EDIT - feature request #2625

[E+M Overview] Data aggregation functionalities for E+M (TransmissionEngine)

09/27/2011 01:45 PM - Andreas Müller

Status:	Closed		**Start date:**	
Priority:	New		**Due date:**	
Assignee:	Andreas Müller		**% Done:**	100%
Category:	cdmlib		**Estimated time:**	0:00 hour
Target version:	Euro+Med Portal Release			
Severity:	normal			

Description

Implement data aggregation workflows such as aggregating distribution data from sub areas to larger areas as well as lower taxa to higher taxa.

According to existing E+M aggregation functionality.

The latest code of the transmission engine is available here:
Y:\BDI\PESI\EM_MCL\EM_MCL_DataAndProgramming\Anton_Programming\TransmissionEngineOccurrence (only BGBM internal access)

The latest version of the transmission engine code with additional comments is attached to this ticket
TransmissionsEngineOccurrence_V14.bas

Important notes:

source references

~~source references of the accumulated distributions are also accumulated into the new distribution, this has been especially implemented for the EuroMed Checklist Vol2 and might not be a general requirement~~ This was a mistake, we now have a special ticket for handling the aggregation of source references: #4366

summaryStatus

Each distribution information has a summaryStatus (emOccurSumCat.xlsx), this is an summary of the status codes (emOccurStatCat.xlsx) as stored in the fields of emOccurrence native, introduced, cutivated, ...

The summaryStatus seems to be equivalent to the distribution status @PresenceAbsenceTermBase@.

Each summary status has a priority field which specified the preference of one status over another.

These priorities have been defined an a long intensive process by Anton Güntsch and Eckhard Raab-Straube. It is <u>questionable if these priorities are project specific or if they are applicable in general</u>. This leads to the **requirement that the priorities must not be hard coded, they must be stored in the database** and in order to allow configuring them.

map generation

When generating maps from the accumulated distribution information some special cases have to be handled:

1. if a entered or imported status information exist for the same area for which calculated (accumulated) data is available, the calculated data has to be given preference over other data.

2. If there is an area with a sub area and both areas have the same status only the subarea status should be shown in the map, whereas the super area should be ignored. see #5050

The *TransmissionEngineDistribution can be triggered via a REST service*:

./description/accumulateDistributions?mode=[byAreas|byRanks|byAreasAndRanks]&frontendBaseUrl=<server-instance-base-URL>& priority=[1...7,DEFAULT:3]

This REST service is still a special implementation for the Euro+Med project. The parameters for the superAreas, the areas to which the subordinate areas should be projected, lowerRank, upperRank are hardcoded to: TDWG_LEVEL3 areas, SUBSPECIES, GENUS

---

check if #2083 (CICHORIEAE implement hierarchy for distribution status) is fixed by this once the transmission engine has been run on the cichorieae data

---

email discussion on how to treat references in the transmission engine: Discussion-Transmission_Referenzen.txt

Subtasks:	
bug # 4134: Transmissionengine Distribution seems to miss distributions for higher Taxa	**Closed**
feature request # 4366: Transmissionengine Distribution: implement rules for source ref...	**Duplicate**

Related issues:	
Related to EDIT - feature request #8677: Add distribution aggregation to set ...	**Closed**
Related to EDIT - bug #8312: Test, fix, improve and run Transmission engine f...	**Closed**
Related to EDIT - task #8651: Unify description aggregation methods (distribu...	**Closed**
Related to EDIT - task #8679: Further unify description aggregation methods	**Closed**
Related to EDIT - task #8871: Remaining issues to unify description aggregati...	**In Progress**
Related to EDIT - task #8811: Open issues for "Add distribution aggregation t...	**New**

History

#1 - 09/27/2011 02:15 PM - Andreas Müller

- Aggregation of distribution data according to ranks and regions (#2630)

#2 - 09/27/2011 02:27 PM - Andreas Müller

- Priority changed from Priority10 to Priority14

#3 - 02/18/2013 05:23 PM - Andreas Kohlbecker

@Andreas Müller:

what do you think is the best place to put the priorities into the database, Extensions?

#4 - 02/25/2013 08:47 AM - Andreas Kohlbecker

- Status changed from New to In Progress

- Assignee changed from Andreas Müller to Andreas Kohlbecker

I decided to use Extensions to store priorities.

#5 - 02/25/2013 09:56 AM - Andreas Müller

Replying to a.kohlbecker:

> I decided to use Extensions to store priorities.

Sorry for answering only now. I am not sure if extensions are the best choice. They are string based and therefore usually not the first option to express an order. However, we do have already an ExtensionType "Order" (uuid = "ecb7770d-a295-49ee-a88f-e9e137a7cabb") which we could use for it.

For me the more natural choice would be the order of the vocabulary itself. Atleast as long as there is no other reason for having "PresenceAbsenceTerms" ordered. But if we do so we have to discuss 2 things:

- Reorder the current PresenceTerm and AbsenceTerm vocabulary according to the "priority" in E+M

- Merge the presence and the absence terms into 1 class (this has been discussed long time ago), instead use a absence flag maybe.

