

Impordi moodul

Juhend

Tööversioon seisuga 12.03.2017

PlutoF

Sisukord

1. Sissejuhatus

2. Impordi etapid

3. ETAPP 1: Alusfaili loomine

Ava impordi moodul

Alusta uut importi

Vali andmete tüüp (moodul; sisestusvormi nimi)

Täpsusta väljad

Laadi alla fail

Tühi alusfail

4. ETAPP 2: Töötamine andmefailiga

Andmete ettevalmistamine

Enne üleslaadimist (levinumad probleemid)

Kasulikud arvutustabeli käsud

5. ETAPP 3: Andmefaili üleslaadimine/import

6. ETAPP 4: Parandamise tööetapp

Parandamise režiimi aktiveerimine

7. Kasulikud näpunäited

Sisestusvormi valimine

Geenijärjestuste sidumine allika/vanema kirjetega

Geenijärjestusele määrangu / taksoni nime lisamine

Andmete sidumine projektiga

Leiukohtade taaskasutamine

Ülemalade eel-defineerimine

1. Sissejuhatus

PlutoF impordi tööriist põhineb CSV formaadil. Hetkel ei ole toetatud import MS Excel, OpenOffice, jt failitüüpidest. Kasutaja peab salvestama enda andmed CSV formaadis kasutades väljade eraldajana koma. Vastav funktsionaalsus on toetatud kõikides tabelarvutus tarkvarades.

Andmetüübid, mida hetkel importida lubatakse, on järgmised: kogude eksemplarid, loodusvaatlused, proovid ja geenijärjestused. Erinevate andmetüüpide impordi spetsiifika ja võimalused on täpsemalt lahti seletatud peatükis 7 (Kasulikud näpunäited).

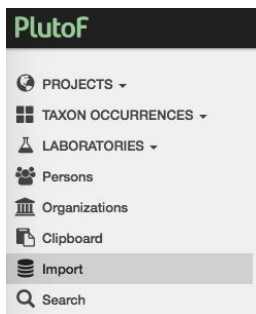
2. Impordi etapid

1. Alusfaili loomine PlutoF töölaual (CSV fail salvestatakse kasutaja arvutisse)
2. Alusfaili täitmine (andmete kopeerimine eelmises etapis salvestatud alusfaili)
3. Faili salvestamine CSV formaadis
4. CSV faili üleslaadimine PlutoF töölaual
5. Parandamise tööetapp (vajalik vaid juhul kui andmed ei ole süsteemi poolt äratuntavad)

3. ETAPP 1: Alusfaili loomine

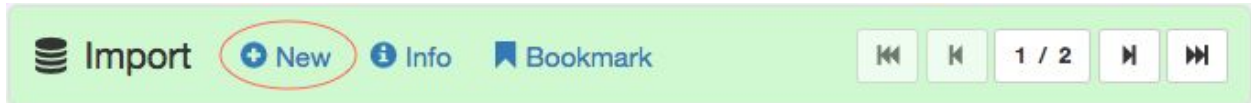
Enne andmete importimist PlutoFi peab kasutaja looma alusfaili, mis on süsteemi poolt äratuntav. Fail peab olema CSV formaadis. Alusfaili salvestamine on sarnaste andmete impordil vajalik vaid üks kord - sama faili saab järgmiste üleslaadimiste juures taaskasutada. Selle protsessi käigus luuakse fail, mis vastab kasutaja andmetele ning on süsteemi poolt äratuntav. PlutoF töölaual kuvatakse nimekiri kõikidest andmeväljadest, mis on konkreetse andmetüübi puhul toetatud, ning kasutaja saab valida need väljad, mis tema andmetele vastavad. Loodavas alusfailis on veergude nimed salvestatud kujul, mis on süsteemi poolt ära tuntavad. **NB! Palume veergude nimesid mitte muuta.**

