USER MANUAL

HOW TO IMPORT A XML ABCD FORMATTED FILE IN DARWIN

TABLE OF CONTENTS

1. UPDATES - DECEMBER 2014

After extensive tests made within the IT, it was shown that the "create missing taxon" tool could lead to unforeseen errors. It was therefore decided that the creation of taxonomy had to be made in a previous step to the specimens import. A template for importing of taxonomy has therefore been prepared.

This template for taxonomy comes in an excel format. Each taxonomic level corresponds to a column. An export in xml format produces an xml file that can be imported directly into DaRWIN via an import interface for taxonomy, replacing the "Create missing taxon" button that is not available anymore.

For more information about this template, you can read the revised user manual for the template specimens and for the taxonomy import template.

2. INTRODUCTION

Since January 2014, DaRWIN allows you to import your data via a pre-formatted XML ABCD template. This tool is available for curators and encoders in the collections for which they have rights, when logged in DaRWIN (http://darwin.naturalsciences.be).

The size limit for an upload file is set at 20M.

You can access to this functionality by clicking on Import > Specimens in the Administration menu.

WARNING: The import tool has been developed to allow the transfer of large amount of new data into DaRWIN in a simpler manner. Nevertheless, the import tool was not designed for updating of specimens already present in DaRWIN. If you import a file with specimens already present in DaRWIN, duplicates will be created. Similarly, if you import 2 times the same file, duplicates will be created. Additionally, once you've uploaded a file within the import tool in DaRWIN, it is not possible anymore to edit specimens. If adaptations have to be implemented for some specimens in your uploaded file, you will have to make them in DaRWIN when import is over.
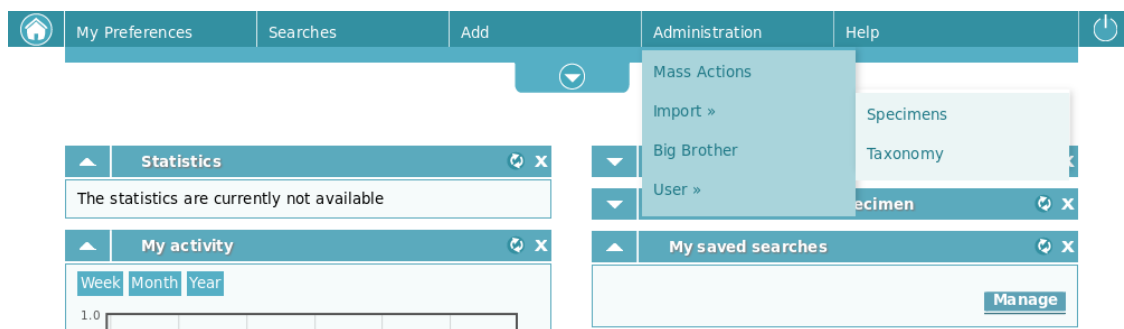
Figure 1 - DaRWIN Homepage, Administration menu

The Import page consists of a search engine for ongoing imports, as well as a summary table of imports. You can see imports (ongoing or finished) in collections on which you have edition rights, that is the files you downloaded and the files downloaded by other encoders as well.
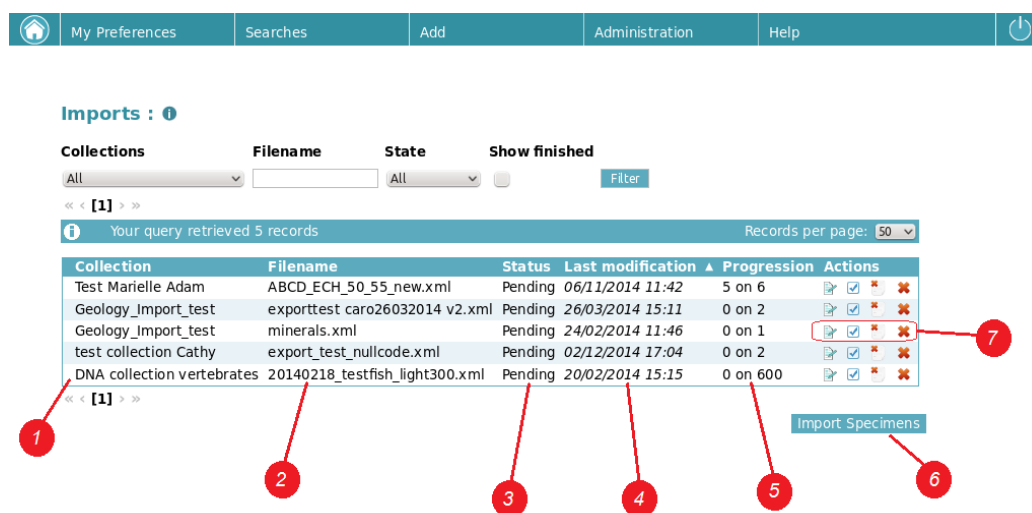