However, if you can think about any other semantics for the "ordered" attribute of the presenceAbsence vocabularies we should rethink this solution.

#6 - 02/25/2013 11:24 AM - Andreas Kohlbecker

Replying to a.mueller:

Replying to [a.kohlbecker](#):

> I decided to use Extensions to store priorities.

Sorry for answering only now. I am not sure if extensions are the best choice. They are string based and therefore usually not the first option to express an order. However, we do have already an ExtensionType "Order" (uuid = "ecb7770d-a295-49ee-a88f-e9e137a7cabb") which we could use for it.

For me the more natural choice would be the order of the vocabulary itself. Atleast as long as there is no other reason for having "PresenceAbsenceTerms" ordered. But if we do so we have to discuss 2 things:

- Reorder the current PresenceTerm and AbsenceTerm vocabulary according to the "priority" in E+M

- Merge the presence and the absence terms into 1 class (this has been discussed long time ago), instead

use a absence flag maybe.

> However, if you can think about any other semantics for the "ordered" attribute of the presenceAbsence vocabularies we should rethink this solution.

Anton was not sure that the priorities are necessarily the same for all projects, so we need some flexibility here, so using the term order is not really an option. Furthermore there are also terms which are omitted during the aggregation process, these terms do not have a priority at all, maybe a negative one? Where should these terms be put at the top of the list at the bottom? Would a term select list still be useful for users or rather confusing since the terms order looks a bit arbitrary?

I think we should not try superimposing a "secret" meaning into the term order. For the moment using the Extensions is a really good choice since this is not causing a model change.

#7 - 04/04/2013 03:36 PM - Andreas Kohlbecker

check if #2083 (CICHORIEAE implement hierarchy for distribution status) is fixed by this once the transmission engine has been run on the cichorieae data

#8 - 04/18/2013 12:55 PM - Andreas Kohlbecker

dataportal sie implemented: r17640

#9 - 04/18/2013 01:35 PM - Andreas Kohlbecker

additional work on the library side [17642:17665]

#10 - 01/20/2014 05:07 PM - Andreas Kohlbecker

adding text of email discussion on how to treat references in the transmission engine: [Discussion-Transmission_Referenzen.txt](#)

#11 - 01/20/2014 05:45 PM - Andreas Kohlbecker

- Status changed from In Progress to Resolved

- Assignee changed from Andreas Kohlbecker to e.raab-straube -

as far as I remember this is ticket is completed and can now be reviewed

#12 - 04/10/2014 06:25 PM - Andreas Kohlbecker

- Keywords set to Euro+Med,Migration

#13 - 08/18/2014 02:19 PM - Andreas Kohlbecker

- Subject changed from [E+M Overview] Data aggregation functionalities for E+M to [E+M Overview] Data aggregation functionalities for E+M (TransmissionEngine)

#14 - 07/16/2015 03:04 PM - Andreas Müller

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

#15 - 06/21/2016 11:28 AM - Andreas Kohlbecker

- Assignee changed from e.raab-straube - to Andreas Kohlbecker

#16 - 06/27/2016 02:47 PM - Andreas Kohlbecker

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

#17 - 05/10/2017 01:38 PM - Andreas Kohlbecker

- Private changed from Yes to No

#18 - 11/11/2019 11:40 PM - Andreas Müller

- Description updated

- Assignee changed from Andreas Kohlbecker to Andreas Müller

#19 - 11/12/2019 06:12 PM - Andreas Müller

- Related to feature request #8677: Add distribution aggregation to set subtree menu added

#20 - 12/07/2019 06:11 PM - Andreas Müller

- Related to bug #8312: Test, fix, improve and run Transmission engine for E+M added

#21 - 03/01/2021 04:37 PM - Andreas Müller

- Related to task #8651: Unify description aggregation methods (distribution and structured descriptive data) added

#22 - 03/01/2021 04:38 PM - Andreas Müller

- Related to task #8679: Further unify description aggregation methods added

#23 - 03/01/2021 04:38 PM - Andreas Müller

- Related to task #8871: Remaining issues to unify description aggregation methods added

#24 - 03/01/2021 04:39 PM - Andreas Müller

- Related to task #8811: Open issues for "Add distribution aggregation to set subtree menu" added

#25 - 03/01/2021 04:43 PM - Andreas Müller

- Status changed from Resolved to Closed

This should be fixed with the transmission engine implementations (see subtasks) and with the implementation of description aggregations (see related tickets) as well as the integration of description aggregation into the TaxEditor navigator menu (#8677).

Files

TransmissionsEngineOccurrence_V14.bas	21.1 KB	02/18/2013	Andreas Kohlbecker
emOccurStatCat.xlsx	9.44 KB	02/18/2013	Andreas Kohlbecker
emOccurSumCat.xlsx	9.97 KB	02/18/2013	Andreas Kohlbecker
status-mapping-eumed-cdm.ods	19.2 KB	02/27/2013	Andreas Kohlbecker
Discussion-Transmission_Referenzen.txt	3.66 KB	01/20/2014	Andreas Kohlbecker