

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	Iuliia Burdun (0000-0002-1436-2550)
Contributor name(s) (+ ORCID) & roles	Gabrielle De Lannoy (0000-0002-6743-7122) supervisor Michel Bechtold (0000-0002-8042-9792) supervisor
Project number ¹ & title	PEATWATER: Estimating ground water depths using hydrological simulations and multiple satellite observations
Funder(s) GrantID ²	12A5O24N
Affiliation(s)	<input type="checkbox"/> KU Leuven <input type="checkbox"/> Universiteit Antwerpen <input type="checkbox"/> Universiteit Gent <input type="checkbox"/> Universiteit Hasselt <input type="checkbox"/> Vrije Universiteit Brussel <input type="checkbox"/> 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	<p>Peatlands are wetlands with great long-term carbon sequestration potential. Today, peatlands store twice as much carbon as the world's forests. The storage of this enormous amount of carbon is enabled by waterlogged conditions that facilitate the accumulation of plant remnants as a carbon-rich peat layer. However, recently alarming drying trends were reported in Northern Hemisphere, where most peatlands are located. The drawdown of the water table in peatlands transforms peatlands into a carbon source through peat oxidation, leaching of dissolved organic carbon, and peat fires. Given the peatlands' carbon sink response to water table drawdown, monitoring the water table at high spatial and temporal resolutions is urgently needed. Nonetheless, large-scale water table monitoring remains challenging because the water table is "hidden" by the land surface and cannot be observed directly. PEATWATER will address this challenge by combining indirect indicators of the water table, such as vegetation (using optical remote sensing) and peat moisture status (using active microwave remote sensing), with advanced peatland-specific land surface modelling in a data assimilation framework. The three major outcomes of PEATWATER will be (1) novel techniques for estimating water table, (2) water table estimates over the Northern Hemisphere at fine spatial resolutions, and (3) detection of the anthropogenic influence in European peatlands for the recent ~40 years.</p>
--	---

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.

Dataset Name	Description	New or Reused	Digital or Physical	ONLY FOR DIGITAL DATA	ONLY FOR DIGITAL DATA	ONLY FOR DIGITAL DATA	ONLY FOR PHYSICAL DATA
				Digital Data Type	Digital Data Format	Digital Data Volume (MB, GB, TB)	Physical Volume
Modern Era Retrospective Analysis for Research and Applications (MERRA-2)	Atmospheric reanalysis of the satellite era produced by NASA	Reuse	Digital	Numerical (Spatio-temporal)	netCDF	<1 TB	
ERA5	Reanalysis for the global climate and weather by ECMWF	Reuse	Digital	Numerical (Spatio-temporal)	netCDF	>5 TB	
PEATMAP	GIS shapefile dataset shows a distribution of peatlands	Reuse	Digital	Binary (spatial)	Shapefile	< 100 GB	
The Global Peatland Database	Integrates data on location, extent and	Reuse	Digital	Binary (spatial)	Shapefile	< 100 GB	

³ Add rows for each dataset you want to describe.

	drainage status of peatlands and organic soils						
Landsat 5, Landsat 7, Landsat 8, Landsat 9	Earth observation mission from NASA with the longest-running enterprise for acquisition of optical satellite imagery	Reuse	Digital	Numerical (Spatio-temporal)	netCDF, GeoTIFF	< 5 TB	
Sentinel-1	Earth observation mission from the Copernicus Programme that acquires radar imagery	Reuse	Digital	Numerical (Spatio-temporal)	netCDF, GeoTIFF	< 5 TB	
Sentinel-2	Earth observation mission from the Copernicus Programme that acquires optical imagery	Reuse	Digital	Numerical (Spatio-temporal)	netCDF, GeoTIFF	< 5 TB	
In-situ water table	Dataset for intact and	Reuse	Digital	Textual	csv	< 1 GB	

measurements	drained peatlands						
Model outputs	Model-only and data assimilation water table depth	Generate new data	Digital	Numerical (Spatio-temporal)	netCDF, GeoTIFF	< 5 TB	
Peer-reviewed publications	Three articles in open-access journals	Generate new data	Digital	Textual	MS Word	< 1 GB	
Code	Scripts used for modelling and data analysis	Generate new data	Digital	Textual	R and Python script	< 1 GB	
Catalogue of optical and radar data	List of satellite products used for data assimilation	Generate new data	Digital	Textual	csv	< 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 be described under documentation/metadata.

