DATA MANAGEMENT PLAN

PROJECT	
Project number:	**101079234**
Project acronym:	**BIOLAWEB**
Project name:	**Boosting Institute of Chemistry, Technology and Metallurgy in Water Biomonitoring**

DATA MANAGEMENT PLAN	
Date:	**29.03.2024**
Version:	**V2.0**

Funded by
the European Union

Table of Contents

Funded by
the European Union

1. Data Summary

1.1. Re-use of existing data

The BIOLAWEB project will strongly rely on DNA sequences available in international OA databases such as the reference barcoding library Diat.barcode and GenBank. These data are neccessary to compare the results generated in the BIOLAWEB project.

1.2. Generation of new data

The BIOLAWEB project will generate new DNA sequences of diatoms, phytoplankton and macrophyte species. BIOLAWEB will collect data on water chemistry, macrophyte, benthic diatoms and phytoplankton assemblages from study lakes. BIOLAWEB will also collect data from internal surveys (questionnaires and interviews).

1.3. Data types, purpose, formats, size, origin, and utility outside the project

BIOLAWEB datasets will include (i) quantitative and qualitative physical, chemical, hydrological and species data (species names and abundances of algae and macrophytes, including mapping data), (ii) DNA sequences from metabarcoding of the samples collected, and (iii) relevant third-party data from stakeholders. Table 1 summarizes data sets linked with project tasks/objectives, their purpose and format in which the data will be stored during the BIOLAWEB project.

2. FAIR data

2.1. Making data findable, including provisions for metadata

Persistent identifier

The BIOLAWEB consortium will store generated and collected data in an open online research data repository. The BIOLAWEB team decided to use a digital repository of the Institute of Chemistry, Technology and Metallurgy, University of Belgrade, CeR - Central Repository ICTM (https://cer.ihtm.bg.ac.rs/) as its data archive, based on the compliance of the repository structure and facilities with the FAIR data principles. This CeR repository allows researchers to deposit both publications and data, providing tools to link them to these through **persistent identifiers** and data citations.

Quality-assured sequence data for diatoms, phytoplankton, and macrophytes will be submitted to GenBank, an annotated collection of all publicly available DNA sequences.

In addition to the CeR repository, the consortium has agreed that certain data (publications, abstracts, maps, etc.) generated from the BIOLAWEB project will be deposited on the Zenodo platform (https://zenodo.org/).

Metadata types and standards

The CeR repository provides the possibility to add extensive metadata about researchers' publications, posters, and presentations (e.g., around 29 metadata for each publication including file name, size, format, date accessioned, date available, rights, license, for mapping results – lake name, organism group, method used for mapping/counting/measuring, coordinates, sampling date, etc.).

The Zenodo repository provides the possibility to add extensive metadata pertaining to researchers' publications, presentations, etc. This includes more than 30 metadata points for each publication, encompassing details such as file name, size, format, accession date, availability date, rights, and licensing information. Additionally, metadata relevant to mapping results - such as lake name, organism group, mapping/counting/measuring methods, coordinates, and sampling date - are also included.

Quality-assured macrophyte sequence data will be submitted to GenBank. All data will contain the necessary metadata required by GenBank, including species name, sampling location, collection date, publication reference and a persistent identifier for each entry. GenBank is an annotated collection of all publicly available DNA sequences so that the quality-assured data of macrophytes generated in BIOLAWEB can be easily found and accessed.

Sequence data will be also integrated into the reference barcoding library Diat.barcode. For each sequence, all necessary metadata are given and follow the CEN TR 17244 (CEN, 2018. Water quality - CEN/TR 17244 - Technical report for the management of diatom barcodes 1–11.). Diat.barcode is an open-access library available at https://doi.org/10.15454/TOMBYZ.

