot be understood without taking into account the pre-historical changes. For information on the distant past, we rely on palaeoenvironmental proxy data. However, the currently applied proxies are not providing sufficient spatial and taxonomic detail to disentangle regional upland vegetation changes from local floodplain vegetation changes, which is essential for understanding the sensitivity of river valleys to (long-term) anthropogenic land cover change in their surroundings. By adding the novel palaeoecological proxy 'sedaDNA' to the array of proxies these issues can be overcome. By integrating information from pollen, macrofossils and sedaDNA, we will reconstruct what lowland river systems looked like in the past, how they function in their (semi-) natural state and to what extent they have changed. Moreover, this integration will allow to assess to what degree floodplain change was driven by anthropogenic land cover change, and whether transformations in floodplain properties happen abrupt when thresholds levels in land use change are crossed or rather gradual. Understanding the relationship between regional land use and local floodplain dynamics is essential for evaluating the viability of restoration initiatives and ensuring the preservation of the few remaining semi-natural floodplain ecosystems in temperate Europe.

[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.

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
Pollen data	Pollen, spores and NPPs	☒ Generate new data ☒ Reuse existing data	☒ Digital ☒ Physical	☐ Audiovisual ☒ Images ☐ Sound ☒ Numerical ☐ Textual ☐ Model ☐ Software ☐ Other:	Excel files with both the raw absolute data and processed relative data, LRA in- and output, and NMDS in- and output (.xlsx) R scripts (.R) Diagrams (.svg)	☒ < 1 GB ☐ < 100 GB ☐ < 1 TB ☐ < 5 TB ☐ > 5 TB ☐ NA	Pollen extracts Pollen slides Counting sheets
Macrofossil data	Plant macrofossils and molluscs	☒ Generate new data ☒ Reuse existing data	☒ Digital ☒ Physical	☐ Audiovisual ☒ Images ☐ Sound ☒ Numerical ☐ Textual ☐ Model ☐ Software ☐ Other:	Excel files with both the raw absolute data and processed EIV data, and NMDS in- and output(.xlsx) R scripts (.R) Diagrams (.svg)	☒ < 1 GB ☐ < 100 GB ☐ < 1 TB ☐ < 5 TB ☐ > 5 TB ☐ NA	Sieved fractions of both 500 and 125 µm, and selected macrofossils from these fractions Counting sheets

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

sedaDNA data	P6 and 16S	☒ Generate new data ☐ Reuse existing data	☒ Digital ☒ Physical	☐ Audiovisual ☒ Images ☐ Sound ☒ Numerical ☐ Textual ☐ Model ☐ Software ☐ Other:	Excel files with both the raw and filtered DNA sequencing data (.xlsx and .tsv) R scripts (.R) Diagrams (.svg)	☐ < 1 GB ☐ < 100 GB ☐ < 1 TB ☒ < 5 TB ☐ > 5 TB ☐ NA	DNA extracts
Radiocarbon data	C14	☒ Generate new data ☒ Reuse existing data	☒ Digital ☐ Physical	☐ Audiovisual ☒ Images ☐ Sound ☒ Numerical ☐ Textual ☐ Model ☐ Software ☐ Other:	Excel files with the uncalibrated data, and the Bacon in- and output (.xlsx) R scripts (.R) Graphs (.svg)	☒ < 1 GB ☐ < 100 GB ☐ < 1 TB ☐ < 5 TB ☐ > 5 TB ☐ NA	/
Geochemical data	LOI, humification, and C and N	☒ Generate new data ☒ Reuse existing data	☒ Digital ☐ Physical	☐ Audiovisual ☒ Images ☐ Sound ☒ Numerical ☐ Textual ☐ Model ☐ Software ☐ Other:	Excel files with the raw and processed data (.xlsx) R scripts (.R) Graphs (.svg)	☒ < 1 GB ☐ < 100 GB ☐ < 1 TB ☐ < 5 TB ☐ > 5 TB ☐ NA	/

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*	
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.	All existing data were gathered as part of the PhD project of Renske Hoevers, and are available at KU Leuven (Sharepoint online and Personal network drive (I-drive)).
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: S69797** ☐ Yes, animal data; provide ECD reference number: ☐ Yes, dual use; provide approval number: ☐ No Additional information: **For sedaDNA data**: The field of environmental DNA (eDNA) is rapidly advancing. While this could lead to many novel beneficial applications of eDNA, negative consequences of eDNA approaches might arise as well. A recently described dilemma is human genetic bycatch (HGB): the phenomenon that deep-sequencing-based eDNA approaches capture genomic information from humans (Homo sapiens) just as readily as that from the intended target species. While this phenomenon could enable various valuable applications in fields like medicine, forensics, and environmental research, it also introduces a range of ethical dilemmas. These encompass issues related to consent, privacy, surveillance, and data ownership, requiring careful examination and potentially the development of new regulations.

