PlutoF

eDNA data management and mobilisation

User tutorial

Document information

Date and version no.	Author(s)	Comments/Changes
03.02.2026, v0.1	Kessy Abarenkov	Initial draft.
17.02.2026, v0.2	Kessy Abarenkov, Sten Anslan	Updates regarding sequence data processing and subsequent analysis.
30.03.2026, v1.0	Kessy Abarenkov	First version published on the PlutoF Manuals page.

Table of Contents

Introduction

PlutoF is a web-based research data management platform designed to support a wide range of biodiversity data types and projects across multiple disciplines, including ecology, taxonomy, environmental monitoring, nature conservation, agriculture, and genetics. It provides modules for managing taxon occurrences (e.g., specimens in scientific collections, observations by citizen scientists, material samples, and DNA sequences from barcoding and metabarcoding studies), taxonomic classifications, measurements and traits, molecular experiments, and more. It allows users to manage datasets throughout the entire data lifecycle — from project design and data collection to publishing and archiving — in compliance with FAIR principles.

PlutoF adheres to key data standards in the field, such as TDWG's Darwin Core (DwC)[1], the Genomic Standard Consortium's Minimum Information about any (x) Sequence (GSC MIxS)[2] environmental packages, the Global Genome Biodiversity Network (GGBN) Data Standard[3], and standard compilations like the DNA derived data extension to DwC[4], maintained by the Global Biodiversity Information Facility (GBIF)[5]. This ensures seamless data exchange with global repositories and data portals like GBIF, European Nucleotide Archive (ENA)[6], and DataCite[7].

The current user manual describes the PlutoF workflow for eDNA data management and mobilisation, including project planning, eDNA sample collection, metadata capture, quality control, and the publication of data to ENA and GBIF.

[1] https://dwc.tdwg.org
[2] https://www.gensc.org/pages/projects/mixs-gsc-project.html
[3] https://wiki.ggbn.org/ggbn/GGBN_Data_Standard_v1
[4] https://rs.gbif.org/terms/1.0/DNADerivedData
[5] https://www.gbif.org
[6] https://www.ebi.ac.uk/ena
[7] https://datacite.org/

Terms

ASV – Amplicon Sequence Variant

DOI – Digital Object Identifier

DwCA – Darwin Core Archive

ENA – European Nucleotide Archive

GBIF – Global Biodiversity Information Facility

HTS – High-Throughput Sequencing

INSDC – International Nucleotide Sequence Database Collaboration

LIMS – Laboratory Information Management System

OTU – Operational Taxonomic Unit

SH – Species Hypothesis

1. Creating a PlutoF account

Register as a PlutoF user at https://app.plutof.ut.ee/register. After completing the registration, a confirmation email will be sent to your mailbox. To log in to PlutoF, you must **click the confirmation link** and accept the PlutoF Terms and Conditions.

Note: If you do not receive the confirmation email, please check your Spam folder.

The same PlutoF account should be used to log in to the PlutoF GO app.

Password recovery: If you forget your password, you can recover it at https://app.plutof.ut.ee/recover-password.

2. Creating a project

> Projects in PlutoF allow users to group and organise data for efficient bulk management, set up pre-defined sampling areas, manage user access rights and roles within a project, and prepare data for publishing. Projects can be public, private, or shared with collaborators.

A new project can be created via the **Main menu** → **Projects** → **New** (Figure 1).

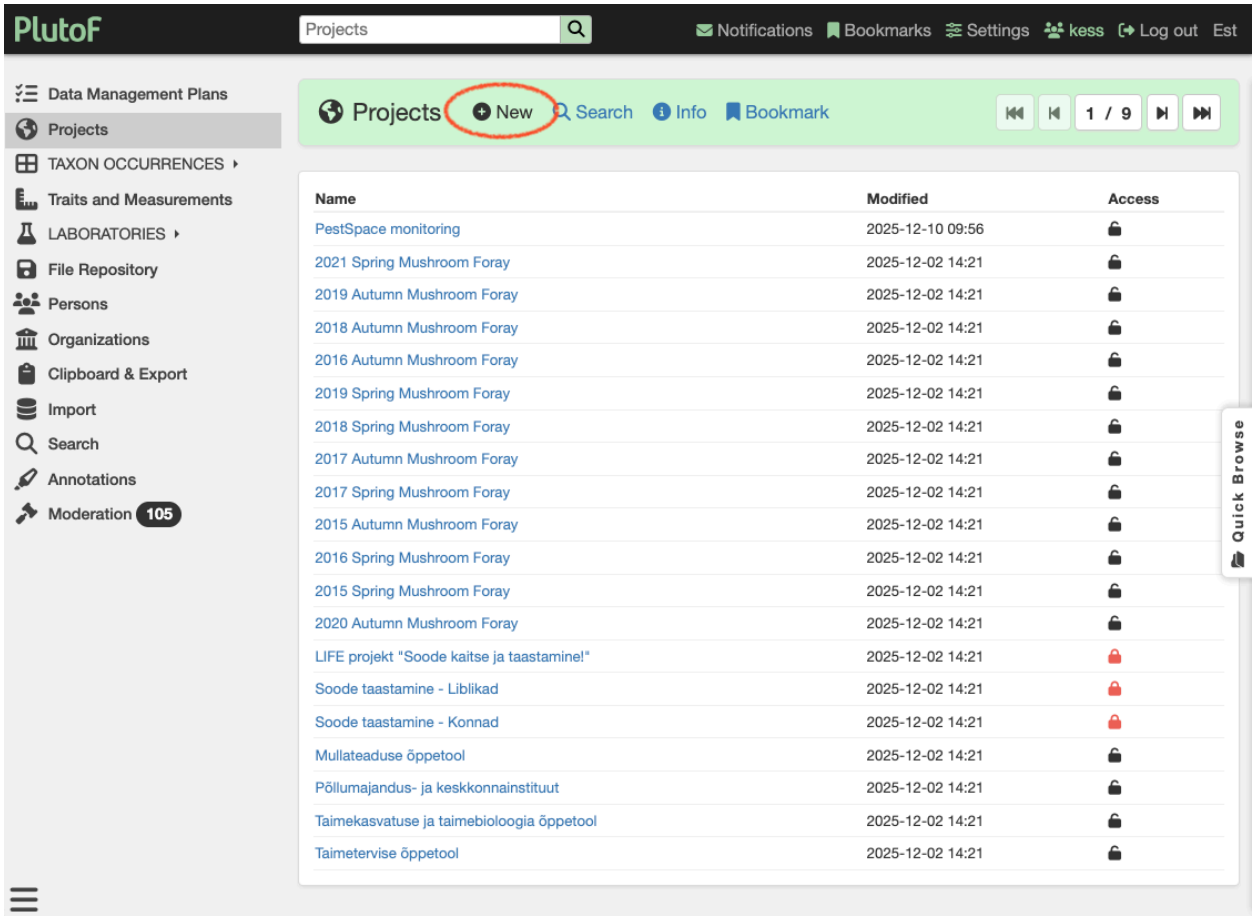


Figure 1. Creating a new project.

3. Assigning users and access rights

Each project has a managing group (Figure 2) in which users can be assigned one of three roles: 1. **Regular users** – can add and edit their own data; 2. **Moderators** – can add data and edit all data within the project; 3. **Owners** – in addition to the above, can delete the project and its data and add new moderators.

