

# Import module

User manual

(Last updated: 12.03.2017)

**PlutoF**

## Table of Contents

### [1. Introduction](#)

### [2. General flow](#)

### [3. STEP 1: Generating template file](#)

[Open Import module](#)

[Start new import](#)

[Select data type \(module; input form name\)](#)

[Select fields](#)

[Download file](#)

[Empty template file](#)

### [4. STEP 2: Working with data file](#)

[Prepare data](#)

[Before upload \(common problems\)](#)

[Useful spreadsheet commands](#)

### [5. STEP 3: Uploading/Importing data file](#)

### [6. STEP 4: Fixing process](#)

[Opening fixing mode](#)

### [7. Tips & Tricks](#)

[Choosing input form \(mainform\)](#)

[Linking sequences to source records](#)

[Specifying taxon name for sequences](#)

[Linking data to project](#)

[Re-using localities](#)

[Pre-defining parent areas](#)

## 1. Introduction

Import tool is CSV format based. Currently PlutoF does not support import for Excel or OpenOffice, etc. file types. User should export his/her data to CSV format with commas used as field delimiters. This functionality is available in all spreadsheet programs.

Data types that can be currently imported include: collection specimens, observations, material samples, and DNA sequences. For further details on the specifics of importing each data type, please refer to paragraph 7 (Tips & Tricks).

## 2. General flow

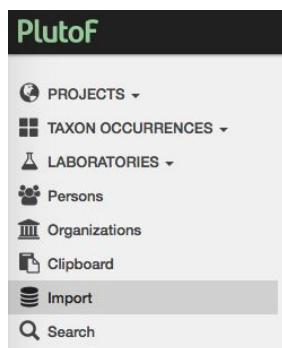
1. Generating template file in PlutoF (CSV file will be downloaded to your computer)
2. Filling out the template file (copy data into newly generated template file)
3. Saving file as CSV
4. Importing new CSV data file to PlutoF
5. Fixing process (only needed if parts of data are not recognized by PlutoF)

## 3. STEP 1: Generating template file

Before importing data into PlutoF, user has to create a file that is understandable to the system. File has to be in CSV format. This step is needed only once – same template can be reused in the future. Generating template means creating a file that matches user's data needs and is understandable to PlutoF.

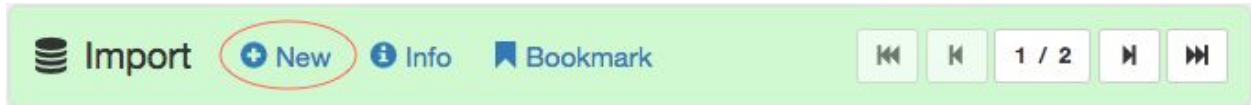
PlutoF provides a list of all the available fields (depending on the data type to be imported) and user can select the ones that are relevant to his dataset. Newly generated template file has column names that are understandable to PlutoF. **NB! Please do NOT change these names.**

### Open Import module



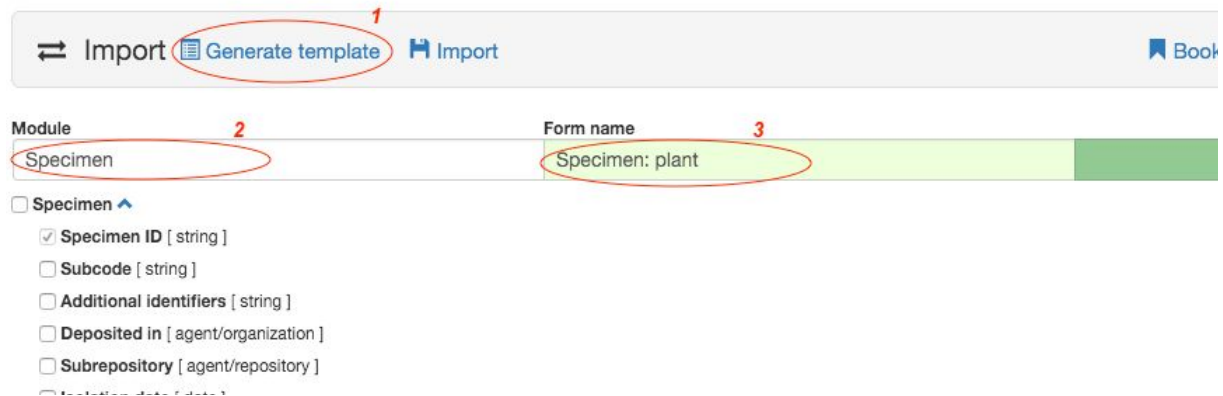
**Figure 1.** Choose *Import* from the main menu to enter the import module.

