

PhycoBank - task #6009

Import the IAPT database into a cdm instance

07/25/2016 02:10 PM - Andreas Kohlbecker

Status:	Closed	Start date:	
Priority:	New	Due date:	
Assignee:	Andreas Kohlbecker	% Done:	100%
Category:	Import	Estimated time:	0:00 hour
Target version:	IAPT Import ready		
Description			
Import aller Algen-Namen in 2 CDM Instanzen:			
1. Eine Instanz als Ersatz für die alte IAPT Anwendung: Komplettimport - DONE http://test.e-taxonomy.eu/cdmserver/iapt/			
2. Eine Instanz als Basis für die Algenregistrierung: Gefilterter Import, SQL Filter für alle Algennamen: higherRank='%PHYCEAE%' - DONE http://test.e-taxonomy.eu/cdmserver/phycobank/portal/classification			
Columns in the csv file:			
<ul style="list-style-type: none">• RegistrationNo_Pk• HigherTaxon• FullName• AuthorsSpelling• LitString• Registration• Type ==> Needs to be parsed, see comment• Caveats• FullBasionym• FullSynSubst• NotesTxt• RegDate• NameString• BasionymString• SynSubstStr• AuthorString			
CDM instances:			
<ul style="list-style-type: none">• http://test.e-taxonomy.eu/cdmserver/phycobank/• http://test.e-taxonomy.eu/cdmserver/iapt/			
Subtasks:			
task # 6018: Diatomeen-Namensbestand in IAPT Registrierung prüfen			Closed
task # 6093: Algen in 'Incertae sedis' & 'No group assigned'			Closed
task # 6034: DataCleaning after final import			Closed
bug # 6276: Nomenclatural references show parsed date string instead of publication year			Closed
task # 6277: use nomenclatural reference as sec. references instead of IAPT			Closed
task # 6279: remove all 'nom. val' status from name			Closed
task # 6294: remove punctuation marks from end of Locality language strings			Closed
task # 6171: Higher Rank			Duplicate
task # 6172: Higher Rank			Duplicate

History

#1 - 07/25/2016 02:12 PM - Andreas Kohlbecker

- Tracker changed from bug to task

#2 - 07/25/2016 04:30 PM - Andreas Kohlbecker

- Description updated

#3 - 07/26/2016 05:50 PM - Andreas Kohlbecker

- Description updated

- Status changed from New to In Progress

#4 - 07/26/2016 05:53 PM - Wolf-Henning Kusber

Andreas Kohlbecker wrote:

Import aller Algen-Namen in 2 CDM Instanzen:

1. Eine Instanz als Ersatz für die alte IAPT Anwendung: Komplettimport
2. Eine Instanz als Basis für die Algenregistrierung: Gefilterter Import, SQL Filter für alle Algennamen: `higherRank='PHYCEAE'`

Columns in the csv file:

- RegistrationNo_Pk
- HigherTaxon
- FullName
- AuthorsSpelling
- LitString
- Registration
- Type
- Caveats
- FullBasionym
- FullSynSubst
- NotesTxt
- RegDate
- NameString
- BasionymString
- SynSubstStr
- AuthorString

siehe auch Subtask

#5 - 08/01/2016 04:17 PM - Andreas Kohlbecker

- Description updated

Die Typus Informationen liegen nur als Freitext vor und enthalten teilweise recht viel Informationen, darunter auch Holotyp, Isotyp, Location. Genügt es für den Prototypen diese Daten als Freitext zu übernehmen oder brauchen wir diese in atomisierter Form?

Henning:

Darüber haben Eckhard und ich heute gesprochen. Dort wo ein Holotyp oder Isotyp vorkommt sollte eine Atomisierung möglich sein, beginnend jeweils mit der Sammlung, gefolgt vom Barcode, bzw. der Nummer des Sammlungsobjekts. Dort wo nur "Typus" und "Locality" steht sind Texte aus den Publikationen übernommen, d.h. nicht atomisierbar weil nicht standardisiert eingegeben

#6 - 09/16/2016 04:30 PM - Andreas Kohlbecker

- Target version set to IAPT Import ready

#7 - 09/19/2016 02:01 PM - Andreas Kohlbecker

- Description updated

for details on the filtered import, see #6026

#8 - 09/20/2016 12:03 PM - Andreas Kohlbecker

- Status changed from In Progress to Resolved

The import is fully implemented and has been successfully run.

#9 - 10/26/2016 05:37 PM - Andreas Kohlbecker

- Category set to Import

#10 - 10/28/2016 03:19 PM - Andreas Kohlbecker

- Description updated

#11 - 10/28/2016 03:20 PM - Andreas Kohlbecker

- Description updated

#12 - 11/03/2016 04:45 PM - Andreas Kohlbecker

- Description updated

#13 - 09/12/2017 09:56 AM - Andreas Kohlbecker

- Status changed from Resolved to Closed

We will no longer work on the import to improve data quality, so this issue can be closed.

#14 - 02/27/2018 11:52 AM - Andreas Kohlbecker

- File *cdm_algea_registry-import-errors-09-20-16.xls* added

- File *cdm_algea_registry-import-stats-09-20-16.txt* added

- File *cdm_iapt-import-errors-09-20-16.xls* added

- File *cdm_iapt-import-stats-09-20-16.txt* added

- File *IAPT_Import-Analyse-Henning-2016-09-12.xls* added

adding reports on the final imports as attachment

Files

cdm_algea_registry-import-stats-09-20-16.txt	561 Bytes	02/27/2018	Andreas Kohlbecker
cdm_algea_registry-import-errors-09-20-16.xls	224 KB	02/27/2018	Andreas Kohlbecker
cdm_iapt-import-stats-09-20-16.txt	570 Bytes	02/27/2018	Andreas Kohlbecker
IAPT_Import-Analyse-Henning-2016-09-12.xls	28 KB	02/27/2018	Andreas Kohlbecker
cdm_iapt-import-errors-09-20-16.xls	2.2 MB	02/27/2018	Andreas Kohlbecker