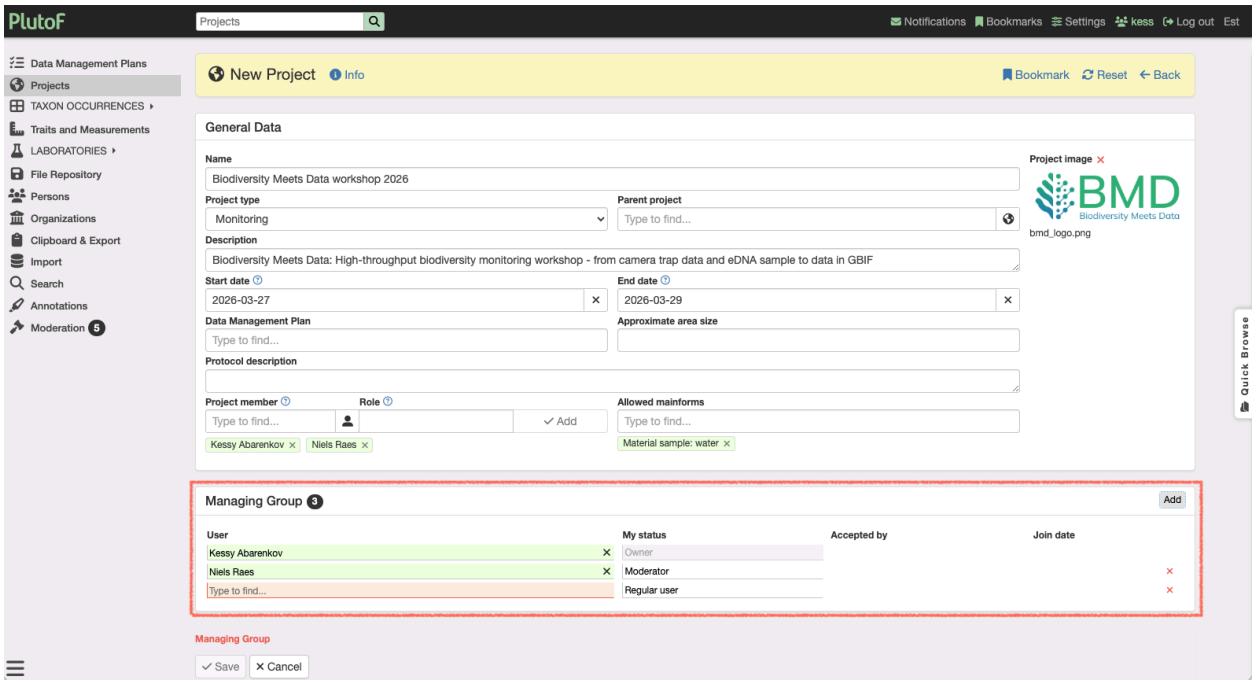


Figure 2. Adding users to the project.

4. Assigning QR identifier blocks

Sample identifiers with a user-defined format can be set up, and specific identifier blocks can be reserved for project members. Identifiers with QR codes can be downloaded, printed out, and used for registering samples in the field with the PlutoF GO app[8].

Sample identifiers can be created within the **Occurrence identifiers** panel in the **Project Add/Edit** form. First, the identifier format must be defined (Figure 3). After that, a block of identifiers matching this format can be reserved for a specific user. Alternatively, the block can be made available for all project members by selecting the corresponding checkbox (Figure 4).

[8] https://plutof.ut.ee/go

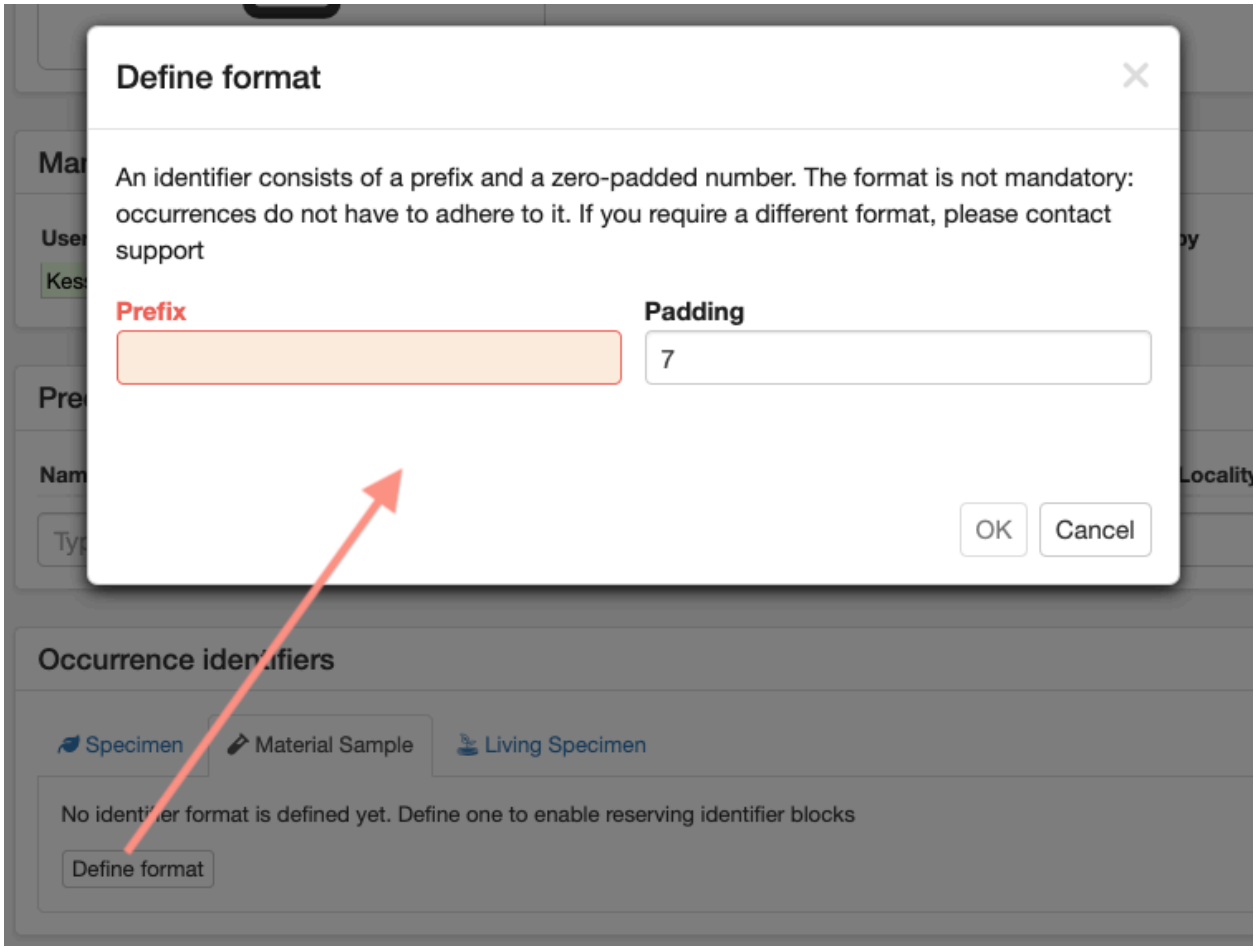


Figure 3. Defining sample identifier format.

If the block is assigned to a specific user, only that user can upload records using identifiers from this block. If the block is made available for all users, any user with write access to the project can upload records using these identifiers.

Note: Uniqueness of the codes is enforced at the project level; duplicate entries with the same code are not allowed.

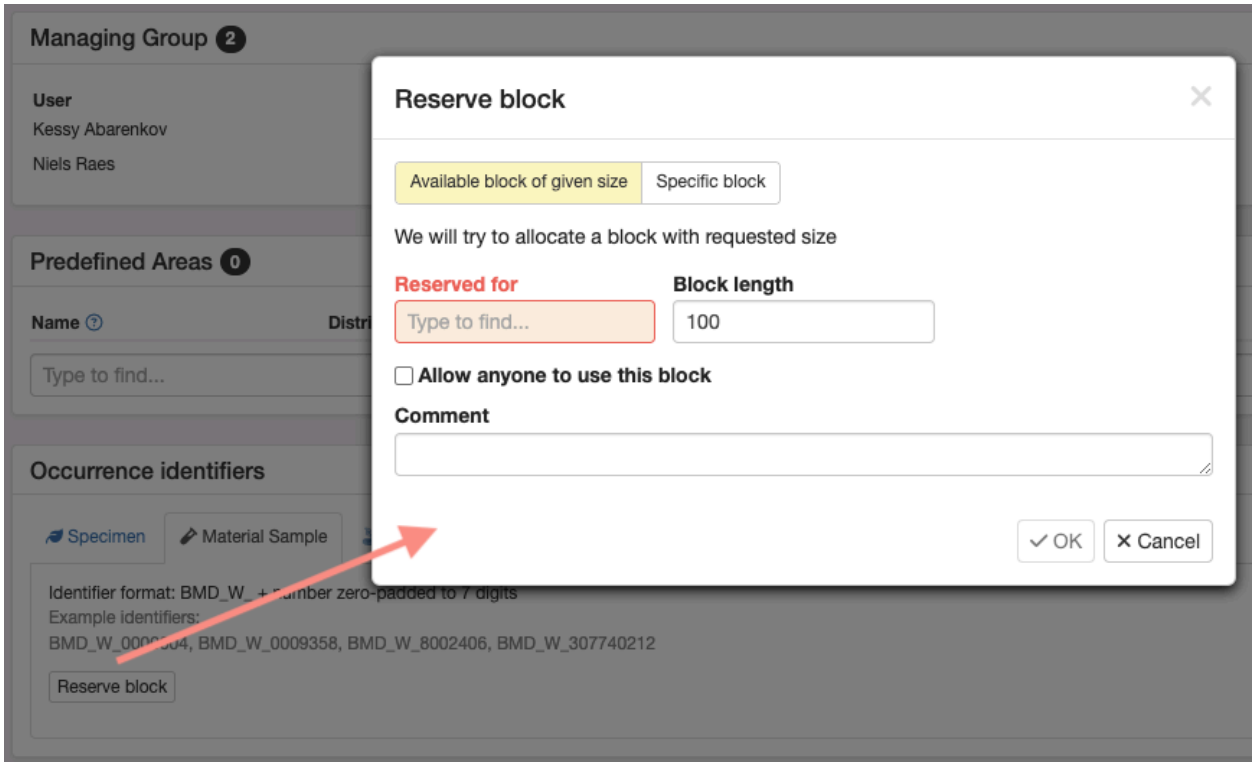


Figure 4. Creating a new reserved identifier block.

