PhycoBank - task #6173

Concept for a useful algae registry taxon classification

11/01/2016 05:32 PM - Andreas Kohlbecker

| Status: | Closed | Start date: | | |
|-----------------|--------------------|-----------------|-----------|--|
| Priority: | New | Due date: | | |
| Assignee: | Andreas Kohlbecker | % Done: | 100% | |
| Category: | Model | Estimated time: | 0:00 hour | |
| Target version: | | | | |

Description

A taxonomic classification is required for the algae registry for grouping taxa logically and semantically. These groups often are not equal to the commonly used higher classification. E.g. It would be scientifically correct (according to some scientist) to use the phylum *Heterokontophyta*, whereas for practicals means it would be beneficial to use the classes *Xanthophyceae* or *Bacillariophyceae* which are included into the former phylum. Also under discussion is the option to use vernacular higher classification names.

A major obstacle in finding a suitable classification are the different opinions in the phycological community on a valid and acceptable classification. Any choice for a specific classification would cause irritation and rejection in parts of the community. Therefore it would be wise not to have a classification at all. But still is there a need for something like a classification to support the below use cases:

The algae registry actually has no need for a real classification except for the purpose of making it easier to find names. The complete reference classification will be provided by the name index which is a separate cdm instance. The classification for a given name will be retrieved from the index by sending a query to REST service.

These considerations lead to the following draft concept:

Use cases and requirements

Use cases:

- 1. **Higher ranked names as keywords**: Choose a higher taxon name to find all names that are covered by this name independently from the position of this name in a specific classification. (e.g. "A user wants to find the latest registrations in a specific taxon group."). Phycobank will offer a couple of prepared **default classifications** as a backbone to which registered name can be attached. ==> Requirement 1.
- 2. **Higher classification data for each name registered**: For each name registered it must be possible to add and remove the ordered list of taxon names as they occur in the higher classification. Each of this higher classifications must be associated with the according reference for this information. To each name multiple higher classification name lists + reference can be assigned. (See #???? ticket for ui implementation):
 - 1. coming from the registered names it must be possible to find the path up to the higher taxa as published in a specific reference.
 - 2. A name added to one higher classification must also be removable without breaking classification information associated with another name. Also must the removal not modify any higher default classification (see point 1.). ==> Requirement 2.

Requirements & conclusions and from the use cases:

- 1. The phycobank quasi classification should therefore be the union of all classifications relevant for the registrations.
- 2. Each name registered must have associations to all the names mentioned in any of its higher classifications. As newly created higher classification names can be reused in by other taxa and users it is necessary to manage the ability to delete a higher ranked name.
- 3. Homotypy relations will only be modeled via the basionym name relationship since phycobank is clearly only focused on nomenclatural information. Expressing basionyms also via taxon relations to the basionym name as synonym would impose a level of taxonomic option we must avoid.

And the winner is ... N1T under point 6.), see below for details)

1) Higher classification with concept relations

We will have one classification to which the taxa belong to. Each taxon will be added to this classification. In the beginning this classification will only consist of the highest ranklevel (phylum, class).

All lower ranked taxa will only be added as needed see 2) and 2) below.

•

04/19/2024 1/12

2) New registration of suprageneric taxa

classified within

suprageneric taxa are added to the classification when a registration for the according name is created:

- 1. Firstly by the author
- 2. secondly by the data curation

3) Species or infra generic taxa

These taxa are added to the genus (automatically on base of the uninomial). In case the genus is not yet in the system it will be creates as post registration in progress. The data curation will then validate this new name.

4) Higher taxonomy as published by 'TaxonInteractions'