## Start new import



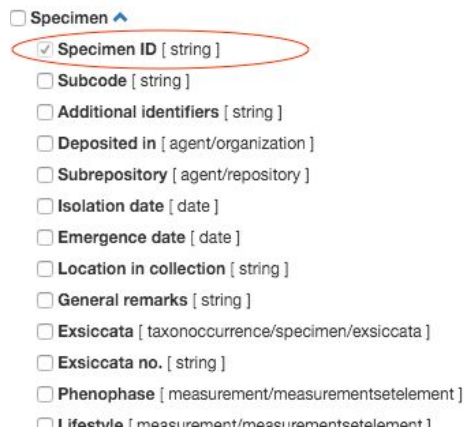
**Figure 2.** Click on the “New” link to start new import. Your previous imports are listed on the page.

## Select data type (module; input form name)



**Figure 3.** Click on the “Generate template” link (1) to start creating your own CSV template file. Available fields depend on the module (2) and input form (3).

## Select fields



**Figure 4.** Some fields are required and marked with checked boxes.

## Download file

Storage location [ string ]

Remarks [ string ]

Preparing agents [ agent/person ]

Substrate ^

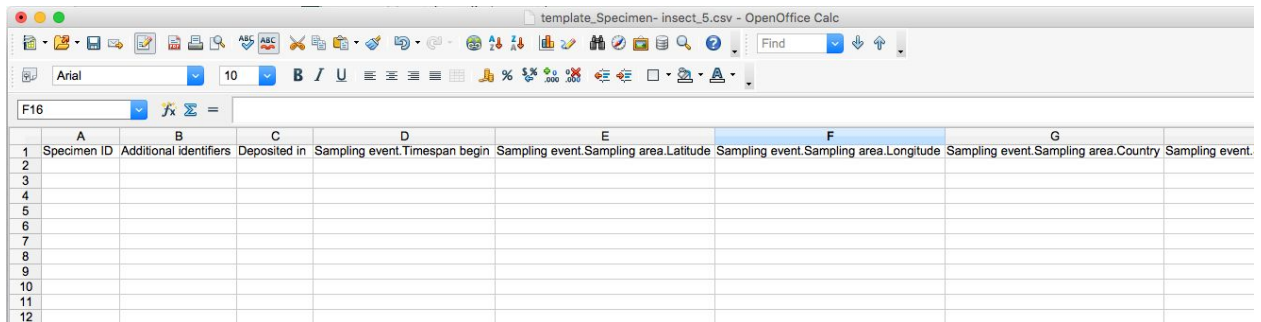
Substrate [ taxonomy/taxonnode ]

Type [ substrate/type ]

Text [ string ]

**Figure 5.** Downloading customized template file.

## Empty template file



The screenshot shows a spreadsheet application window titled "template\_Specimen- insect\_5.csv - OpenOffice Calc". The spreadsheet has 12 rows and 10 columns. The first row contains the following headers: "Specimen ID", "Additional identifiers", "Deposited in", "Sampling event.Timespan begin", "Sampling event.Sampling area.Latitude", "Sampling event.Sampling area.Longitude", "Sampling event.Sampling area.Country", and "Sampling event.". The remaining rows are empty.

	A	B	C	D	E	F	G		
1	Specimen ID	Additional identifiers	Deposited in	Sampling event.Timespan begin	Sampling event.Sampling area.Latitude	Sampling event.Sampling area.Longitude	Sampling event.Sampling area.Country	Sampling event.	
2									
3									
4									
5									
6									
7									
8									
9									
10									
11									
12									

**Figure 6.** Example CSV template file, opened in OpenOffice.

## 4. STEP 2: Working with data file

User should now have 2 files. The first one, without data, is PlutoF generated CSV file (created in the previous step). The second one is user's raw data file - this is the file that contains the data user wants to import.