Figure 2 - Import page

Be aware that an uploaded file is not automatically imported into DaRWIN. Several steps are required before you are allowed to import your file into a collection.

Finished imports are not displayed by default in the table of imports. You can see them by selecting "Show finished" and clicking on the "Filter" button.

The table of imports contains several columns:

(1) **Collection**: the collection affiliated to the import file
    You must define it before the import process.
(2) **Filename:** the name of your uploaded xml file
(3) **Status:** actual status of your file
    More information about different status is available by clicking on the **info button (i)**. Note : Some status need an automatic update made by the system (such as processing or loading). In these cases, you have to wait until the status of your import change and actions are available at the end of the line.
(4) **Last modification:** date of the last modification made for this import process
(5) **Progression:** the number of already imported records in DaRWIN on the total number of records in your file
    In the example above, 6 records have already been imported while 994 records still waiting to be imported.
(7) **Actions buttons:** the four actions available for each import (if you put your cursor on a button, its name will appear)
    More information about available buttons are available by clicking on the **info button (i)**.

Below the table with summary of imports is the "**Import specimens**" button (6). This button allows you to download a new XML ABCD formatted file.

You can consequently edit a file, import lines without errors, cancel or delete an import. If you wish to read more information about the available actions, you can click on the **info button (i)** next to the title of the page. *See Appendix 1.*

Remember also that you can cancel/delete your import anytime during the process, but specimens already imported into DaRWIN will remain. If you cancel the import, the database will keep track of the import. If you delete the import, the database will not keep track of the import. During an import or a check, it is not possible to cancel/delete an import.

3. UPLOAD AN XML ABCD FORMATTED FILE

To import a new XML ABCD formatted file, click on the "**Import a file**" button (Figure 2).
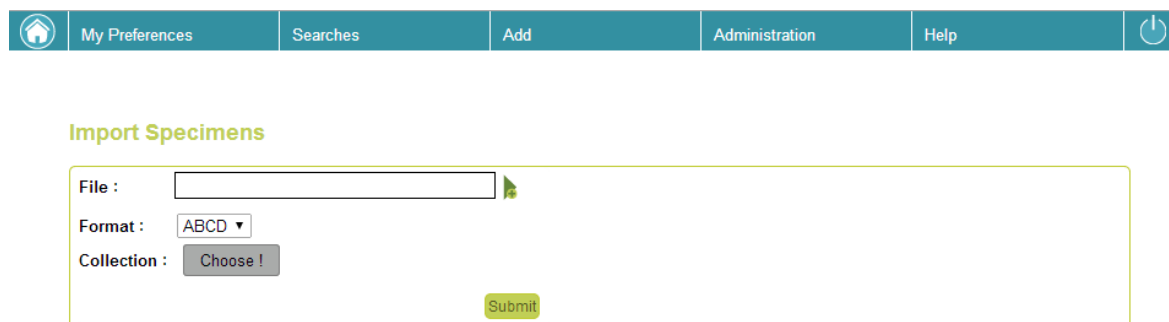

Figure 3 - Import a file

You are asked to select your file, the associated collection and the format of your xml file. The only available format is 'ABCD' standing for all ABCD(DNA-EFG) formatted files. In the list, the 'DNA' format is still displayed but it corresponds to the previous import developed for JEMU in the past.