Ava impordi moodul



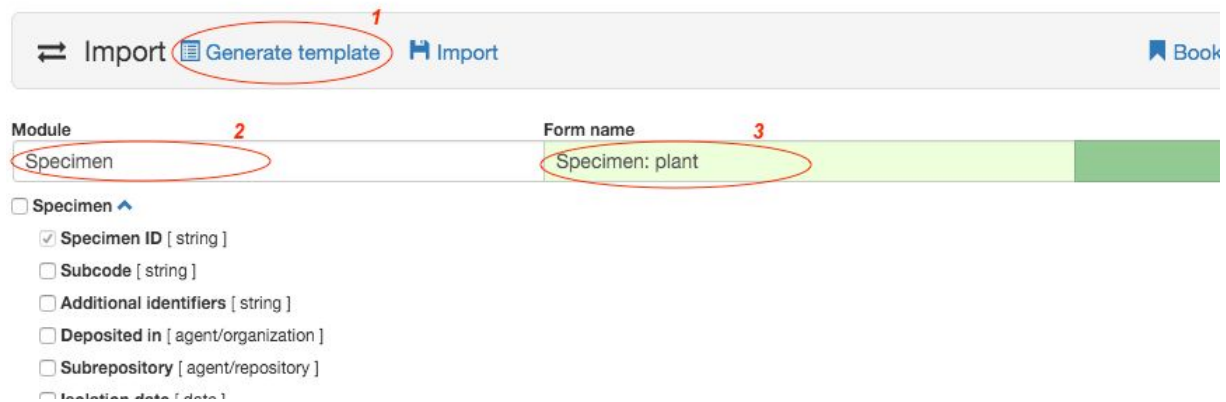
Joonis 1. Sisnemaks impordi moodulisse vali peamenüüst *Import*.

Alusta uut importi



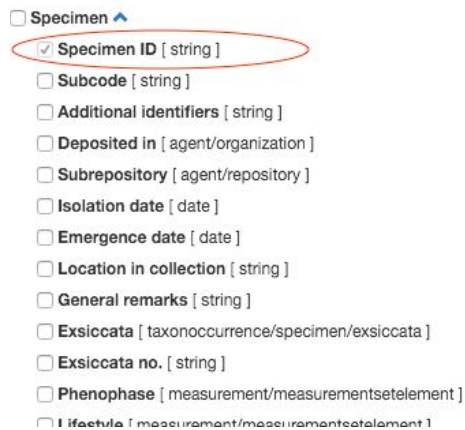
Joonis 2. Klikates lingile “Uus” alustatakse uut impordi protsessi. Varasemad protsessid on kuvatud lehel nimekirjana.

Vali andmete tüüp (moodul; sisestusvormi nimi)



Joonis 3. Uue CSV alusfaili loomiseks tuleb klikata lingil “Generate template” (1). Sisestusväljade nimekiri sõltub moodulist (andmetüübist) (2) ja sisestusvormist (3).

Täpsusta väljad



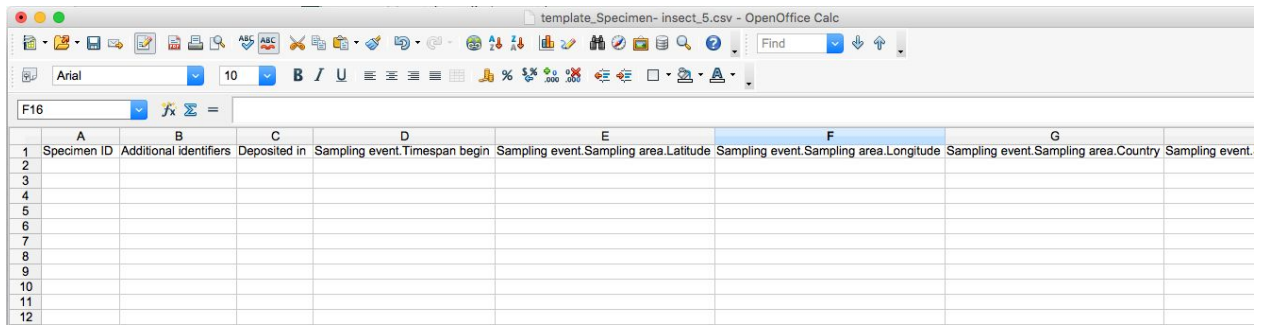
Joonis 4. Kohustuslikud väljad on tähistatud linnukesega välja nime ees.

Laadi alla fail

- Storage location [string]
- Remarks [string]
- Preparing agents [agent/person]
- Substrate ^
 - Substrate [taxonomy/taxonnode]
 - Type [substrate/type]
 - Text [string]

Joonis 5. Kohandatava alusfaili allalaadimine.

Tühi alusfail



The screenshot shows a window titled "template_Specimen- insect_5.csv - OpenOffice Calc". The spreadsheet has a header row (row 1) with the following columns: A (Specimen ID), B (Additional identifiers), C (Deposited in), D (Sampling event.Timespan begin), E (Sampling event.Sampling area.Latitude), F (Sampling event.Sampling area.Longitude), G (Sampling event.Sampling area.Country), and H (Sampling event.). Rows 2 through 12 are empty.