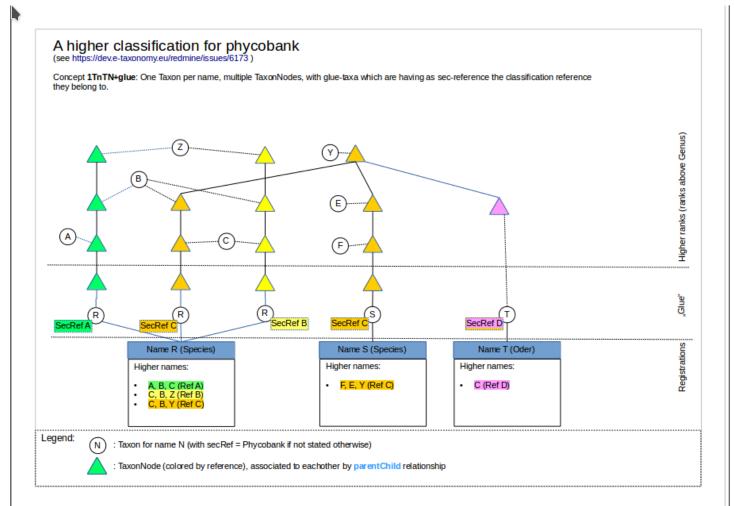
The publications also provide a higher taxonomy for the new names. I makes no sense to create a new separate classification for each case. It is sufficient and much more elegant to add theses taxa as TaxonInteractions to the newly registered name. The label for this feature type could be "Classification as published"

5) Multiple higher classification with TaxonNode parentChild relationships

5.a) [N1TnTN+glue] One Taxon per name, multiple TaxonNodes, with glue-taxa which are having as sec-reference the classification reference they belong to.

This is the idea which was devised after the original idea of modeling the higher 'classification' as taxon graph was mistakenly rejected.

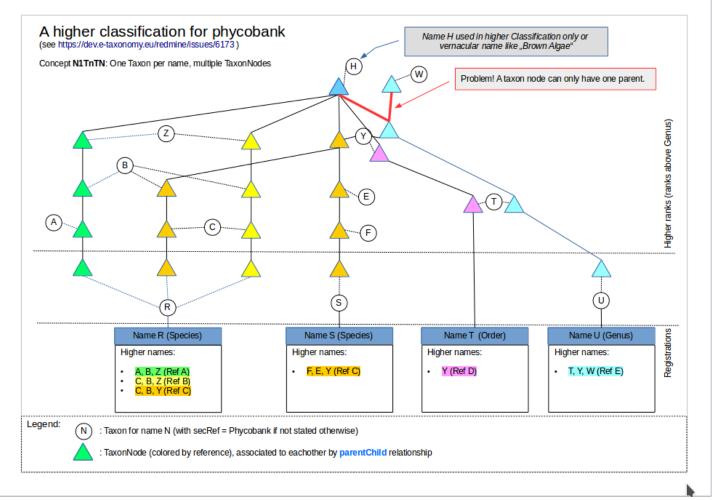
04/19/2024 2/12



Adding the secReference to the Taxon entities, which are created for the registered names in this case, is not needed, since the reference is also associated with the TaxonNode via the Classification. Therefore this idea can be simplified by and we come to the concept named **NnT1TN**

5.b) [N1TnTN] One Taxon per name, multiple TaxonNodes

04/19/2024 3/12

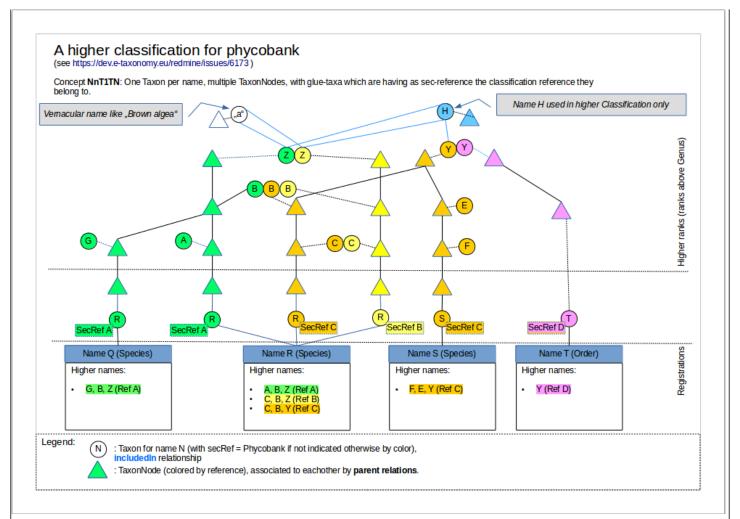


In this diagram a problem is pointed out which would occur in each graph build on TaxonNode relations. A Classification which is only defined for higher ranks can not be linked in a branch of another classification, since this would require that a TaxonNode can have multiple parents. I am not sure if this theoretical problem exists with any proper classification relevant for phycobank. However if we want to use vernacular names like "Brown algae" this situation would become reality.