QR codes can be printed in a variety of formats by clicking on the **Print labels** button in the **Reserved identifier blocks** panel (Figure 5).

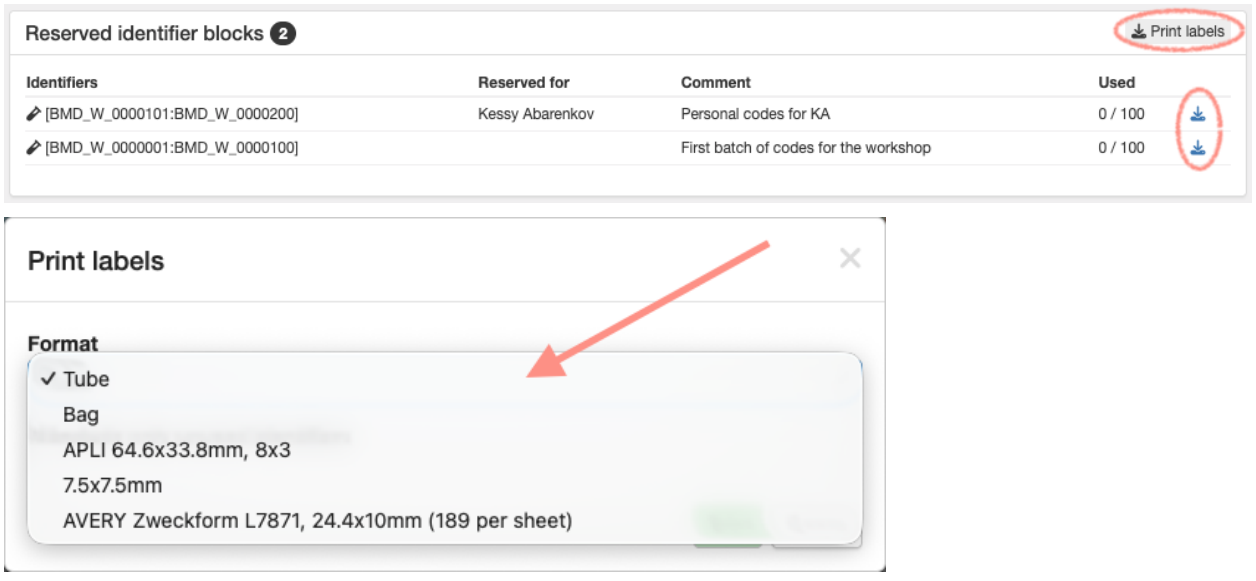


Figure 5. Downloading QR code labels.

A common use case is printing the QR codes onto sticker labels, which can then be attached to sample bags, tubes, envelopes, and similar containers (Figure 6). These can then be easily scanned using **PlutoF GO** when entering sampling metadata in the field.

Figure 6. QR code sticker on a sample bag, ready for scanning with the **PlutoF GO** app. *Photo credit: Inga Hiiesalu-Vahter.*

5. Registering samples in the field using the PlutoF GO app

> The **PlutoF GO** app is a tool for gathering biodiversity data. It supports offline data collection with built-in geo-positioning and datetime pre-filling, as well as scanning sample QR codes from printed labels. Once the data is saved in the app, it can be uploaded to the PlutoF workbench. This can be done immediately or later, in case the data needs amending or an internet connection is not readily available.

To create a new sample record in the **PlutoF GO** app, click the **Material sample** button and select the appropriate input form (Figure 7).

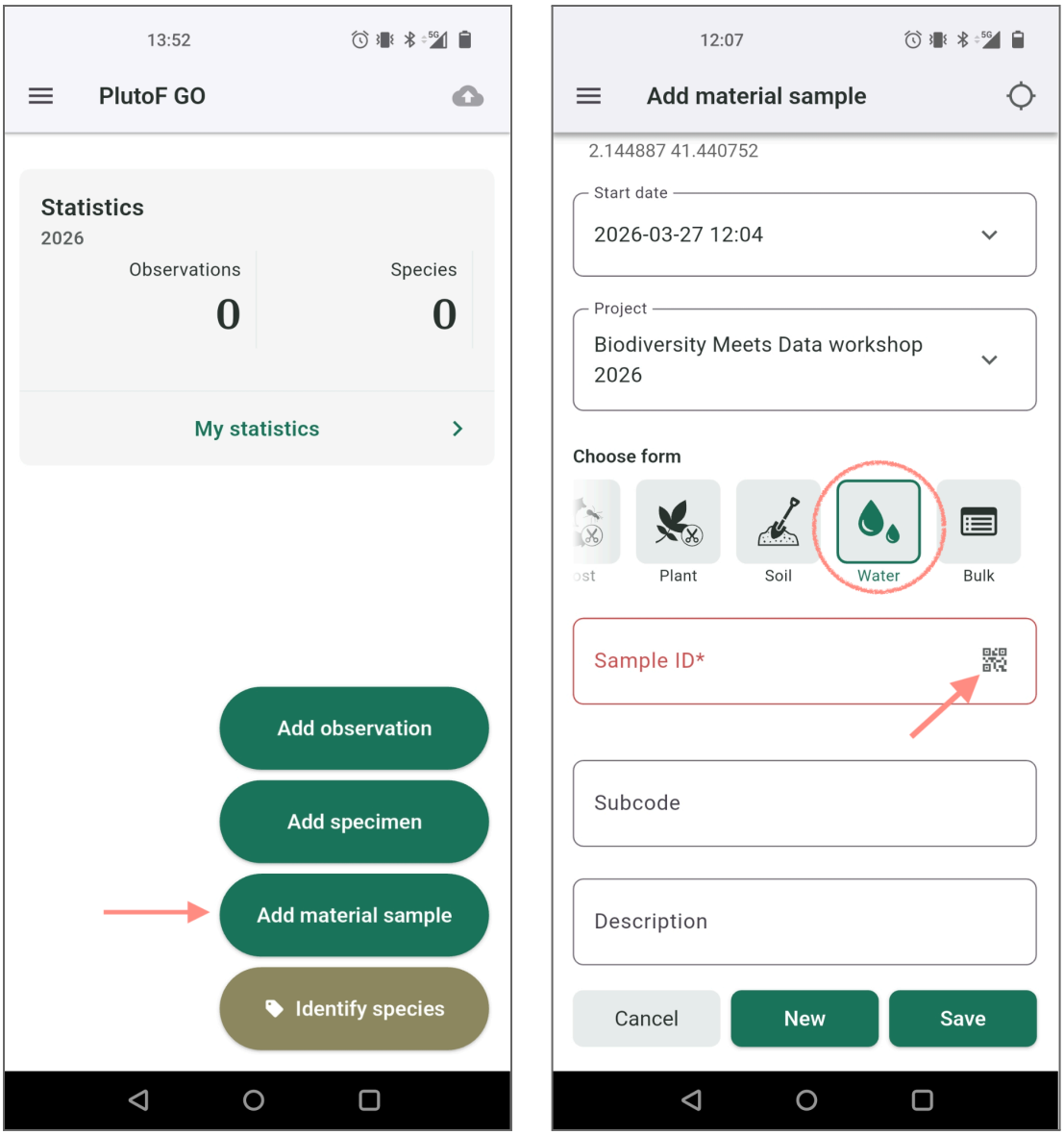


Figure 7. Creating a new sample record in the **PlutoF GO** app. Sample identifiers can be scanned by clicking the QR icon in the **Sample ID** field.

Note: Before adding new sample records, a project must be selected. This can be done using the autocomplete field when entering individual samples or by setting a **Default project** in the app **Settings**. All mandatory fields in the input form are indicated in red.

Records can be uploaded to the **PlutoF workbench** by clicking on the **Upload** icon ( ) at the top of the app header.

