Remapping Problematic LCSH in TRLN Discovery

Kristina M. Spurgin

2018-11-15





library.unc.edu

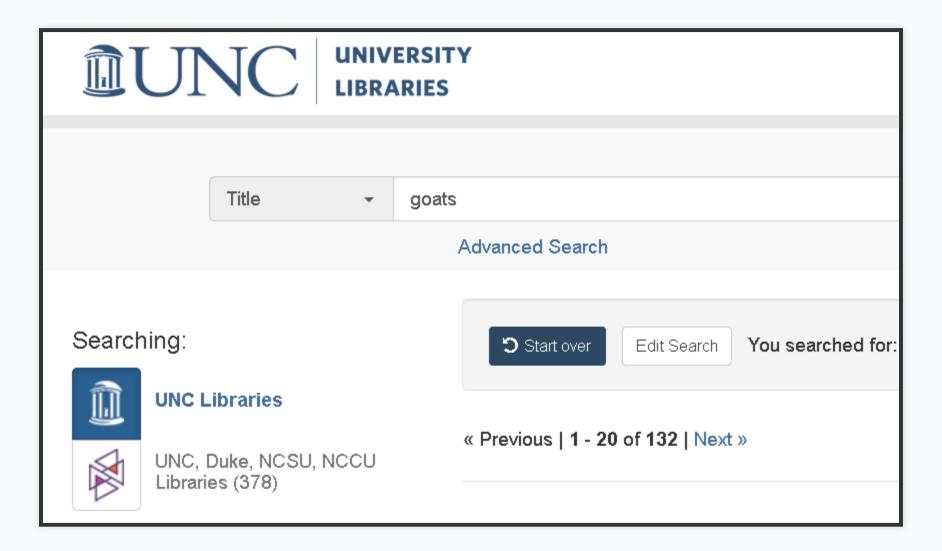
Introduction (About me)

- Library Data Strategist, Head of Library Data Strategy & Services
- TRLN Discovery Developer Team

Introduction (About the shared online catalog)

- TRLN initiative since 2007
- Legacy to now Endeca
 - http://search.trln.org/
 - https://search.lib.unc.edu/
 - https://search.lib.unc.edu/filmfinder/
- Coming soon TRLN Discovery
 - https://disco-qa.lib.unc.edu/

"Expand to TRLN"



TRLN Discovery: Shared index: Solr

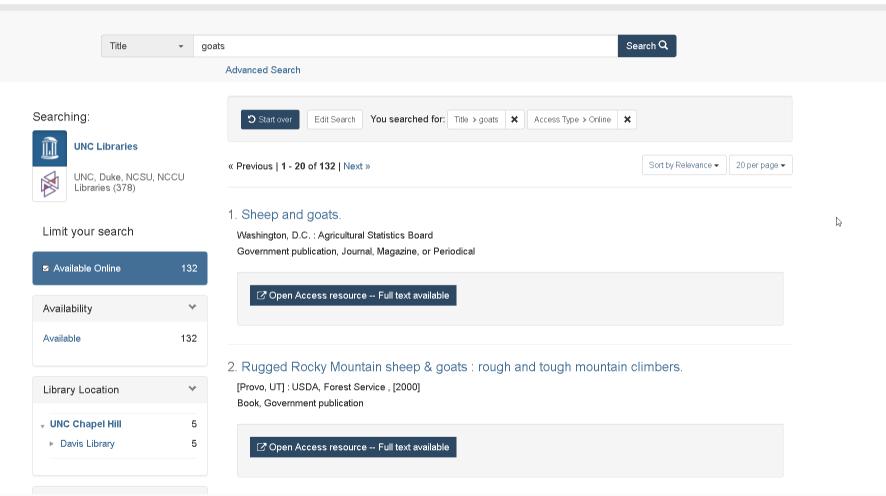
- Solr
 - open source
 - index + search

```
ISON
       Raw Data
                  Headers
Save Copy Pretty Print
  "responseHeader":{
    "zkConnected":true.
    "status":0.
    "OTime": 357,
    "params":{
      "q":"title main t:goats AND access type f:Online AND institution a:unc"}},
  "response":{"numFound":129, "start":0, "docs":[
        "id": "UNCb3547657",
        "url a":["{\"href\":\"http://purl.access.gpo.gov/GPO/LPS4781\",\"type\":\"fulltext\",\"text\":\"Open Access resource -- Full text available\"}"],
        "author suggest":["United States. Agricultural Statistics Board",
          "United States. National Agricultural Statistics Service"],
        "names a":["{\"name\":\"United States. Agricultural Statistics Board\"}",
          "{\"name\":\"United States. National Agricultural Statistics Service\"}"],
        "owner a":["unc"],
        "misc_id_a":["LCCN: 00230086",
                                                                                            1
          "GPO Item Number: 0021-N (online)"],
        "language_a":["English"],
        "local_id":"b3547657",
        "available a":["Available"],
        "frequency current a":["Annual"].
        "lang code a":["eng"].
        "rollup_id":"0CLC44616828",
        "this_work_a":["{\"title\":[\"Sheep and goats (Online)\"]}"],
        "title_suggest":["Sheep and goats."],
        "title_main": "Sheep and goats.",
        "title sort_ssort_single": "Sheep and goats.",
        "access_type_a":["Online"],
        "institution_a":["unc",
          "duke",
          "nccu".
          "ncsu"l.
        "oclc number": "44616828",
        "imprint main a":["{\"type\":\"imprint\",\"value\":\"Washington, D.C. : Agricultural Statistics Board\"}"],
        "note_general_a":["\"MtAn 5-2,\" Jan. 29, 1999-",
          "Description based on: Jan. 1, 1995; title from screen caption.",
          "Latest issue consulted: Feb. 1, 2008."],
        "note summary a":["This full-text file contains the number of sheep, value and classes; Angora goat and all goat inventory and value; lamb crop;
number of operations keeping sheep; operations and inventory by size groups, regions, and U.S.; total number on feed for slaughter. Statistics by state,
region or U.S."1.
```