	A	B	C	D	E	F	G	H
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								

Joonis 6. Näidis alusfail avatuna OpenOffice programmis.

4. ETAPP 2: Töötamine andmefailiga

Esimese etapi lõppedes peaks kasutajal olema 2 faili. Esimene, ilma andmeteta, on PlutoF töölaual loodud CSV fail. Teine on kasutaja enda andmefail, mis sisaldab üleslaadimiseks ette valmistatud andmeridu.

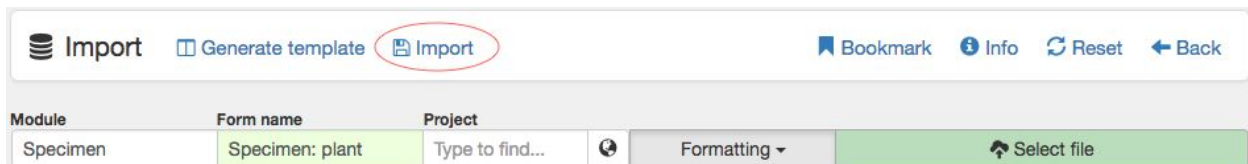
Järgmiseks tuleb need kaks faili ühendada. Seda on võimalik teha kahel viisil:

1. Kopeerida veergude nimed PlutoF alusfailist kasutaja andmefailis õigetele veergudele. Selline lahendus on soovitatud suuremate andmefailide puhul.
2. Kopeerida andmed kasutaja failist PlutoF alusfaili vastavatele väljadele.

Mõlemal juhul tuleb töö lõppedes salvestada fail CSV formaadis. CSV formaadis faili saab MS Excel, OpenOffice jt tabelarvutus tarkvarade abil lihtsasti luua. Reeglina leiab vastava funktsionaalsuse programmi "Fail" menüüst (*Fail => Salvesta kui*. Faili tüübiks tuleb valida CSV, kus väljade eraldajana kasutatakse koma).

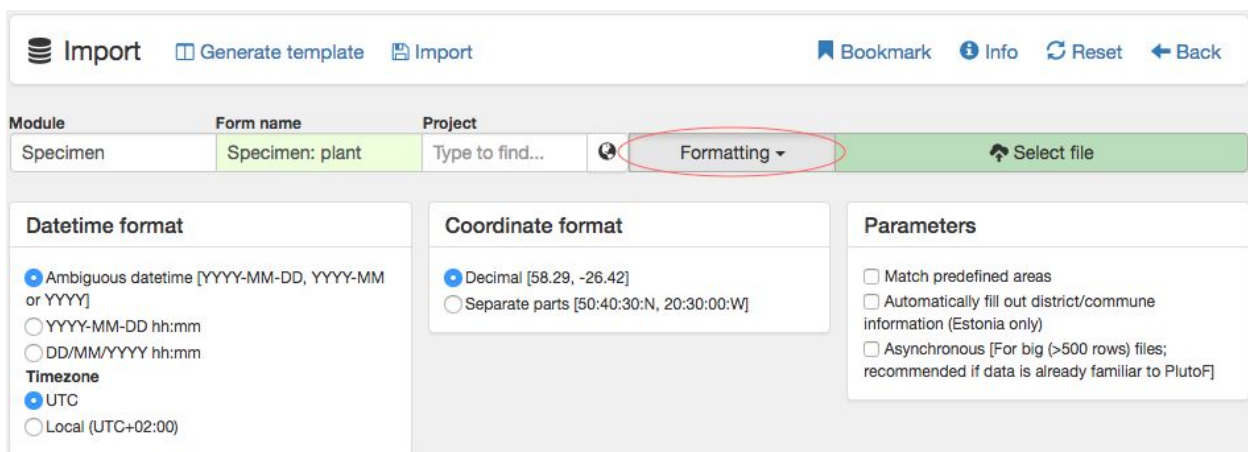
Andmete ettevalmistamine

Pärast kahe faili kokkuviiimist (PlutoF alusfaili veergude nimed ja kasutaja andmed) tuleks andmereal üle kontrollida ning andmete formaadid ühtlustada. Süsteemi poolt aktsepteeritavad formaadid avanevad impordi moodulis (Joonis 7) "Formatting" lingi alt (Joonis 8).



