

## EDIT - bug #4134

feature request # 2625 (Closed): [E+M Overview] Data aggregation functionalities for E+M (TransmissionEngine)

### Transmissionengine Distribution seems to miss distributions for higher Taxa

03/14/2014 10:53 AM - Andreas Kohlbecker

<b>Status:</b>	Closed	<b>Start date:</b>	
<b>Priority:</b>	Highest	<b>Due date:</b>	
<b>Assignee:</b>	Andreas Müller	<b>% Done:</b>	100%
<b>Category:</b>	cdmlib	<b>Estimated time:</b>	0:00 hour
<b>Target version:</b>	Release 5.12	<b>Found in Version:</b>	
<b>Severity:</b>	normal		
<b>Description</b>			
reported by Norbert:			
wie kommt es eigentlich, dass für das Taxon "Cichorieae" die Transmission Engine Distribution nur so extrem fragmentarischen Inhalt hat ( <a href="http://wp6-cichorieae.e-taxonomy.eu/portal/cdm_dataportal/taxon/c34eb662-6bba-4ad7-b84e-dd98f2fd9788">http://wp6-cichorieae.e-taxonomy.eu/portal/cdm_dataportal/taxon/c34eb662-6bba-4ad7-b84e-dd98f2fd9788</a> )??			
Was im Portal angezeigt wird, entspricht dem tatsächlichen Inhalt des features "Transmission Engine Distribution ".			
This may also affect the euro+med project!			

#### History

##### #1 - 08/18/2014 12:15 PM - Andreas Kohlbecker

- Keywords set to Euro+Med,Migration

##### #2 - 08/18/2014 02:09 PM - Andreas Kohlbecker

- Subject changed from Transmissionengine Ditrubution seems to miss distributions for higher Taxa to Transmissionengine Distrubution seems to miss distributions for higher Taxa

##### #3 - 06/29/2015 12:05 PM - Andreas Kohlbecker

- Subject changed from Transmissionengine Distrubution seems to miss distributions for higher Taxa to Transmissionengine Distribution seems to miss distributions for higher Taxa

##### #4 - 06/30/2015 11:27 AM - Andreas Kohlbecker

- Status changed from New to In Progress

- Target version changed from Euro+Med Migration to Release 3.8

##### #5 - 09/01/2015 12:08 PM - Andreas Kohlbecker

waiting for [#5176](#) to be solved before starting with this issue.

##### #6 - 09/01/2015 12:22 PM - Andreas Kohlbecker

- Keywords changed from Euro+Med,Migration to Euro+Med,Migration,TransmissionEngineDistribution

##### #7 - 09/02/2015 03:25 PM - Andreas Kohlbecker

- Status changed from In Progress to Worksforme

- Resolution set to worksforme

- % Done set to 0

- Target version deleted (Release 3.8)

wie kommt es eigentlich, dass für das Taxon "Cichorieae" die Transmission Engine Distribution nur so extrem fragmentarischen Inhalt hat ( [http://wp6-cichorieae.e-taxonomy.eu/portal/cdm\\_dataportal/taxon/c34eb662-6bba-4ad7-b84e-dd98f2fd9788](http://wp6-cichorieae.e-taxonomy.eu/portal/cdm_dataportal/taxon/c34eb662-6bba-4ad7-b84e-dd98f2fd9788))

Derzeit gibt es für diese Taxon keinerlei DescriptionElements und insofern auch keine Verbreitungsangaben.

Die Aggregation über die Rangstufen wurde nur für die Rangstufen UNRANKED\_INFRASPECIFIC bis GENUS durchgeführt. Dies entspricht exakt dem Verhalten der originalen [BerlinModel](#) TransmissionEngine. Hier ein Zitat aus dem Code: "" 6 rank levels to be calculated between microspecies and genus ...\_"

Bei der Aggregation über die Area-Hierarchie werden dagegen alle Taxa berücksichtigt, Daher hat das Taxon [Cichorieae Lam. & DC.](#) eine [TransmissionEngineDistribution](#) die aber leer ist da dieses Taxon keine editierten Verbreitungsangaben enthält.

Das original taxon in der [BerlinModel](#) datenbank hat ebenfalls keine Verbreitungsangaben  
<http://www2.bgbm.org/euroPlusMed/PTaxonDetail.asp?NameCache=Cichorieae&PTRefFk=7000000>

Insofern ist offensichtlich alles korrekt.