[RDM Guidance on data](#)

<p>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.</p>	<p>MERRA-2 - 10.5067/VJAFPLI1CSIV ERA5 - 10.24381/cds.e2161bac PEATMAP - https://doi.org/10.5518/252 The Global Peatland Database - https://greifswaldmoor.de/global-peatland-database-en.html Landsat 5, Landsat 7, Landsat 8, Landsat 9 - https://www.usgs.gov/ Sentinel 1, Sentinel 2 - https://www.esa.int/Applications/Observing_the_Earth/Copernicus In-situ water table measurements – dataset was created by Dr Bechtold and utilised in https://doi.org/10.1029/2018MS001574</p>
<p>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.</p>	<p><input type="checkbox"/> Yes, human subject data; provide SMEC or EC approval number: <input type="checkbox"/> Yes, animal data; provide ECD reference number: <input type="checkbox"/> Yes, dual use; provide approval number: <input checked="" type="checkbox"/> No Additional information:</p>
<p>Will you process personal data⁴? 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).</p>	<p><input type="checkbox"/> Yes (provide PRET G-number or EC S-number below) <input checked="" type="checkbox"/> No Additional information:</p>
<p>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.</p>	<p><input type="checkbox"/> Yes <input checked="" type="checkbox"/> No If yes, please comment:</p>

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.	<input type="checkbox"/> Yes <input checked="" type="checkbox"/> 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.	<input checked="" type="checkbox"/> Yes <input type="checkbox"/> No If yes, please explain: Data sharing agreements with collaborating researchers that provide in-situ water table measurements for validation will be established. The read/write permissions for those datasets on the Vlaams Supercomputer Centrum will be restricted to the research team.

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). <i>RDM guidance on documentation and metadata.</i>	The model outputs (netCDF files) will be accompanied by machine-readable metadata that is available inside the files. A clear hierarchical directory/file structure and README files will also be created. This will be based on the common practice of the research team of my promoter, Gabrielle De Lannoy.

<p>Will a metadata standard be used to make it easier to find and reuse the data?</p> <p>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.</p> <p><i>REPOSITORIES COULD ASK TO DELIVER METADATA IN A CERTAIN FORMAT, WITH SPECIFIED ONTOLOGIES AND VOCABULARIES, I.E. STANDARD LISTS WITH UNIQUE IDENTIFIERS.</i></p>	<p><input checked="" type="checkbox"/> Yes <input type="checkbox"/> No</p> <p>If yes, please specify (where appropriate per dataset or data type) which metadata standard will be used:</p> <p>The output of the data assimilation and PEATCLSM modelling will follow the NetCDF CF Metadata Conventions (http://cfconventions.org/)</p> <p>If no, please specify (where appropriate per dataset or data type) which metadata will be created:</p>
---	--

4. Data Storage & Back-up during the Research Project	
<p>Where will the data be stored?</p> <p><i>Consult the interactive KU Leuven storage guide to find the most suitable storage solution for your data.</i></p>	<p> <input type="checkbox"/> Shared network drive (J-drive) <input type="checkbox"/> Personal network drive (I-drive) <input type="checkbox"/> OneDrive (KU Leuven) <input type="checkbox"/> Sharepoint online <input type="checkbox"/> Sharepoint on-premis <input checked="" type="checkbox"/> Large Volume Storage <input type="checkbox"/> Digital Vault <input type="checkbox"/> Other: </p>
<p>How will the data be backed up?</p> <p><i>WHAT STORAGE AND BACKUP PROCEDURES WILL BE IN PLACE TO PREVENT DATA LOSS?</i></p>	<p> <input type="checkbox"/> Standard back-up provided by KU Leuven ICTS for my storage solution <input type="checkbox"/> Personal back-ups I make (specify) <input checked="" type="checkbox"/> Other (specify) For the large volume storage, we will use Tier-1 data for which two copies are stored in two different locations: https://docs.vscentrum.be/data/tier1data/introduction.html </p>

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.	<input checked="" type="checkbox"/> Yes <input type="checkbox"/> No If no, please specify:
How will you ensure that the data are securely stored and not accessed or modified by unauthorized persons? <i>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.</i> Guidance on security for research data	Password and key-protected access to Vlaams Computer Centrum High Performance Computing, including Tier-1 Data.
What are the expected costs for data storage and backup during the research project? How will these costs be covered?	HPC Tier-2 storage: 1500 EUR/year, covered by the bench fee of the FWO grant. Tier-1 Data (large volume storage) will be financed by the Tier-1 grants of the Storage4Climate consortium, a network of five modelling groups in Flanders.