TRLN Discovery: Shared catalog application: Blacklight

- Blacklight
 - open source
 - library-created
 - discovery interface





```
ISON
       Raw Data
                  Headers
Save Copy Pretty Print
  "responseHeader":{
    "zkConnected":true.
    "status":0.
    "OTime": 357,
    "params":{
      "q":"title main t:goats AND access type f:Online AND institution a:unc"}},
  "response":{"numFound":129, "start":0, "docs":[
        "id": "UNCb3547657",
        "url a":["{\"href\":\"http://purl.access.gpo.gov/GPO/LPS4781\",\"type\":\"fulltext\",\"text\":\"Open Access resource -- Full text available\"}"],
        "author suggest":["United States. Agricultural Statistics Board",
          "United States. National Agricultural Statistics Service"],
        "names a":["{\"name\":\"United States. Agricultural Statistics Board\"}",
          "{\"name\":\"United States. National Agricultural Statistics Service\"}"],
        "owner a":["unc"],
        "misc_id_a":["LCCN: 00230086",
                                                                                            1
          "GPO Item Number: 0021-N (online)"],
        "language_a":["English"],
        "local_id":"b3547657",
        "available a":["Available"],
        "frequency current a":["Annual"].
        "lang code a":["eng"].
        "rollup_id":"0CLC44616828",
        "this_work_a":["{\"title\":[\"Sheep and goats (Online)\"]}"],
        "title_suggest":["Sheep and goats."],
        "title_main": "Sheep and goats.",
        "title sort_ssort_single": "Sheep and goats.",
        "access_type_a":["Online"],
        "institution_a":["unc",
          "duke",
          "nccu".
          "ncsu"l.
        "oclc number": "44616828",
        "imprint main a":["{\"type\":\"imprint\",\"yalue\":\"Washington, D.C. : Agricultural Statistics Board\"}"],
        "note_general_a":["\"MtAn 5-2,\" Jan. 29, 1999-",
          "Description based on: Jan. 1, 1995; title from screen caption.",
          "Latest issue consulted: Feb. 1, 2008."],
        "note summary a":["This full-text file contains the number of sheep, value and classes; Angora goat and all goat inventory and value; lamb crop;
number of operations keeping sheep; operations and inventory by size groups, regions, and U.S.; total number on feed for slaughter. Statistics by state,
region or U.S."1.
```

TRLN Discovery: Shared data model: Argot

- Argot is our name for our shared data model
 - MARC -> Argot
 - Digital collections -> Argot
 - ICPSR DDI metadata -> Argot
- Argot specification is publicly available

Representing subject data in Argot

- subject_headings
- subject_suggest
- subject_topical
- subject_chronological
- subject_geographic
- subject_headings_remapped

Subject data in Argot: subject_headings

- indexed for keyword and subject search
- displayed in full record view with adaptive hyperlinking

=650 \0\$aAbolitionists\$zNorth Carolina\$xHistory\$y19th century\$vSources.

```
"subject_headings": [
     "Abolitionists -- North Carolina -- History -- 19th century -- Sources"
]
```

This item is about

- Antislavery movements > North Carolina > Sources
- Abolitionists > North Carolina > History >
 19th century > Sources