Table 1. BIOLAWEB project data summary

Data type	Task	Description	Data utility*
Dataset 1	Task 1.2: Internal communication	Purpose: To collect information for preparing consortium and internal meetings, presentations and minutes from the meetings Data origin: List of participants, presentations, pictures, promo video materials Format: .docx, .pdf,.xlsx, .ppt, .png, .jpg, .mp4 Expected size: biannual consortium meetings (non-video files: 6×150MB=**900 MB**; video files (.mp4): 2×600 MB=**1.2 GB**), weekly UB-ICTM internal meetings (120×1.75 MB=**210 MB**)	All files from internal UB-ICTM meetings are restricted to the BIOLAWEB consortium. Pictures and promo videos from kick-off/consortium meetings will be made openly available.
Dataset 2	Task 2.1 Expert visit	Purpose: To collect data on the internal structure of UB-ICTM (Objective 1 – Scientific strategy for UB-ICTM) Data origin: Attendance list, pictures, presentations, expert visit report Format: .pdf, .ppt, .png, .jpg Expected size: .pdf and .ppt files=**30 MB,** .jpg meeting pictures=**25 MB**	Restricted to BIOLAWEB consortium.
Dataset 3	Task 2.2 Systematic survey of current research excellence indicators of UB-ICTM	Purpose: To collect information on UB-ICTM research performance in the field of aquatic ecology (Objective 1 – Scientific strategy for UB-ICTM) Data origin: Questionnaire Format: .pdf, .docx Expected size: pdf and .docx files=**2 MB**	Restricted to BIOLAWEB consortium.
Dataset 4	Task 2.3 Identification of short and long-term research needs of UB-ICTM	Purpose: To collect data on UB-ICTM existing knowledge gaps and needs in the field of biomonitoring (Objective 1 – Scientific strategy for UB-ICTM) Data origin: Attendance list related to the interview with researchers, pictures Format: .pdf,.docx, .jpg Expected size: .pdf and .docx files=**1 MB,** .jpg pictures=**1.5 MB**	Restricted to BIOLAWEB consortium.

Dataset 5	Task 2.4 Formulating Strategy and Action plan	Purpose: To collect information for creating Science and Innovation Strategy of UB-ICTM (Objective 1 – Scientific strategy for UB-ICTM) Data origin: Science and Innovation Strategy of UB-ICTM in the field of biomonitoring, pictures Format: .pdf, .docx, .jpg Expected size: pdf and .docx files=**25 MB,** .jpg meeting pictures=**2 MB**	Restricted to BIOLAWEB consortium.
Dataset 6	Task 3.1: Methods for developing ecological status indices	Purpose: teaching material; no new data will be generated within this task Data origin: n.a. Format: .ppt, .pdf Expected size: **100 MB**	Primarily restricted to BIOLAWEB consortium. A somewhat reduced version (to respect copyrights) will be made publicly available.
Dataset 7	Task 3.2: Metabarcoding of diatoms and phytoplankton	Purpose: teaching material, HTS sequencing data Data origin: n.a. Format: .pdf + fastq Expected size: **1.2 GB**	Teaching material : Accessible to any participants of the courses. Sequencing data : Primarily restricted to BIOLAWEB consortium. Then, open-access publication.
Dataset 8	Task 3.3: eDNA of macrophytes	Purpose: teaching material; no new data will be generated within this task Data origin: n.a. Format: .ppt, .doc, .xls Expected size: **200 MB**	Teaching material: Accessible to any participants of the courses.
Dataset 9	Task 4.1: eDNA for ecological status assessment (ESA) using diatoms	Purpose: To collect information on diatom composition in the sampled lakes using morphological and genetic approaches. Data origin: sampling reports, list of species Format: .xls Expected size: **10-20 MB**	Primarily restricted to the BIOLAWEB consortium. Then,open-access publication.
Dataset 10	Task 4.2: eDNA for ecological status assessment using phytoplankton	Purpose: To collect information on phytoplankton composition in the sampled lakes using a morphological and genetic approach Data origin: sampling reports, list of species Format: .xls, .pdf, .docx, .jpg Expected size: **10-20 MB**	Primarily restricted to the BIOLAWEB consortium. Then, open-access publication.
Dataset 11	Task 4.3: eDNA for ecological status assessment using macrophytes	Purpose: to develop a novel method to detect the occurrence of macrophytes from a water sample Data origin: macrophyte DNA sequences from 4 lakes in Serbia, and macrophyte mapping data from the same lakes Format: .xls, raw sequence reads (fastq, fasta) Expected size: **1 to 10 GB** (depending on number of sequences)	The data from “traditional” and genetic macrophyte mapping will be published in scientific publications and in the digital repository of UB-ICTM and the Zenodo platform.

