



ProCleanLakes

Data management plan

Deliverable No: 6.6
Work package: WP6

Official delivery date: 30.11.2024
Actual delivery date: 27.11.2024

Dissemination level: Public



**Funded by
the European Union**

Project: 101157886 | HORIZON-MISS-2023-OCEAN-01-04 | www.procleanlakes.eu

Disclaimer

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting authority can be held responsible for them.

Version	Date	Author(s)
V1	30.11.2024	BOKU



Contents

Executive Summary	5
1 Data Summary.....	6
1.1 Data types and formats which ProCleanLakes will re-use.....	6
1.2 Data types and formats which ProCleanLakes will generate	6
1.3 The purpose of the data generation and its relation to the objectives of the project	7
1.4 Expected size of the data that we intend to generate or re-use	8
1.5 Origin/ provenance of the re-used data	9
1.6 Origin/ provenance of the generated data	11
1.7 To whom the data might be useful outside the project	12
2 FAIR data	12
2.1 Making data findable, including provisions for metadata.....	12
2.2 Making data accessible	12
2.3 Making data interoperable.....	13
2.4 Increase data re-use.....	13
3 Other research outputs.....	14
4 Allocation of resources	14
5 Data security	15
6 Ethics	15
7 Other issues	16

List of Abbreviations and Acronyms

Acronym	Meaning
BC CAS	Biology Centre of Czech Academy of Sciences
BOKU	BOKU university (coordinator)
CC0	Creative Commons licenses
CC BY 4.0	Creative Commons Attribution 4.0 International
DS	Demonstration site
ENL	European natural lakes
ICRA-CERCA	Catalan Institute for Water Research
MS	Monitoring site
NbS	Nature-based solutions
NKUA	National and Kapodistrian University of Athens
NPVJ	Nature Park Vransko Lake
PSB	Paris School of Business
SO1 – SO4	Specific objective 1 - 4
UDJG	Dunarea de Jos University
UiT	Arctic University of Norway
WP	Work package

Executive Summary

This Data management plan outlines the ways in which data is collected, generated and/or processed throughout the lifespan of the project ProCleanLakes. Being a living document, a new version will be created every six months to update the initial assumptions on handling data. The overall aim is to keep all used and generated data as open as possible under the FAIR principle, which is the ultimate goal of this and the future ProCleanLakes data management plans.

1 Data Summary

The project ProCleanLakes will use historical data for quality of lake water at the following Demonstration and Monitoring sites (DS and MS):

DS1: Trichonis lake, Greece

DS2: Brates lake (Romania)

DS3: Langvatnet lake (Norway)

MS1: Vransko lake, Croatia

MS2: Medard lake (Czechia)

MS3: Zaghen lake (Romania)

The purpose of the use of these data is to create a database containing both historical and new data of water quality which will be used for development of possibly best-fitting nature-based solutions for remediation and protection of European natural lakes. All the data will be part of the deliverable No D1.1 "Database of all available data on ecological and chemical status of each DS and MS"

For assessment of water quality by remote sensing, data provided by Copernicus open access hub will be used to develop a reliable tool for remote sensing. This data will be complemented with the historical and newly produced data (see previous paragraph) and processed within the project.

Besides the measurement data, the project will generate and collect new data within the co-creation activities with stakeholders. These data will contain database of all relevant stakeholders at the demonstration sites and results of surveys, interviews and round tables at demonstration sites. Similar activities will be done later at replication sites.

The last type of generated data will be software for digital tools which will be used to raise awareness of the project results and to increase the uptake of the final solutions by end users. These tools will be: virtual business incubator and accelerator, knowledge hub, e-learning courses and citizen science app.

1.1 Data types and formats which ProCleanLakes will re-use

Historical data of water quality will be standardized to common units and saved in a in the .csv format.