5.c) [N1TnTN] Multiple Taxon per Name, one TaxonNodes each

Anreas Müller: "Hatten wir nicht in der längeren Diskussion darüber, wie wir das mit den verschiedenen Klassifikationen bzw. includedIn Beziehungen handeln beschlossen, includedIn Beziehungen zu erzeugen, damit Suchen der Form gebe mir alle Namen die zu folgendem Familiennamen gehören, durchführen zu können mittels der getIncludedInTaxa Methode oder wie die heißt, die ich für die Roten Listen geschrieben habe. Das würde dann nicht funktionieren, wenn wir keine separaten Taxa anlegen pro Name in einer Klassifikation."

04/19/2024 4/12



The includedIn taxon relation being used in this idea avoid the problem pointed out in 5.b).

Only inter-classification relations which are not implicit by usage of the same name in different taxa need to be modeled as **includedIn** taxon relation. So only the taxa ("a") and (H) need to be connected to the graph this way. This must be done manually by the curation.

Disadvantage: if a name is used in multiple taxa and they all have the same includedIn relationship to a higher taxon of another classification multiple includeIn relationships need to be created.

Search example (search for name "Mastolgoiales", this name and names found under this name in the hierarchy are marked green):

| phylum | Bacillariophyta | | | phylum | Ochrophyta | | |
|------------|-------------------|----------------|----|------------|-------------------|-----------------------|--|
| classis | Bacillariophyceae | | • | classis | Bacillariophyceae | | |
| subclassis | Bacillarion | ohycidae | Ť. | subclassis | Bac | Bacillariophycidae | |
| ordo | Masto | gloiales | | ordo | | Achnanthales | |
| familia | Ac | chnanthaceae | | familia | | Achnanthaceae | |
| genus | | Achnanthes | | genus | | Achnanthes | |
| familia | M. | astogloiaceae | | genus | | Amphicocconeis | |
| genus | | Aneumastus | | genus | | Haloroundia | |
| genus | | Caspedostauros | | genus | | Platessa | |
| genus | | Decussata | | ordo | | Mastogloiales | |
| genus | | Mastogloia | | familia | | Mastogloiaceae | |
| | | | | genus | | Aneumastus | |
| | | | | genus | | Mastogloia | |
| | | | | genus | | Mastogloiopsis | |
| | | | | genus | | Skeletomastus | |

6) [N1T] Higher taxon-graphs with includedIn relations taxon relationships

This strategy has won the competition - congratulations!!!

Das ursprüngliche Konzept war ein Taxon-Graph, bei dem pro Name jeweils ein Taxon angelegt wird. Die Taxa werden per includedIn TaxonBeziehung miteinander verknüpft. Die Suche nach Namen von höheren Taxa zu niedrigeren Rängen funktioniert

04/19/2024 5/12

hier wunderbar. Was aber verloren geht, ist die Information zu welcher Referenz das Taxon gehört. Die Pfade vom registrierten Namen durch die höheren Rangstufen sind uneindeutig und lassen sich nicht mehr klar einer Referenz zuordnen. Mehrere Sec-Referenzen pro Taxon sind nun mal nicht möglich. Es gibt daher 2 weitere Möglichkeiten:

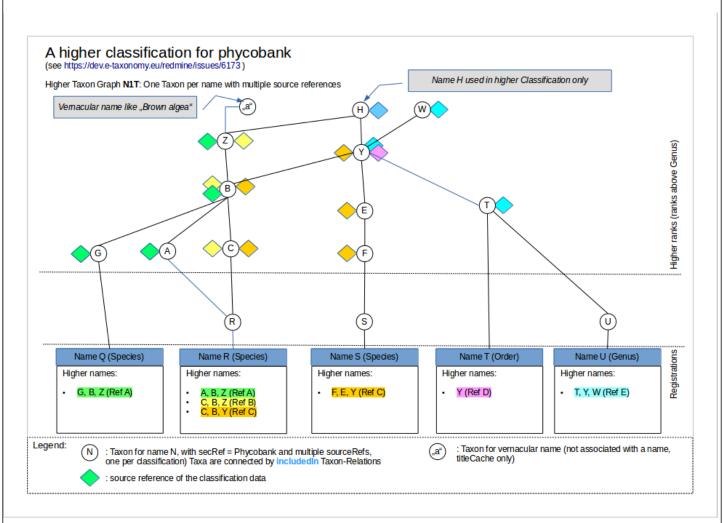
- 1. Ein Taxon pro Name, wobei die SecReferenz immer Phycobank ist aber pro Klassifikation eine Source-Referenz an das Taxon gehängt werden.
- 2. Mehrere Taxa pro Namen wobei die SecReferenz immer der Referenz der Klassifikation entspricht. Würde die top-down Suche mit der listIncludedTaxa Methode in diesem Fall funktionieren? Ich denke nicht, denn diese berücksichtigt nur die Relationen und bezieht identische Namen der Taxa nicht mit ein. (Note AM: if the concepts are expected to be equivalent you may add an isCongruentTo relationship, this way the search may work again)