6. Managing sample data on the PlutoF workbench

Once the data has been uploaded from the **PlutoF GO** app or imported via CSV files, it can be validated, organised, and further edited on the **PlutoF workbench**. Users can update individual records or use the bulk update functionality in the Clipboard to modify multiple records simultaneously.

All data linked to a project are available in the project view (Figure 8), which includes shortcuts for searching associated material samples and for sending records to the **Clipboard**. In the **Clipboard**, records can be visualised on a map, exported in various formats, modified in bulk, and sent to different types of analyses.

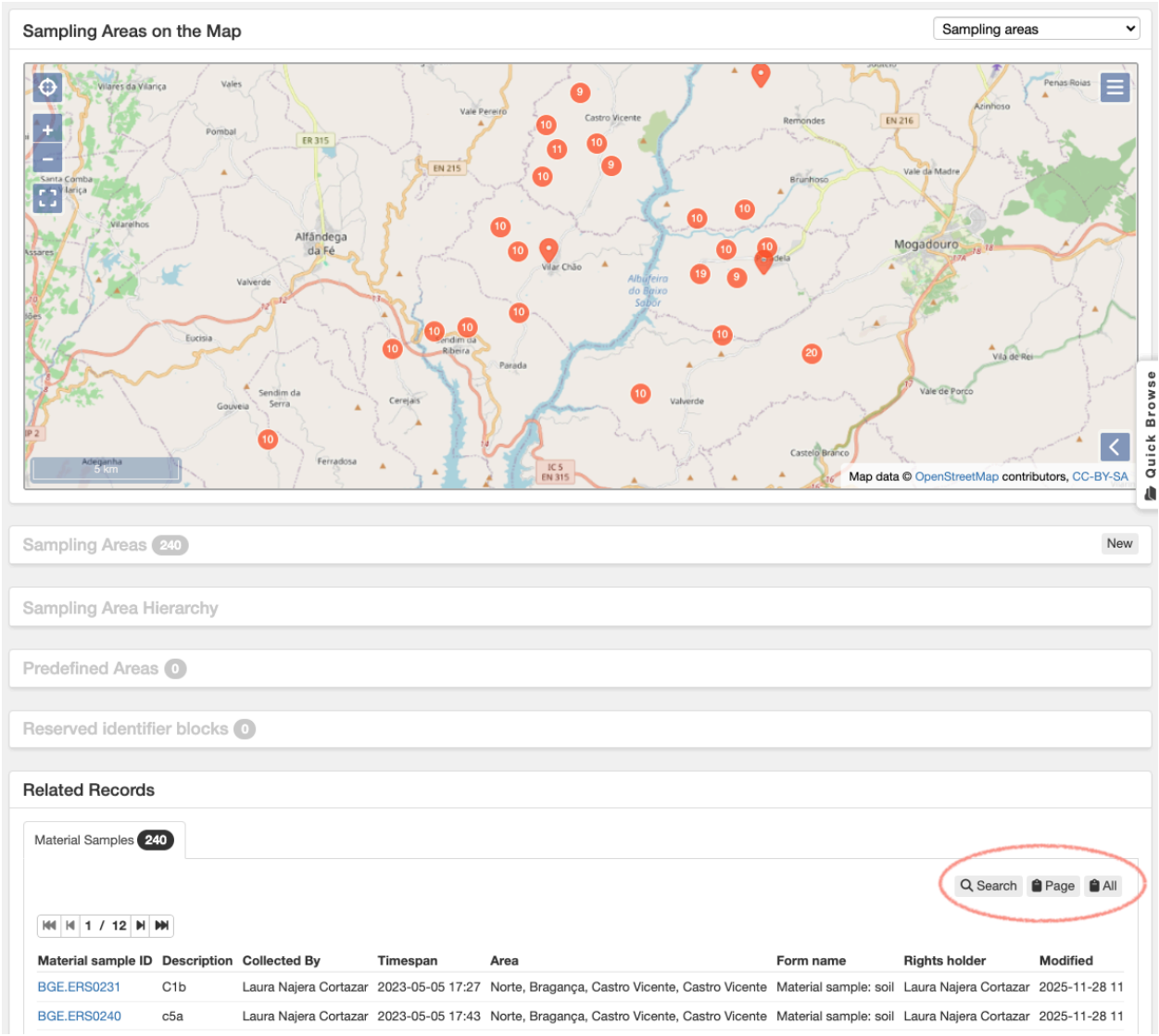


Figure 8. Project view with links to project-related search and Clipboard functionalities.

7. Uploading sample records to ENA/Biosamples

> Using the Publishing module in PlutoF, users can submit sample records to the ENA and BioSamples[9] databases. The PlutoF platform acts as a broker for ENA, utilising its programmatic Webin submission service for sample data submission. The resulting ENA identifiers are stored in PlutoF alongside the original sample records.

Publishing material sample data from PlutoF to ENA and BioSamples is project-based, meaning all samples within a user-selected study will be published as a single submission. After publication, material samples will receive ENA and BioSamples

[9] https://www.ebi.ac.uk/biosamples/

identifiers. It is also possible to update sample data in ENA by re-publishing the dataset at later stages.

To publish your dataset in **ENA**, go to **Main menu** → **Laboratories** → **Publishing Lab** → **ENA Datasets** → **New** (https://app.plutof.ut.ee/publishing/ena/add, Figure 9).

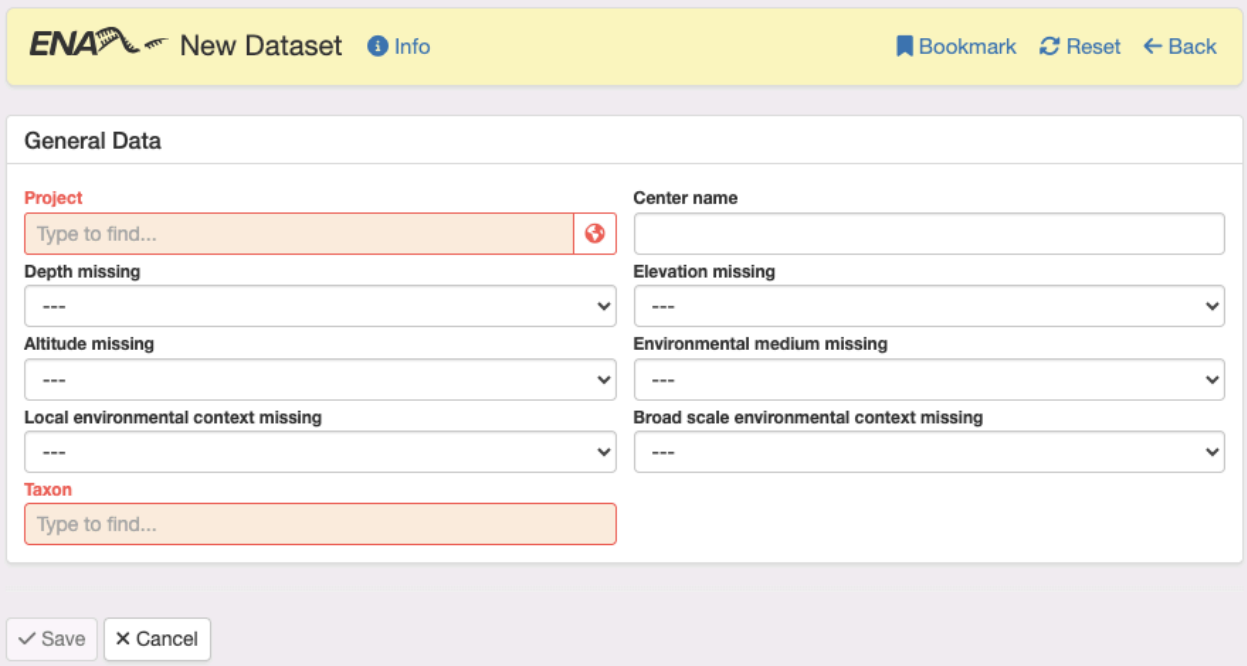


Figure 9: Creating a new dataset for publishing to ENA.

Steps to publish:

1. Select the project name from the autocomplete list (project moderator rights are required).
2. Fill in any missing mandatory field values as required by ENA.
3. Save the dataset.
4. After administrator approval, publish the dataset to ENA.

Note: Samples uploaded to ENA/BioSamples are treated as independent samples (i.e., they are not linked to a BioProject). They will be linked to a BioProject when the raw sequence data are associated with the samples (step 8 below).

