

BiodivERsA

BiNatUr

Bringing nature back – biodiversity-friendly nature-based solutions in cities

Data Management Plan

Accepted version
December 2022



History of changes			
Version	Author	Date	Description
1 st draft	Maria Söderholm (Syke)	19.9.2022	The preliminary version of the DMP based on Biodiversa instructions
2 nd draft	Dagmar Haase (Humboldt Universität), Michael Monaghan (IGB Berlin), Pedro Pinho (University of Lisbon), Dörthe Tetzlaff (IGB Berlin), Kati Vierikko (Syke)	27.10.2022	Internal revision
3 rd draft	Maria Söderholm (Syke)	2.12.2022	Updated draft after researchers' contribution
4 th draft	Kati Vierikko (Syke)	20.12.2022	Final internal revision
Final version	Maria Söderholm	20.12.2022	The 1 st BiNatUr DMP

Contents

1. PROJECT DESCRIPTION.....	3
2. DATA MANAGERS.....	3
3. DATA IDENTIFICATION & DESCRIPTION.....	4
4. DATA STORAGE, ORGANISATION AND EXCHANGE WITH THE PROJECT PARTNERS.....	6
5. DATA RESTRICTIONS.....	7
6. DATA SHARING, STANDARDS & METADATA.....	7
7. DATA PUBLISHING, LICENSING & ARCHIVING.....	8
8. COSTS.....	9

1. PROJECT DESCRIPTION

The BiNatUr project will explore the role of biodiversity and its linkages with regulating ecosystem services (ES) in urban aquatic nature-based solutions (aquaNBS), and its overall aim is to improve the planning, building, restoration, and management of aquaNBS, supporting the transformation to climate-smart, biodiversity-friendly, and sustainable cities. BiNatUr will explicitly focus on four main research questions:

- How are biodiversity and ES of aquaNBS mediated by social, ecological, and technological factors?
- Does this vary among cities in different regions of Europe?
- How does biodiversity influence the regulating ES provided by aquaNBS?
- How can urban planning effectively design, manage, and monitor the biodiversity and regulating ES of aquaNBS?

BiNatUr is a consortium of six European research organisations, funded by BiodivRestore RDI in Biodiversa and Water JPI joint call 2021. The project started on 1 April 2022 and will end on 31 March 2025. The project coordinator is the Finnish Environment Institute (Syke) and the project partners are Humboldt University of Berlin (UBER), Leibniz Institute of Freshwater (IGB), Poznań University of Life Sciences, University of Antwerpen and University of Lisbon.

2. DATAMANAGEMENT RESPONSIBILITIES

Data management support is organised as follows:

- The project level: The project data manager is Maria Söderholm, appointed by Syke. The responsibilities of the data managers include, for example, drafting and updating the data management plan (DMP) and drawing on the expertise of the WPs in planning. The data manager also supports the implementation of data management as agreed in the project and reports publications and published research data in the Water JPI JSP database. Finnish Environment Institute as the project coordinator has a general responsibility for data management and compliance with common guidelines.
- Work package (WP) level: The WPs ensure proper data management of the data under their responsibility, provide their expertise to support others in the collection/use of these data and contribute to the joint DMP. The practical responsibility for implementing the data management lies with everyone who produces and/or processes data. At WP level, the working group leader has overall responsibility for ensuring compliance with and implementation of the data management plan. The WP leader is also responsible for the overall quality assurance of the data in the work package. WP leaders appoint people to carry out data management tasks as required.
 - **WP1:** In this WP we use secondary data consisting of published literature. Every partner collects and manage publications related data and shares the data analysis with the partners. Responsible person: Kati Vierikko (Syke).
 - **WP2:** We will conduct individual and group interviews in this WP. Each partner collects, manages and processes local interview data itself and shares the anonymised data between partners. The data will be strictly handled in accordance with GDPR. Responsible person: Hanna Nieminen (Syke).

- **WP3:** In this WP we will use geo-data and a respective geo-database (using a Geographical Information System and open-source data such as OSM, Urban Atlas, Urban Audit). Every partner collects and manage local geo-data including remote sensing images by themselves and shares the data with other partners. Responsible persons: Dagmar Haase and Thilo Wellmann (UBER).
- **WP4:** In the WP we will sample precipitation, surface water, water from aquaNBSs and groundwater in each partner city. Every partner collects local water samples by themselves and sends them to IGB Berlin for analysis in the isotope laboratory. IGB has a dedicated data manager to assist data handling of the isotope data. Responsible person: Dörthe Tetzlaff (IGB).
- **WP5:** In this WP we are sampling ecological components of selected aquaNBS. Every partner collects and manage local data by themselves. A common database will be provided to be completed by all partners, and that database data will be shared among partners by using Teams. For eDNA analyses every partner collects local water samples by themselves and sends them to IGB, Berlin. IGB in Berlin will process water samples (DNA extraction, PCR, next-generation sequencing, bioinformatics). Responsible person: Pedro Pinho (University of Lisbon) and Michael Monaghan (IGP).