When all information is filled in, click on the "**Submit**" button. You will automatically be redirected to the Import search page. Your file is now being loaded and checked. It is not possible to see the progression of the loading/checking of your file. The system checks every 15 minutes if there is a file to check or there are records to import into DaRWIN.

Note: If you wish to read more information about the state of your file visible in the import table, you can click on the **info button (i)** next to the title of the page. *See Appendix 1.*

4. CORRECT YOUR DATA FOR IMPORT

Your file has been downloaded but nothing has been imported into DaRWIN yet.

DaRWIN first checks your records one by one to look for possible problems (e.g., a people or taxon name not yet encoded in DaRWIN). When your file is in a *pending* state, DaRWIN has finished checking for problems and is waiting for you to correct remaining errors in your records. In order to correct every errors detected by DaRWIN, you can edit your import, by clicking on the edit button at the end of the line of your file.

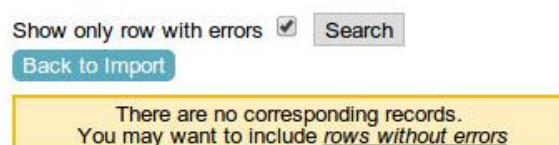### 4.1. First case: no problem was found (Figure 5)

Figure 4 - No errors in your file

Well done! Your file fits perfectly. Click on "**Import "OK" lines**" to import your data into DaRWIN.

4.2. Second case: some problems were detected (Figure 6)

Back to Import
Show only row with errors ☑
⦿ Zoology
◯ Geology

Search

« ‹ [1] › »

ⓘ Your query retrieved 6 records · Records per page: 50 ▾

| Actions	Error(s) found	Status	Linked Info	Codes	Category	Expedition	Sampling Location	Taxon.	I.G.	Acquisition	Part	Status	Institution	Number
📝 ✖	4	Error	37	addID AB1234567 code RBINS_AVES_00001 AVES	specimen	expédition à Nouméa	test_dna_20140908	Epinephelus mera submera	12345	donation	specimen		Royal Belgian Institute of natural Sciences	10
📝 ✖	3	Error	11	AB42715495 RBINS_AVES_00001_sample01		expédition à Nouméa	test_dna_20140908	Epinephelus mera submera			clip, blood		Royal Belgian Institute of natural Sciences	
📝 ✖	5	Error	20	BOLD Systems Accession number AB356604 AGB95530 RBINS_AVES_00004_sample02_DNA01		expédition à Nouméa	test_dna_20140908	Epinephelus mera submera			DNA extract		RMCA	
📝 ✖	3	Error	11	AB42715495 RBINS_AVES_00004_sample01		expédition à Nouméa	test_dna_20140908	Epinephelus mera submera			clip, blood		Royal Belgian Institute of natural Sciences	
📝 ✖	5	Error	20	BOLD Systems Accession number AB356604 AGB95531 RBINS_AVES_00001_sample01_DNA01		expédition à Nouméa	test_dna_20140908	Epinephelus mera submera			DNA extract		RMCA	
📝 ✖	5	Error	20	BOLD Systems Accession number AB356604 AGB95532 RBINS_AVES_00004_sample01_DNA01		expédition à Nouméa	test_dna_20140908	Epinephelus mera submera			DNA extract		RMCA	

Back to Import   Import "Ok" lines

Figure 5 – Summary of your records

When you edit your import, a page is displayed where all records in your file are listed. Above this table, you have a filter "Show only row with errors" that allows you to display only records with errors (by default) or all records (Figure 5). You can also choose between the "Zoology" or the "Geology" view depending on your data. The columns will vary depending on the view. Click on "Search" to validate your choice.

You can click on "Import 'OK' lines" anytime during the process to import lines without errors.

Some fields could be highlighted in the table:

- Orange highlighting means that DaRWIN detected a problem for this value
- Green highlighting means that DaRWIN recognized the value and found a related record in its database.

Remember that some lines can be flagged as containing error(s), even if you don't see orange cell in the summary: the error is in another field than those displayed in the table.

For each line you can decide to delete or edit information for this record.