8. Managing molecular lab experiments

The Molecular Laboratory in PlutoF acts as a Laboratory Information Management System (LIMS), allowing users to store and manage information about DNA extraction, PCR, and Library prep. & DNA sequencing experiments carried out on the material samples registered in PlutoF during step 4 of this workflow.

Note: Specific experiments available in this module are not covered in this tutorial.

9. Uploading raw sequence data

> Using the ENA Submission Portal[10], users are required to create a project and upload raw sequence data directly to ENA. During the submission process, sample records submitted to ENA in step 6 of this workflow can be used and linked to the raw sequence files.

Note: Uploading raw sequence data to ENA is not covered in this tutorial. However, a detailed tutorial was prepared as part of the Bioscan Europe project (Uploading raw sequence data to ENA). Please also refer to the ENA tutorial (How to Submit Raw Reads) for a comprehensive overview of the process.

10. Sequence data processing and analysis

> Process raw sequence data to obtain Amplicon Sequence Variants (ASVs) or Operational Taxonomic Units (OTUs) by using a sequence analysis pipeline selected by the user. Sequence data processing and analysis to generate ASVs or OTU representative sequences are performed outside the PlutoF platform.

Note: Processing and analysing raw sequence data is not covered in this tutorial. Please refer to the following (non-exhaustive) list of resources for tutorials on processing and analysing raw sequence data:

- **Pipecraft2**[11] – a cross-platform graphical user interface (GUI) tool that implements various popular tools for metabarcoding data analyses. It implements ready-to-run (pre-defined) pipelines as well as an option to run a variety of individual steps outside of a full pipeline. For the most commonly applied DNA markers in metabarcoding studies, dedicated raw sequence data processing and analysis workflows have been developed as part of the Biodiversity Genomics Europe[12] project and are accessible at https://bioscanflow.readthedocs.io. These bioinformatic workflows have also been implemented in PipeCraft2 (https://pipecraft2-manual.readthedocs.io).
- **QIIME 2**[13] – a popular suite of plugins and hundreds of bioinformatics methods, allowing its users to perform extensive microbiome analysis with interactive visualisations and robust statistical tools.
- **DADA2**[14] – a powerful and fast bioinformatics software package that infers exact ASVs from raw amplicon sequencing data.

---

[10] https://www.ebi.ac.uk/ena/submit/webin/login
[11] https://pipecraft2-manual.readthedocs.io/en/latest/
[12] https://biodiversitygenomics.eu
[13] https://qiime2.org
[14] https://benjjneb.github.io/dada2

11. Uploading OTU/ASV representative sequences to PlutoF

OTU representative sequences and ASVs can be uploaded to PlutoF using the PlutoF import module. A template file for the import can be created and downloaded via **Main menu → Import → Generate template** by selecting the Model (Sequence) and Form name (Sequence: HTS representative), as shown on Figure 10. A template file with the selected field labels can be downloaded by clicking the **Download file** button.

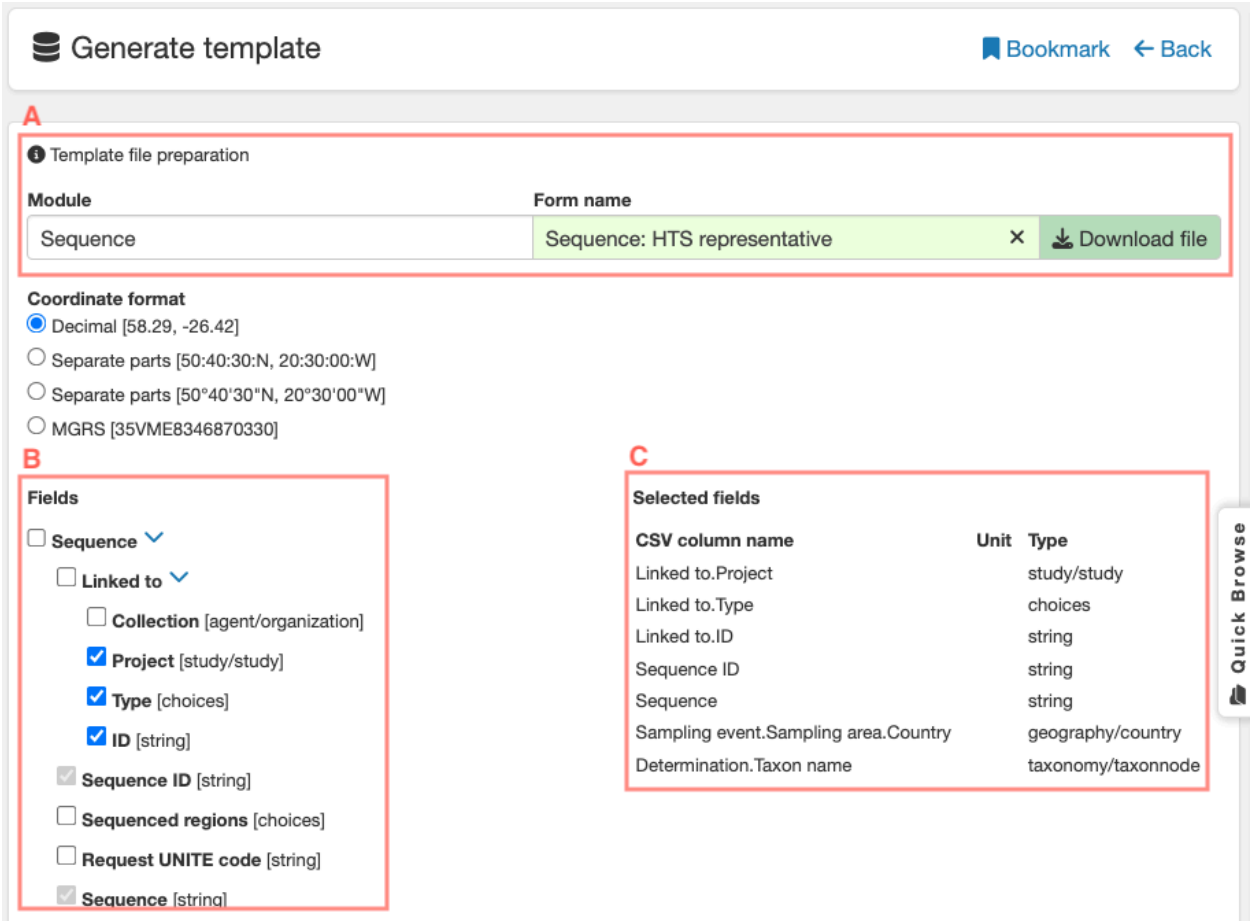


Figure 10. Generating a template file for representative sequence import. **A:** Select the module and form name. **B:** Select the data fields to be imported. **C:** List of data fields selected for import.

Once the template file has been filled with data, it can be imported via **Main menu → Import → New** (Figure 11).

Users must:

1. Upload the file,

2. Select the **module** name (*Sequence*),
3. **Form** name (*Sequence: HTS representative*),
4. **Project** name, and
5. Choose the appropriate import settings (e.g., check the box for “Use source record’s area and event if event columns are empty”).

After clicking the **Save and start** button, the interactive import process will begin and guide the user through the entire procedure. The user will be notified if any errors occur during the import.

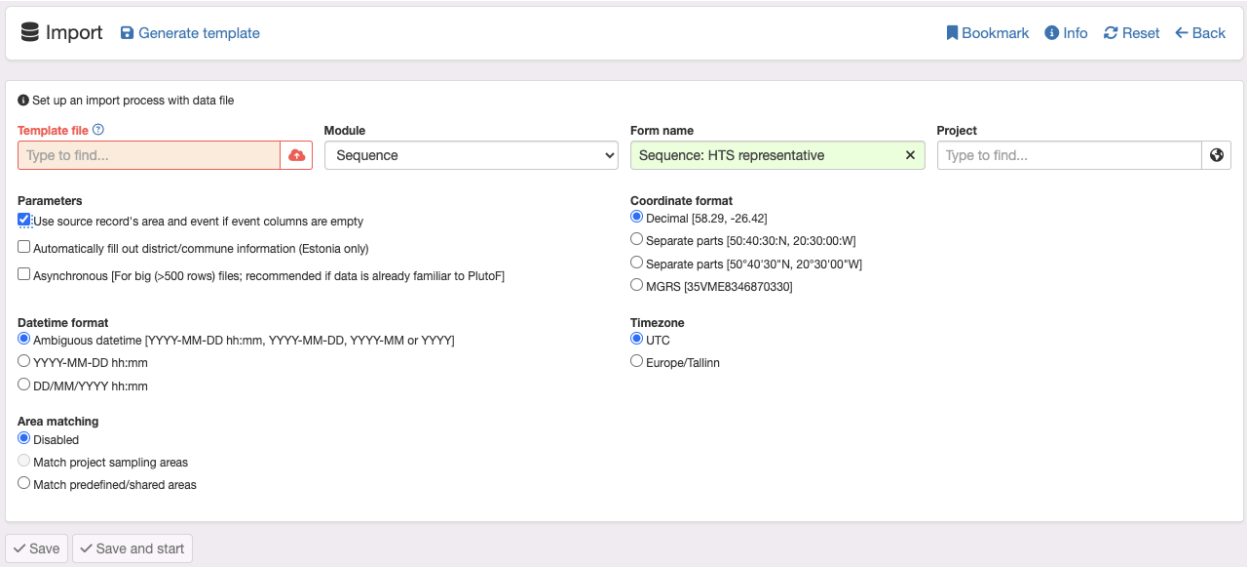


Figure 11. Starting the sequence import process.