The DMP will be updated when there are significant changes in data collection or management. The plan is available to all project participants on the Syke BiNatUr Teams.

3. DATA IDENTIFICATION & DESCRIPTION

1) Interview data (recordings, notes and transcriptions)

We will conduct interviews and organize stakeholder meetings to bring the local and expert understanding and knowledge to our project. Regarding the stakeholder engagement and expert interviews, each partner will select and invite local stakeholders and experts to be involved or interviewed in the project. The obtained data will be personal, as it will initially include relevant contact information. Expert interviews will be recorded, and original records are saved and managed by every partner themselves. Every partner is collecting the interview data and storing them locally with restricted access. Individual recordings are transcribed to the text format by each partner. Transcribed interviews are categorized and summarized by each partner into the excel file followed by the protocol and guidelines made by Syke researchers.

2) Isotope data within water samples

Stable isotope data will be measured in different waters (precipitation, surface water, groundwater) through a combination of autosamplers (higher temporal resolution) but mainly grab sampling. Samples collected locally by each partner will be analysed in the stable isotope lab of IGB. These are numerical data. In terms of number of samples: each body water in all cities will be sampled at least once every season, i.e., 4 times a year. Berlin, Antwerp and Helsinki will produce a higher number of samples as groundwater and precipitation will also be sampled.

3) Data on ecological components of aquaNBS

The following components are sampled from the selected aquaNBS: vegetation in the adjacent terrestrial habitats, aquatic macrophytes, aquatic macroinvertebrates, diatoms and algae, and habitat quality. Filtered water and sediment samples are collected by each

partner and shipped to IGB for eDNA-metabarcoding analysis. Filters are kept frozen (-20 deg C) in 2-ml plastic tubes until processing, after which case no physical samples are kept. Extracted DNA is stored long-term in a deep-freezer (-70 deg C with liquid-N2 backup in case of power loss) at IGB. We expect ca. 250 DNA samples kept in long-term storage. Once sequenced, eDNA data are stored as image files (raw data from the Illumina sequencer) on the Hyper-store file system at the genomics facility (<https://begendiv.de/work-with-us/>) where data are processed. We predict ca. 4 GB of (g-zipped) image data. Processed images are text sequence files (fastq format). We predict ca. 10 GB of (g-zipped) sequence data.

4) GIS data and remote sensing images

Geographical Information Systems e.g. GIS usable data and remote sensing images to create a base for integrating spatial analyses. Local GIS data will be provided by the case study teams and stored during the project lifetime at the BiNatUr partner's servers to facilitate data processing and modelling calculations. After the project's end, all GIS data except own ones by the cases will be deleted. Jointly produced new data layers will be shared among the partners and can be further used in case original data access and use rights allow this.

5) Taxonomic data

Taxonomic data of aquatic and surrounding vegetation by using standardized methods and sample aquatic macroinvertebrates, using WFD standardized methods adapted to local conditions. Samples of macroinvertebrates are identified by taxonomic experts and in some partner cities identifications are made by consultants. Filtered water and sediment samples are collected by each partner and shipped to IGB for DNA-metabarcoding analysis. Taxonomic data of phytoplankton and benthic diatoms will be obtained through sequence matches to reference databases (e.g., PR2). Sampling will allow calculating several taxonomic and trait-based biodiversity-metrics, including those suggested by the WFD. Taxonomic data (species observations) will be stored to national and international research databases (e.g. FinBIF) to make it available for public use and other research purposes.

6) Documents related data

The project will also utilize already published literature as data, such as scientific papers, reports and other written documents.

7) Data obtained elsewhere

The project uses an open access geo-data and a respective geo-database provided by others such as OSM, Urban Atlas, Urban Audit Urban Audit, and Corine Land Cover and others available at the European Data Portals (e.g., Copernicus). In addition, we will gather data from local collaborators and stakeholders including water quality data and unpublished planning documents provided by local authorities and research organisations.

8) Other types of outputs produced by the project

The project will produce up to 10 scientific papers. The project also aims to publish seven blog writings for general public on our own project webpage (www.bringingnatureback.com).

