EDIT - task #10232

Import Finnish common names

01/23/2023 03:48 PM - Andreas Müller

Status:	Closed	**Start date:**	
Priority:	Highest	**Due date:**	
Assignee:	Andreas Müller	**% Done:**	100%
Category:	cdmadapter	**Estimated time:**	0:00 hour
Target version:			
Severity:	normal		

Description

File to be found at:

ERS:

ja genau, in erster Linie geht es bei dieser Liste nur um den Common Name Import, da hatte ich mich nicht ganz klar ausgedrückt. Andere Daten aus dieser Liste zu importieren/abzugleichen ist auch denkbar, aber jetzt nicht prioritär.

AM:

also es geht um die Spalten FINNISH_NAME und SWEDISH_NAME in der Tabelle „All“ nehme ich dann mal an.
Ein Common Name besteht ja i.d.R. aus Name, Sprache und Areal, in dem der Name Verwendung findet (z.B. Deutschland, Schweiz, …). Für die finnischen Namen nehme ich da sicherlich Finnland, für die schwedischen auch? Oder Schweden, beides, gar nichts?

ERS:

ja, um diese Namen geht es. Finnland ist klar. Zu den schwedischen Namen heißt es:

„Almost exclusively they correspond to Swedish vernacular names used in Sweden. However, dozens of plants in Finland have traditionally been called by local Swedish vernacular names, which deviate from those used in Sweden.”

Welche das sind, müsste man aber im Einzelfall klären – wenn überhaupt, machen wir das im Nachgang. Bis dahin würde ich für die schwedischen Namen beides – also Schweden und Finnland – angeben.

AM:

ok. Es gibt 2 Möglichkeiten, die 2 Areale zu realisieren. Einem einzigen Common Name Datensatz kann man (derzeit) nicht 2 Areale zu weisen.
Es geht nur, entweder 2 Datensätze zu erzeugen, die sich nur im Areal unterscheiden oder ein Areal „Sweden and Finnland“ anzulegen, was semantisch beide Areale umfasst.
Ich tendiere stark zu der zweiten Lösung, da sie datensparsamer ist und semantisch auch exakter, da wir bislang ja nur für das Gesamtareal eine Aussage machen, nicht für jedes Areal einzeln.

Würdest du da zustimmen?

ERS:

ja, sehr gut. Einverstanden.

====

Source: Kurtto, A., Lampinen, R., Piirainen, M. & al., Checklist of the vascular plants of Finland. Suomen putkilokasvien luettelo. 2019

Related issues:

Related to EDIT - task #6137: Urgent imports	**In Progress**

Associated revisions

Revision d5797bc3 - 01/24/2023 11:28 PM - Andreas Müller

ref #10232 add Finnish common name activator

Revision 33078d01 - 01/27/2023 08:41 AM - Andreas Müller

ref #10232 first implementation for E+M finnish common name import

Revision 051d16ee - 01/30/2023 10:04 AM - Andreas Müller

ref #10232 add Finnish common name activator (cont.)

Revision 077baea5 - 01/30/2023 03:08 PM - Andreas Müller

ref #10232 finalize CommonNameExcelImport and some cleanup

Revision 5d83c972 - 02/01/2023 07:55 PM - Andreas Müller

ref #10232 fix issue in common name export

History

#1 - 01/23/2023 04:33 PM - Andreas Müller

- Description updated

#2 - 01/23/2023 05:29 PM - Andreas Müller

- Description updated

#3 - 01/23/2023 09:01 PM - Andreas Müller

- Related to task #6137: Urgent imports added

#4 - 01/24/2023 10:15 PM - Andreas Müller

- Subject changed from Import finnish common names to Import Finnish common names

#5 - 01/27/2023 04:56 PM - Andreas Müller

- % Done changed from 0 to 60

AM:

ich habe jetzt den Commons Name Import für die Finnischen Daten weitgehend fertig.
Das Hauptproblem ist dabei das Namematching. Anbei eine Logdatei eines Testimports anhand der man sehen kann, welche Namen nicht 100% gematched haben.

Es gibt dabei v.a. 3 Fälle:

1.) No match: es konnte auf Anhieb kein Taxon gefunden werden, welches matched; der Common Name wurde ausgelassen. Häufig sind das Hybride, aber auch andere Taxa, die so in E+M nicht vorkommen.
2.) differs in author: ein exakter Match konnte lediglich auf dem Namen ohne Autor gefunden werden. Der Common Name wurde an diesen Namen gehängt, auch wenn der Autor nicht stimmt. Meist sind das zusätzliche Autoren (z.B. ex-Autoren) oder andere Schreibweisen.
Es könnten aber auch Homonyme darunter sein. Das habe ich nicht näher untersucht.
Der Match kann hier auch für ein Synonym stattgefunden haben.
In 2 Fällen wurde mehr als ein match gefunden, da müsste geprüft werden, ob die Zuordnung stimmt.
3.) synonym match: es wurde ein exakter Match in E+M gefunden, aber für ein Synonym. Der Common Name wurde an das akzeptierte Taxon des Synonyms gehängt. Ein einigen Fällen ist es auch kein Synonym sondern ein fehlender Cache-Update, der zu dieser Einstufung führte.