The screenshot shows the top navigation bar with 'Import' circled in red. Below it is a table with columns 'Module', 'Form name', and 'Project'. The 'Module' is 'Specimen', 'Form name' is 'Specimen: plant', and 'Project' is 'Type to find...'. To the right of the table is a 'Formatting' dropdown menu and a 'Select file' button.

Joonis 7. Impordi seadete avamine.



The screenshot shows the 'Formatting' dropdown menu expanded. It contains three sections: 'Datetime format' with options for 'Ambiguous datetime [YYYY-MM-DD, YYYY-MM or YYYY]', 'YYYY-MM-DD hh:mm', and 'DD/MM/YYYY hh:mm'; 'Timezone' with options for 'UTC' (selected) and 'Local (UTC+02:00)'; 'Coordinate format' with options for 'Decimal [58.29, -26.42]' (selected) and 'Separate parts [50:40:30:N, 20:30:00:W]'; and 'Parameters' with options for 'Match predefined areas', 'Automatically fill out district/commune information (Estonia only)', and 'Asynchronous [For big (>500 rows) files; recommended if data is already familiar to PlutoF]'.

Joonis 8. Formaadi valikute avamine.

Enne üleslaadimist (levinumad probleemid)

Kuupäevad - kontrolli, et failis olevad kuupäevad oleksid süsteemi poolt äratuntavas formaadis (Joonis 8).

Koordinaadid - kontrolli, et koordinaatide olemasolul kattuks nende formaat süsteemi poolt toetatud formaadiga (Joonis 8).

Mitu väärtust ühel väljal - mitme väärtuse samal väljal sisestamiseks tuleks need eraldada semikooloniga “;” (nt. isikud: John Smith; Jonas Smith, või teisesed koodid: A1; A2).

Koodid – eksemplari andmete importimisel peavad eksemplaride koodid olema kogu/asutuse piires unikaalsed.

Teiseste koodide “Allika” väli - peab olema koodist eraldatud kooloniga “:” (nt. A1:koguja number).

Geenijärjestusele määrangu / taksoni nime lisamine - taksoni nimi on impordis kohustuslik väli. Juhul kui üleslaetavad geenijärjestused seotakse (elus-)eksemplaridega, on lubatud taksoni nime väli tühjaks jätta, kuid see peab imporditavas failis väljana kirjas olema. Sel juhul pärib geenijärjestus oma määrangu vanemobjektilt ning süsteemi ei tekitata sama taksoni nimega topeltkirjet.

Kasulikud arvutustabeli käsud

CONCATENATE() - liidab kokku mitmes eraldi veerus asuvad väärtused (nt. taksoni nime perekond + liigi epiteet).

TRIM() - eemaldab üleliigsed tühikud välja alguses ja lõpus.

5. ETAPP 3: Andmefaili üleslaadimine/import

Andmete impordiks:

1. Vali moodul (andmetüüp)
2. Vali sisestusvorm
3. Täpsusta kuupäeva formaat
4. Täpsusta koordinaatide formaat
5. Vali "Sobita eeldefineeritud aladega" kui failis on kasutatud eelnevalt süsteemi sisestatud "eeldefineeritud" tüüpi alad
6. Juhul kui fail sisaldab geo-koordinaate, mis asuvad Eestis, ent failist puuduvad maakonna, linna/valla ja täpse asukoha väljad, on võimalik vastavad väljad automaatselt täita, lülitades lisaseadete alt sisse "Täida haldusjaotuse väljad automaatselt (Ainult Eesti)"
7. Vali "Asünkroonne" sisestus kui andmefail on suur (>500 rida) ja enamik andmeid on juba süsteemile tuttavad (vajadus andmete parandamiseks tõenäoliselt puudub).

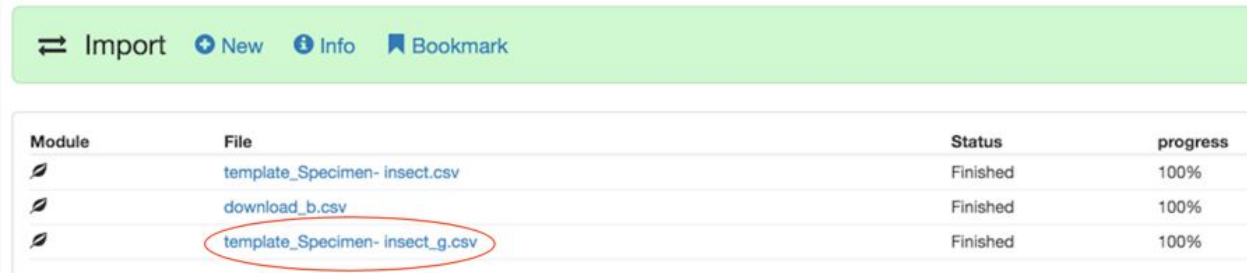
Joonis 9. Kontrolli formaatide valikuid - need peaksid minema kokku andmefailis kasutatud formaatidega.