4. DATA STORAGE, ORGANISATION AND EXCHANGE WITH THE PROJECT PARTNERS

In general, data is stored in accordance with secured procedures: The ICT services from partner organisations provide the safe storage and backup services for the project. Backup is mainly done automatically on a regular scheduled basis. Data and all other materials are stored on password-protected servers, accessible only to authorised persons. No special storage solution is therefore needed for personal data, such as interviews.

The storage and proper management of the original data is the responsibility of the partner who collects the data in question or is responsible for processing the samples.

Once the original data has been classified, analysed, summarised, or otherwise processed into an agreed common format, it is reported to and shared with the partners. Data can also be shared with partners using other services such as data archives/repositories and databases as appropriate.

The primary data collected and generated in the project will be stored locally by partners, but pre-analysed results, or common format and aggregated data are shared with the partners. The collaboration platform for the project is Teams, provided by Syke ICT. This Microsoft SharePoint cloud-based service will be used in the project to manage, share, and store non-sensitive documents. Teams can also be used to share (processed) data as applicable. In Teams, each WP has its own folder to disseminate documents and data of their choice. All partners are granted access to the data and documentation in the Syke Teams. Access to other data storage/sharing services used by the project partners will be according to the partner's instructions and in a manner enabled by the system.

To keep oversight of data and ensure the quality, findability and usability, we follow good data storage practices: The data will be organised according to the data specific procedures and associated with the necessary information/ metadata to understand the data, taking advantage of the features of each storage/sharing system. In case of data files proper file naming procedures are used to distinguish between the files.

The following primary data will be collected and stored locally. Primary data that has been classified in a uniform format and/or processed is shared with the project partners using Teams or, depending on the nature of the data, by any other appropriate means:

- Geo-data including remote sensing images:
 - Jointly produced new data layers will be shared among the partners and can be further used in case original data access and use rights allow this.
 - Remote sensing data will be shared within the project's lifetime.
- Expert interview data
- Qualitative data based on document analysis.

The following data will be stored and shared through open data archives/repositories:

- Data on ecological components of aquaNBS and isotopic data based on water samples are stored and shared using the open data repository Freshwater Research and Environmental Database (FRED)*.
- The Sequence Read Archive (SRA)* is used to store and share sequencing data based on local water samples.

* For the descriptions of services, see chapter 7.

5. DATA RESTRICTIONS

The data produced by the project will be made openly available always when possible and useful for re-users. This will be done in accordance with ethical guidelines and applicable legislation, such as the EU's General data protection regulation (GDPR), and respecting ownership and terms of use.

So far, we have recognised that only interview data and remote sensing images cannot be made publicly available in our project.

As the interview recordings and transcriptions are personal data, special attention will be paid to the collection, processing and sharing of this data:

- The collection of interview data will be planned and carried out with respect for the rights of participants: The interview data is collected and processed only for scientific research purposes based on consent, and the collection of direct and indirect identifiers will be minimized. The participation in the study is voluntary and informed consent will be sought from participants. The participants are given adequate information about the research and project, re-use of data and all legal rights to the data concerning them, e.g., to review the data and stop participating in the data collection at any point in time. The GDPR-compliant record of processing activities sets out the processing of interview data over its lifecycle.
- The data are anonymized, classified, and quantified before being shared among partners. All personal data (emails, addresses etc.) are stored to the institution's local network drives with restricted access.
- The recordings are not published because speech is a direct identifier and cannot be reliably anonymised.
- If it is possible to anonymise the interview data, it can be made available to other researchers on request, as long as an agreement is signed. If this is not possible, the data will be described in accordance with good scientific practice and care will be taken not to disclose personal data when reporting the results.

The Remote sensing data will be shared within the project's lifetime, but all original data rights remain at the respective partner institution which bought the images.

6. DATA SHARING, STANDARDS & METADATA

All data will be collected/generated using standardized, data specific methods and mixed-method approach combining qualitative and quantitative methods defined in the research plan. Each WP ensures that the data collection/generation for which it is responsible is carried out in compliance with standards and good scientific practice. The consistency and reliability of the data is also ensured by creating common instructions and practices for collection/generation and analysis under the leadership of the WP responsible for the data. Instructions and procedures have been outlined so far as follows:

- WP1: Based on the previous literature, IUCN Global Standard for Nature-Based Solutions and SET (social, ecological, technological) framework developed by project researchers Syke researchers will develop a guideline and a template for the analysis of documents related to the projects' aquaNbS cases.

