EDIT - task #7316

[E+M] Handle name relations correctly

03/16/2018 09:56 AM - Andreas Müller

Status:	Resolved	**Start date:**	
Priority:	Highest	**Due date:**	
Assignee:	Andreas Müller	**% Done:**	90%
Category:	data	**Estimated time:**	0:30 hour
Target version:	Euro+Med Portal Release		
Severity:	normal		

Description

Decision:

- we export ALL fromNames not yet in taxon export (ERS: also later homonyms might be of interest)
- we export all toNames with qualifier 2, 4, 5, 13, 14, 15, 17, 18, 37, 62 (4,5,13,14,37 should not exist anyway, 62 should be selfreferencing)

====

Currently in v_cdm_exp_namesRelatedTo we export all name relations of type

1 is basionym for => questionable as all new combinations should be also synonyms (but maybe similar names were mixed up sometimes?)
2 is later homonym of !! => anyway check if there is not maybe a duplicate
3 is replaced synonym for => questionable as all nom. nov. should be also synonyms (but maybe similar names were mixed up sometimes?)
9 is first parent of => questionable if hybrids of accepted taxa should be exported too if not being an explicit taxon
10 is second parent of => see 9)
16 is orthographic variant of => this should not happen as the taxa/synonyms should not be variants

The following relations we do not export because: looks like they all do not have data that do not yet exist as taxa

4 is validation of
5 is later validation of
6 is type of
7 is conserved type of
8 is rejected type of

11 is female parent of
12 is male parent of
13 is conserved against
14 is rejected in favour of
15 is treated as later homonym of
17 is alternative name for
18 has same type as
37 is later isonym of
61 is lectotype of
62 type not designated

```
SELECT rel.RelNameQualifierFk, count(*) as n
FROM RelName rel
-- WHERE
--(rel.NameFk1 IN (SELECT PTNameFk AS NameId
--                FROM  dbo.v_cdm_exp_taxaAll))
GROUP BY rel.RelNameQualifierFk
ORDER BY  rel.RelNameQualifierFk

SELECT n1.FullNameCache fromName, n1.Created_Who, n1.nameId, n2.NameCache toName, n2.FullNameCache
 toNameAuthor, n2.created_Who, n2.nameId, rel.RelNameId, rel.RefFk relRef, rel.RelNameQualifierFk,
  ta.RIdentifier, ta.PTRefFk
FROM RelName rel INNER JOIN v_cdm_exp_taxaAll ta ON ta.PTNameFk = rel.NameFk1
INNER JOIN Name n1 ON n1.nameId = rel.NameFk1
```

```
INNER JOIN Name n2 ON n2.nameId = rel.NameFk2
WHERE n2.nameId NOT IN (SELECT ta2.PTNameFk FROM v_cdm_exp_taxaAll ta2)
-- AND n2.NameCache IN (SELECT nn.NameCache FROM v_cdm_exp_taxaAll ta2 INNER JOIN Name nn ON ta2.P
TNameFk = nn.NameId )
ORDER BY RelNameQualifierFk, ta.PTRefFk, n1.FullNameCache
```

v_cdm_exp_namesRelatedFrom:

```
SELECT n1.FullNameCache fromName, n1.Created_Who, n1.nameId, n2.NameCache toName,  n2.
FullNameCache nomNovtoNameAuthor,  n2.created_Who, n2.nameId, rel.RelNameId, rel.RefFk relRef, rel
.RelNameQualifierFk, q.RelNameQualifier,  ta.RIdentifier, ta.PTRefFk
FROM RelName rel INNER JOIN v_cdm_exp_taxaAll ta ON ta.PTNameFk = rel.NameFk2
INNER JOIN Name n1 ON n1.nameId = rel.NameFk1
INNER JOIN Name n2 ON n2.nameId = rel.NameFk2
INNER JOIN RelNameQualifier q ON q.RelNameQualifierId = rel.RelNameQualifierFk
WHERE n1.nameId NOT IN (SELECT ta2.PTNameFk FROM v_cdm_exp_taxaAll ta2)
ta2.PTNameFk = nn.NameId )
-- AND RelNameQualifierFk NOT IN (2, 16)
-- AND RelNameQualifierFk IN (3)
ORDER BY RelNameQualifierFk, ta.PTRefFk, n1.FullNameCache
```