Dataset 12	Task 4.3: Maps for purpose of publishing articles, abstracts, ppt, poster, etc.	Purpose: to visualize areas of research. Data origin: different maps for 4 lakes in Serbia. Maps include different types of layers beside the lakes themselves. Different thematic maps are created depending on the purpose and research goals. Format: .jpg, .pdf, .png, .tiff Expected size: **100 MB to 1 GB** (depending on the number of maps, size and required resolution etc.)	These data are useful for the scientific community (within freshwater ecology). The data from mapping will be published in scientific publications and in the digital repository of UB-ICTM and Zenodo platform.
Dataset 13	Task 5.1 Expert visit for need analysis in project management	Purpose: Bilateral discussion to estimate needs in project management. Data origin: Attendance list, pictures, presentations, expert visit report Format: .docx, .pdf, .ppt, .jpg Expected size: .docx, .pdf and .ppt files=**10 MB,** .jpg meeting pictures=**75 MB**	Restricted to UB-ICTM staff and BIOLAWEB consortium.
Dataset 14	Task 5.2 Setting up a new Research Management Office	Purpose: To collect information on potential candidates for the new office Data origin: application material, evaluation material Format: .pdf, .docx Expected size: pdf and .docx files=**4 MB**	Strictly and confidentially used by UB-ICTM (Human Resources Department).
Dataset 15	Task 5.4 Organization of workshops, on-site and virtual trainings in project management	Purpose: To collect data for workshops and trainings Data origin: Attendance list, presentations, pictures, video materials Format: .docx, .xlsx, .pdf, .ppt, .jpg Expected size: .docx, .xlsx, .pdf, .ppt=**13 MB,** .jpg event pictures=**15 MB**	Materials (except the attendance list) will be made openly available.
Dataset 16	Task 5.5 Support of ICPO at UB-ICTM to researchers in project submission and implementation	Purpose: To collect data for info days and trainings Data origin: Presentations, training material, pictures Format: .docx, .xlsx, .pdf, .ppt, .jpg Expected size: .docx, .xlsx, .pdf, .ppt= **25 MB,** .jpg event pictures=**15 MB**	Restricted to UB-ICTM staff and BIOLAWEB consortium.
Dataset 17	Task 6.2: Communication resources and tools	Purpose: To collect quantitative data on visitors of BIOLAWEB website, social media, Festival of Science and Researchers' night Data origin: Web site, Google Analytics Format: .html Expected size: for .html=**10 GB**	Restricted to UB-ICTM staff and BIOLAWEB consortium.
Dataset 18	Task 6.3: Generation of roundtables with relevant stakeholders	Purpose: To collect information on roundtable participants and stakeholders Data origin: Attendance list, pictures, presentations, roundtable minutes Format: .docx, .pdf, .ppt, .jpg Expected size: .docx, .pdf and .ppt files=**25 MB,** .jpg event pictures=**15**	Restricted to BIOLAWEB consortium.

		MB	
Dataset 19	Task 6.4: Field work with stakeholders	Purpose: Field work with stakholders Data origin: Attendance list, pictures, reports Format: .pdf, .jpg Expected size: .pdf files=**5 MB,** .jpg event pictures=**30 MB**	Restricted to BIOLAWEB consortium.
Dataset 20	Task 6.5: Scientific dissemination	Purpose: To disseminate project results Data origin: peer-reviewed scientific articles and conference papers Format: .pdf, .docx, .ppt Expected size: **100 MB**	The journal articles and conference abstracts will be will be made publicly available (Open Access and digital repositories CER and the Zenodo platform).

*outside the project; **International Cooperation and Project Office (ICPO) is a new and officially accepted name (with acronym) for the established office. IRMO was an initial acronym used in the project proposal.

2.2. Making data accessible

Repository

The BIOLAWEB consortium identified CeR as a trusted repository suitable for storing datasets generated during the project. It is based on the platform DSpace (https://duraspace.org/dspace/) which is one of the most popular platforms for open science. Within CeR, the BIOLAWEB project data are deposited and organized in a specific part of the repository (catalog) intended for international projects to facilitate data reuse (https://hdl.handle.net/21.15107/rcub_cer_7507). The CeR repository ensures that the data is assigned an identifier and resolves the identifier into a digital object.

Zenodo has become an important resource in the academic community as it facilitates the sharing and access of scientific data, contributing to transparency, collaboration, and research advancement. Data published on Zenodo receive permanent digital identifiers (DOIs), making it easier to cite and track them in academic works. Zenodo integrates with other tools and services such as GitHub (for coding), ORCID (for researcher identification), and OpenAIRE (for promoting open access). Zenodo allows for the publication of various data formats, including text documents, images, video recordings, datasets, and other types of digital content.

To avoid potential duplication of entries on the Zenodo platform, a Data Protection Officer (DPO) will be responsible for entering data into the database.

GenBank is an annotated collection of all publicly available DNA sequences, a well-established international database so that the quality-assured data generated in BIOLAWEB can be easily accessed.

Embargo and restricted access

Following the HORIZON EUROPE guidelines, scientific data will be made available for re-use (the sequence data, the mapping data, and the water chemistry) as well as workshops materials. On the other hand, in case personal data is needed they will be anonymized, and restricted to the BIOLAWEB consortium (Table 2).