Also bleibt nur Möglichkeit 1. Diese hatten Henning und ich aber auch aus irgend einem Grund ausgeschlossen. Ich kann mich aber nicht mehr erinnern weshalb.

Im Moment fällt mir nichts ein was gegen diese Möglichkeit spricht. Darüberhinaus löst diese Variante auch das Problem der TaxonNode Graphen in denen es nicht möglich ist einen Knoten mit mehreren Eltern zu verknüpfen.

Note AM: General Problem with 1 is that different concepts are not reflected, some searches may lead to too large results. A solution might be to use >1 taxon per name where differences in concepts are known and relationships to parents and children are easy to define.

Note2 AM: Why not using the nomenclatural reference as sec reference for the taxa to indicate that we are only talking about names here, phycobank as secundum is somehow misleading



Outcome of the final discussion:

The winner is **N1T* under point 6.)*

We (Andreas K, Andreas M & Henning) decided to also create the TaxonNode graph for each standard classification to be imported. Taxa will rather be reused for multiple TaxonNodes than creating one Taxon per TaxonNode.

The strategy **N1T** allows to discover names found under a specific higher name in a broad manner. It might turn out that this strategy includes too many unwanted names in some cases. These situations could be relaxed by creating multiple taxa for a single name whereas the source references are distributed to the taxa and the TaxonNodes will also reference only one of these taxa of course.

04/19/2024 6/12

The search process primarily aims in finding genera, all species and subspecies which fall under each genus found are to be displayed independently of the higher classification information associated with the individual names. Practically this will be solved by creating for each TaxonName an *includenIn* relation to the Taxon for the name used in the uninomialOrGenus and in the specificEpithet field.

Ranks like e.g. *variety* and *subgenus* can not be associated automatically, so their relation to the genus or species respectively must be created manually if this information is available or if it is required in a specific case. It also should be considered to associate e.g. species to the according subgenera.

We should also consider to **reactivate the formerly rejected TaxonReationship** *taxomicallyInduldedIn* which has been removed in the past. Using this relation type would be semantically more correct than using *includedIn* which rather expresses that a group of taxa is included into a bigger group of taxa of all taxa of the same rank.

Related issues:

| Related to PhycoBank - task #7298: Associate TaxonNames with higher ranks 'ke | Rejected |
|---|----------|
| Related to EDIT - feature request #7419: Provide example NormalImplicit impor | Closed |
| Related to EDIT - task #7420: Import for higher taxon graph for phycobank | Closed |
| Related to EDIT - feature request #7648: Create taxonrelation to genus or spe | Resolved |
| Related to EDIT - feature request #7791: Visualize the phycobank higher taxon | Closed |

History

#1 - 11/01/2016 05:32 PM - Andreas Kohlbecker

- Description updated

#2 - 11/01/2016 05:36 PM - Andreas Kohlbecker

- Description updated

#3 - 03/14/2017 12:17 PM - Andreas Kohlbecker

- Status changed from New to Resolved

Hi Henning,

I think this issue is completely resolved now. Do you agree?

Then, please add or reference the final decisions on the "classification" in an appropriate way and close the ticket.

Andreas

#4 - 09/19/2017 12:14 PM - Andreas Kohlbecker

- Description updated

#5 - 09/20/2017 11:45 AM - Wolf-Henning Kusber

Higher ranks:

PhycoBank can not use classes as published in ICN, ING, NCU because this system differs very much from current classifications. PhycoBank checked two modern resources and provides a compilation including classes and divisions most likely accepted by the majority of PhycoBank users as search criteria. PhycoBank do not want to establish an own classification system as such.