**#8 - 09/03/2015 08:33 AM - Andreas Kohlbecker**

- Status changed from *Workforme* to *Feedback*

- Resolution deleted (*workforme*)

- % Done changed from 0 to 0

Hallo Andreas,

das verstehe ich nicht ganz: OK, [http://cichorieae.e-taxonomy.net/portal/cdm\\_dataportal/taxon/c34eb662-6bba-4ad7-b84e-dd98f2fd9788](http://cichorieae.e-taxonomy.net/portal/cdm_dataportal/taxon/c34eb662-6bba-4ad7-b84e-dd98f2fd9788) hat keine eigene Distribution, aber die Transmission-Engine-Distribution ist ja nicht leer, sondern hat 8 Einträge, die ja auch angezeigt werden.

Wie auch immer, heißt dies, dass wenn dieses Taxon eine editierte Distribution hat, dann würden die volle Aggregation über die Area-Hierarchie laufen?

Beste Grüße,

Norbert

**#9 - 09/03/2015 09:56 AM - Andreas Kohlbecker**

Hi Norbert,

< das verstehe ich nicht ganz: OK, hat keine eigene Distribution, aber die Transmission-Engine-Distribution ist ja nicht leer, sondern hat 8 Einträge, die ja auch angezeigt werden.

das ist tatsächlich etwas seltsam. Ich kann mir das nur so erklären, dass die Transmission-Engine-Distribution für die Cichorieae wohl doch so gestartet worden ist, dass sie über alle Rangstufen läuft. Denn bei Euro+Med, wo der die Standarteinstellung (UNRANKED\_INFRASPECIFIC bis GENUS) benutzt worden ist ist alles wie erwartet.

... wenn dieses Taxon eine editierte Distribution hat, dann würden die volle Aggregation über die Area-Hierarchie laufen?

Genau so ist es, die Verbreitungsangaben zu den Arealen werden derzeit für jedes Taxon aufaggregiert, das editierte Verbreitungsangaben hat. Dieser Aggregationsschritt läuft zuerst. Danach startet die Aggregation über die Taxon-Hierarchie. Bei diesem Schritt werden sowohl editierte als auch berechnete Angaben berücksichtigt, wobei die berechneten Vorrang haben.

Der erste Aggregationsschritt über die Areal-Hierarchie wird für alle Taxa durchgeführt. Der zweite Schritt kann auf ein bestimmtes Rangintervall beschränkt werden, er bezieht also nur diese Taxa die in diesem Intervall liegen. Dies entspricht exakt dem Verhalten der Transmission-Engine im Berlin-Model. Ich fände es logischer und auch das Ergebnis der Aggregation wäre konsistenter wenn die Beschränkung auf bestimmte Taxa auch für die Aggregation über die Areal-Hierarchie gelten würde. Ansonsten gibt es Taxa beider die Verbreitungsangaben der Unter-Taxa mir einbezogen wurde und bei anderen wurde nur die Areal-hierarchie berücksichtigt. Mir leuchtet nicht ein warum man so etwas wollen kann.

Andreas

**#10 - 09/28/2015 02:07 PM - Andreas Kohlbecker**

Hallo Andreas,

ich finde es logisch, was Du schreibst und sehe das genauso. Ehrlich gesagt hätte ich auch erwartet, dass sowohl die Areal-Aggregation als auch die Taxa-Aggregation nur bis zu bestimmten Rangstufen durchgeführt werden. Dass dies nicht der Fall ist, wundert mich etwas. Andererseits: in Euro+Med hatten wir editierte Verbreitungsangaben grundsätzlich nur für Unterarten und Arten, da dürfte es also keine Rolle spielen, oder?

Schönen Gruß, Eckhard

**#11 - 09/29/2015 02:50 PM - Andreas Kohlbecker**

- Status changed from *Feedback* to *Closed*

- Resolution set to *fixed*

- % Done changed from 0 to 100

Hallo,

ich habe mittlerweile angefangen den code für die Transmissionengine so umzuschreiben, dass alle Aggregationen nur die Taxa berücksichtigen die vom definierten Rang-Intervall erfasst werden. Dies führt nicht nur zu einer besseren Konsistenz sondern verbessert auch die Performance, daher sind all diese commits im ticket [#5177](#) (improve performance of TransmissionEngineDistribution) zu finden.