Table 2. Description of internal data created during the BIOLAWEB project

Data type	**Utility**	**Description and purpose**
Internal strategy datasets	Datasets were used to develop Science and Innovation Strategy and Action Plan (Objective 1)	All generated data from expert interviews and filled questionnaires remained restricted to the BIOLAWEB consortium.
Personal datasets	To make a list of potential workshops, trainings and roundtables attendees and their expectations (Objectives 2, 3 and 5).	Attendance lists and filled questionnaires (workshops, trainings, roundtables, etc.) remained restricted to the BIOLAWEB consortium.

BIOLAWEB researchers will publish their work in scientific journals with large audiences and aim to publish Gold Open Access (OA) wherever possible (research data immediate availability). According to the GA, the option for using repository-based or Green OA for some project publications is also foreseen. In addition, the list of project publications with links to the publications and the selected supplementary data will be uploaded to the BIOLAWEB website. According to the established access protocol within the CeR repository, the BIOLAWEB research data will either be freely accessible (biological, chemical and sequence data) or only accessible to registered users (e.g., research staff of the institute). The identity of the person accessing the data is linked with her/his institutional e-mail. At the moment, there is no need for a data access committee.

Metadata

In the CeR repository, metadata associated with scientific publications, project presentations and reports will be openly available and licenced under a public domain (CC0) during the project lifetime and at least five years after the project ends.

GenBank is an annotated collection of all publicly available DNA sequences, and metadata are provided in a standardized way. This ensures that all macrophyte sequence data will be provided with the necessary metadata including species name, sampling location, collection date, publication reference and a persistent identifier for each entry. The BIOLAWEB team will ensure that all data will be made publicly available in GenBank immediately after publication.

2.3. Making data interoperable

The BIOLAWEB team will deposit data and related metadata in the open repository CeR which fulfils the criterion of interoperability. This repository uses general standard Dublin Core which will make BIOLAWEB data interoperable with other services that provide or offer metadata about scientific publications. The use of uncommon or the creation of specific vocabularies is not foreseen for the project data, but this point could be discussed or reassessed in the consortium meetings.

All data submitted to Genbank follow the community-endorsed standard and therefore are in an interoperable format.

The sequencing data will be analysed using reference barcoding databases developed in former research programs (Diat.barcode and Phytool). These reference libraries are used by a worldwide audience of scientists working on algal metabarcoding.

2.4. Increase data reuse

Bioinformatic pipelines and statistical analyses will be published on Open Access platforms (GitHub, GenBank, entrepot.recherche.data.gouv.fr). For example, GenBank accession numbers will be provided in all publications using these entries, thereby ensuring maximum reuse of the data.

BIOLAWEB will make research publications available through the deposition of a copy of the published, or final, peer-reviewed version, in identified suitable institutional repositories. Datasets uploaded in the CeR or Zenodo repository will be freely accessible, immediately in case of Gold Open Access (OA), or after an embargo period - Green Open Access. Potential users are expected to adhere with the CeR or Zenodo terms of Use.

BIOLAWEB will also publish the project results (list of scientific publications, reports and other results) on the project website as soon as possible, taking into account the EC deadline.

Datasets generated in the BIOLAWEB project and deposited in the CeR or Zenodo repository under an Open Access license can be used without restrictions by third parties at least 5 years after the end of the project. Data deposited in GenBank are freely available “for eternity “.

The provenance of research data generated in the BIOLAWEB project will be tracked and those responsible for its generation will be adequately credited for their work using appropriate Creative

Commons copyright licenses. Data submitted to GenBank follow a community-endorsed standard and will therefore be thoroughly documented.

Data quality assurance

The members of the Project Executive Board approve data collection and deposition in the open repository by checking the following critical data quality dimensions: 1. **Integrity** (data stability during the lifecycle); 2. **Validity** (data items can be traced); 3. **Accuracy** (data fits reality); 4. **Consistency** (values in former and latter datasets are consistent); 5. **Timeliness** (the time between expected and actual data availability); 6. **Currency** (data is up-to-date) and 7. **Completeness** (no missing values in table/box fields). After the end of the project, this role will be transferred to the repository administrator at UB-ICTM.

In general, only quality-assured data will be published and submitted to GenBank, thereby ensuring that only valid data will be made publicly available. Metabarcoding and sequence data will be screened by using suitable software.

3. Other research outputs

Diatom samples collected in the BIOLAWEB project will be deposited in the Diatom Collection of Serbia (DCSR) at the University of Belgrade, Institute of Chemistry, Technology and Metallurgy, National Institute of the Republic of Serbia.

Teaching material generated in WP3 will be made available on our website and as project deliverables. In some cases, we will make a slightly reduced version available, to avoid copyright issues (generally related to illustrations which may be used for teaching purposes only).