When you want to edit, all detected errors are displayed for the record (Figure 6). Each field that was flagged during the system check is displayed and can be corrected, by searching corresponding DaRWIN catalogue. For people name (identifier and collector), you also have the possibility to delete the value for the record by clicking on the red cross next to the value. It could happen for example if the name is not explicit enough to be able to create a new corresponding people in the DaRWIN catalogues or if the value is not relevant.

Beside each unrecognized field, a small magnifier allows you to correct/precise your entry by browsing the DaRWIN Catalogues.

Figure 6 - Errors for one record

It is possible that an existing taxon/chrono/litho/mineral was not recognized during the check and is therefore displayed in errors.

If you use the template with only the "taxonFullName" for identification, it is probably because of a misspelling or because the taxon isn't present in the Taxonomy Catalogue in DaRWIN.

If you use the template with the various classification columns, it's probably because the specified taxonomical level or parent(s) you've referenced doesn't match what already is in the Taxonomy/Chronostratigraphy/Lithology/Mineralogy Catalogues of DaRWIN. Even when you've specified exactly the same name, if the level or the parent doesn't match, DaRWIN won't consider it as the same value. When a level is recognized, you can see its full taxonomical tree by clicking on the info button in front of the name and compare with the levels that were referenced in the template.

You can click on the **info button (i)** to see to full taxonomical tree for each recognized (green) level.

Figure 7 - Taxon error : taxonomy info

5. IMPORT YOUR DATA INTO DaRWIN

When no errors remain, you can import your data.

You could also have decided to import lines that were already corrected at anytime of the process. Indeed, once a record is corrected it can be imported into DaRWIN. You can therefore interrupt your work for a file, import "OK" lines and come back later to continue the process.

Remember that it is only once you have clicked on the **"Import "OK" lines"** button that the system will actually import your records into DaRWIN. Before this step, your records are stored in a temporary table and are not visible in DaRWIN.

When you click on "**Import "OK" lines**" button, the file status turns into "*Processing".* You won't be able to edit your file until this process is over.

If every line was imported, your file status will change to "*Finished"*. A finished file isn't editable anymore but is still displayed for information purpose.

6. APPENDIX

Appendix 1 – Imports info button (i)

<u>Help on available State</u>

To be loaded : This file is ready to be loaded, an automatic task will be activated to load lines.
Loaded : Your file has been loaded, but still need to be checked
Checking : Your file has been loaded and is being checked
Pending : Your file has been loaded and checked, you can edit line in errors or import corrects lines
Processing : Your 'Ok' lines are beeing imported in DaRWIN
Finished : This file has been completely imported in DaRWIN
Aborted : This file has been aborted. This line will remain for a limited time in the summary list just for information purposes only.
Error : Errors appeared during import, you can either check these errors with the error icon, and you can continue the import process or you can delete the entry and re-import your corrected file
Being processed : The file is actually being loaded in database

<u>Help on available button</u>

 : This button (only visible when your file is on Pending state) will allow you to edit or remove data
 : This button will cancel the importing of the file, your file will have the Aborted state, it will allow you to keep a trace of this import, even if the data is not imported
 : This will delete this line. Already imported lines will remain in the database. All other lines will be deleted without a trace.
 : This button will show you all errors that occurred during import, it's your choice to decide whether to continue the import process or to delete your import, correct the file and re-import it

Appendix 2 - List of supported ABCD(DNA) concepts in DaRWIN (needs to be updated for EFG extension) – in Datasets/Dataset/Units/Unit/:

- SourceInstitutionID
- SourceID
- UnitID
- UnitReferences/UnitReference/TitleCitation
- Identifications/Identification/Result/TaxonIdentified/HigherTaxa/HigherTaxon/HigherTaxonName
- Identifications/Identification/Result/TaxonIdentified/HigherTaxa/HigherTaxon/HigherTaxonRank
- Identifications/Identification/Result/TaxonIdentified/ScientificName/NameAtomised/Zoological/GenusOrMonomial
- Identifications/Identification/Result/TaxonIdentified/ScientificName/NameAtomised/Zoological/Subgenus
- Identifications/Identification/Result/TaxonIdentified/ScientificName/NameAtomised/Zoological/SpeciesEpithet
- Identifications/Identification/Result/TaxonIdentified/ScientificName/NameAtomised/Zoological/SubspeciesEpithet