Kannst du mal drüber schauen, ob das für dich so stimmt oder ob im Einzelfall etwas anders gemacht werden soll? Wir können dann entscheiden, ob wir das vor oder nach dem Import korrigieren

ERS:

für mich sieht das alles sehr logisch und korrekt aus.

1. die no match-Fälle sind ja vor allem Hybriden und Varietäten. Die übrigen sind nützlich, weil ich sie möglicherweise in E+M ergänzen muss, meistens sind es introduced taxa, die noch fehlen.
2. das ist auch sinnvoll, oft sind es ex-Autoren oder mehr als zwei. Homonyme müsste man irgendwie finden können.
3. Synonyme ok, aber wie behandelt man die Fälle mit fehlendem Cache-Update? Letzteren Punkt verstehe ich nicht so ganz.

#6 - 01/27/2023 06:19 PM - Andreas Müller

Andreas Müller wrote in #note-5:

> AM:
>
> ich habe jetzt den Commons Name Import für die Finnischen Daten weitgehend fertig.
> Das Hauptproblem ist dabei das Namematching. Anbei eine Logdatei eines Testimports anhand der man sehen kann, welche Namen nicht

100% gematched haben.

Es gibt dabei v.a. 3 Fälle:

1.) No match: es konnte auf Anhieb kein Taxon gefunden werden, welches matched; der Common Name wurde ausgelassen. Häufig sind das Hybride, aber auch andere Taxa, die so in E+M nicht vorkommen.
2.) differs in author: ein exakter Match konnte lediglich auf dem Namen ohne Autor gefunden werden. Der Common Name wurde an diesen Namen gehängt, auch wenn der Autor nicht stimmt. Meist sind das zusätzliche Autoren (z.B. ex-Autoren) oder andere Schreibweisen.
Es könnten aber auch Homonyme darunter sein. Das habe ich nicht näher untersucht.
Der Match kann hier auch für ein Synonym stattgefunden haben.
In 2 Fällen wurde mehr als ein match gefunden, da müsste geprüft werden, ob die Zuordnung stimmt.
3.) synonym match: es wurde ein exakter Match in E+M gefunden, aber für ein Synonym. Der Common Name wurde an das akzeptierte Taxon des Synonyms gehängt. Ein einigen Fällen ist es auch kein Synonym sondern ein fehlender Cache-Update, der zu dieser Einstufung führte.

Kannst du mal drüber schauen, ob das für dich so stimmt oder ob im Einzelfall etwas anders gemacht werden soll? Wir können dann entscheiden, ob wir das vor oder nach dem Import korrigieren

ERS:

für mich sieht das alles sehr logisch und korrekt aus.

1. die no match-Fälle sind ja vor allem Hybriden und Varietäten. Die übrigen sind nützlich, weil ich sie möglicherweise in E+M ergänzen muss, meistens sind es introduced taxa, die noch fehlen.
2. das ist auch sinnvoll, oft sind es ex-Autoren oder mehr als zwei. Homonyme müsste man irgendwie finden können.
3. Synonyme ok, aber wie behandelt man die Fälle mit fehlendem Cache-Update? Letzteren Punkt verstehe ich nicht so ganz.

AM:

ok. Die Frage ist, wie es weiter geht.

1. Scheint von meiner Seite aus bearbeitet. Die restliche Auswertung des Log-Files liegt bei, korrekt?
2. Was meinst du mit „Homonyme müsste man irgendwie finden können“? Soll das noch vor dem Import geschehen? Und wer sollte es machen?
3. Den Cache Updater Fall schaue ich mir an. Ich denke aber nicht, dass das kritische Fälle sind. In dem Kontext schaffe ich es vielleicht auch noch, die Caches nochmal komplett zu updaten. Das steht bei E+M noch aus, da es da Out-Of-Memory Probleme gab.

Wichtig wäre jetzt, ob noch irgendwas geklärt werden muss, bevor wir den Import starten.

#7 - 01/30/2023 12:20 PM - Andreas Müller

ERS:

aus meiner Sicht kann es losgehen, scheint alles geklärt. Die wenigen eventuellen Homonyme kann ich sicher nach dem Import finden, die Zahl dürfte minimal sein oder gegen Null gehen

#8 - 01/30/2023 02:39 PM - Andreas Müller

- Status changed from New to Resolved

- Priority changed from New to Highest

- % Done changed from 60 to 80

We still need to run the final import for the Swedish names. Afterwards, the ticket can be closed.

#9 - 02/01/2023 10:29 PM - Andreas Müller

- Status changed from Resolved to Closed

- % Done changed from 80 to 100

I guess this is fully implemented. The open issues that common names are not found in (advanced) search does not seem to be related to the import.

#10 - 02/03/2023 09:13 PM - Andreas Müller

Maybe we still want to set the preferred flag.

#11 - 03/20/2023 03:37 PM - Andreas Müller

- Target version changed from Release 5.44 to Release 5.36

#12 - 03/22/2023 02:54 PM - Andreas Müller

- Target version deleted (Release 5.36)