Automating the Authority Control Process

Stacey Wolf
Cataloging and Metadata Librarian
May 13, 2019
Outline:

• Me and my library
• Different ways to automate authority control processes
• Some ideas on how to automate the authority control maintenance process
• Tips
• Resources and Questions
About Me

6 years at UNT

Started as a para-professional staff copy cataloger

- juvenile books and authority control
- batch e-resource record editing and loading

Promoted to Cataloging and Metadata Librarian in 2016

- supervise the staff person who does most of our authority work
- working towards using an outside vendor
University of North Texas
Denton, Texas

Map source: google maps
University of North Texas Libraries' Catalog

<table>
<thead>
<tr>
<th>Sierra ILS</th>
<th>Nearly 3 million records</th>
<th>Add about 100,000 records annually</th>
</tr>
</thead>
<tbody>
<tr>
<td>Decentralized cataloging</td>
<td>7 catalogers do their own authority work</td>
<td>1 person does all of the library's authority work</td>
</tr>
</tbody>
</table>
Local authority control priorities (Then)

• Search OCLC for every new name and subject
• Download the authority record to the correct index if it was established and it had a cross reference or see also reference (4xx and 5xx)
• Validate new names and subjects
• Update authority records with closed dates and LCSH updates
Local authority control priorities (Now)

• Utilize the "preceded by" and "followed by" lines to see if there is a potential conflict and only investigate those.
• Don't load new subject headings records.
• Don't update authority records as the national authority file changes.
• General collection catalogers do their own authority work.
Automate!

MARCIVE® INC.

sierra

Notepad++ v7.5.9

Backstage LIBRARY WORKS

MarcEdit 7.1.102 By Terry Reese

Tools  Plug-ins  Help

I would like to:

- Edit a MARC File
- Translate Excel to MARC
- Harvest OAI Data
- Work with Linked Data
- Merge MARC Files
- Recently Run Programs

MARC Tools
Delimited Text Translator
Harvest OAI Records
MARC SQL Explorer
Z39.50/SRU Client
Sierra (or your local ILS)

Heading Reports can identify new names and subjects, invalid uses, and duplicate records.

Bibliographic Maintenance feature automatically updates bib records to a match new heading over night.
MarcEdit

- validate names and subjects to the right format before ever loading records
- download authority records
- add URI to $0
Vendors

**Automated process for**
- cleaning up records,
- validating headings, and
- providing the library with authority records

**On-going maintenance**
- process new bibliographic records
- notify the library if authorities change or are newly created from backfile
Batch-loading New Authority Records
Load records into Sierra, run Heading Reports, copy new names to Excel
| 1112 | aAMS Special Session on Spectral Calculus |
| 1112 | aResearch Workshop of the Israel Science and Technology Council |
| 7001 | aBeceanu, Marius, editor. |
| 7001 | aChen, Geng, d1980- editor. |
| 7001 | aKrause, Dan. |
| 7001 | aLee, Ellie, d1970-4pro, 4drt |
| 7001 | aMiller, Florence, 4nrt |
| 7001 | aPhan, Tuoc, q(Tuoc Van), editor. |
| 7001 | aPlotkin, Eugene, d1955- editor. |

Clean up copied information so just the name remains
Validate names in MarcEdit, adding $0

- $aBeceanu, Marius, $eeditor.
- $aChen, Geng, $d1980-$eeditor.
- $aKrause, Dan.
- $aLee, Ellie, $d1970-$4pro$4drt
- $aMiller, Florence, $4nrt
- $aPhan, Tuoc$q(Tuoc Van), $eeditor.
- $aPlotkin, Eugene, $d1955-$eeditor.
$0 was found 17 times.