1.2 Data types and formats which ProCleanLakes will generate

- Biological data collected in fish surveys will be saved in the .csv format
- Water quality data: .csv format
- Pictures of lakes provided by citizens will be stored as .jpeg or .png data
- Database of business-related stakeholders will be stored as a table in the .csv format
- Interviews with stakeholders: .csv datasets
- Surveys with stakeholders: audio and transcripts (.doc, .pdf)
- Round tables with stakeholders: minutes (.doc, .pdf)
- Remote sensing: we will use Copernicus Data (Sentinel-2 satellite Images) either aggregated in the format "csv" or as raw data in "netcdf" format.
- Citizen science app, Virtual business incubator and accelerator, Knowledge hub and E-learning courses: the input and output types of data used in our software are as follows:
 - Numbers: integer and real
 - Texts: pdf, doc, txt, strings
 - Audio: mp3, wav
 - Video: mp4
 - Images: png, jpg
 - Dates (calendar days)

1.3 The purpose of the data generation and its relation to the objectives of the project

- **Water quality data and fish survey data** will be used for the development of nature-based solutions (NbS) for restoration and remediation of European natural lakes (ENL). This will directly contribute to the Specific objective 1: Design, optimization, demonstration and evaluation of integrated NbS for restoration and protection of ENL.
- **Pictures of lakes provided by citizens** will be generated by citizens while using the citizen science app which will be developed as a part of citizen engagement and also as a source of data for evaluation of a lake status based on the lake colour. This will directly contribute to the Specific objective SO4: Inform, educate and build-capacity of relevant stakeholders on the co-develop and co-design NbS &

business opportunities and engage citizens toward restoration and protection of ENL.

- **Database of business-related stakeholders** is a preparatory activity to develop market analysis for the three demonstration sites, contributing to the Specific objective SO2: Identify innovative business models for sustainable exploitation of healthy European natural lakes towards circular economy.
- **Interviews, surveys and round tables with stakeholders** will be carried out as a part of co-creation approach for development of solutions tailored for the stakeholders. This will directly contribute to the Specific objective SO3: Establish strategies for an effective and targeted spreading and uptake of the developed Action Plan & business opportunities to relevant local-, national-, and European authorities and policy makers.
- **Software data:** development of digital tools (Citizen science app, Virtual business incubator and accelerator, Knowledge hub and E-learning material). This will contribute directly to the Specific objective SO2: Identify innovative business models for sustainable exploitation of healthy European natural lakes towards circular economy (virtual business incubator), SO3: Establish strategies for an effective and targeted spreading and uptake of the developed Action Plan & business opportunities to relevant local-, national-, and European authorities and policy makers (knowledge hub), and SO4: Inform, educate and build-capacity of relevant stakeholders on the co-develop and co-design NbS & business opportunities and engage citizens toward restoration and protection of ENL (citizen science app)
- **Copernicus data (Sentinel-2 satellite Images)** will be used for development of remote sensing tool for assessment of water quality of ENL. This will contribute directly to the Specific objective SO1: Design, optimization, demonstration and evaluation of integrated NbS for restoration and protection of ENL, specifically to the development of remote sensing for assessment of water quality of ENL based on satellite data

1.4 Expected size of the data that we intend to generate or re-use

- .csv data are expected not to exceed 2,000 KB
- .doc data are expected up to 2,000 KB
- .jpeg and .png files might differ 2,000 and 8,500 KB, depending on the quality of the figures. The expected number of the citizen science app is >50, so the final size of the collected pictures van be assessed to $50 \cdot 5,000 \text{ KB} = 250 \text{ MB}$
- Software: the audio type of data an amount of 50 – 100 GB, for the videos 100 – 200 GB and the images 50 - 100 GB.

- Copernicus data: 10 – 50 GB

1.5 Origin/ provenance of the re-used data

Sources of the historical data are as follows:

Type of source	Author(s)	Title	Source	Year	Lake	Unique identifier	Publicly open?
Sci publication	Overbeck, J., Anagnostidis, K. & Economou-Amilli, A.	A limnological survey of three Greek lakes: Trichonis, Lyssimachia and Amvrakia.	Arch. Hydrobiol. 95: 365-394.	1982	Trichonis	n/a	yes
PhD thesis	Tafas, T.	The microflora of Lake Trichonis (Aetolo-Akarnania, Western Greece): Qualitative and quantitative phytoplankton analysis	Dissertation thesis. Biology Department. National and Kapodestrian University of Athens	1991	Trichonis	n/a	No
Sci publication	George Kehayias, Evangelia Doulka	Trophic State Evaluation of a Large Mediterranean Lake Utilizing Abiotic and Biotic Elements	<u>Journal of Environmental Protection</u>	2013	Trichonis	DOI: 10.4236/jep.2014.51003	yes
Sci publication	Mavromati, E., Kemitzoglou, D., Tsiaoussi, V., & Lazaridou, M.	A new WFD-compliant littoral macroinvertebrate index for monitoring and assessment of Mediterranean lakes (HeLLBI)	<u>Environmental Monitoring and Assessment</u>	2021	Trichonis	DOI: https://doi.org/10.1007/s10661-021-09493-1	yes
Sci publication	Nikolaidis, N. P., Koussouris, T., Murray, T. E., Berthahas, I., Diapoulis, A., & Gritzalis, K	Seasonal variation of nutrient and heavy metals in Phragmites australis of Lake Trichonis, Greece	Lake and Reservoir Management, (12)3, 364-370.	1996	Trichonis	n/a	no
Sci publication	Perivolioti, T., M., Frouzova, J., Tuser, M., & Bobori, D.	Assessing the fish stock status in lake trichonis: A hydroacoustic approach	Water	2020	Trichonis	DOI: https://doi.org/10.3390/w12061823	yes
Technical report	T. Koussouris, I. Bertahas, A. Diapoulis, V. Pakos	Limnological and Hydrobiological Characteristics of Lake Trichonis	In: Daoulas, C., Ed., Limnological, Ichthyological and Fisheries	1993	Trichonis	n/a	no

	and K. Gritzalis		Investigation of Lake Trichonis, NCMR Technical Report (TEE/144), Athens, 1993, pp. 1-37				
Sci publication	T. Tafas, D. Danielidis, J. Overbeck and A. Economou Amilli.	Limnological Survey of the Warm Monomictic Lake Trichonis (Central Western Greece)	The Physical and Chemical Environment," Hydrobiologia, Vol. 344, No. 1-3, 1997, pp. 129-144	1997	Trichonis	DOIhttps://doi.org/10.1023/A:1002914629984	yes
Museum	n/a	n/a	The Goulandris Natural History Museum Greek Biotope/Wetland Centre	?	Trichonis	n/a	no
Sci publication	Vlastos, D., Dailianis, S., Kindou, A., Antonopoulou, M., Gianni, A., & Zacharias, I.	Assessing the environmental/human risk of potential genotoxicants in water samples from lacustrine ecosystems: The case of lakes in Western Greece.	Science of The Total Environment	2017	Trichonis	DOI: https://doi.org/10.1016/j.scitotenv.2016.09.042	yes
Sci publication	Zacharias, I., Bertachas, I., Skoulikidis, N. & Koussouris, T.	Greek lakes: Limnological overview	Lakes & Reservoirs: Research & Management, 7(1), 55-62.	2002	Trichonis	DOI: 10.1046/j.1440-1770.2002.00171.x	yes
Sci publication	Ajeagah A., Praisler M., Cioroi M., Constantin O., Palea M., Bahrim G.E.	Biological and Physico-Chemical Evaluation of the Eutrophication Potential of a Highly Rated Temperate Water Body in South – Eastern Romania	Journal of Environment and Ecology	2014	Brates	DOI: 10.5296/je.v5i2.6512	yes
Conference paper	Ciubotariu A.C., Istrate G.G.	PHYSICO – CHEMICAL PARAMETERS OF WATER FROM GALATI AREA (ROMANIA)	16th International Multidisciplinary Scientific GeoConference SGEM 2016	2016	Brates	DOI: 10.5593/SGEM2016/B31/S12.073	yes
Unpublished data	n/a	n/a	Measurements (confidential)	2020-2023	Brates	n/a	no
Conference paper	Iticescu C., Georgescu P.L., Calmuc M.,	CLIMATIC AND ANTHROPOGENIC INFLUENCES ON WATER QUALITY IN		2023	Brates	DOI: 10.5593/sgem2023V/3.2/s12.07	no

