

EDIT - task #9501

Find inconsistent higher area distribution status data in E+M

03/02/2021 07:03 PM - Andreas Müller

Status:	New	Start date:	
Priority:	New	Due date:	
Assignee:	Andreas Müller	% Done:	0%
Category:	data	Estimated time:	0:00 hour
Target version:	Euro+Med post migration		
Severity:	normal		
Description			
The distribution status of a higher area (area with subareas) should usually not have a higher status then the highest status of it's subareas if subarea information exists.			
This can be found by the following sql			
<pre>SELECT n.titleCache nameTitle, n.nameCache, a.titleCache area, a.idInVocabulary areaId, st.titleCache statusTitle, sec.titleCache sec FROM DescriptionElementBase deb INNER JOIN DescriptionBase db ON db.id = deb.inDescription_id INNER JOIN TaxonBase tb ON tb.id = db.taxon_id INNER JOIN TaxonName n ON n.id = tb.name_id INNER JOIN DefinedTermBase a ON a.id = deb.area_id INNER JOIN DefinedTermBase st ON st.id = deb.status_id INNER JOIN Reference sec ON sec.id = tb.sec_id WHERE tb.publish = 1 AND deb.DTYPE = 'Distribution' AND EXISTS (SELECT * FROM DescriptionElementBase deb2 INNER JOIN DescriptionBase db2 ON db2.id = deb2.inDescription_id INNER JOIN DefinedTermBase a2 ON a2.id = deb2.area_id LEFT JOIN DefinedTermBase st2 ON st2.id = deb2.status_id WHERE a.id = a2.partOf_id AND tb.id = db2.taxon_id AND st.orderIndex < st2.orderIndex AND st2.id <> 2409) AND NOT EXISTS (SELECT * FROM DescriptionElementBase deb3 INNER JOIN DescriptionBase db3 ON db3.id = deb3.inDescription_id INNER JOIN DefinedTermBase a3 ON a3.id = deb3.area_id LEFT JOIN DefinedTermBase st3 ON st3.id = deb3.status_id WHERE a.id = a3.partOf_id AND tb.id = db3.taxon_id AND st.orderIndex >= st3.orderIndex) AND a.id <> 2269 -- EM area ORDER BY n.titleCache, a.titleCache</pre>			
For areas with status "undefined" change to "st2.id = 2409"			
===			
other data cleaning issues concerning status:			
<ul style="list-style-type: none">undefined: #4822#note-22			
Related issues:			
Related to EDIT - bug #8297: Fix condensed distribution string for E+M		Closed	
Related to EDIT - feature request #9500: Allow removing certain distribution ...		Feedback	
Related to EDIT - task #8671: Distribution in E+M (BM) on different levels		Rejected	

History

#1 - 03/02/2021 07:03 PM - Andreas Müller

The results of the query were sent to ERS today.

#2 - 03/02/2021 07:04 PM - Andreas Müller

- Related to bug #8297: Fix condensed distribution string for E+M added

#3 - 03/02/2021 07:04 PM - Andreas Müller

- Related to feature request #9500: Allow removing certain distribution status from distribution publication added

#4 - 03/02/2021 07:34 PM - Andreas Müller

- Related to task #8671: Distribution in E+M (BM) on different levels added

#5 - 03/02/2021 08:12 PM - Andreas Müller

AM:

Ein ähnlicher Fall, der mir nicht ganz klar ist, ist der, wenn es für das Überareal einen Status mit höherer Priorität gibt, als für alle Unterareale zusammen. Beispiel: *Prunus domestica* (http://www.europusmed.org/cdm_dataportal/taxon/1abe7c21-314a-4148-97d0-1d7fe617ca57) für die Balearen, mit introduced: doubtfully introduced (perhaps cultivated only) für das Gesamteareal, während alle Unterareale lediglich cultivated haben. Solche Fälle mit höherem Gesamtstatus gibt es derzeit 1575 Fälle. Soll dieser höhere Status dann trotzdem vernachlässigt werden? Oder sind das alle Fälle für die Datenreinigung bzw. müssen wir mit diesen Inkonsistenzen leben, weil die Statusangaben eben mit vielen Fragezeichen behaftet sind?

Ich hänge jedenfalls mal eine Übersicht an, um welche Fälle es sich handelt, sowohl für die undefined, als auch für die defined Fälle.

#6 - 03/07/2021 11:45 AM - Andreas Müller

- Description updated