- WP2: For the document analyses and expert interviews Syke researchers will develop common guidelines and protocol included with e.g., structured questions and a template to report the results. The interview data will be conducted by the experienced researchers from each partner organization.
- WP4: IGB has existing guidelines and protocols for handling isotope samples, as well as well-established quality control and data pre-processing approach.
- WP5: Sampling guides and strategies are provided by the researchers of University of Lisbon and Poznan: The common criteria for the selection of the sampling sites; Procedure and guidelines to carry out the manual sampling, storing and sending stable water isotope sample vials to IGB; A metadata template, originally based on RHS, to document the characteristics of the river habitat.

Data description and metadata generation will be started at the time of data production/collection to ensure the (re)usability of data. The proper documentation and metadata will ensure the smooth repository/archive storage of data and the findability of published data. Data(sets) will be documented with descriptive metadata as required, e.g., each data set will include metadata about data ownership, creators and access rights. Data files are also named using the necessary information to distinguish between the files when necessary. The extra documentation, e.g., read-me-files in case of isotope data, will also be created about the data and processing of data whenever it is necessary to ensure the usability and understandability of the data. The usability and interoperability of data intended to be openly available or to be shared with third parties in another way is guaranteed by using commonly used or non-proprietary formats. It is also essential from the point of view of sharing, findability and understandability of data that the repositories/archives where we will publish open data support the provision of comprehensive metadata and documentation.

The instructions to collect/generate data (samples) on standard bases, metadata production, documentation and data opening will be provided as necessary and will be available on the Syke Teams.

7. DATA PUBLISHING, LICENSING & ARCHIVING

Repositories/archives are used as data publications channels. Publication mainly takes place at the same time as the results are published in the scientific or data journals.

We will follow publishing practices that support immediate and permanent open access to our research data whenever possible. Data we cannot open in repositories/archives is kept for at least 5 years after the project to ensure the exploitation of the data and the verification of the results.

Data that can be made openly accessible is published in repositories which ensure findability, understandability, and long-term availability of data. Thus, we prefer systems where data can be published under the CC-BY or more permissive licence and assigned with a persistent identifier such as DOI number. In these repositories/archives data can be documented with sufficient metadata and separate descriptive documents, such as read-me-files, can be attached to the data. The services we use for publishing are also well-known and run by established organisations. The following repositories/archives where the open data is published have been defined so far:

- Isotope data on water samples will be published in the IGB [Freshwater Research and Environmental Database](#) (FRED). It is a special repository for environmental data from observations of lakes, rivers, peatlands and other freshwater habitats. FRED is

maintained by the Deutsche Forschungsgemeinschaft and the Leibniz-Institute for Freshwater Ecology and Inland Fisheries. The database metadata standard schema is based on DCC and the metadata specification language is EML (Ecological Metadata Language). FRED provides DOI minting service.

- Sequencing data on water samples will be made publicly available on the [Sequence Read Archive](#) (SRA) which is the largest publicly available repository of high-throughput sequencing data and particularly for metagenomic and environmental surveys. SRA is maintained by the National Institutes of Health, United States National Library of Medicine, National Center for Biotechnology Information. The database metadata standard schema is based on DCC. The metadata specification language is unique developed for the purpose of SRA. The associated environmental data will be deposited to FRED.
- Taxonomic data of aquatic and surrounding vegetation will be published in the Finnish Biodiversity Information Facility (FinBIF, <https://laji.fi/en>) or in a similar database.

The results of the project will be published mainly in the scientific publications. The research plan indicates which scientific papers are planned to be published in the WPs and which data the results will mainly relate to. The scientific articles are preferably published in open access journals. We only use hybrid journals as a publication channel if a suitable open journal is not available. Articles that cannot be published open access are self-archived as (final) accepted manuscripts. The long-term accessibility of articles is granted by storing them in the institutional publication archives provided by partner organisations, e.g., [Syke's Helda archive](#). We also publish scientific papers in other series than in scientific journals. In this case, we will have to comply with the terms and conditions set by the publisher for open access/archiving. Other publications, such as the Factsheets on biodiversity values of aquaNBS and SET factors influencing on biodiversity and ESs in aquaNBS, will be published in archives/repositories, where they are given a persistent identifier. The blog writings for general public will be published on our own project webpage (www.bringingnatureback.com).

8. COSTS

The day-to-day management practices are mainly integrated into research and mostly the facilities are free of charge or provided by the host organisations and thus special costs are difficult to define. However, we also have allocated a budget to cover the costs to make our research available to the scientific and planners' public.