	Calmuc V., Topa C.	LAKE BRATES, ROMANIA					
Sci public ation	Vasile A., Zara M., Paltenea E.	THE EVOLUTION OF CERTAIN PHYSICAL- CHEMICAL PARAMETERS OF THE WATER USED FOR THE PROVISIONING OF THE AQUATIC ECOSYSTEM BRATEȘ	Scientific Study and Research	2008	Brates	ISSN 1582- 540X	yes
Measu rement s	n/a	n/a	Measurements at lake stations	1987 - 2008	Langvatn et	n/a	no
Measu rement s	n/a	n/a	Unpublished data (confidential)	2012	Medard	n/a	no
Measu rement s	n/a	n/a	Unpublished data (confidential)	2012	Zaghen	n/a	no
Mesur ement s	n/a	n/a	Croatian Waters	2007 - 2023	Vransko	n/a	no

1.6 Origin/ provenance of the generated data

- **Fish survey data** will be collected in the demonstration lakes by the partner BC CAS
- **New data of water and sediments quality:** direct measurements at the above-mentioned demonstration- and monitoring lakes. Different parameters for water and sediments samples were measured on-site and in laboratories of different project partners: BOKU, BC-CAS, PMF, UiT, NPVJ, ICRA-CERCA, UNIVE, NKUA, AUA and UDJG.
- **Pictures of lakes:** generated by citizen with the use of citizen science app (to be developed in the project)
- **Database of business-related stakeholders:** on-line and on-site recherche at the demonstration sites
- **Interviews and round tables with stakeholders:** protocols will be developed by the project; data will be collected at demonstration and replication lakes as a part of co-creation approach: UDJG, PSB
- **Surveys with stakeholders:** surveys will be sent to identified stakeholders on-line and collected as anonymised responses: UDJG, PSB
- **Software data** will be developed by the project partner Altfactor
- **Satellite data:** Copernicus Earth's observation system

1.7 To whom the data might be useful outside the project

Data of water quality and fish surveys data (both historical and newly produced): researchers and experts in the field of water quality and restoration and remediation of water sources, lake managers, authorities dealing with water quality, biodiversity experts

Pictures of lakes: water quality experts, general public interested in water quality

Database of business-related stakeholders: entrepreneurs who wish to start business activities at the demonstration lakes, use for scale-up of the solutions in replication regions

Software data: experts in software development

2 FAIR data

2.1 Making data findable, including provisions for metadata

All generated data will be identified by a persistent identifier (e.g., DOI). To allow discovery of data, rich metadata will be provided for each dataset which will be deposited in a repository. Each dataset will be accompanied by a Readme file which will summarize the basic information of the uploaded dataset. A template for Readme file for the project ProCleanLakes has been created to harmonize the recorded information.

Beside Readme file, the following scheme for creating of datasets titles and their versions will be used:

YYYYMMDD-ProCleanLakes-Title-VXX (number of version)

Due to the multidisciplinary nature of the project, Dublin Core metadata standard will be used for the description of metadata. To optimize the possibility for discovery and potential re-use, up to five search keywords will be provided as a part of metadata for each dataset. All Data will be described with rich metadata, so they can be harvested and indexed.

2.2 Making data accessible

Repository:

All data will be centrally deposited in Zenodo, ProCleanLakes-community which was created by the consortium. Data related to water quality and water biodiversity will be deposited also in Freshwater Biodiversity Data Portal:

<https://data.freshwaterbiodiversity.eu/>. Remote sensing-related data will be deposited also in GitHub.

In the above-mentioned repositories, we have created accounts which will be used for upload of the data by the project Data manager or by project partners. All the above-mentioned repositories provide the opportunity to assign to each dataset a persistent identifier.

Data:

If some of the data will be used in a publication, they will be made available immediately after the publication gets accepted. In case some of the data provided by the project beneficiaries cannot be openly shared due to restricted access to them, this part of the data will be removed from the respective dataset. All relevant data will be openly available through the Zenodo repository after an internal check of the consortium. Central data sets will be made openly available after the end of the project at the latest.

All data will be stored on a cloud with a restricted access only to the consortium members. Besides, all partners will keep data which they generated and re-used on their personal computers as a back-up. The access rights will be managed by Lada Fialova and Erika Yessenia Cuida (both BOKU).

Metadata:

Metadata will be made openly available and licenced under a public domain dedication CC0, as per the Grant Agreement. Metadata will contain information to enable the user to access the data, the metadata standard Dublin Core will be used. In Zenodo, as our central repository, data and metadata will stay 20 years (see Zenodo Policies <https://about.zenodo.org/policies/>). All software produced by ProClenLakes will be generated as open-source code with rich metadata which can be used later for e.g., extensions of the software for more outputs.

