EDIT - task #9291

Clear all unnecesary TypeDesignation.citation entities and citationDetails

11/10/2020 05:04 PM - Andreas Kohlbecker

Status:	Closed		**Start date:**		
Priority:	Highest		**Due date:**		
Assignee:	Andreas Kohlbecker		**% Done:**		100%
Category:	data		**Estimated time:**		0:00 hour
Target version:					
Severity:	normal				

Description

This ticket should be solved after #9290 is being released and deployed to phycobank

This is the query to which has been used to create the report for #9290. The update statement can be based on that:

```
SELECT
    tn.uuid as TaxonName_uuid, tn.nameCache as TaxonName_nameCache,
    tn_r.titleCache as TaxonName_nomRef, tn_r.refType as TaxonName_nomRef_type, tn_r.uuid as
TaxonName_nomRef_uuid,
    ts.titleCache as TypeStatus_label, ts.id as TypeStatus_id,
    tdb_r.titleCache as TypeDesignation_citation, tdb_r.refType as TypeDesignation_citation_type,
tdb_r.uuid as TypeDesignation_citation_uuid,
    tdb.uuid as TypeDesignation_uuid
FROM TypeDesignationBase tdb
LEFT JOIN DefinedTermBase ts ON tdb.typeStatus_id = ts.id
JOIN TaxonName_TypeDesignationBase tntdb on tdb.id = tntdb.typedesignations_id
JOIN TaxonName tn on tn.id = tntdb.TaxonName_id
LEFT JOIN Reference tn_r on tn_r.id = tn.nomenclaturalReference_id
LEFT JOIN Reference tdb_r on tdb_r.id = tdb.citation_id
WHERE tdb.citation_id IS NOT NULL
    AND ts.id NOT IN (853 ,839 ,840 ,843 ,850 ,846 ,852 ,847 ,839 ,2108 ,2110 ,2111 ,2112)
    AND tdb.citation_id <> tn.nomenclaturalReference_id
ORDER BY tdb_r.uuid
```

Related issues:

Related to EDIT - bug #9290: stop setting Registration.citation as citation f...	**Closed**

History

#1 - 11/10/2020 05:05 PM - Andreas Kohlbecker

- Related to bug #9290: stop setting Registration.citation as citation for TypeDesignations in all type editors by default added

#2 - 11/13/2020 04:17 PM - Andreas Kohlbecker

- File TaxonName and Typedesignation citations unequal_diff.xlsx added

- Description updated

#3 - 11/13/2020 04:36 PM - Andreas Kohlbecker

the Excel file belongs to the other ticket - cleaning up here ....

#4 - 11/13/2020 04:36 PM - Andreas Kohlbecker

- File deleted (TaxonName and Typedesignation citations unequal_diff.xlsx)

#5 - 01/27/2021 03:14 PM - Andreas Kohlbecker

- Target version changed from Release 5.19 to Release 5.21

#6 - 02/01/2021 10:54 AM - Andreas Kohlbecker

- Priority changed from Priority14 to Highest

#7 - 03/10/2021 02:55 PM - Andreas Müller

- Target version changed from Release 5.21 to Release 5.22

#8 - 05/03/2021 02:10 PM - Andreas Kohlbecker

select statement simplified and updated to current model version

```
SELECT
    tn.uuid as TaxonName_uuid, tn.nameCache as TaxonName_nameCache,
    ts.titleCache as TypeStatus_label, ts.id as TypeStatus_id,
    tdb_r.titleCache as TypeDesignation_citation, tdb_r.refType as TypeDesignation_citation_type, tdb_r.uuid
as TypeDesignation_citation_uuid,
    tdb.uuid as TypeDesignation_uuid
FROM TypeDesignationBase tdb
LEFT JOIN DefinedTermBase ts ON tdb.typeStatus_id = ts.id
JOIN TaxonName_TypeDesignationBase tntdb on tdb.id = tntdb.typedesignations_id
JOIN TaxonName tn on tn.id = tntdb.TaxonName_id
LEFT JOIN OriginalSourceBase osb on osb.id = tdb.designationSource_id
LEFT JOIN Reference tdb_r on tdb_r.id = osb.citation_id
WHERE osb.citation_id IS NOT NULL
    AND ts.id NOT IN (853 ,839 ,840 ,843 ,850 ,846 ,852 ,847 ,839 ,2108 ,2110 ,2111 ,2112)
ORDER BY ts.titleCache
```