- Identifications/Identification/Result/TaxonIdentified/ScientificName/NameAtomised/Zoological/AuthorTeamOriginalAndYear
- Identifications/Identification/Result/TaxonIdentified/ScientificName/NameAtomised/Zoological/AuthorTeamParenthesisAndYear
- Identifications/Identification/Result/TaxonIdentified/ScientificName/NameAtomised/Botanical/GenusOrMonomial
- Identifications/Identification/Result/TaxonIdentified/ScientificName/NameAtomised/Botanical/FirstEpithet
- Identifications/Identification/Result/TaxonIdentified/ScientificName/NameAtomised/Botanical/InfraspecificEpithet
- Identifications/Identification/Result/TaxonIdentified/ScientificName/NameAtomised/Botanical/AuthorTeamParenthesis
- Identifications/Identification/Result/TaxonIdentified/ScientificName/NameAtomised/Botanical/AuthorTeam
- Identifications/Identification/Result/TaxonIdentified/ScientificName/NameAddendum
- Identifications/Identification/Result/TaxonIdentified/NameComments
- Identifications/Identification/Identifiers/Identifier/PersonName/FullName
- Identifications/Identification/References/Reference/TitleCitation
- Identifications/Identification/Date/ISODateTimeBegin
- Identifications/Identification/Method
- Identifications/Identification/Notes
- Identifications/IdentificationHistory
- Identifications/Identification/Result/Extension/efg:MineralRockIdentified/efg:MineralRockGroup/efg:MineralRockGroupName
- Identifications/Identification/Result/Extension/efg:MineralRockIdentified/efg:MineralRockGroup/efg:MineralRockClassification
- Identifications/Identification/Result/Extension/efg:MineralRockIdentified/efg:ClassifiedName/efg:FullScientificNameString
- Identifications/Identification/Result/Extension/efg:MineralRockIdentified/efg:ClassifiedName/ efg:MineralRockNameAtomised/efg:ScientificNameString
- Identifications/Identification/Result/Extension/efg:MineralRockIdentified/efg:InformalNameString
- RecordBasis
- KindOfUnit
- SpecimenUnit/Acquisition/AcquisitionDate
- SpecimenUnit/Acquisition/AcquisitionType
- SpecimenUnit/Acquisition/AcquiredFrom/Organisation/Name/Representation/Text
- SpecimenUnit/Acquisition/AcquiredFrom/Person/FullName
- SpecimenUnit/Accessions/AccessionCatalogue
- SpecimenUnit/Accessions/AccessionNumber
- SpecimenUnit/Preparations/Preparation/PreparationType
- SpecimenUnit/Preparations/Preparation/PreparationMaterials
- SpecimenUnit/NomenclaturalTypeDesignations/NomenclaturalTypeDesignation/TypeStatus
- ZoologicalUnit/PhasesOrStages/PhaseOrStage
- MultiMediaObjects/MultiMediaObject/FileURI
- Associations/UnitAssociation/AssociatedUnitSourceInstitutionCode
- Associations/UnitAssociation/AssociatedUnitSourceName
- Associations/UnitAssociation/AssociatedUnitID
- Associations/UnitAssociation/AssociationType
- Gathering/Code
- Gathering/DateTime/DateTime/ISODateTimeBegin
- Gathering/DateTime/DateTime/TimeOfDayBegin
- Gathering/DateTime/DateTime/ISODateTimeEnd
- Gathering/DateTime/DateTime/TimeOfDayEnd
- Gathering/Agents/GatheringAgent/Person/Fullname
- Gathering/Project/ProjectTitle
- Gathering/Method
- Gathering/LocalityText
- Gathering/NamedAreas/NamedArea/AreaClass
- Gathering/NamedAreas/NamedArea/AreaName
- Gathering/SiteCoordinatsSets/SiteCoordinates/CoordinatesLatLong/LongitudeDecimal
- Gathering/SiteCoordinatsSets/SiteCoordinates/CoordinatesLatLong/LatitudeDecimal
- Gathering/Altitude/MeasurementOrFactAtomised/LowerValue
- Gathering/Altitude/MeasurementOrFactAtomised/UnitOfMeasurement
- Gathering/Depth/MeasurementOrFactAtomised/LowerValue
- Gathering/Depth/MeasurementOrFactAtomised/UnitOfMeasurement
- Gathering/SiteMeasurementsOrFacts/SiteProperty/MeasurementOrFactAtomised/Parameter
- Gathering/SiteMeasurementsOrFacts/SiteProperty/MeasurementOrFactAtomised/LowerValue
- Gathering/SiteMeasurementsOrFacts/SiteProperty/MeasurementOrFactAtomised/UnitOfMeasurement
- Gathering/Biotope/Text