Potential homonym duplicates:

```
SELECT n1.FullNameCache laterHomonym, n1.Created_Who laterHomonymCreated, n1.nameId laterHomonymId
, n2.NameCache earlierName,  n2.FullNameCache earlierNameAuthor,  n2.created_Who
earlierNameCreated, n2.nameId earlierNameId, rel.RelNameId, rel.RefFk relRef, rel.
RelNameQualifierFk, q.RelNameQualifier,  ta.RIdentifier, ta.PTRefFk
FROM RelName rel INNER JOIN v_cdm_exp_taxaAll ta ON ta.PTNameFk = rel.NameFk2
INNER JOIN Name n1 ON n1.nameId = rel.NameFk1
INNER JOIN Name n2 ON n2.nameId = rel.NameFk2
INNER JOIN RelNameQualifier q ON q.RelNameQualifierId = rel.RelNameQualifierFk
WHERE n1.nameId NOT IN (SELECT ta2.PTNameFk FROM v_cdm_exp_taxaAll ta2)
 AND EXISTS (SELECT nn.NameCache FROM v_cdm_exp_taxaAll ta2 INNER JOIN Name nn ON ta2.PTNameFk =
nn.NameId WHERE nn.nameId <> n1.nameId AND nn.nameId <> n2.nameId AND nn.NameCache = n2.NameCache
AND nn.nameCache <> nn.FullNameCache)
 AND RelNameQualifierFk IN (2)
--  AND (n1.FullNameCache > 'Geranium laevigatum DC.' OR PTRefFk > 7400000) AND PTRefFk > 7300000
ORDER BY RelNameQualifierFk, ta.PTRefFk, n1.FullNameCache
```

Homonym candidates:

```
SELECT  NameId,   rd.FullNomRefCache, FullNameCache, pt.PTRefFk,  rd.FullRefCache, at.
AuthorTeamCache, at.FullAuthorTeamCache, RankFk, NameCache, UnnamedNamePhrase, n.PreliminaryFlag,
SupragenericName, Genus, GenusSubdivisionEpi, SpeciesEpi, InfraSpeciesEpi, AuthorTeamFk,
ExAuthorTeamFk,
                        BasAuthorTeamFk, ExBasAuthorTeamFk, HybridFormulaFlag, MonomHybFlag,
BinomHybFlag, TrinomHybFlag, CultivarGroupName, CultivarName, NomRefFk, NomRefDetailFk,
NameSourceRefFk, Source_Acc,
                        n.Created_When, n.Created_Who, n.Notes, ParsingComments, OldNomRefFk,
OldNomRefDetailFk, n.Updated_Who, OrthoProjection, HybridFormula, TempKewId, n.uuid
FROM            Name n
LEFT OUTER JOIN RefDetail rd ON rd.RefDetailId = n.NomRefDetailFk
LEFT OUTER JOIN AuthorTeam at ON at.AuthorTeamId = n.AuthorTeamFk
LEFT OUTER JOIN PTaxon pt ON pt.PTNameFk = n.NameId
WHERE        (NameId IN (-1)) OR
                        (FullNameCache LIKE 'Taraxacum laciniatum%')

ORDER BY FullNameCache, PTRefFk, NameID
```

Update:

```
SELECT        TOP (200) RelNameId, NameFk1, NameFk2, RelNameQualifierFk, RefFk, RefDetailFk,
Created_When, Updated_When, Created_Who, Updated_Who, Notes
FROM            RelName
WHERE        (RelNameId IN (0)) OR
                        (NameFk2 IN (7534221)) OR
                        (NameFk1 IN (0))
```

Alternative name:

Der eine "is alternative name for" Fall wird in #7317 abgehandelt.