2.3 Making data interoperable

All data should be interoperable by using Dublin Core as standard for metadata.

2.4 Increase data re-use

Data re-use will be increased by the use of readme files which will contain information on creator, reason(s) for data analysis, creation date of file(s), used method(s), used software (incl. versions and add-ons) and tools, data (title(s) and content), code (title(s) and content), additional files, licences, notes.

Our data will be licensed under Creative Commons Attribution 4.0 International (CC BY 4.0) and will be freely available for a possibly widest re-use. We strive for openness and re-usability of the data even after the project end. Their presence in the trusted repositories will enable their use by interested parties. Provenance of the respective data will be secured through the use of the readme files and Dublin Core metadata standards.

Quality assurance of the generated data concerns mostly the measurements of water quality. Following quality assurance has been implemented:

- Collecting samples: the consortium has developed protocols for harmonized collection of water samples which allows generation of comparable data and which will enable development of best fitting nature-based solutions for each demonstration lake
- Laboratory analyses: samples collected at demonstration lakes will be sent to BOKU, UNIVE, AUA, ICRA-CERA and UDJG. For each analysis, standard quality assurance tests have been developed which are binding for each of the laboratories

Quality assurance of data generated in surveys and interviews with stakeholders will be secured through a double check of the interview protocol and survey questionnaire by two consortium members different than the creators of the protocol and questionnaires.

Besides, all generated data are part of project deliverables which have to undergo a double check by one consortium member and the project coordinator.

3 Other research outputs

Besides research data, other project outputs such as presentations, surveys, protocols, audio records of interviews and their transcripts, software will be deposited in Zenodo using the above-described procedures (titles, versioning, readme files, metadata standard).

4 Allocation of resources

For the storage of the data, OneDrive cloud will be used. No additional costs are necessary. Costs for Data Manager who takes care of creating proper Readme files and metadata, are fully covered by the project as part of personnel costs of the beneficiary BOKU.

For creating an open access to the data, the repositories Zenodo, GitHub and Freshwater Biodiversity Data Portal will be used which are free of charge.

Person responsible for data management in the project ProCleanLakes is Lada Fialova (BOKU). Datasets must be uploaded first in Zenodo, which can be done by each project partner and saved as a draft, keep in all procedures for data management established in the consortium (readme file, title, metadata etc.). Data Manager will do a check of each uploaded dataset before making it publicly available.

Preservation of data will depend on the storage length of Zenodo and GitHub. For the time being, Zenodo declares 20 years of preservation time (see policy <https://about.zenodo.org/policies/>).

5 Data security

ProCleanLakes will make sure that the all the data collected/generated is safely stored for long term preservation. As stated above, the general principle will be a central archival on the Zenodo platform for the whole consortium plus a storage of each partner's data on their institutional servers. This issue will be discussed later in the project and the proper measurements and actions will be taken toward this objective.

6 Ethics

During engagement with stakeholders (interviews, surveys, round tables, WP2, 3, 7), we will collect anonymous data, which cannot be in any way related to a specific person. Questions during in-person interviews will not be of a sensitive nature and are not foreseen to cause harm to participants and any personal data will be collected and handled in accordance with the General Data Protection Regulation [GDPR - Regulation (EU) 2016/679]. Each participant will be informed on these facts before start of a respective activity.

Collection of biological data during fish surveys: the project will conduct fish community survey and assessment of (predominantly) species composition, abundance and biomass. The collection of these data will cause no harm of discomfort to the fish and will be handled in accordance with the EU Regulation on protection of animals used for scientific purposes [Directive 2010/63/EU], and national legislation and licencing requirements, to ensure that the high standards of animal welfare and scientific research.

Use of artificial intelligence during development of digital tools: in our project, we are committed to upholding the principles of ethical artificial intelligence as mandated by relevant legislation, including the AI Ethics Act of 2023 and the Responsible AI Development and Deployment Act of 2023. Our approach to the integration of artificial intelligence ensures the responsible and ethical use of AI technologies throughout the project lifecycle.

7 Other issues

Not applicable.