Die Frage die zur Eröffnung dieses Tickets geführt hat ist beantwortet, also werde ich das ticket schließen. Die Antwort nochmal in Kurzform: Die TransmissionEngine hat keine Taxa mit höheren Rängen ausgelassen, sondern ein Aggregationsschritt war nicht auf die Taxa im Rangintervall beschränkt, daher die scheinbar unvollständigen Distributions.

Andreas

**#12 - 05/10/2017 01:39 PM - Andreas Kohlbecker**

- Private changed from Yes to No

**#13 - 11/08/2019 11:43 PM - Andreas Müller**

- Description updated

- Status changed from Closed to Feedback

- Assignee changed from Andreas Kohlbecker to Norbert Kilian

- Target version set to Release 5.12

Das ist für mich nicht logisch. Auch wenn die Aggregation über die Areale auch auf höheren Rangstufen erfolgte (was im BM wirklich so war, weil dort die Aggregation NUR für E+M geschrieben war und in E+M es keine Verbreitungen mit Rang >Species gibt), erklärt das nicht, woher die Verbreitungen in "Cichorieae Lam. & DC." stammen. Schließlich gibt es eigentlich keine nativ eingegebene Verbreitung, dann hätte da doch auch nichts aggregiert worden sein können entlang der Arealhierarchie. Das sollten wir nach der Neuimplementierung/Korrektur der Verbreitungsaggregation nochmal checken.

Anm.: Später wurde wohl doch noch 1 native Verbreitung (Germany, TDWGL4) von dir eingegeben, Norbert. Ohne Quelle. Soll die so sein oder ist das ein Versehen.

Und: können wir diese ganzen Berechneten Verbreitungen auf oberer Ebene (Tribus, Subtribus) löschen? Dann ließe sich die nächste Verbreitungsaggregation besser testen, die dann ja nur bis zur Gattung soll(?).

**#14 - 11/08/2019 11:43 PM - Andreas Müller**

- Due date deleted (09/29/2015)

**#15 - 11/11/2019 09:28 AM - Andreas Müller**

- Status changed from Feedback to In Progress

- Assignee changed from Norbert Kilian to Andreas Müller

- % Done changed from 100 to 70

NK:

Vor eine Neuaggregation können und sollten die Neuberechnungen auf Subtribus- und Tribus-Ebene gelöscht werden. Aber: die Verbreitungsdatenaggregation sollte im Cichorieae-Portal schon bis auf Tribusebene einschließlich erfolgen. In absehbarer Zeit wird sich die Klassifikation insoweit ändern, als eine höhere Rangstufe, die Unterfamilie, dazukommt (und wahrscheinlich 2 Taxa als neue Tribus ausgegliedert werden), deshalb wäre es sinnvoll, wenn die Aggregation bis zum höchsten verwendeten Taxon berechnet wird.

**#16 - 11/11/2019 09:37 AM - Andreas Müller**

AM:

Ja, ich habe dann später auch festgestellt, dass die Aggregation bis zum Tribus in dem Fall gewünscht war. Daher stellt sich jetzt aber andersherum die Frage, warum nur so wenige Areale bis auf diese Stufe aggregiert wurden. Da stimmt auf jeden Fall etwas noch nicht. Das müssen nach der Überarbeitung der Distribution Aggregation testen.

**#17 - 11/12/2019 06:13 PM - Andreas Müller**

- Status changed from In Progress to Resolved

- % Done changed from 70 to 50

I set to resolved as I guess it is fixed. But needs to be reviewed.

**#18 - 11/19/2019 05:36 PM - Andreas Müller**

- Priority changed from New to Highest

**#19 - 12/03/2019 02:14 PM - Andreas Müller**

- Assignee changed from Andreas Müller to Norbert Kilian

- % Done changed from 50 to 90

I think this is fully fixed with the new implementation of the transmission engine (which allows the user to decide to which final subtree parent and/or rank he/she wants to aggregate).

Norbert, can you please have a final look if any of the issues discussed here might be missing in the current implementation or in the data currently aggregated with the new algorithm on cichorieae?  
Afterwards we may close the ticket.

**#20 - 07/14/2020 04:13 PM - Andreas Müller**

- Status changed from Resolved to Closed

- Assignee changed from Norbert Kilian to Andreas Müller

**#21 - 07/14/2020 04:14 PM - Andreas Müller**

- % Done changed from 90 to 100