Resources consulted: Ruggiero M.A., Gordon D.P., Orrell T.M., Bailly N., Bourgoin T., Brusca R.C., et al. 2015: Correction: A Higher Level Classification of All Living Organisms. PLoS ONE 10(6): e0130114. doi:10.1371/journal.

pone.0130114Syllabus of Plant Families 13th edition 2012, 2015, 2017 (ed. Frey, W.) - Borntraeger, Stuttgart).

Main difference are the alternative usage of Bacillariophyta and its Classed as agreed by the majority of diatomists according to Cox in Frey (2015) instead of using subclasses because of consistency within the Ochrophyta.

#6 - 02/27/2018 11:49 AM - Andreas Kohlbecker

- Related to task #7298: Associate TaxonNames with higher ranks 'keyword' taxa added

#7 - 03/28/2018 06:02 PM - Andreas Kohlbecker

- File picture 764-1.png added

[a.mueller]

ich habe das gerade mit AK nochmal telefonisch besprochen. Wir haben die entsprechenden Sammel-Taxonknoten mit "unplaced" getagged. Dadurch erscheinen sie im Editor jetzt am Ende der Liste (jedenfalls wenn man nach Rank + alphabetisch sortiert, ab dem nächsten Release auch bei rein alphabetischer Sortierung)

04/19/2024 7/12

- ▶ Rhodophyta
- ? Order Names
- > ? Family Names
- ? Tribe Names
- > ? No group assigned

Zudem habe ich Camel Case aus den Namen entfernt und den Sammel Taxa den jeweiligen Superrang gegeben, also z.B. Order Names wurde zu Superorder. No group assigned habe ich auf "unranked" gelassen.

Den nicht mehr genutzen Rang "Families incertis sedis" habe ich gelöscht.

Man könnte jetzt die einzelnen Pseudotaxa innerhalb der Gruppen auch noch mit unplaced taggen. Da sie sich aber alle in einem getaggten Container befinden, ist das nicht unbedingt notwendig, außer man will das unplaced/incertis sedis später auf einer Taxonseite mal explizit angezeigt bekommen (bislang nicht implementiert im Portal).

#8 - 05/02/2018 02:30 PM - Andreas Kohlbecker

- File Phycobank-higher-classification.pdf added

Diagram for new concept added

#9 - 05/02/2018 02:31 PM - Andreas Kohlbecker

- File Phycobank-higher-classification.odg added

#10 - 05/02/2018 02:32 PM - Andreas Kohlbecker

- File picture866-1.png added
- Description updated

#11 - 05/03/2018 09:40 AM - Andreas Kohlbecker

- File v1-1TnTN+glue.png added
- File v2-NnT1TN.png added
- File v3-N1TnTN.png added
- Description updated

adding more classification concept variants and updating existing ones

#12 - 05/03/2018 09:41 AM - Andreas Kohlbecker

- File deleted (Phycobank-higher-classification.pdf)

#13 - 05/03/2018 09:41 AM - Andreas Kohlbecker

- File deleted (picture866-1.png)

#14 - 05/03/2018 09:41 AM - Andreas Kohlbecker

- File deleted (Phycobank-higher-classification.odg)

#15 - 05/03/2018 09:44 AM - Andreas Kohlbecker

- Description updated

#16 - 05/03/2018 09:46 AM - Andreas Kohlbecker

- File deleted (v3-N1TnTN.png)

#17 - 05/03/2018 09:47 AM - Andreas Kohlbecker

- File NnT1TN.png added

#18 - 05/03/2018 09:48 AM - Andreas Kohlbecker

- File v3-N1TnTN.png added

#19 - 05/03/2018 09:48 AM - Andreas Kohlbecker

- File deleted (NnT1TN.png)

04/19/2024 8/12

#20 - 05/03/2018 09:49 AM - Andreas Kohlbecker

- File Phycobank-higher-classification-v1-N1TnTN+glue.odg added
- File Phycobank-higher-classification-v2-NnT1TN.odg added
- File Phycobank-higher-classification-v3-N1TnTN.odg added