Joonis 10. Laadi üles andmefail.

6. ETAPP 4: Parandamise tööetapp

Andmete parandamise etapp aktiivse impordi ajal on vajalik vaid siis kui süsteem tuvastab lisatavates andmetes vigu (nt. kuupäev vales formaadis) või kui osa andmeid on süsteemile tundmatud (nt. taksoni nimi puudub PlutoF klassifikatsioonist).

Parandamise režiimi aktiveerimine



Module	File	Status	progress
	template_Specimen-insect.csv	Finished	100%
	download_b.csv	Finished	100%
	template_Specimen-insect_g.csv	Finished	100%

Joonis 11. Parandamise režiim avaneb impordi mooduli nimekirja-vaates faili nimel klikates.

Kasutajale kuvatakse tabel parandamist vajavate väljadega, milles probleemne väärtus on tähistatud punase värviga (Joonis 12). Kasutaja saab probleemset väljad kohe parandada ja impordi protsessiga jätkata. Juba parandatud kirjed ei peaks probleemsete kirjetena enam tabelisse ilmuma.

Status		Legend										
File: Alex_test.csv	Importing: 14	Add new record (if not found in PlutoF)	Fixing recommended (click to copy value)	<input checked="" type="checkbox"/> Saves fixed data and continues with the import	Back to imported files							
Created: 1	Rows ignored: 0	<input type="checkbox"/> Ignore full row or individual field (ignored info will stay in file for later download)	<input type="checkbox"/> Fix required	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	Alexander Ordynets	fruitbody	Inocutis rheades
2	CWU6475	CWU	2010-5-5	Alexander O.	49.18	37.01	Ukraine	West from Chervonyi Shakhtar village	Sistotrema oblongisporum	Alexander Ordynets	fruitbody	Corylus avellana

Joonis 12. Kasutajale kuvatav parandamist vajavate kirjetabel.

Punane hoiatus kuvatakse juhul kui süsteem ei tunne ära väljal olevat väärtust. Kui tegemist on kirjaveaga, tuleks väärtuse valimiseks kasutada välja *autocomplete* funktsionaalsust. Juhul kui väärtus süsteemist puudub, on see võimalik lisada kasutades "Lisa" funktsionaalsust vastava veeru nime kõrval (Joonis 13). Pärast uue väärtuse lisamist on võimalik see valida uuesti välja *autocomplete* funktsionaalsust kasutades.

Üksikuid parandamist vajavaid andmevälju ja terveid andmeridu võib ka eirata, et nendega hiljem tegeleda. Seda saab teha kasutades "[-]" linki andmevälja kõrval (tühjendades üksiku andmevälja) või rea alguses (eiramaks tervet rida). Joonisel 13 on näidatud levinumad parandamise tegevused.

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	⊕ Add Timespan begin	Collected by	⊕ Add Latitude	Longitude	Country	Locality text	Taxon name	⊕ Add Determiners	⊕ Add Type	Substrate	⊕ Add Text
0	CWU5618	CWU	2010-11-6	▲ Alexander O. Type to find...	49.18	37.01	Ukraine	West from Chervonyi Shakhtar village	▲ Sistotrema bbrinkmannii sistotrema bfr	Alexander Ordynets	fruitbody	Ganoderma applanatum	
1	CWU5810	CWU	2010-11-6	▲ Alexander O. Type to find...	49.18	37.01	Ukraine	West from Chervonyi Shakhtar village	Sistotrema bbrinkmannii (Bres.) J. Erikss., 1948 (spe) Sistotremataceae (fam) Sistotremastrum J. Erikss., 1958 (gen)		fruitbody	Inocutis rheades	

Joonis 13. Parandamise töötapp: 1-uuu taksoni nime lisamine; 2-uuu isiku nime lisamine; 3-väljal oleva väärtuse kopeerimine; 4-rea eiramine; 5-väärtuse valimine *autocomplete* väljalt.

