# Plan Overview

*A Data Management Plan created using DMPTuuli*

**Title:** Biodiversity loss in freshwaters under changing environment - lessons from island diatoms

**Creator:**Janne Soininen

**Principal Investigator:** Janne Soininen

**Data Manager:** Janne Soininen

**Affiliation:** University of Helsinki

**Funder:** Academy of Finland

**Template:** Academy of Finland data management plan guidelines

**Project abstract:**

One of the most stunning features of Earth is how biodiversity varies in space and time. Quantification of biodiversity at multiple scales is extremely timely as biodiversity crisis calls for studies about how humans affect ecosystems, and about how resilient biodiversity is to land use changes, for example. Biodiversity scaling with area is, however, still poorly understood especially for functional diversity and for small unicellular taxa (e.g. diatoms) in freshwaters even if these taxa maintain critical ecosystem functions and services. Such small taxa also offer great possibilities to conduct field experiments in islands, not feasible for many larger taxa.

Our aim is to examine freshwater diatom biodiversity patterns and drivers in islands and continents and explain and predict biodiversity variation using key ecosystem characteristics including isolation, land use change and climate. We use observational regional and global data sets and conduct field experiments at smaller spatial scales in the Baltic Sea islands. Specific research questions are: 1) Do diatom functional and taxonomic biodiversity patterns differ between islands and continents at regional and at global scales. 2) What is the mechanism behind diatom biodiversity differences between islands. 3) Do islands and continent differ in the balance between deterministic (environmental selection) and stochastic (dispersal processes) factors driving communities and do such processes lead to differences in community composition and functional trait structure between islands and continent. Finally, 4) we forecast how regional diatom diversity decreases with decreasing area and increasing land use intensity and assess how biodiversity loss affects critical ecosystem functions. Our research goes well beyond earlier biodiversity studies by combining highly standardized field experiments with observations and investigating functional aspects of island biogeography.

The questions we ask are fundamental - what controls diatom biodiversity in communities at different spatial scales and how communities are assembled in islands and continent. Perhaps most importantly, we will answer how ecosystem area and environmental features drive functional and taxonomic biodiversity and how to predict changes in biodiversity. Such findings are essential in the modern era at which ecosystem fragmentation and homogenization stress biota to an unprecedented degree.

**ID:** 19236

**Start date:** 01-09-2022

**End date:** 30-08-2026

**Last modified:** 09-06-2022

**Grant number / URL:** 346812

**Copyright information:**

# Biodiversity loss in freshwaters under changing environment - lessons from island diatoms

## 1. General description of data

**1.1 What kinds of data is your research based on? What data will be collected, produced or reused? What file formats will the data be in? Additionally, give a rough estimate of the size of the data produced/collected.**

Our project will assemble (i) quantitative data of diatoms (species occurrences and abundances) and trait data, (ii) water chemistry, climate, and land use data and (ii) quantitative data of ecosystem functioning and services in the study areas. Data will be collected in approximately 6000 study sites globally including five continents and 18 islands.

- We collect following data in the field and laboratory ourselves: Diatom data comprise species data (number of individuals per sample and species occurrence per sample) and trait data which include morphological traits such as cell size and cell dimensions and shape. Water chemistry data will include results of water chemistry analyses per each sample comprising variables such as total P, conductivity, water temperature and water pH. Ecosystem function data comprise measurement of productivity such as biomass of primary producers, here, benthic algae.
- Following data are collected from open sources: Land use data comprise relative proportions of different land use types (e.g. forest cover, agricultural land, wetland, urban areas) for each sampling site drawn from CORINE database (https://land.copernicus.eu/pan-european/corine-land-cover) and climate data (e.g. mean annual temperature and mean annual precipitation) are drawn from Worldclim database (www.worldclim.org).

Data will be in Excel format and will be 10-50 MB. The written R codes will be published in GitHub repository.

**1.2 How will the consistency and quality of data be controlled?**

Data management person (s) PI Janne Soininen and post-doctoral researchers will oversee the data consistency (e.g. units) and quality of the data. We will use SI units for water chemistry variables and use the newest taxonomic names for diatom species based on AlgaeBase database. Data will be continuously screened for possible outliers/errors.

## 2. Ethical and legal compliance

**2.1 What legal issues are related to your data management? (For example, GDPR and other legislation affecting data processing.)**

No legal issues are expected. We do not except any data security, privacy and intellectual property restrictions. We do not collect data about humans, animals or e.g. protected plants.

**2.2 How will you manage the rights of the data you use, produce and share?**

As agreed with project partners and collaborators, we will not place any restrictions for the data usage and data which we collect and share within the project will be freely downloadable. The researchers transfer parallel rights to all the results achieved during the research to the University of Helsinki. Research data produced in this project will be open by CCO license and R codes MIT license.

## 3. Documentation and metadata

**3. How will you document your data in order to make the data findable, accessible, interoperable and re-usable for you and others? What kind of metadata standards, README files or other documentation will you use to help others to understand and use your data?**

Before the data stored in IDA can be opened and used further, descriptive information of the data (that is, metadata such as title, descriptive details and other basic data information, keywords, field of science, sampling dates, spatial and temporal coverage, corresponding persons, and appropriate references) is linked with the data using the metadata tool Qvain (qvain.fairdata.fi). We use EML (Ecological Metadata Language) metadata standard, which is suitable for ecological data. We use following README file template to better understand and use of the data https://cornell.app.box.com/v/ReadmeTemplate

## 4. Storage and backup during the research project

**4.1 Where will your data be stored, and how will the data be backed up?**

During the project, field and laboratory data will be stored in university network disc drives and cloud services such as Microsoft OneDrive at least monthly. Data will also be stored in Figshare repository. For published papers, data for each paper will be stored in Dryad repository.

**4.2 Who will be responsible for controlling access to your data, and how will secured access be controlled?**

PI Janne Soininen and post-doctoral researchers will be responsible for controlling access to data. During the project, all relevant data are shared among the project partners according to written and signed agreements in the beginning of the project and stored in cloud services such as OneDrive. The hired post-doctoral researchers will be mainly responsible for data management actions in practice.

## 5. Opening, publishing and archiving the data after the research project

**5.1 What part of the data can be made openly available or published? Where and when will the data, or its metadata, be made available?**

Prior to the end of the project, data and any other digital outputs will be stored into IDA storage service for a long-term use. The research datasets published with Qvain and IDA get persistent identifiers and landing pages in Etsin search tool (*etsin.fairdata.fi*). Data will also be stored in Figshare and Zenodo repositories and data for each research article in Dryad repository.

**5.2 Where will data with long-term value be archived, and for how long?**

Prior to the end of the project, data and any other digital outputs will be stored into IDA storage service for a long-term use for at least ten years. The research datasets published with Qvain and IDA get persistent identifiers and landing pages in Etsin search tool (etsin.fairdata.fi). This makes the dataset findable for others and enables re-use of the openly downloadable data and creating a

scientific reference and relevant back-ups. Data will be stored also in Figshare repository.

## 6. Data management responsibilities and resources

**6.1 Who (for example role, position, and institution) will be responsible for data management?**

PI Janne Soininen and hired post-doctoral researchers in University of Helsinki will be responsible for data management.

**6.2 What resources will be required for your data management procedures to ensure that the data can be opened and preserved according to FAIR principles (Findable, Accessible, Interoperable, Re-usable)?**

Data management procedures require some time investment. No special financial resources are needed for the data management procedures except for Dryad repository, which costs ca. 120 € per published article.