<table>
<thead>
<tr>
<th>Found Text</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>=1001 $aDonner, Ezra,$d1986-$ecomposer,$0h...</td>
<td>Jump to Record #: 0</td>
</tr>
<tr>
<td>=1001 $aGilheany, Kevin P.,$d1963-$author,$0...</td>
<td>Jump to Record #: 0</td>
</tr>
<tr>
<td>=1001 $aSeitz, Paul,$d1951-$ecomposer,$0http...</td>
<td>Jump to Record #: 0</td>
</tr>
<tr>
<td>=1112 $aResearch Workshop of the Israel Sciences...</td>
<td>Jump to Record #: 0</td>
</tr>
<tr>
<td>=7001 $aBeceau, Marius,$eseeditor,$0<a href="http://id.loc.gov">http://id.loc.gov</a>...</td>
<td>Jump to Record #: 0</td>
</tr>
<tr>
<td>=7001 $aChen, Geng,$d1980-$eseeditor,$0<a href="http://id.loc.gov">http://id.loc.gov</a>...</td>
<td>Jump to Record #: 0</td>
</tr>
<tr>
<td>=7001 $aLee, Ellie,$d1970-$4pro$4drt$0<a href="http://id.loc.gov">http://id.loc.gov</a>...</td>
<td>Jump to Record #: 0</td>
</tr>
<tr>
<td>=7001 $aMiller, Florence,$4nrt$0<a href="http://id.loc.gov">http://id.loc.gov</a>...</td>
<td>Jump to Record #: 0</td>
</tr>
<tr>
<td>=7001 $aPhan, Tuoc$sq(Tuoc Van),$eseeditor,$0http...</td>
<td>Jump to Record #: 0</td>
</tr>
<tr>
<td>=7001 $aPlotkin, Eugene,$d1955-$eseeditor,$0http...</td>
<td>Jump to Record #: 0</td>
</tr>
<tr>
<td>=7001 $aRodriguez, Maria Teresa$sc(Filmmaker)...</td>
<td>Jump to Record #: 0</td>
</tr>
<tr>
<td>=7001 $aSoffer, Avy$sq(Avraham),$eseeditor,$0http...</td>
<td>Jump to Record #: 0</td>
</tr>
<tr>
<td>=7001 $aTumanov, Vladimir,$eseeditor,$esranslate...</td>
<td>Jump to Record #: 0</td>
</tr>
<tr>
<td>=7001 $aWeaver, Jonathan,$4cng$0<a href="http://id.loc.gov">http://id.loc.gov</a>...</td>
<td>Jump to Record #: 0</td>
</tr>
<tr>
<td>=7001 $aZheng, Shijun,$eseeditor,$0<a href="http://id.loc.gov">http://id.loc.gov</a>...</td>
<td>Jump to Record #: 0</td>
</tr>
<tr>
<td>=7102 $aNational Minority Consortia (U.S.),$0http...</td>
<td>Jump to Record #: 0</td>
</tr>
<tr>
<td>7102 $aOne World Group of Broadcasters,$eoo</td>
<td>Jump to Record #: 0</td>
</tr>
</tbody>
</table>
Compare against new names from Heading Report

Donner, Ezra, $d1986-$ecomposer.
Gilheany, Kevin P., $d1963-$eauthor.
Seitz, Paul, $d1951-$ecomposer.

Research Workshop of the Israel Science Foundation on Group
Beceau, Marius, $eeditor.
Chen, Geng, $d1980-$ecomposer.

Krause, Dan.
Lee, Ellie, $d1970-$4pro$4drt
Miller, Florence, $4nrt
Phan, Tuco$qq(Tuoc Van), $eeditor.
Plotkin, Eugene, $d1955-$ecomposer.
Rodriguez, Maria Teresa$Sc(Filmmaker)
Soffer, Avy$qq(Avraham), $eeditor.
Tumanov, Vladimir, $ecomposer.$etranslator
Weaver, Jonathan.$4ng
Zheng, Shijun,$eeditor.

National Minority Consortia (U.S.)
One World Group of Broadcasters,$eproduction company.

AMS Special Session on PDE Analysis on Fluid Flows$d(2017; $cAtlanta, Georgia),$eauthor.
Batch search OCLC for LCCNs and export the record to ILS
Keeping Current

Maintaining authority files
Authority records for 20190428

Authority Records updates posted after January 19, 2008 require Arial Unicor software packages or can be purchased separately. [Learn more about Arial I

<table>
<thead>
<tr>
<th>n</th>
<th>Old</th>
<th>New</th>
</tr>
</thead>
<tbody>
<tr>
<td>83218565</td>
<td>100 1 Akhmatova, R. S.</td>
<td>100 1 Akhmatova, R. S.</td>
</tr>
<tr>
<td></td>
<td>#d 1928-</td>
<td>#d 1928-1992</td>
</tr>
<tr>
<td>81059040</td>
<td>100 1 Amiss, John Milton</td>
<td>100 1 Amiss, John M.</td>
</tr>
<tr>
<td></td>
<td>#d 1867-</td>
<td>#d 1887-1968</td>
</tr>
<tr>
<td>93064201</td>
<td>100 1 Anderson, Marie</td>
<td>100 1 Anderson, Marie</td>
</tr>
<tr>
<td></td>
<td>#d 1916-</td>
<td>#d 1916-1996</td>
</tr>
<tr>
<td>95118310</td>
<td>100 1 Arnold, Johann Christoph</td>
<td>100 1 Arnold, Johann Christoph</td>
</tr>
<tr>
<td></td>
<td>#d 1940-</td>
<td>#d 1940-2017</td>
</tr>
<tr>
<td>200905966</td>
<td>100 1 Atkinson, David</td>
<td>100 1 Atkinson, David</td>
</tr>
<tr>
<td></td>
<td>#d 1921-</td>
<td>#d 1921-2012</td>
</tr>
</tbody>
</table>

https://www.loc.gov/aba/cataloging/subject/weeklylists/  
https://www.oclc.org/authority-records.html
Automate the update
Batch-updating Subject Headings
Extract the LCSH record number from the list of changed headings

- Copy the subject change list into Excel
- Highlight everything and unmerge cells
- Use (ctrl+f) Find/replace to replace [sp with ^]
Extract the LCSH record number from the list of changed headings (cont.)

• Use (ctrl+f) Find/replace to replace ] with nothing

• Use the Text to columns feature to separate ^ to its own column.
  • Delimited -->
  • Other: ^ -->
  • Finish
Build the JSON query in Notepad++
Build the JSON query in Notepad++ (cont.)

