

SQL Queries for authorities in Evergreen

--- Bib records with unauthorized subject headings, useful even if authority_control_fields.pl has not been run recently, but it's pretty slow

```
SELECT bib_heading FROM (
SELECT DISTINCT
public.naco_normalize(unnest(
xpath('//m:datafield[@tag="650" and
@ind2="0"]/m:subfield[@code="a"]/text()', marc::xml,
ARRAY[ARRAY['m','http://www.loc.gov/MARC21/slim']]))::varchar[]))
AS bib_heading
FROM biblio.record_entry
WHERE NOT deleted
AND create_date > (now() - interval '1 month')) bibs
WHERE bib_heading NOT IN (SELECT DISTINCT sort_value
FROM authority.simple_heading)
AND bib_heading NOT LIKE '%fictitious character%'
LIMIT 100;
```

--- Recently modified records with a particular thesaurus

```
SELECT sort_value FROM authority.record_entry
INNER JOIN authority.simple_heading
ON simple_heading.record=record_entry.id
WHERE thesaurus='L'
AND NOT record_entry.deleted
AND record_entry.edit_date < (CURRENT_DATE - interval '2
months')
ORDER BY heading ASC;
```

--- Bib records with unauthorized name headings, only useful if authority_control_fields.pl has been run recently and its links are up to date

```
SELECT DISTINCT ent.value FROM metabib.browse_entry
ent, metabib.browse_entry_def_map map,
config.metabib_field f, biblio.record_entry b
WHERE ent.id=map.entry
AND map.def=f.id AND map.source=b.id AND NOT b.deleted
AND f.field_class = 'author'
AND ent.id NOT IN
(SELECT DISTINCT entry FROM
metabib.browse_entry_def_map WHERE authority IS NOT
NULL);
```

--- Duplicate authority records

```
SELECT dups.id, heading AS duplicate_authority_heading,
occurrence,
rec_descriptor.thesaurus from (
SELECT id, heading,
ROW_NUMBER() OVER(PARTITION BY heading ORDER BY
id asc) AS occurrence
FROM authority.record_entry
WHERE NOT deleted
) dups
INNER JOIN authority.rec_descriptor on
rec_descriptor.record=dups.id
WHERE occurrence > 1
AND occurrence < 4
ORDER BY occurrence DESC
LIMIT 20;
```