#21 - 05/03/2018 09:51 AM - Andreas Kohlbecker

- Description updated

#22 - 05/03/2018 10:57 AM - Andreas Kohlbecker

- File v4-N1T.png added

#23 - 05/03/2018 11:01 AM - Andreas Kohlbecker

- Description updated

#24 - 05/03/2018 11:03 AM - Andreas Kohlbecker

- Description updated

#25 - 05/03/2018 11:06 AM - Andreas Kohlbecker

- File v4-N1T.png added

#26 - 05/03/2018 11:07 AM - Andreas Kohlbecker

- File Phycobank-higher-classification-v4-N1T.odg added

#27 - 05/03/2018 05:46 PM - Andreas Kohlbecker

- Description updated

#28 - 05/03/2018 05:47 PM - Andreas Kohlbecker

- Description updated

#29 - 05/03/2018 05:53 PM - Andreas Kohlbecker

- Description updated

#30 - 05/04/2018 08:05 AM - Andreas Kohlbecker

- Description updated

#31 - 05/04/2018 08:08 AM - Andreas Kohlbecker

- Description updated

#32 - 05/04/2018 08:10 AM - Andreas Kohlbecker

- File v2-N1TnTN.png added

#33 - 05/04/2018 08:11 AM - Andreas Kohlbecker

- File deleted (Phycobank-higher-classification-v2-NnT1TN.odg)

#34 - 05/04/2018 08:11 AM - Andreas Kohlbecker

- File Phycobank-higher-classification-v2-N1TnTN.odg added

#35 - 05/04/2018 08:36 AM - Andreas Kohlbecker

- Description updated

#36 - 05/04/2018 09:32 AM - Andreas Kohlbecker

- File N1TnNT.png added

#37 - 05/04/2018 09:39 AM - Andreas Kohlbecker

- File Phycobank-higher-classification-v2-N1TnTN.odg added
- Description updated

04/19/2024 9/12

#38 - 05/04/2018 09:40 AM - Andreas Kohlbecker

- File deleted (Phycobank-higher-classification-v2-N1TnTN.odg)

#39 - 05/04/2018 09:41 AM - Andreas Kohlbecker

- Description updated

#40 - 05/04/2018 09:43 AM - Andreas Kohlbecker

- File v4-N1T.png added

#41 - 05/04/2018 09:50 AM - Andreas Kohlbecker

- File v3-NnT1TN.png added

#42 - 05/04/2018 09:50 AM - Andreas Kohlbecker

- Description updated

#43 - 05/04/2018 09:55 AM - Andreas Kohlbecker

- Description updated

#44 - 05/04/2018 10:49 AM - Andreas Kohlbecker

- Description updated

#45 - 05/04/2018 10:57 AM - Andreas Kohlbecker

- File v3-NnT1TN.png added

#46 - 05/04/2018 11:08 AM - Andreas Kohlbecker

- Description updated
- Status changed from Resolved to In Progress

#47 - 05/04/2018 12:03 PM - Andreas Kohlbecker

- File v3-NnT1TN.png added

#48 - 05/04/2018 12:21 PM - Andreas Kohlbecker

- File v3-NnT1TN.png added

#49 - 05/04/2018 12:22 PM - Andreas Kohlbecker

- File deleted (v3-NnT1TN.png)

#50 - 05/04/2018 12:22 PM - Andreas Kohlbecker

- File deleted (v3-NnT1TN.png)

#51 - 05/04/2018 12:22 PM - Andreas Kohlbecker

- File deleted (v3-NnT1TN.png)

#52 - 05/04/2018 12:23 PM - Andreas Kohlbecker

- File deleted (v2-NnT1TN.png)

#53 - 05/04/2018 12:23 PM - Andreas Kohlbecker

- File deleted (v3-N1TnTN.png)

#54 - 05/04/2018 12:23 PM - Andreas Kohlbecker

- File deleted (v4-N1T.png)

#55 - 05/04/2018 12:23 PM - Andreas Kohlbecker

- File deleted (v4-N1T.png)