12. Assigning UNITE SH identifiers

> rDNA ITS sequences can be assigned to UNITE SHs using the **SH MATCHING** analysis tool (Abarenkov et al., 2022), implemented in PlutoF, to link ASVs and OTU representative sequences obtained in step 9 to their corresponding SH identifiers. These SHs are mapped to the GBIF taxonomic backbone, enabling the publication of DNA-derived taxon occurrences in GBIF (Abarenkov et al., 2023).

PlutoF sequence records can be copied to the **Clipboard** either in the Project view (within the Related Records panel) or from the Search results by clicking the clipboard icon ( ).

SH Matching analysis for query sequences can be started from the sequence clipboard by clicking the “New SH matching (all)” link on the bottom of the page (Figure 12).

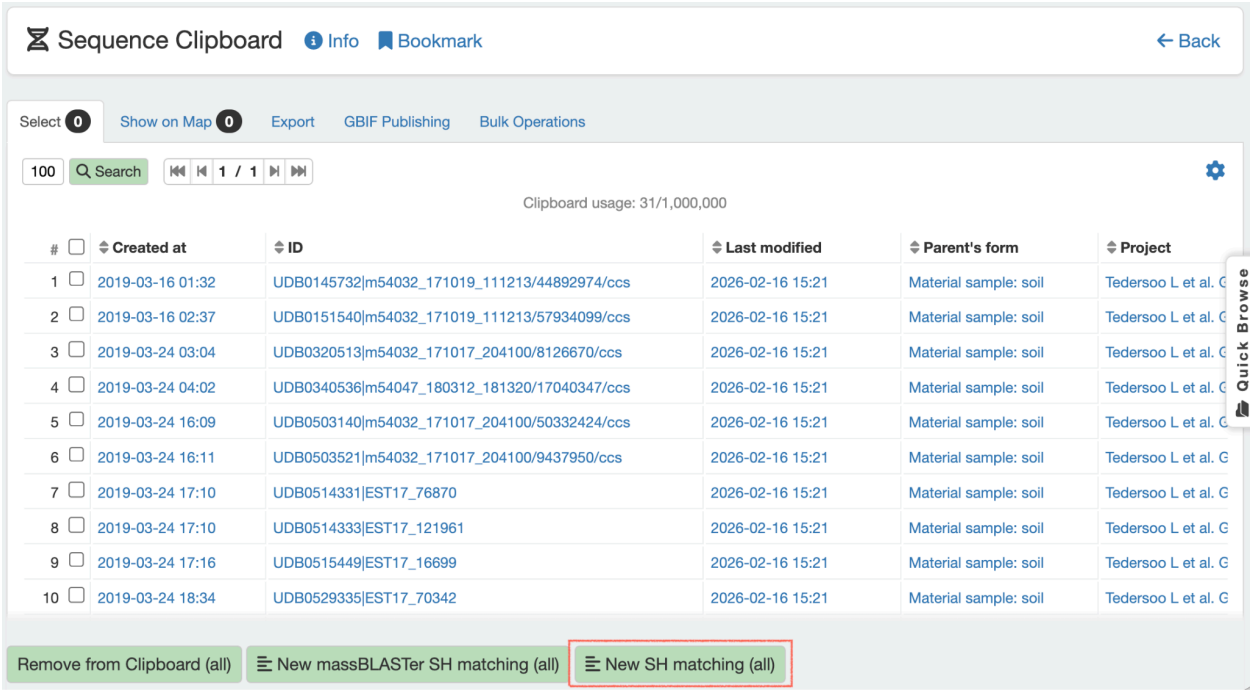


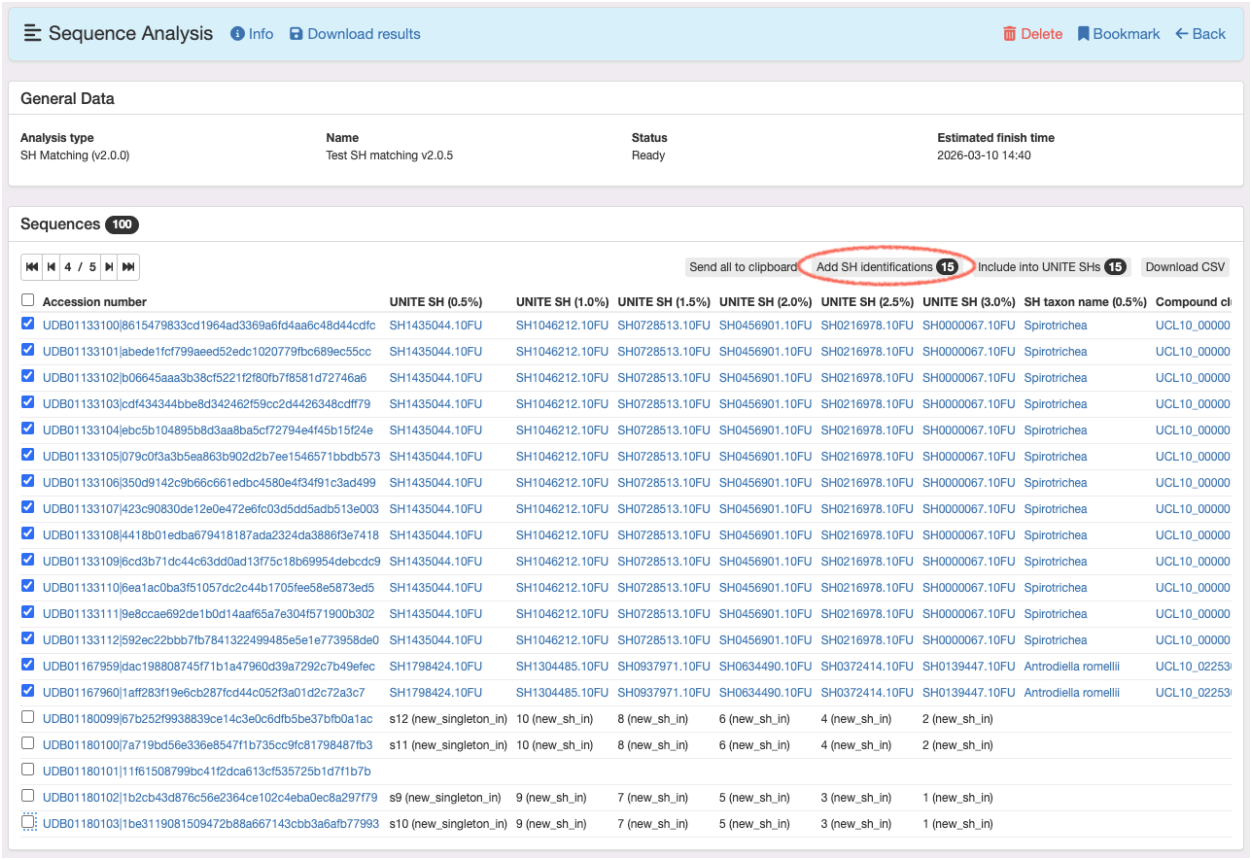
Figure 12. Starting the SH Matching analysis on the Sequence Clipboard.

Analysis results will be available under the **Main menu → Laboratories → Analysis Lab**. The SH Matching analysis provides a table listing the query sequences and their corresponding SH assignments (Figure 13).

For sequences with a strong match (e.g., an SH assigned at 0.5% and a similarity score above 99%), SH identifications can be added by selecting the desired query sequences and clicking the **Add SH identifications** button. These identifications can also be used when publishing your sequence data in GBIF as a DNA-derived taxon occurrence dataset.

Sequences that form new SHs can be incorporated into UNITE by selecting them and clicking **Include into UNITE SHs**. These sequences will be added to the UNITE SH system during the next regular update and will become available for matching thereafter.