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:
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 If yes, please explain:
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:

[4] See Glossary Flemish Standard Data Management Plan

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*.	All Excel files mentioned under "2. Research Data Summary" have a final tab in which the metadata is included. The original source of the data can be found here, along with a description of how to interpret the presented data and what changes have been made compared to the original/raw data. All R scripts for data processing have annotation for each step and are accompanied by example files, including both input files (.xlsx and .tsv) and output files (.xlsx and .svg).

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 If yes, please specify (where appropriate per dataset or data type) which metadata standard will be used: All datasets (pollen, macrofossils, sedaDNA, radiocarbon and geochemical) will be uploaded on **Neotoma (https://www.neotomadb.org/)**, and metadata will be generated following their standard: "Metadata stored in Neotoma include site locations and descriptions, sediment descriptions, information about the original workers, publications, geochronologic data, and age models. Data and metadata standards have been established by working groups for the various data types. A full description of the Neotoma relational database structure and the base tables used to store data and metadata is available from the Neotoma website. Data can be searched and viewed at the Neotoma website. Individual datasets can be downloaded as .csv files, which are easily imported into Microsoft Excel or other spreadsheet programs. The entire database is also available for download in either Microsoft Access for SQL Server format." If no, please specify (where appropriate per dataset or data type) which metadata will be created:

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) ☐ Teams ☒ Sharepoint online ☐ Sharepoint on-premis ☐ Large Volume Storage ☐ ManGO ☐ Digital vault ☐ Other:
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): **annual back-up on an external drive stored at KU Leuven** ☐ 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*	Shared and personal network drives are suitable for public, confidential and strictly confidential data. SharePoint online is suitable for strictly confidential data, as long as multifactor authentication with the KU Leuven Authenticator app is activated.
What are the expected costs for data storage and backup during the research project? How will these costs be covered?	Shared and personal network drives: € 450.76 / TB / year SharePoint online: free (up to 1 TB) Covered in the annual contribution to SET-IT (foreseen in our budget).

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): **Neotoma (https://www.neotomadb.org/)** The Neotoma Database is stored on an enterprise-class network storage appliance that is mirrored to a secondary off-site appliance for disaster-recovery purposes. Additionally, snapshots of the file system on which the database is stored are taken hourly and rolled into daily, weekly, and monthly snapshots. Separate quarterly snapshots of the full database are archived to tape and held for a minimum of 5 years.
What are the expected costs for data preservation during the expected retention period? How will these costs be covered?	€ 95.14 / TB / year on the large volume storage network drive of KU Leuven, covered by Gert Verstraeten. Neotoma combines a centralized database structure, thereby reducing IT development and maintenance costs, and a decentralized scientific governance structure, in which expert working groups are charged with oversight of data quality, data accession, and setting development priorities for particular data types. Recent Neotoma development is funded by the NSF Geoinformatics Program.

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 **For pollen, macrofossil, radiocarbon and geochemical data** ☐ Yes, as embargoed data (temporary restriction) ☐ Yes, as restricted data (upon approval, or institutional access only) ☐ No (closed access) ☒ Other, please specify: **For sedaDNA data**: To ensure compliance and transparency, our research will adhere to the standard practice of sharing filtered datasets when publishing research. The filtering steps will be detailed in our publications, but will broadly consist of excluding non-target DNA, such as HGB, food contaminants based on 'blacklists' built up from previous research at Trømso Museum and positives relating to common PCR errors. We will keep up with the discussions related to the establishment of appropriate filtering criteria and act in accordance with the best practices and regulations agreed upon by the scientific community and the GDPR at the time of publication.
If access is restricted, please specify who will be able to access the data and under what conditions.	**For sedaDNA data:** Given the GDPR's emphasis on protecting sensitive data, any samples or datasets containing HGB will only be shared with ethical approval for subsequent studies.
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:

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): **Neotoma (https://www.neotomadb.org/)** ☐ 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) **All data available through the Neotoma Paleoecology Database are free to use with attribution under a CC BY 4.0 license (https://www.neotomadb.org/data/data-use-and-embargo-policy).**
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?	Neotoma combines a centralized database structure, thereby reducing IT development and maintenance costs, and a decentralized scientific governance structure, in which expert working groups are charged with oversight of data quality, data accession, and setting development priorities for particular data types. Recent Neotoma development is funded by the NSF Geoinformatics Program.

7. Responsibilities	
Who will manage data documentation and metadata during the research project?	**Renske Hoevers**
Who will manage data storage and backup during the research project?	**Renske Hoevers**
Who will manage data preservation and sharing?	**Renske Hoevers**
Who will update and implement this DMP?	**Renske Hoevers**