#9 - 05/03/2021 02:33 PM - Andreas Kohlbecker

All type designations that do not have a state like below must not have a designationSource:

```
// NameTypeDesignationStatus
LECTOTYPE
SUBSEQUENT_DESIGNATION
SUBSEQUENT_MONOTYPY
PRESENT_DESIGNATION
// SpecimenTypeDesignationStatus
SECOND_STEP_NEOTYPE
SECOND_STEP_LECTOTYPE
PARALECTOTYPE
ISOEPITYPE
ISONEOTYPE
ISOLECTOTYPE
EPITYPE
NEOTYPE
LECTOTYPE

select DTYPE, id, titleCache from DefinedTermBase dtb
    where id in (853 ,839 ,840 ,843 ,850 ,846 ,852 ,847 ,848 ,839 ,2108 ,2110 ,2111 ,2112)
```

#10 - 05/03/2021 02:58 PM - Andreas Kohlbecker

- Status changed from New to Feedback

- Assignee changed from Andreas Kohlbecker to Andreas Müller

- % Done changed from 0 to 40

@Andreas M.: Could you please double check this sql script?

```
CREATE TEMPORARY TABLE temp_ds_cleanup(
    type_designation_id int(11),
    original_ource_id int(11)
);

INSERT INTO temp_ds_cleanup(type_designation_id, original_ource_id)
    SELECT tdb.id, osb.id FROM TypeDesignationBase tdb
    LEFT JOIN OriginalSourceBase osb on osb.id = tdb.designationSource_id
    WHERE tdb.designationSource_id IS NOT NULL
        AND tdb.typeStatus_id NOT IN (853 ,839 ,840 ,843 ,850 ,846 ,852 ,847 ,848 ,839 ,2108 ,2110 ,2111 ,2112
);

-- SELECT * FROM temp_ds_cleanup;

UPDATE TypeDesignationBase tdb SET tdb.designationSource_id = NULL
    WHERE tdb.id IN (select tdc.type_designation_id FROM temp_ds_cleanup AS tdc);


DELETE FROM OriginalSourceBase WHERE id in (select tdc.original_ource_id FROM temp_ds_cleanup AS tdc);
```

```
DROP TEMPORARY TABLE temp_ds_cleanup;
```

#11 - 05/03/2021 11:34 PM - Andreas Müller

- Assignee changed from Andreas Müller to Andreas Kohlbecker

If the condition from #note-9 is correct the query looks ok except for the fact that it does not include AUD tables which makes AUD and NON-AUD tables inconsistent.

Probably it is enough to run the UPDATE and the DELTE statement also on AUD tables and therefore pretend that the false citations never existed. But you know the specifics of this ticket better and maybe I am wrong.

#12 - 05/04/2021 01:24 PM - Andreas Kohlbecker

- Assignee changed from Andreas Kohlbecker to Andreas Müller

wouldn't it be anyway much better to do such cleanup tasks in general via the DAO layer?

#13 - 05/04/2021 01:36 PM - Andreas Müller

- Assignee changed from Andreas Müller to Andreas Kohlbecker

Andreas Kohlbecker wrote:

> wouldn't it be anyway much better to do such cleanup tasks in general via the DAO layer?

This depends from case to case. In general I agree that if possible using hibernate and not SQL is preferred to be on the save side with AUD tables and other caching issues. So, of cours, you could also write a script using service or persistence layer.
In some cases, like model updates, it is very difficult to do so as the model is out-of-sync then.

For the given case it is definetely possible to write such a script if you want.

#14 - 05/04/2021 03:00 PM - Andreas Kohlbecker

- Assignee changed from Andreas Kohlbecker to Andreas Müller

Ok, for model updates it may be problematic. I rather thought on all other taks, like data cleaning and so on, when I said "in general".
What is your personal policy for these taks? Do you accept inconsistent AUD tables or do you write java clean up apps in almost all cases?

I am asking since we should have a common strategy.

#15 - 05/08/2021 01:35 AM - Andreas Müller

- Assignee changed from Andreas Müller to Andreas Kohlbecker

- Target version changed from Release 5.22 to Release 5.25

we have discussed this

#16 - 05/31/2021 04:58 PM - Andreas Kohlbecker

- Status changed from Feedback to Closed

- % Done changed from 40 to 100

1760 cases cleaned

#17 - 06/01/2021 10:18 AM - Andreas Müller

- Target version changed from Release 5.25 to Release 5.24

#18 - 06/01/2021 06:54 PM - Andreas Müller

- Target version deleted (Release 5.24)

"Data" tickets are usually not related to a target version once they are fixed. Therefore I remove it.