Note: SH identifications can also be added from CSV files using the PlutoF import module.



Accession number	UNITE SH (0.5%)	UNITE SH (1.0%)	UNITE SH (1.5%)	UNITE SH (2.0%)	UNITE SH (2.5%)	UNITE SH (3.0%)	SH taxon name (0.5%)	Compound cl
UDB01133100|8615479833cd1964ad3369a6fd4aa6c48d44cdfc	SH1435044.10FU	SH1046212.10FU	SH0728513.10FU	SH0456901.10FU	SH0216978.10FU	SH0000067.10FU	Spirotrichea	UCL10_00000
UDB01133101|abede1fcf799aeed52edc1020779fbc689ec55cc	SH1435044.10FU	SH1046212.10FU	SH0728513.10FU	SH0456901.10FU	SH0216978.10FU	SH0000067.10FU	Spirotrichea	UCL10_00000
UDB01133102|b06645aaa3b38cf5221f2f80fb7f8581d72746a6	SH1435044.10FU	SH1046212.10FU	SH0728513.10FU	SH0456901.10FU	SH0216978.10FU	SH0000067.10FU	Spirotrichea	UCL10_00000
UDB01133103|cdf434344bbe8d342462f59cc2d4426348cdff79	SH1435044.10FU	SH1046212.10FU	SH0728513.10FU	SH0456901.10FU	SH0216978.10FU	SH0000067.10FU	Spirotrichea	UCL10_00000
UDB01133104|ebc5b104895b8d3aa8ba5cf72794e4f45b15f24e	SH1435044.10FU	SH1046212.10FU	SH0728513.10FU	SH0456901.10FU	SH0216978.10FU	SH0000067.10FU	Spirotrichea	UCL10_00000
UDB01133105|079c0f3a3b5ea863b902d2b7ee1546571bbdb573	SH1435044.10FU	SH1046212.10FU	SH0728513.10FU	SH0456901.10FU	SH0216978.10FU	SH0000067.10FU	Spirotrichea	UCL10_00000
UDB01133106|350d9142c9b66c661edbc4580e4f34f91c3ad499	SH1435044.10FU	SH1046212.10FU	SH0728513.10FU	SH0456901.10FU	SH0216978.10FU	SH0000067.10FU	Spirotrichea	UCL10_00000
UDB01133107|423c90830de12e0e472e6fc03d5dd5adb513e003	SH1435044.10FU	SH1046212.10FU	SH0728513.10FU	SH0456901.10FU	SH0216978.10FU	SH0000067.10FU	Spirotrichea	UCL10_00000
UDB01133108|4418b01edba679418187ada2324da3886f3e7418	SH1435044.10FU	SH1046212.10FU	SH0728513.10FU	SH0456901.10FU	SH0216978.10FU	SH0000067.10FU	Spirotrichea	UCL10_00000
UDB01133109|6cd3b71dc44c63dd0ad13f75c18b69954debcdc9	SH1435044.10FU	SH1046212.10FU	SH0728513.10FU	SH0456901.10FU	SH0216978.10FU	SH0000067.10FU	Spirotrichea	UCL10_00000
UDB01133110|6ea1ac0ba3f51057dc2c44b1705fee58e5873ed5	SH1435044.10FU	SH1046212.10FU	SH0728513.10FU	SH0456901.10FU	SH0216978.10FU	SH0000067.10FU	Spirotrichea	UCL10_00000
UDB01133111|9e8ccae692de1b0d14aaf65a7e304ff571900b302	SH1435044.10FU	SH1046212.10FU	SH0728513.10FU	SH0456901.10FU	SH0216978.10FU	SH0000067.10FU	Spirotrichea	UCL10_00000
UDB01133112|592ec22bbb7fb7841322499485e5e1e773958de0	SH1435044.10FU	SH1046212.10FU	SH0728513.10FU	SH0456901.10FU	SH0216978.10FU	SH0000067.10FU	Spirotrichea	UCL10_00000
UDB01167959|dac198808745f71b1a47960d39a7292c7b49efec	SH1798424.10FU	SH1304485.10FU	SH0937971.10FU	SH0634490.10FU	SH0372414.10FU	SH0139447.10FU	Antrodiella romellii	UCL10_02253
UDB01167960|1aff283f19e6cb287fcd44c052f3a01d2c72a3c7	SH1798424.10FU	SH1304485.10FU	SH0937971.10FU	SH0634490.10FU	SH0372414.10FU	SH0139447.10FU	Antrodiella romellii	UCL10_02253
UDB01180099|67b252f9938839ce14c3e0c6dfb5be37bfb0a1ac	s12 (new_singleton_in)	10 (new_sh_in)	8 (new_sh_in)	6 (new_sh_in)	4 (new_sh_in)	2 (new_sh_in)		
UDB01180100|7a719bd56e336e8547f1b735cc9fc81798487fb3	s11 (new_singleton_in)	10 (new_sh_in)	8 (new_sh_in)	6 (new_sh_in)	4 (new_sh_in)	2 (new_sh_in)		
UDB01180101|11f61508799bc41f2dca613cf535725b1d7f1b7b								
UDB01180102|1b2cb43d876c56e2364ce102c4eba0ec8a297f79	s9 (new_singleton_in)	9 (new_sh_in)	7 (new_sh_in)	5 (new_sh_in)	3 (new_sh_in)	1 (new_sh_in)		
UDB01180103|1be3119081509472b88a667143cbb3a6afb77993	s10 (new_singleton_in)	9 (new_sh_in)	7 (new_sh_in)	5 (new_sh_in)	3 (new_sh_in)	1 (new_sh_in)		

Figure 13. SH Matching analysis results.

13. Publishing project dataset in GBIF

> The full project dataset can be published in GBIF as a dataset of DNA-derived taxon occurrences, making eDNA-based biodiversity data publicly discoverable and reusable (the data becomes citable with a DOI).

GBIF — the Global Biodiversity Information Facility — is an international network and research infrastructure funded by the world's governments and aimed at providing anyone, anywhere, open access to data about all types of life on Earth.

"It is important to realize that a sequence with coordinates and a timestamp is a valuable biodiversity occurrence which is useful in a much broader context than its original purpose. To realize this potential, DNA-derived data needs to be discoverable through biodiversity data platforms" (Abarenkov et al., 2025).

Recommended guidelines for publishing DNA-derived data through biodiversity data platforms are presented in Abarenkov et al., 2025[15].

[15] https://doi.org/10.35035/doc-vf1a-nr22

A GBIF dataset can be created via **Main menu → Publishing Lab → GBIF Datasets → New.**

Steps in PlutoF for preparing the dataset for publication in GBIF include:

- Creating a new GBIF dataset and filling in the **general metadata** (Figure 14)
- Adding a **logo**, if available (a representative image of the project or organisation)
- **Project:** describing the project under which the data was collected
- **Contacts:** providing information about creators, metadata providers, and general contacts
- **Bibliography:** adding relevant citations from printed literature or web sources