- Gathering/Notes
- MeasurementsOrFacts/MeasurementOrFact/MeasurementOrFactAtomised/Parameter
- MeasurementsOrFacts/MeasurementOrFact/MeasurementOrFactAtomised/LowerValue
- MeasurementsOrFacts/MeasurementOrFact/MeasurementOrFactAtomised/UnitOfMeasurement
- Sex
- Notes
- RecordURI
- UnitExtension/efg:EarthScienceSpecimen/efg:UnitStratigraphicDetermination/efg:LithostratigraphicAttributions/efg:LithostratigraphicAttribution/efg:Group
- UnitExtension/efg:EarthScienceSpecimen/efg:UnitStratigraphicDetermination/efg:LithostratigraphicAttributions/efg:LithostratigraphicAttribution/efg:Formation
- UnitExtension/efg:EarthScienceSpecimen/efg:UnitStratigraphicDetermination/efg:LithostratigraphicAttributions/efg:LithostratigraphicAttribution/efg:Member
- UnitExtension/efg:EarthScienceSpecimen/efg:UnitStratigraphicDetermination/efg:LithostratigraphicAttributions/efg:LithostratigraphicAttribution/efg:Bed
- UnitExtension/efg:EarthScienceSpecimen/efg:UnitStratigraphicDetermination/efg:LithostratigraphicAttributions/efg:LithostratigraphicAttribution/efg:InformalLithostratigraphicName
- UnitExtension/efg:RockUnit/efg:RockType
- UnitExtension/efg:RockUnit/efg:RockPhysicalCharacteristics/efg:RockPhysicalCharacteristic/MeasurementOrFactAtomised/…
- UnitExtension/efg:RockUnit/efg:Petrology/efg:PetrologyDescriptiveText
- UnitExtension/efg:MineralogicalUnit/efg:MineralMeasurementsOrFacts/efg:MineralMeasurementOrFact
- UnitExtension/efg:MineralogicalUnit/efg:MineralDescriptionText
- UnitExtension/storage:Storage/Localisation/Institution
- UnitExtension/storage:Storage/Localisation/Building
- UnitExtension/storage:Storage/Localisation/Floor
- UnitExtension/storage:Storage/Localisation/Room
- UnitExtension/storage:Storage/Localisation/Row
- UnitExtension/storage:Storage/Localisation/Column
- UnitExtension/storage:Storage/Localisation/Shelf
- UnitExtension/storage:Storage/Localisation/Box
- UnitExtension/storage:Storage/Localisation/Rack
- UnitExtension/storage:Storage/Localisation/Tube
- UnitExtension/storage:Storage/Localisation/Position
- UnitExtension/storage:Storage/Localisation/Barcode
- UnitExtension/storage:Storage/storage:Codes/storage:Code/storage:Type
- UnitExtension/storage:Storage/storage:Codes/storage:Code/storage:Value
- UnitExtension/dna:DNASample/dna:Tissue
- UnitExtension/dna:DNASample/dna:Preservation
- UnitExtension/dna:DNASample/dna:ExtractionStaff
- UnitExtension/dna:DNASample/dna:DNAExtractionDate
- UnitExtension/dna:DNASample/dna:DNAExtractionMethod
- UnitExtension/dna:DNASample/dna:RatioOfAbsorbance260_280
- UnitExtension/dna:DNASample/dna:Concentration
- UnitExtension/dna:DNASample/dna:Amplifications/dna:Amplification/dna:Sequencings/dna:Sequencing/dna:GenBankNumber