Example mappings: subject_headings

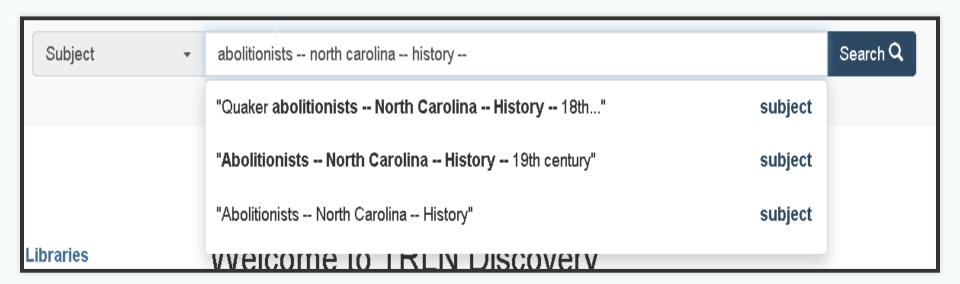
element/field 🕞	subelement/subfield(s)	constraints	processing_type
653	a	i2 =~ /[012345]/	subelement_to_value
662	abcdfgh	none	concat_subelements
656	akvxyz	none	concat_subelements
657	avxyz	none	concat_subelements
600	abcdfghjklmnopqrstu vxy z	none	concat_subelements
610	abcdfghklmnoprstuvxyz	none	concat_subelements
650	abcdgvxyz	none	concat_subelements
611	acdefghklnpqstuvxyz	none	concat_subelements
647	acdgvxyz	none	concat_subelements
630	adfghklmnoprstvxyz	none	concat_subelements
651	agvxyz	none	concat_subelements
648	avxyz	none	concat_subelements

Subject data in Argot: subject_suggest

- searched when you start typing in keyword or subject search box
- displayed in auto-suggest dropdown

=650 \0\$aAbolitionists\$zNorth Carolina\$xHistory\$y19th century\$vSources.

```
"subject_suggest": [
   "Abolitionists -- North Carolina -- History -- 19th century -- Sources"
]
```



Subject data in Argot: subject_topical

populates "About Topic" facet

=650 \0\$aAbolitionists\$zNorth Carolina\$xHistory\$y19th century\$vSources.

```
"subject_topical": [
    "Abolitionists", "History"
]
```

About Topic	*
Abolitionists	47
History	35
Antislavery movements	28
Slavery	23
African Americans	11
Fugitive slaves	9
Quakers	9
Slaves	8
Hedrick, Benjamin Sherwood, 1827-1886	7
Politics and government	7
more »	

Example mappings: subject_topical

element/field 🗊	subelement/subfield(s)	constraints	processing_type
600	abcdfghjklmnopqrstu	none	concat_subelements
600	x	none	subelement_to_value
610	abcdfghklmnoprstu	none	concat_subelements
610	X	none	subelement_to_value
611	acdefghkInpqstu	none	concat_subelements
611	х	none	subelement_to_value
630	adfghklmnoprst	none	concat_subelements
630	х	none	subelement_to_value
647	acdg	none	concat_subelements
647	х	none	subelement_to_value
648	х	none	subelement_to_value
650	abcdg	none	concat_subelements
650	х	none	subelement_to_value
651	X	none	subelement_to_value

	653 a	i2 = 4	subelement_to_value
ı	653 a	i2 =~ /[0123]/	subelement_to_value
ı	656 a	none	subelement_to_value
ı	656 x	none	subelement_to_value

Subject data in Argot: subject_chronological, subject_geographic

populates "About Time Period" and "About Places" facets

=650 \0\$aAbolitionists\$zNorth Carolina\$xHistory\$y19th century\$vSources.

```
"subject_chronological": [
    "19th century"
],
"subject_geographic": [
    "North Carolina"
]
```

About Time Period	*
19th century	22
Civil War, 1861-1865	7
1775-1865	4
1865-1950	3
1861 - 1865	2
18th century	2
1861-1865	1
John Brown's Raid, 1859	1
Revolution, 1775-1783	1

About Places	*
North Carolina	52
United States	27
Southern States	5
Ohio	4
Arkansas	2
Cabarrus County (N.C.)	2
Confederate States of America	2
Cottage Home Plantation (Lin- coln County, N.C.)	2
Dallas County (Ala.)	2
Indiana	2
more »	

Subject data in Argot: subject_headings_remapped

- NOT displayed in record, facets, or auto-suggest
- indexed for keyword and subject search

=650 \0\$alllegal aliens\$zEurope.

```
"subject_topical": [
    "Undocumented immigrants"
],
    "subject_headings": [
        "Undocumented immigrants -- Europe"
],
    "subject_headings_remapped": [
        "Illegal aliens -- Europe"
]
```

This item is about

Undocumented immigrants > Europe

Undocumented immigrants > Senegal

Behavior in TRLN Discovery

- Search for subject: "illegal aliens"
 - 1993 results; not seen in About Topic facet, record; autosuggest issue
- Search for subject: "undocumented immigrants"
 - 1974 results; seen in About Topic facet, record

Not perfect yet... Proof of concept stage! Some weird glitches to look into

What to remap and what to map it to?

- Initial list for proof of concept
- Not complete
- Need TRLN-wide agreement
- Governance for this still undecided
- Your input needed

So what?

- This doesn't make our catalog unbiased or fully inclusive (c.f. Emily Drabinski)
- Can be seen as confusing to users Is CU's approach better?
- Technically, not very difficult
- Small step = good step?