

Edit - task #7446

Check if all extensions are needed

06/03/2018 03:41 PM - Andreas Müller

Status:	New	Start date:	06/03/2018
Priority:	Priority14	Due date:	
Assignee:	Andreas Müller	% Done:	0%
Category:	cdmadapter	Estimated time:	5.00 hours
Target version:	Euro+Med Migration		
Severity:	normal		

Description

about 560.000 Extensions are currently imported.

We need to check if they are all needed or if some of them are duplicates. Also check if they should better run into alternative identifiers.

Once done we need to check if they should be shown on the dataportal.

```
SELECT e.type_id, et.titleCache, ab.DTYPE, r.refType, dtb.DTYPE, count(*) as n
FROM Extension e INNER JOIN DefinedTermBase et ON et.id = e.type_id
LEFT OUTER JOIN AgentBase_Extension abMN ON abMN.extensions_id = e.id
LEFT OUTER JOIN AgentBase ab ON ab.id = abMN.AgentBase_id
LEFT OUTER JOIN DefinedTermBase_Extension dtbMN ON dtbMN.extensions_id = e.id
LEFT OUTER JOIN DefinedTermBase dtb ON dtb.id = dtbMN.DefinedTermBase_id
LEFT OUTER JOIN Reference_Extension rMN ON rMN.extensions_id = e.id
LEFT OUTER JOIN Reference r ON r.id = rMN.Reference_id
```

```
GROUP BY type_id, et.titleCache, ab.DTYPE, r.refType, dtb.DTYPE
ORDER BY et.titleCache, ab.DTYPE, r.refType, dtb.DTYPE, n
```

```
SELECT e.type_id, e.value, et.titleCache, tb.DTYPE, r.refType, r.id, r.titleCache, dtb.DTYPE
FROM Extension e INNER JOIN DefinedTermBase et ON et.id = e.type_id
LEFT OUTER JOIN AgentBase_Extension abMN ON abMN.extensions_id = e.id
LEFT OUTER JOIN AgentBase ab ON ab.id = abMN.AgentBase_id
LEFT OUTER JOIN DefinedTermBase_Extension dtbMN ON dtbMN.extensions_id = e.id
LEFT OUTER JOIN DefinedTermBase dtb ON dtb.id = dtbMN.DefinedTermBase_id
LEFT OUTER JOIN Reference_Extension rMN ON rMN.extensions_id = e.id
LEFT OUTER JOIN Reference r ON r.id = rMN.Reference_id
LEFT OUTER JOIN TaxonBase_Extension tbMN ON tbMN.extensions_id = e.id
LEFT OUTER JOIN TaxonBase tb ON tb.id = tbMN.TaxonBase_id
LEFT OUTER JOIN TaxonName_Extension nMN ON nMN.extensions_id = e.id
LEFT OUTER JOIN TaxonName n ON n.id = nMN.TaxonName_id
WHERE et.titleCache like '%Nomenclatural Standard%'
ORDER BY e.value, r.id
```

Done:

- common name reference language deduplicated

TODO:

- remove experts and species experts, check if same as sec ref now and adapt PESI import to use sec in case of E+M
- Berlin Model IdInSource currently only for Synonyms, where do they come from, can they be moved to alternative identifiers, why do accepted taxa not have this id, is the semantics always the same, aren't there similar fields in other BM tables?
- very few references have DateString extensions, this should be unified with similar information in Reference.notes, RefDetail notes, and other fields with similar information
- Nomenclatural Standard can often be handled as BPH or TL/2 (alternative) identifier. Check how "-" should be handled (remove

or does it mean no BPH or TL/2 entry exists)? what to do with BPH/S? Few exceptions exist.

- Check what Source_Acc means for names (74.702x, 4x for taxa)
- Handle IsoCode and TDWG code extension as term relationship, once this relationship exists

Associated revisions

Revision b06d238e - 06/03/2018 04:16 PM - Andreas Müller

ref #7446 add extension already exists check to IdentifiableEntity

Revision acc45005 - 06/03/2018 04:19 PM - Andreas Müller

ref #7446 deduplicate common name language references in E+M

Revision 62e9b5fa - 06/03/2018 05:05 PM - Andreas Müller

ref #7446 handle initials as initials in BM/EM import, not extension anymore

Revision 23b77c49 - 06/03/2018 09:53 PM - Andreas Müller

ref #7446 cleanup extension handling a bit

History

#2 - 06/03/2018 04:26 PM - Andreas Müller

- Description updated

- Estimated time set to 5.00 h

#3 - 06/03/2018 04:28 PM - Andreas Müller

- Description updated

#4 - 06/03/2018 04:36 PM - Andreas Müller

- Description updated

#5 - 06/03/2018 05:08 PM - Andreas Müller

- Description updated

#6 - 06/03/2018 05:21 PM - Andreas Müller

- Description updated

#7 - 06/03/2018 05:35 PM - Andreas Müller

- Description updated

#8 - 09/20/2018 03:55 PM - Andreas Müller

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