In this step we will merge these two files into one. This can be done in two ways:

1. Copy column headers from PlutoF template file and paste them into user's data file (to match relevant data). This is easier solution if user has long data file.
2. Copy data from user's data files and paste them into PlutoF template file (under relevant column name).

Both solutions require user to save file as CSV file. CSV file is easily created using Excel, OpenOffice or any other spreadsheet type application. Usually this functionality can be found using File menu (*File => Save As* and set "File type" to CSV with comma used as field separator).

### Prepare data

By now user should have merged two files into one (with PlutoF column headers and user's data). Before uploading this file to PlutoF, it is recommended to take a look at your data and fix some obvious mistakes. To understand what kind of data format is ok for PlutoF, open Import (Figure 7) and check available formatting options (Figure 8).

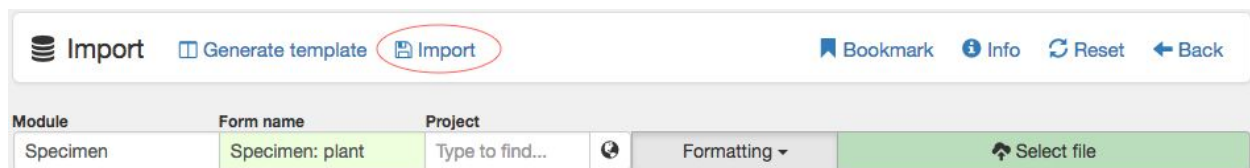


Figure 7. Opening Import settings.

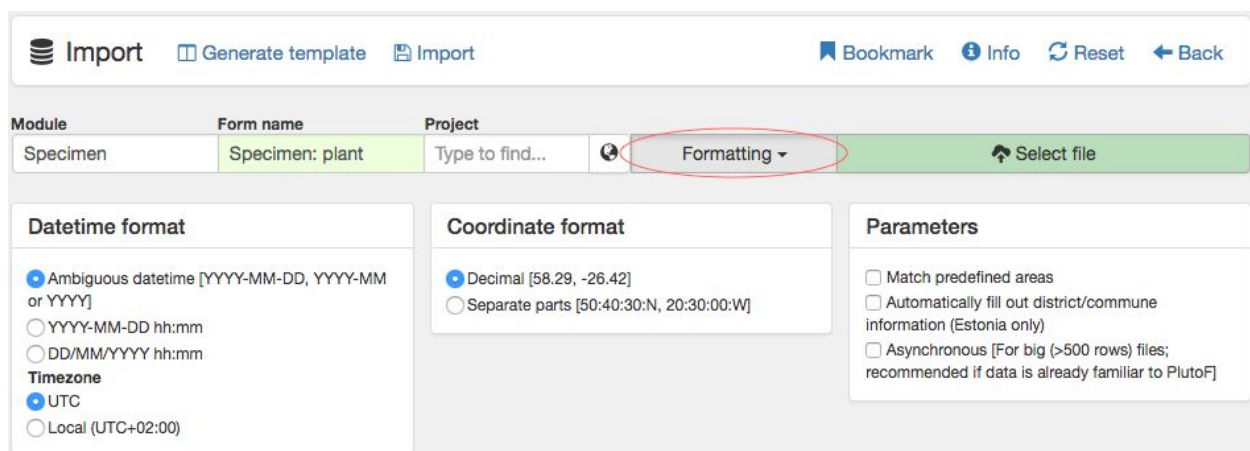


Figure 8. Opening formatting options.

## Before upload (common problems)

**Dates** - make sure that your date fields are the same as "Datetime format" (Figure 8).

**Coordinates** - verify that your coordinates, if present, are in one of the supported format (Figure 8).

**Multiple values** - to save multiple values for one field, they have to be separated with semicolon ";" (e.g. persons: John Smith; Jonas Smith, or additional identifiers: A1; A2).