#56 - 05/04/2018 12:36 PM - Andreas Kohlbecker

04/19/2024 10/12

#57 - 05/04/2018 01:26 PM - Andreas Müller

- Description updated

#58 - 05/04/2018 01:33 PM - Andreas Müller

- Description updated

#59 - 05/04/2018 01:46 PM - Andreas Müller

- Description updated

#60 - 05/04/2018 02:12 PM - Andreas Kohlbecker

- File search-5c.png added

#61 - 05/04/2018 02:13 PM - Andreas Kohlbecker

- Description updated

#62 - 05/04/2018 02:13 PM - Andreas Kohlbecker

- Description updated

#63 - 05/04/2018 04:18 PM - Andreas Kohlbecker

- Description updated

#64 - 05/04/2018 04:29 PM - Andreas Kohlbecker

- Description updated

#65 - 05/04/2018 04:52 PM - Andreas Kohlbecker

- Description updated

#66 - 05/04/2018 05:59 PM - Andreas Kohlbecker

- Description updated
- % Done changed from 0 to 30

#67 - 05/04/2018 06:01 PM - Andreas Kohlbecker

- Description updated

#68 - 05/04/2018 06:10 PM - Andreas Kohlbecker

- Description updated

#69 - 05/04/2018 06:31 PM - Andreas Kohlbecker

- Description updated

#70 - 05/15/2018 03:51 PM - Andreas Kohlbecker

- Category set to Model
- Assignee changed from Wolf-Henning Kusber to Andreas Kohlbecker
- % Done changed from 30 to 40

Almost complete. As last task in this ticket we should create a diagram to illustrate the final decision.

#71 - 05/15/2018 03:57 PM - Andreas Kohlbecker

- Related to feature request #7419: Provide example NormalImplicit import file for phycobank higher taxon graph imports added

#72 - 05/15/2018 04:10 PM - Andreas Kohlbecker

- Related to task #7420: Import for higher taxon graph for phycobank added

#73 - 08/09/2018 12:37 PM - Andreas Müller

04/19/2024 11/12

#74 - 08/10/2018 02:21 PM - Andreas Müller

How should species (and other names below genus) be related to genus taxa. There might be multiple genus taxa per name. Is this relationship automatically created or are they only related by special search algorithms?

#75 - 08/13/2018 03:28 PM - Andreas Kohlbecker

- Related to feature request #7648: Create taxonrelation to genus or species when subordinate names are created added

#76 - 08/13/2018 03:32 PM - Andreas Kohlbecker

Andreas Müller wrote:

How should species (and other names below genus) be related to genus taxa. There might be multiple genus taxa per name. Is this relationship automatically created or are they only related by special search algorithms?

This relation between genus and subordinate names needs to be created on the fly when creating new Species and also when changing the genus for a species. The implementation of this mechanism is handled in #7648. All the details, which are actually found in this issue description are copied as extract to the issue #7648. The summary is that there should only be one taxon per name.

#77 - 09/28/2018 08:32 AM - Andreas Kohlbecker

- Related to feature request #7791: Visualize the phycobank higher taxon graph added

#78 - 05/28/2019 12:59 PM - Andreas Kohlbecker

- Status changed from In Progress to Closed
- % Done changed from 40 to 100

completed

#79 - 04/03/2020 12:28 PM - Andreas Müller

- Related to task #8886: Import Lichenes Genera added

#80 - 04/03/2020 12:30 PM - Andreas Müller

- Related to deleted (task #8886: Import Lichenes Genera)

Files

| picture764-1.png | 1.76 KB | 03/28/2018 | Andreas Kohlbecker |
|--|---------|------------|--------------------|
| v1-1TnTN+glue.png | 82.9 KB | 05/03/2018 | Andreas Kohlbecker |
| Phycobank-higher-classification-v1-N1TnTN+glue.odg | 19.2 KB | 05/03/2018 | Andreas Kohlbecker |
| Phycobank-higher-classification-v3-N1TnTN.odg | 19.7 KB | 05/03/2018 | Andreas Kohlbecker |
| Phycobank-higher-classification-v4-N1T.odg | 16.8 KB | 05/03/2018 | Andreas Kohlbecker |
| v2-N1TnTN.png | 86 KB | 05/04/2018 | Andreas Kohlbecker |
| N1TnNT.png | 109 KB | 05/04/2018 | Andreas Kohlbecker |
| Phycobank-higher-classification-v2-N1TnTN.odg | 22.2 KB | 05/04/2018 | Andreas Kohlbecker |
| v4-N1T.png | 106 KB | 05/04/2018 | Andreas Kohlbecker |
| v3-NnT1TN.png | 117 KB | 05/04/2018 | Andreas Kohlbecker |
| search-5c.png | 52.9 KB | 05/04/2018 | Andreas Kohlbecker |
| | | | |

04/19/2024 12/12