Pärast andmete parandusi on võimalik protsessi jätkata “Salvesta ja jätka” nupule klikates (Joonis 14). “Tagasi” nupp viib kasutaja tagasi impordi mooduli nimekirja-vaatesse. Andmefaili impordi ei pea korraga lõpule viima - kasutaja võib jätta impordi protsessi tööle ning jätkata PlutoF töölauda teiste moodulite kasutamist. Impordil võib olla 4 erinevat staatust:

1. Lõpetatud - 100% failist on töödeldud ning töö lõpetatud
2. Ootel - faili analüüsitakse
3. Sisaldab vigu - süsteem on tuvastanud vigu, mis nõuavad parandamist kasutaja poolt
4. Peatatud - import on kasutaja poolt peatatud (kasutades “Salvesta töötlemata/eiratud read” valikut)

Kasutaja võib impordi igal ajal peatada. Valides “Salvesta töötlemata/eiratud read” tagastatakse kasutajale fail, mis sisaldab vaid neid ridu, mis on süsteemi poolt veel importimata. Kasutaja saab impordi jätkata sama failiga, ilma et peaks muretsema duplikaatide tekkimise pärast.

18	20i	2002-05-06	31.22	63.56	Estonia	▲ agilis Type to find...	▲ Allan Puu Type to find...
19	21i	2002-05-07	32.22	64.56	Estonia	▲ tardus Type to find...	▲ Allan Puu Type to find...


Save and continue
 Back
 Download unprocessed/ignored rows

Figure 14. Impordi protsessi juhtimiseks on võimalik kasutada “Salvesta ja jätka”, “Tagasi”, ja “Salvesta töötlemata/eiratud read” valikuid.

7. Kasulikud näpunäited

“;” – mitme väärtuse samal väljal sisestamisel kasutatav väärtuste eraldaja

“:” – teiste koodide “Allika” eraldamiseks koodist

 – väljal oleva väärtuse kopeerimiseks

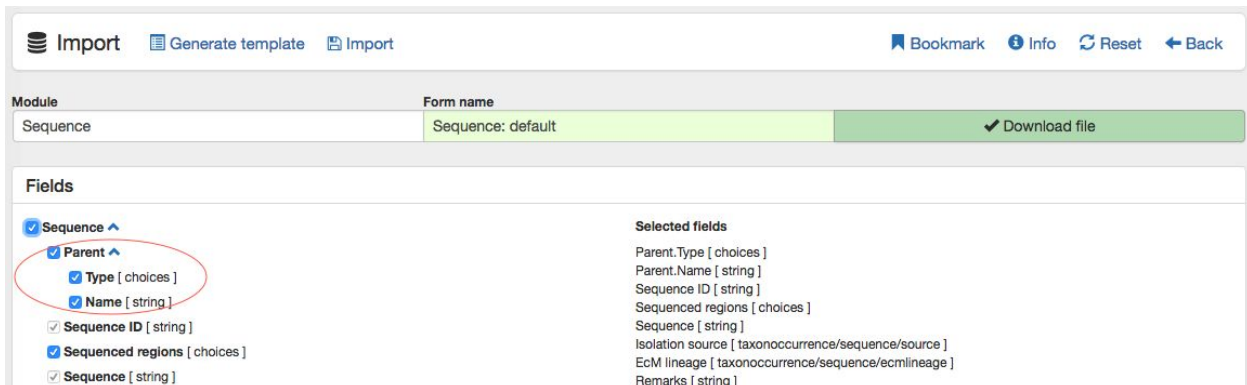
Allpool on ära toodud mõningad näpunäited andmete võimalikult optimeeritud impordiks. Palume need läbi lugeda ning võimalusel alati rakendada - teie andmed on seeläbi edaspidi tunduvalt lihtsamini hallatavad. Küsimuste tekkimisel palume meiega julgesti ühendust võtta saates e-maili aadressile support@plutof.ut.ee.

Sisestusvormi valimine

Vorm peab sobima imporditavate andmetega. Kirjeid saab üles laadida vaid ühe vormi kaupa. Näiteks, kui hoiate lindude, putukate ja taimede vaatluseid ühes failis, siis PlutoFi impordimiseks tuleb need kolmeks andmesetiks ja CSV failiks lahku lüüa - üks fail iga erineva sisestusvormi (Vaatlused: linnud, Vaatlused: putukad, Vaatlused: vaikumisi) jaoks.