**IDs** – if importing specimen data, specimen IDs have to be unique to collection.

**Source field for additional identifiers** - has to be separated by colon ":" (e.g. A1:private code).

**Specifying taxon name for sequences** - taxon name is a mandatory field in all imports. If sequences linked to specimens or living specimens are imported, taxon name can be omitted by leaving this field empty (but present in file). This way, sequence will inherit specimen's taxon name and duplicate identification will not be created.

## Useful spreadsheet commands

CONCATENATE() - concatenate two columns (e.g. taxon genus + species epithet).

TRIM() - remove additional spaces.

## 5. STEP 3: Uploading/Importing data file

Setting up import process:

1. Select module
2. Select form
3. Select your datetime format
4. Select coordinate format
5. Select “Match predefined areas” if using predefined areas in your data
6. If your data includes geo-coordinates for Estonia and no data on district, commune and locality text fields, these fields can be automatically filled by the system by selecting “Automatically fill out district/commune information (Estonia only)”
7. Use “Asynchronous” if data file is big (>500 rows) and data is already familiar to PlutoF (requires no manual interaction from the user)

The screenshot shows the 'Import' interface with the following configuration:

- Module:** Specimen (1)
- Form name:** Specimen: plant (2)
- Project:** Type to find... (3)
- Datetime format:** Ambiguous datetime [YYYY-MM-DD, YYYY-MM or YYYY] (3)
- Coordinate format:** Decimal [58.29, -26.42] (4)
- Parameters:**
  - Match predefined areas (5)
  - Automatically fill out district/commune information (Estonia only) (6)
  - Asynchronous [For big (>500 rows) files; recommended if data is already familiar to PlutoF] (7)

Figure 9. Check formatting options – they should correctly describe incoming data.

The screenshot shows the 'Import' interface with the 'Select file' button highlighted, indicating the next step in the process.

Figure 10. Upload data file.



## 6. STEP 4: Fixing process

This step is only needed if import process has detected some errors (e.g. date is in wrong format) or if part of the data is unknown to PlutoF (e.g. taxon name is missing from PlutoF classification).

### Opening fixing mode



Module	File	Status	progress
	<a href="#">template_Specimen-insect.csv</a>	Finished	100%
	<a href="#">download_b.csv</a>	Finished	100%
	<a href="#">template_Specimen-insect_g.csv</a>	Finished	100%

**Figure 11.** Fixing mode can be opened by clicking on the file name in import module's list-view.

Table with problematic fields is presented to the user with the indication of problem type (Figure 12). User can fix the problems and continue with the import. Already fixed records should not appear again.

**Status**  
File: Alex\_test.csv  
Importing: 14  
Created: 1  
Rows ignored: 0

**Legend**  
 Add new record (if not found in PlutoF)  
 Fixing recommended (click to copy value)  
 Ignore full row or individual field (ignored info will stay in file for later download)  
 Fix required

Saves fixed data and continues with the import  
 Back to imported files  
 Downloads .CSV file with only unprocessed/ignored rows (current import will be stopped)

created		unprocessed													
#	Specimen ID	Deposited in	Timespan begin	Collected by	Latitude	Longitude	Country	Locality text	Taxon name	Determiners	Type	Substrate	Text		
0	CWU5618	CWU	2010-11-6	Alexander O. <input type="text" value="Type to find..."/>	49.18	37.01	Ukraine	West from Chervonyi Shakhtar village	Sistotrema bbrinkmannii <input type="text" value="Type to find..."/>	Alexander Ordynets	fruitbody	Ganoderma applanatum			
1	CWU5810	CWU	2010-11-6	Alexander O. <input type="text" value="Type to find..."/>	49.18	37.01	Ukraine	West from Chervonyi Shakhtar village	Sistotrema bbrinkmannii <input type="text" value="Type to find..."/>	Alexander Ordynets	fruitbody	Inocutis rheades			
2	CWU6475	CWU	2010-5-5	Alexander O. <input type="text" value="Type to find..."/>	49.18	37.01	Ukraine	West from Chervonyi Shakhtar village	Sistotrema oblongisporum	Alexander Ordynets	fruitbody	Corylus avellana	branch		

**Figure 12.** Table presented to the user with rows that need fixing.

Red warning is shown in cases where PlutoF does not recognize some data field. User can use autocomplete field to re-check if this info is missing from PlutoF (alternatively, it can just be a typo). If data is truly missing, it can be added to the system using “Add” link next to specific column header. After that, field value can be re-selected using the same autocomplete field.