5. Data Preservation after the end of the Research Project

<p>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...).</p> <p>Guidance on data preservation</p>	<p> <input type="checkbox"/> All data will be preserved for 10 years according to KU Leuven RDM policy <input type="checkbox"/> 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 <input checked="" type="checkbox"/> Certain data cannot be kept for 10 years (explain) Final generated data used in publications will be preserved. Re-used data will not necessarily be preserved unless they remain used or are no longer supplied by the original data providers. </p>
<p>Where will these data be archived (stored and curated for the long-term)?</p> <p><i>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.</i></p>	<p> <input type="checkbox"/> KU Leuven RDR <input checked="" type="checkbox"/> Large Volume Storage (longterm for large volumes) <input type="checkbox"/> Shared network drive (J-drive) <input type="checkbox"/> Other (specify): </p>
<p>What are the expected costs for data preservation during the expected retention period? How will these costs be covered?</p>	<p>As done for the data storage during the project, for the large volume storage, we will use Tier-1 data for which two copies are stored in two different locations: https://docs.vscentrum.be/data/tier1data/introduction.html There is currently no guarantee that the current Storage4Climate grant for large-volume storage will be extended for ten years. If this grant ends, alternative funding will be searched for the largest datasets that cannot be stored on tier-2 storage.</p>

6. Data Sharing and Reuse

<p>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.</p> <p><i>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</i></p>	<p><input checked="" type="checkbox"/> Yes, as open data</p> <p><input type="checkbox"/> Yes, as embargoed data (temporary restriction)</p> <p><input type="checkbox"/> Yes, as restricted data (upon approval, or institutional access only)</p> <p><input type="checkbox"/> No (closed access)</p> <p><input type="checkbox"/> Other, please specify:</p> <p>All newly generated datasets will be shared publicly as part of publications and published at Zenodo.</p>
<p>If access is restricted, please specify who will be able to access the data and under what conditions.</p>	
<p>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.</p>	<p><input type="checkbox"/> Yes, privacy aspects</p> <p><input checked="" type="checkbox"/> Yes, intellectual property rights</p> <p><input type="checkbox"/> Yes, ethical aspects</p> <p><input type="checkbox"/> Yes, aspects of dual use</p> <p><input type="checkbox"/> Yes, other</p> <p><input type="checkbox"/> No</p> <p>If yes, please specify:</p> <p>The in situ water level data that is shared by international researchers for the analyses in PEATWATER will not generally be shared publicly as raw data. All other reuse data is publicly available and all newly generated data will be made publicly available.</p>

<p>Where will the data be made available? If already known, please provide a repository per dataset or data type.</p>	<p> <input type="checkbox"/> KU Leuven RDR <input checked="" type="checkbox"/> Other data repository (specify) <input type="checkbox"/> Other (specify) Zenodo </p>
<p>When will the data be made available?</p>	<p> <input checked="" type="checkbox"/> Upon publication of research results <input type="checkbox"/> Specific date (specify) <input type="checkbox"/> Other (specify) </p>
<p>Which data usage licenses are you going to provide? If none, please explain why.</p> <p><i>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.</i></p> <p>Check the RDR guidance on licences for data and software sources code or consult the License selector tool to help you choose.</p>	<p> <input checked="" type="checkbox"/> CC-BY 4.0 (data) <input type="checkbox"/> Data Transfer Agreement (restricted data) <input type="checkbox"/> MIT licence (code) <input type="checkbox"/> GNU GPL-3.0 (code) <input type="checkbox"/> Other (specify) </p>
<p>Do you intend to add a PID/DOI/accession number to your dataset(s)? If already available, please provide it here.</p> <p><i>INDICATE WHETHER YOU INTEND TO ADD A PERSISTENT AND UNIQUE IDENTIFIER IN ORDER TO IDENTIFY AND RETRIEVE THE DATA.</i></p>	<p> <input checked="" type="checkbox"/> Yes, a PID will be added upon deposit in a data repository <input type="checkbox"/> My dataset already has a PID <input type="checkbox"/> No </p>
<p>What are the expected costs for data sharing? How will these costs be covered?</p>	<p>No additional resources are needed.</p>

7. Responsibilities

Who will manage data documentation and metadata during the research project?	Project PI
Who will manage data storage and backup during the research project?	Project PI, HPC cluster system administrators, local system administrators
Who will manage data preservation and sharing?	Project PI
Who will update and implement this DMP?	Project PI