In the framework of the BIOLAWEB project, protocols that will be used for diatom and phytoplankton metabarcoding will be protocols already developed in other projects of INRAE. BIOLAWEB members will refer in deliverables to these protocols which are open-accessible on platforms of INRAE. If changes are made to these protocols, BIOLAWEB members will notify it and will make these modified protocols open-accessible. The protocols for macrophytes generated in the project will be made openly available via our website and as project deliverable.

4. Allocation of resources

No costs for the curation of data related to DCSR are foreseen. Submission of data to GenBank is part of the publication process and is free of charge. No costs are incurred for further archiving.

Data protection and long-term storage

UB-ICTM is responsible for data management of the macrophyte sequence data. The Data Protection Officer (DPO) will be responsible for overall data management in the BIOLAWEB project.

The Data Protection Officer (DPO) of the BIOLAWEB project is:

Dr Vladimir Petrović,
University of Belgrade
National Institute of the Republic of Serbia
Institute of Chemistry, Technology and Metallurgy
Department of Ecology and TechnoEconomics
Njegoševa 12, 11000 Belgrade, Serbia
Phone: +381 60 5866885
E-mail: petrovic@ihtm.bg.ac.rs

The Data Protection Officer (DPO) of the BIOLAWEB project is responsible for:

- the highest standards of the security measure to prevent unauthorized access to personal data. DPO should follow the data minimization principle, where all data that are not relevant will be immediately deleted from the record;

- the storage of the data, collection, recording, organization, structuring restriction, erasure, availability, and destruction of the data;
- ensuring that essential documents are not destroyed before the end of the required retention of 5 years and that after that they will be destroyed in their entirety making sure that they cannot be recovered. Data retained for auditing processes will be stored securely and further processed for those purposes only;

The coordinating institution of the BIOLAWEB project (UB-ICTM) will use CeR for long-term storage of research data and metadata produced in BIOLAWEB. Data will be machine-readable, citable, published in a registered long-term open access repository and interlinked with other project outcomes. Each dataset will be accompanied by a full metadata description, explaining their contents, origin, methods, licensing, citation, etc.

Data submitted to GenBank are foreseen to be archived “for eternity,” and necessary resources are provided by the international scientific community.

5. Data security

During the implementation of the BIOLAWEB project, datasets, either generated or reused, will be archived on the storage system of the beneficiaries. Each partner involved in collecting or analysing data is responsible for their secure storage and transfer, and follows national and institutional guidelines concerning data security. Datasets that must be kept confidential are BIOLAWEB internal survey data and documents related to the Science Strategy and Action Plan (D2.2). Sensitive data will not be made openly available.

GenBank is part of the International Nucleotide Sequence Database Collaboration. It is a long-standing foundational initiative that operates between DDBJ (DNA Databank of Japan), EMBL-EBI (EMBL’s European Bioinformatics Institute) and NCBI (National Center for Biotechnology Information, USA). Long-term storage and data security are therefore ensured in the best possible way.

After the end of the project, the responsibility for secure long-term preservation and curation will be transferred to the repositories storing the dataset.

6. Ethics

It is planned to collect local resources in the Republic of Serbia that might include samples of endangered flora. The project will comply with Serbian law with respect to sample collection. This activity is related to WP4 (Research and partnership) and WP3 (Training and networking). Samples containing diatoms, phytoplankton, and macrophytes including *Chara* species will be transferred from Serbia to France and from Serbia to Norway. They will be used only for the purpose of the project (material for training, material used for contributing new sequences to the reference barcoding library, material used for scientific publications). Each Partner is responsible for compliance with the Nagoya Protocol on Access to Genetic Resources and the Fair and Equitable Sharing of Benefits Arising from their Utilization to the Convention on Biological Diversity.

Questionnaires and other surveys dealing with personal data used in the project will include a disclaimer informing respondents on the purpose of data collection and the way the information is going to be further used within the scope of the project (e.g., confidentiality).

7. Other issues

No other procedures for data management are foreseen.

HISTORY OF CHANGES		
VERSION	PUBLICATION DATE	CHANGE
1.0	23.03.2023	Initial version
2.0	29.03.2024	Updated version

QUALITY CONTROL		
ACTIVITY	NAME	DATE
V2.0 created	Dr Miloš Ćirić	15.03.2023
V2.0 reviewed	Željka Milovanović, Dr Vladimir Petrović, Dr Srđan Miletić	18.03.2023
V2.0 created	Dr Andreas Ballot, Dr Susanne Schneider, Dr Frédéric Rimet, Dr Danijela Vidaković	25.03.2023

Funded by
the European Union