Fields and rows can also be ignored for now, to deal with them later on. This can be done using “[X]” link next to field (to ignore specific field) or at the beginning of the row (to skip entire row). See Figure 13 for common fixing actions.

**Status**  
 File: Alex\_test.csv  
 Importing: 14  
 Created: 1  
 Rows ignored: 0

**Legend**  
 Add new record (if not found in PlutoF)  
 Fixing recommended (click to copy value)  
 Ignore full row or individual field (ignored info will stay in file for later download)  
 Fix required

✓ Saves fixed data and continues with the import  
 ← Back to imported files  
 Downloads .CSV file with only unprocessed/ignored rows (current import will be stopped)

created		unprocessed										
#	Specimen ID	Deposited in	Timespan begin	Collected by	Latitude	Longitude	Country	Locality text	Taxon name	Determiners	Type	Substrate
0	CWU5618	CWU	2010-11-6	Alexander O.	49.18	37.01	Ukraine	West from Chervonyi Shakhtar village	Sistotrema bbrinkmannii	Alexander Ordynets	fruitbody	Ganoderma applanatum
1	CWU5810	CWU	2010-11-6	Alexander O.	49.18	37.01	Ukraine	West from Chervonyi Shakhtar village	Sistotrema bbrinkmannii (Bres.) J. Erikss., 1948 (spe)		fruitbody	Inocutis rheades

**Figure 13.** Fixing process: 1-add taxon name; 2-add person; 3-copy field value; 4-ignore a row; 5-select correct value from autocomplete field.

If user has fixed the errors, clicking on “Save and continue” will start the file processing again (Figure 14). “Back” button will take user back to his list of imports. Imports don’t have to be finished in one go. User can leave the import running and continue using other parts of PlutoF. Imports can have 3 different states:

1. Finished - file is 100% processed
2. Waiting - file is being analyzed
3. Has errors - waiting for a user to continue in fixing mode
4. Stopped - import has been stopped by user (when “Download unprocessed/ignored rows” is clicked)

User can stop the import process at any time. Selecting “Download unprocessed/ignored rows” will return a file containing only the rows which have NOT been imported, so that user can continue with the same file without worrying about duplicates.

18	20i	2002-05-06	31.22	63.56	Estonia	agilis	Allan Puu
19	21i	2002-05-07	32.22	64.56	Estonia	tardus	Allan Puu

**Figure 14.** Import flow can be directed using the “Save and continue”, “Back”, and “Download unprocessed/ignored rows” buttons.

## 7. Tips & Tricks

- “,” – specify multiple values
- “.” – specify source for additional identifier
- ⚠ – copy value

Below are listed recommendations for importing data in the most optimised way. Please **read them through and apply whenever possible** - your data will be much more easier to manage later if you follow them. In case of any questions or comments please contact [support@plutof.ut.ee](mailto:support@plutof.ut.ee).

### Choosing input form (mainform)

**Choose input form that fits your data.** Records can be imported only with one form at the time. For example, if bird, insect and plant observations are stored in the same file, they need to be split into 3 separate datasets and CSV files - one file for one import with different mainform (Observation: bird, Observation: insect, Observation: default).

### Linking sequences to source records

**DNA sequences can be linked to existing source/parent records already in PlutoF.** These include preserved specimens (type: specimen), living specimens (type: livingspecimen) and material samples (type: materialsample). To make these links, it is needed to fill out “Parent: Type” and “Parent: Name” fields in sequence import file (Figure 15). While type refers to parent type record, name field refers to preserved specimen ID, living specimen Code, or material sample Name.

