

FADA Excel data injection process

Prerequisites

After having validated the uploaded data I preferred to create a resume of the contents of the taxa information in order to simplify a number of stages during the injection. So I created a table `importandsyncfada.minitaxa_table` which contains such a resume for each group that will be processed. After all has been validated all records for a specific group are deleted and new records are created. See file `prepare_minitaxa_table` for the two sql lines used for this.

How we got about it

Find all the authors in the taxa information and add them to the authors table if they are not already present. Unlink publications from group
Unlink `biofresh_key_species` Delete any group element from tables `gpreferences` `groups_taxons` `species_taxons` Unlink `biofresh_key_taxons`
Unlink `biofresh_key_synonyms` Unlink any Synonyms from group

Processing families

Find author using name. Try to find a family element in the information provided by COL (`col.families`) The way to search information is

```
SELECT distinct kingdom, phylum, cclass, oorder
FROM col.families
WHERE family = theFamilyName;
```

If nothing is found the `parent_id` of this family will be null (it happens). If a line is found check that you can find the elements in `fada.taxons`. If they are not found create them. Create and connect the family to the lowest ranking element that you have created/found. When creating the taxon element take out any trailing alphabetic sign from the year part. Convert '1952a' to '1952'

Processing subfamilies

Get the author by name. Find the family type taxon object the subfamily must be related to (you know the group and name of the family and the rank) Make sure you have a "get out procedure" and an error message in case there were several families with the same name in the group (you never know). If one single family has been found, create and connect the subfamily to the family. When creating the taxon element take out any trailing alphabetic sign from the year part. Convert '1952a' to '1952'

Processing tribe, subtribe, genus, subgenus, speciesgroup

Same process as subfamilies but for lower levels.

Processing original_genus, declension_species, syn_original_genus

We will now use

```
taxa_table
```

and a search condition.

```
group_id=theGroup AND (parentheses='Y' OR syn_original_genus IS NOT NULL)
```

When the `original_genus` is blank, use the genus value of the name as original genus. Look for this (name,group,rank) in the taxons. If it is not found create it without a link to the higher element. If `declension_species` has a value try to find it in the taxons. If it is not found create it with a link to the `original_genus`. If `syn_original_genus` is not blank Check that for a "Genus (Subgenus)" value using `/^[A-Z][a-z]\s([A-Z][a-z]*)$/`

Processing species subspecies

Got through the `taxa_table` using condition

```
group_id=theGroup AND rank=theRank AND syn_genus_subgenus IS NULL AND syn_species_subspecies IS NULL
```

where `theRank` will be 'SPECIES' or 'SUBSPECIES'

Find the taxon instance parent Create parameters for the new taxon instance Create the new taxon instance of "SPECIES", "SUBSPECIES" level and link it to it's parent taxon Create the new species instance. You have to look for the already created `original_genus` or `declension_species` taxons, add the parentheses and calculate the `scientific_name` value

Undating biofresh key species

Updating biofresh_key_species

Relink to existing elements using scientific_name Delete unlinked elements Add missing elements, just adding the result of a select with an MINUS/EXCEPT operation

Processing synonyms

Go through the taxa_table using condition

```
group_id=theGroup AND
NOT (syn_genus_subgenus IS NULL AND
syn_species_subspecies IS NULL AND
syn_original_genus IS NULL) AND
syn_authors IS NOT NULL AND
syn_date IS NOT NULL
```

Find the taxon instance to which the synonym will be linked to. Look for the synonym. Remember, the genus_species_name is a concatenation of genus, species separated by a tab character, take out a trailing alpha character in the the year of creation If an instance is found link the synonym to it's taxon instance. If not create it and link. After having processed the table. Delete all the synonyms with no links.

Processing group_taxons

The point of group_taxons is to find a taxon instance that is the root of the group tree. So it should be the taxon element with a group number which has the smallest rank_id. Unfortunately it's not always that easy and you can go a little higher to find a suitable taxon instance. So basically start with the taxon instances of this group with the smallest rank_id. Then from there, find one that would top them all. This is the one you want to put in the group_taxons table.

Processing distribution info.

Loop through table distr_table. Use the info to find the species instance that must be completed. Link the species instance to the regions. Be carefull, now the regions also contain the TDWG level regions. and the regions table is now regions_species_all

Linking back biofresh_key_taxons, biofresh_key_synonyms

Linking biofresh_taxons is done through a hierarchical walk of taxons from group roots to species, subspecies If a taxon is found it is linked. If not it is eliminated at the end. Completely new taxons are added as a result of a select with MINUS/EXCEPT operation Not to forget there is a step to take into account original_genus and declension species elements.

Linking back biofresh synonyms