Suche von ungehandelten Fällen im Import Log:

```
SELECT         NameId as AllNames
FROM             v_cdm_exp_namesAll
WHERE         (NameId IN (95621, 7300008, 58609));

SELECT         NameId as relFrom, toName
FROM             v_cdm_exp_namesRelatedFrom v
WHERE         (NameId IN (95621, 7300008, 58609));

SELECT         NameId as relTo , rel.RelNameId relId, rel.NameFk1 fromId, rel.RelNameQualifierFk
type
FROM             v_cdm_exp_namesRelatedto v LEFT OUTER JOIN RelName rel ON rel.NameFk2 = v.NameId
WHERE         (NameId IN (95621, 58609));

SELECT         NameId as commonName, v.NameId
FROM             v_cdm_exp_namesCommonNameSource v
WHERE         (NameId IN (95621, 58609));

SELECT         ta.PTNameFk AS taxon, ta.PTRefFk, pt.StatusFk
FROM           dbo.v_cdm_exp_taxaAll ta LEFT OUTER JOIN PTaxon pt ON pt.RIdentifier = ta.
RIdentifier
WHERE      ta.PTNameFk IN (95621, 58609, 95623)
;

SELECT *
FROM Name
WHERE NameID IN (95621, 58609, 7300008);

SELECT *
FROM RelName rel
WHERE rel.NameFk1 IN (7300008, 7300009, 58609) OR rel.NameFk1 IN (7300008, 7300009, 58609)
```

Associated revisions

Revision 09539dc4 - 06/02/2018 12:49 PM - Andreas Müller

fix #7316 final fix for name relations import

History

#1 - 03/16/2018 09:57 AM - Andreas Müller

- Description updated

#2 - 03/16/2018 12:31 PM - Andreas Müller

- Description updated

#3 - 03/16/2018 01:27 PM - Andreas Müller

- Description updated

#4 - 03/16/2018 01:50 PM - Andreas Müller

- Description updated

#5 - 03/16/2018 02:10 PM - Andreas Müller

- Description updated

#6 - 03/16/2018 02:11 PM - Andreas Müller

- Description updated

#7 - 03/16/2018 02:13 PM - Andreas Müller

- Subject changed from [E+M] Select correct name relations for export to [E+M] Handle name relations correctly

#8 - 03/16/2018 02:53 PM - Andreas Müller

- Description updated

#9 - 03/16/2018 02:53 PM - Andreas Müller

- Description updated

#10 - 03/16/2018 02:54 PM - Andreas Müller

- Description updated

#11 - 03/19/2018 05:22 PM - Andreas Müller

- Description updated

#12 - 03/19/2018 05:28 PM - Andreas Müller

- Description updated

#13 - 03/19/2018 06:06 PM - Andreas Müller

- Description updated

#14 - 03/19/2018 06:10 PM - Andreas Müller

- Description updated

#15 - 03/19/2018 06:13 PM - Andreas Müller

- Description updated

#16 - 03/23/2018 09:56 PM - Andreas Müller

- Status changed from New to In Progress

#17 - 03/23/2018 09:56 PM - Andreas Müller

- % Done changed from 0 to 20

#18 - 05/09/2018 03:21 PM - Andreas Müller

- % Done changed from 20 to 70

- Estimated time set to 10:00 h

#20 - 05/25/2018 12:23 PM - Andreas Müller

- Estimated time changed from 10:00 h to 2:00 h

#21 - 05/25/2018 12:23 PM - Andreas Müller

- Priority changed from New to Highest

#22 - 05/30/2018 01:45 PM - Andreas Müller

- Description updated

#23 - 05/30/2018 01:45 PM - Andreas Müller

- Description updated

#24 - 06/02/2018 12:10 PM - Andreas Müller

- Description updated

#25 - 06/02/2018 12:10 PM - Andreas Müller

- % Done changed from 70 to 90

- Estimated time changed from 2:00 h to 0:30 h

#26 - 06/02/2018 12:48 PM - Andreas Müller

- Description updated

#27 - 06/02/2018 12:48 PM - Andreas Müller

- Description updated

#28 - 06/02/2018 12:50 PM - Andreas Müller

- Status changed from In Progress to Resolved

- % Done changed from 90 to 50

Applied in changeset cdmlib-apps|09539dc4fba8474bb74c0d0f67fac92412e708f9.

#29 - 06/02/2018 12:51 PM - Andreas Müller

- % Done changed from 50 to 90

#30 - 06/03/2018 07:41 PM - Andreas Müller

- Description updated