The screenshot shows the PlutoF import interface for DNA sequences. At the top, there are navigation buttons: 'Import', 'Generate template', and 'Import'. On the right, there are 'Bookmark', 'Info', 'Reset', and 'Back' buttons. Below this, there is a table with two columns: 'Module' and 'Form name'. The 'Module' is 'Sequence' and the 'Form name' is 'Sequence: default'. A 'Download file' button is visible on the right. Below the table, there is a 'Fields' section. On the left, there is a list of fields with checkboxes. The 'Parent' section is expanded, and 'Type [choices]' and 'Name [string]' are circled in red. On the right, there is a 'Selected fields' list. The selected fields are: 'Parent.Type [choices]', 'Parent.Name [string]', 'Sequence ID [string]', 'Sequenced regions [choices]', 'Sequence [string]', 'Isolation source [taxonoccurrence/sequence/source]', 'EcM lineage [taxonoccurrence/sequence/ecmlineage]', and 'Remarks [string]'.

**Figure 15.** In order to link DNA sequence to existing PlutoF records, parent type and name fields should be filled.

## Specifying taxon name for sequences

Taxon name is a mandatory field in all imports. If sequences linked to specimens or living specimens are imported, taxon name can be omitted by leaving this field empty (but present in file). This way, sequence will inherit specimen's taxon name and duplicate identification will not be created.

## Linking data to project

**All imported data types can be linked to project.** For this, before uploading CSV file, project should be selected by typing project name on autocomplete field (Figure 9). Projects in PlutoF are mainly used to effectively organize, share, search, and publish data as unified datasets.

## Re-using localities

**Localities linked to projects can be re-used when importing records collected from the exact same location.** For this, all unique localities in CSV file should be complemented with unique "Sampling area: Name" fields (see example on Figure 16). Then, the first occurrence of unique area will be added as new, and all subsequent records labeled with this area name will be linked to the same, newly added area. If area with the same name already exists under user's selected project, already existing area will be used instead of adding new one.

N	O	P	Q
Sampling event.Sampling area.Name	Sampling event.Sampling area.Longitude	Sampling event.Sampling area.Latitude	Sampling event.Sampling area.Country
site1	24.6744649	36.9887823	Greece
site2	24.674465	36.9887824	Greece
site1	24.6744649	36.9887823	Greece
site1	24.6744649	36.9887823	Greece
site2	24.674465	36.9887824	Greece
site3	24.6744651	36.9887825	Greece
site3	24.6744651	36.9887825	Greece
	24.6744652	36.9887826	Greece

**Figure 16.** "Sampling area: Name" field should be filled with unique location names in order to re-use them when importing taxon occurrence records under project.

Re-using localities for the data collected from the same place makes it easy to manage your data - in case of a need to change locality details (e.g. geo-coordinates, locality text, coordinate uncertainty, etc), this has to be changed only once for multiple taxon occurrences.

## Pre-defining parent areas

**Parent areas can be used to group areas with point location (but also other types like box, polygon) under larger project areas (polygons).** This functionality contributes to designing sampling schemes where sub-samples are collected from pre-defined areas (sites or plots). For this, "Sampling area: Parent sampling area" field has to be filled with existing area name under the project that imported taxon occurrences will be linked to (Figure 17).

I	J	K	L	M
Sampling event.Sampling area.Parent sampling area	Sampling event.Sampling area.Name	Sampling event.Sampling area.Latitude	Sampling event.Sampling area.Longitude	Sampling event.Sampling area.Country
parent area 1	Morocco1	31.54632224	-7.646282287	Morocco
parent area 1	Morocco1	31.54632224	-7.646282287	Morocco
parent area 1	Morocco1	31.54632224	-7.646282287	Morocco
parent area 1	Morocco1	31.54632224	-7.646282287	Morocco
parent area 1	Morocco1	31.54632224	-7.646282287	Morocco
parent area 1	Morocco1	31.54632224	-7.646282287	Morocco
parent area 1	Morocco1	31.54632224	-7.646282287	Morocco
parent area 1	Morocco1	31.5467588	-7.646788303	Morocco
parent area 1	Morocco2	31.54681366	-7.646766845	Morocco
parent area 1	Morocco2	31.54681366	-7.646766845	Morocco
		31.54478334	-7.614883929	Morocco
		31.5467403	-7.646841779	Morocco
parent area 2	Morocco3	31.40126678	-7.480887473	Morocco
parent area 2	Morocco3	31.40126678	-7.480887473	Morocco

**Figure 17.** “Sampling area: Parent sampling area” fields should be filled with existing area names to link them to predefined parent areas under project.