Geenijärjestuste sidumine allika/vanema kirjetega

Geenijärjestusi on võimalik siduda PlutoF süsteemis olemas olevate allika/vanema kirjetega (objekt, millest geenijärjestus on saadud). Nendeks võivad olla nt kogude eksemplarid (tüüp: specimen), elus eksemplarid (tüüp: livingspecimen) või proovid (tüüp: materialsample). Lingi tekitamiseks geenijärjestuse ja vanema vahel peab ära täitma “Parent: type” ja “Parent: Name” väljad imporditavas CSV failis (Joonis 15). “Name” väli viitab vanemobjekti koodile (eksemplar, eluseksmeplar) või nimele (proov).



The screenshot shows the PlutoF import interface. At the top, there are buttons for 'Import', 'Generate template', and 'Import'. On the right, there are buttons for 'Bookmark', 'Info', 'Reset', and 'Back'. Below this, there is a table with two columns: 'Module' and 'Form name'. The 'Module' column contains 'Sequence' and the 'Form name' column contains 'Sequence: default'. To the right of this table is a green button with a checkmark and the text 'Download file'. Below the table is a section titled 'Fields'. On the left, there is a list of fields with checkboxes: 'Sequence' (checked), 'Parent' (checked), 'Type [choices]' (checked), 'Name [string]' (checked), 'Sequence ID [string]' (checked), 'Sequenced regions [choices]' (checked), and 'Sequence [string]' (checked). A red circle highlights the 'Parent' field and its sub-fields 'Type' and 'Name'. On the right, there is a section titled 'Selected fields' with a list of fields: '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. Sidumaks geenijärjestusi PlutoFis olemasolevate vanemobjektidega, tuleb ära täita vanema tüübi ja nime väljad.

Geenijärjestusele määrangu / taksoni nime lisamine

Taksoni nimi on impordis kohustuslik väli. Juhul kui üleslaetavad geenijärjestused seotakse (elus-)eksemplaridega, on lubatud taksoni nime väli tühjaks jätta, kuid see peab imporditavas failis väljana kirjas olema. Sel juhul pärib geenijärjestus oma määrangu vanemobjektilt ning süsteemi ei tekitata sama taksoni nimega topeltkirjet.

Andmete sidumine projektiga

Kõiki imporditavaid andmetüpe on võimalik siduda projektiga. Selleks tuleb enne imporditava faili üleslaadimist täpsustada *autocomplete* väljal projekti nimi (Joonis 9). Projekte kasutatakse peamiselt andmete efektiivseks grupeerimiseks, jagamiseks, otsimiseks ja publitseerimiseks ühtse andmesetina.

Leiukohtade taaskasutamine

Projektiga seotud leiukohti on võimalik samast kohast kogutud andmete sisestamisel taaskasutada. Selleks tuleb kõikide unikaalsetele leiukohtadele lisada ala nimi ("Sampling area: Name", näidis Joonisel 16). Esimene unikaalse nimega leiukohtades lisatakse uuena, järgnevad sama nimega tähistatud alaga kirjed seotakse aga juba loodud leiukoha kirjega. Kui failis oleva ala nimega leiukoht sama projekti all juba eksisteerib, kasutatakse kirjete lisamisel olemasolevat leiukohta.

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. Projekti alade taaskasutamiseks tuleb unikaalsetele leiukohtadele lisada ala nimi.

Leiukohtade taaskasutamine muudab andmete haldamise lihtsamaks - asukohainfo (nt koordinaadid, täpsustaste, haldusüksuste nimed) muutmiseks tuleb seda teha samast kohast kogutud andmete jaoks vaid üks kord.

Ülemalade eel-definimine

Ülemalaid on võimalik kasutada punktkoordinaadiga (või ka polügoni ja ruuduna) sisestatud alamalade grupeerimiseks ülemala (polügon) alla. Selline funktsionaalsus lubab disainida andmete kogumise skeeme, kus geo-koordinaatidega tähistatud proove kogutakse suurema prooviala seest. Selleks tuleb ülemala väljal ("Sampling area: Parent sampling area") täpsustada importimisel seotava projekti all oleva (olemasolava) ülemala nimi (Joonis 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. Imporditavate leiukohtade grupeerimiseks ülemala alla tuleb imporditavas failis täpsustada projektiga seotud, olemasoleva ülemala nimi.