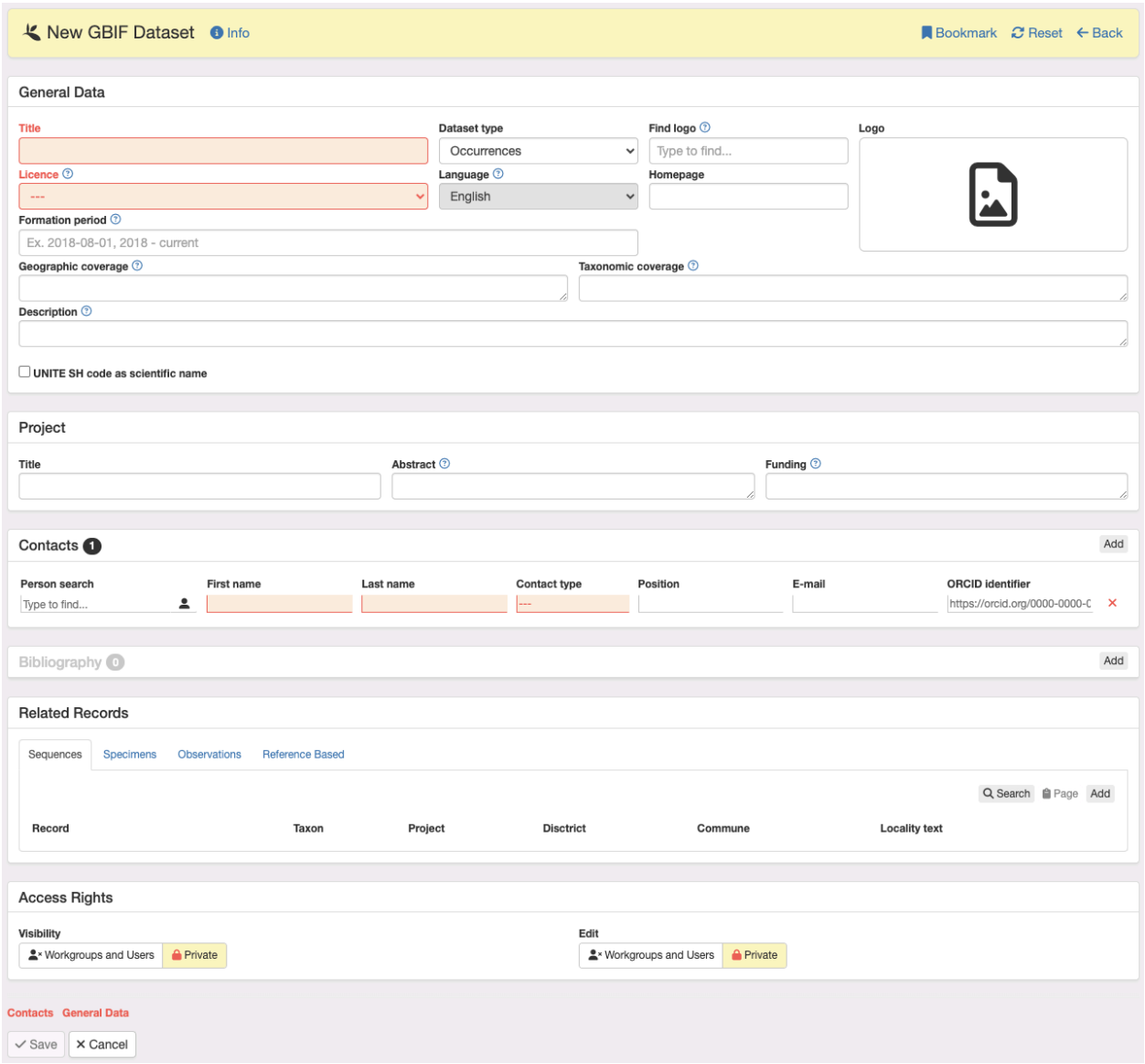


Figure 14. New GBIF dataset creation form with the mandatory metadata fields indicated in red.

Once the dataset is created, sequence records can be associated with it via the Sequence Clipboard. Project sequences can be copied to Sequence Clipboard by clicking the Clipboard icon in the project view under the **Related Records: Sequences** panel (Figure 15).

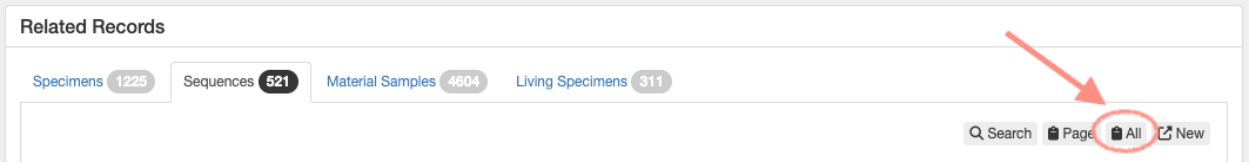


Figure 15. Copying project sequences to Sequence Clipboard.

Sequences can be linked to the GBIF dataset in the GBIF Publishing tab by selecting the dataset name in the autocomplete field (Figure 16).

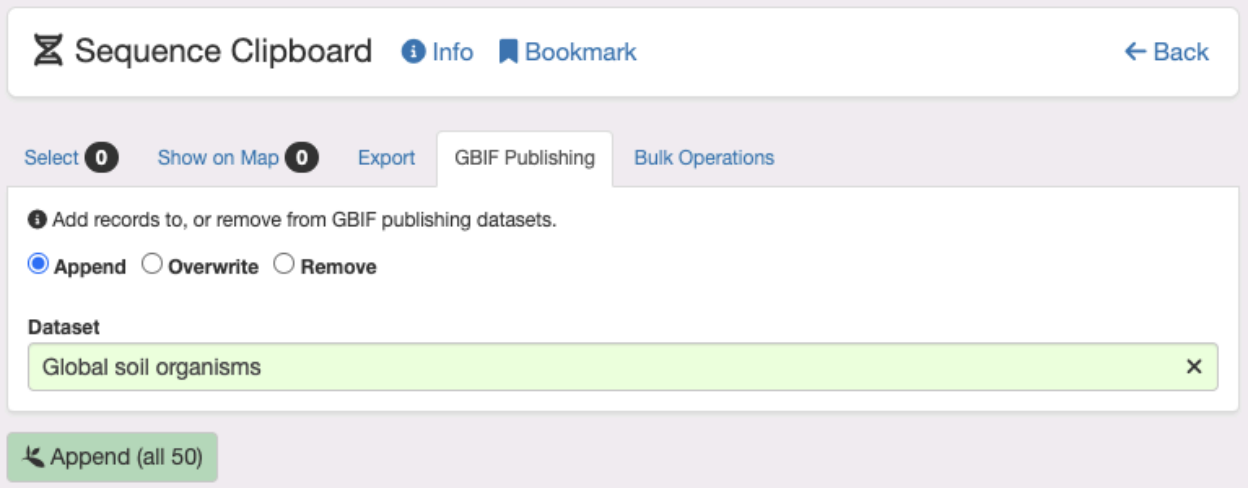


Figure 16. Associating DNA sequence records with the GBIF dataset.

Back on the GBIF dataset page (**Main menu** → **Publishing** → **GBIF Datasets**), review the dataset subjected to publishing in GBIF -

- Click **Generate DwCA** to create the Darin Core Archive (DwCA) file.
- Download the DwCA file from the **History** panel of the selected GBIF dataset and review it to ensure everything looks correct.

If everything looks correct, click the **Publish to GBIF** button.

After successful publishing, the dataset is assigned a GBIF DOI (Figure 17) which directs to the dataset's page in the GBIF portal.

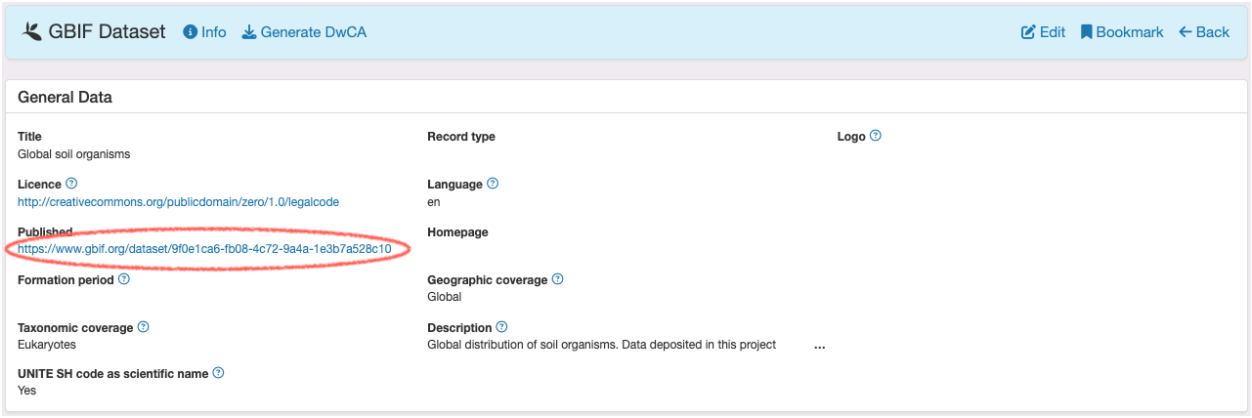


Figure 17. Dataset published to GBIF.

Additional Notes:

- Each GBIF dataset may include data from one PlutoF project only.
- Anyone can create new datasets, but an administrator needs to verify them before the initial publication to GBIF. If your dataset is ready for evaluation, please contact support@plutof.ut.ee. Re-publishing after the first approval does not require a new approval process.
- Previously published datasets can be re-published to add or remove records or to update metadata.
- Although published datasets can be deleted, they never fully disappear. They remain accessible on the GBIF side via a direct link, but they contain no records and no longer appear in new searches performed through the GBIF portal.
- While the dataset is being published, it can still be edited, but it cannot be re-published until the publishing process is complete.
- A published GBIF dataset receives a link to its page on https://www.gbif.org/. Although the page becomes available as soon as the dataset is published, some of its content may take additional time to populate. This depends on GBIF's system load and, in the worst-case scenario, may take up to 24 hours.