- Use Search -> Replace (Ctrl-H) to add the necessary text around the numbers
  - Find what:
    - (^.*)
  - Replace with:
    - {"target": {"record": {"type": "authority"}, "field": {"marcTag": "010"}}, "expr": {"op": "regex","operands": ["sh\s*$1"]}}, "or",
  - Check “Regular expression”
  - Click “Replace All”

This search query finds authority records with 010 that have “sh[as many spaces as needed][one LCCN]”. The “or” on the end of the line indicates that there are more searches to perform.
Build the JSON query in Notepad++ (cont.)

```json
{
    "queries": [
        {"target": "record", "type": "authority", "field": {"marcTag": "010"}, "expr": {"op": "regex", "operands": ["sha:2018003013"], "or":}, "target": "record", "type": "authority", "field": {"marcTag": "010"}, "expr": {"op": "regex", "operands": ["sha:2018002946"]}, "or":,
```
Run query through Sierra Create Lists

- Name your file
- Store record type = authority
- Select JSON
- Highlight the {} and paste your search from Notepad++
- Click Search
Search Results

- We have 14 out of the 220 records on the list.
- From here, export the 010 fields to a text file.
- Keep the list to delete old records later
Format data to batch search OCLC

Dido (legendary character) in fiction was deleted, so the record doesn’t exist anymore.
View your authority records in OCLC
Export updated authority records

- Batch Export from OCLC and then upload using the Subject Load table in Data Exchange
- Delete old headings to prevent duplicates
Resolving problems

- Check your headings reports for duplicate authority records and blind references
- Delete the old authority records that are duplicated by the upload
- Resolve blind references (which come from headings changing, like fictional characters)
- Most of the records were only deleting or adding subfields to the record, not changing the 1XX, so there shouldn't be too many to fix.
Batch-updating Name authorities
Closed Dates in Authority Records

Since the publication of revised LCRI 22.17 in February 2006, catalogers have had the option to add death dates to personal name headings with open dates.

Both the Library of Congress and OCLC are aware that many libraries are interested in incorporating these changes into their catalogs. Libraries that use authority control services for ongoing processing will receive these changes as part of their normal workflows but, for other libraries, OCLC is providing this service to announce headings to which death dates have been added.

Lists for each week since the new policy went into effect on 1 February 2006 are available below and as an RSS feed. Instructions for subscribing to RSS feeds can be found on RSS Feeds and Podcasts. In general, a list of newly changed headings will be added as soon as possible after they have been received.

**NOTE:** Authority Records updates posted after 19 January 2008 require Arial Unicode font to display all extended characters. This font ships with various software packages or can be purchased separately (Learn more about Arial Unicode fonts.) Older entries require the ALA BT Courier font to display properly. Use Internet Explorer for best results when viewing the ALA BT Courier font.

Recent updates

Authority records for 20190428

Authority records for 20190421

https://www.oclc.org/authority-records.html
Build the JSON query in Notepad++

Option 1:
Record number

Option 2:
Old name format

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>n 2008005874</td>
</tr>
<tr>
<td>2</td>
<td>n 2008042334</td>
</tr>
<tr>
<td>3</td>
<td>n 2010039374</td>
</tr>
<tr>
<td>4</td>
<td>n 2019006746</td>
</tr>
<tr>
<td>5</td>
<td>n 50007231</td>
</tr>
<tr>
<td>6</td>
<td>n 50034831</td>
</tr>
<tr>
<td>7</td>
<td>n 50043013</td>
</tr>
<tr>
<td>8</td>
<td>n 50045770</td>
</tr>
<tr>
<td>9</td>
<td>n 78026542</td>
</tr>
</tbody>
</table>

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Almeida,</td>
<td>Fernando d', 1955-</td>
</tr>
<tr>
<td>Averill,</td>
<td>Gerald P., 1919-</td>
</tr>
<tr>
<td>Bennett,</td>
<td>Rodney, 1935-</td>
</tr>
<tr>
<td>Berezowski,</td>
<td>Cezary, 1898-</td>
</tr>
<tr>
<td>Berger,</td>
<td>Klaus, 1901-</td>
</tr>
<tr>
<td>Boggs,</td>
<td>W. Arthur, 1916-</td>
</tr>
<tr>
<td>Bragg,</td>
<td>Don, 1935-</td>
</tr>
</tbody>
</table>
Build the JSON query in Notepad++:
Option 1: Record Numbers

This finds lines that start with [one or more non-numbers]([all of the numbers]). The information inside the parentheses will be retained, as noted by “$1”, leaving just the numbers.
Build the JSON query in Notepad++:
Option 1: Record Numbers (cont.)

Find what: (^.*)

Replace with: 
{"target": {"record": 
{"type": "authority"}, "field": 
{"marcTag": "010"}}, 
"expr": 
{"op": "regex", "operands": ["n.*$1"]}, 
"or", 

This finds everything in that line, or in this case, the LCCN.

This search query finds authority records with 010 that match "[n(zero or more of any character)LCCN]". The "or" on the end of the line indicates that there are more searches to perform.
Build the JSON query in Notepad++:

Option 2: Old name format

This finds `[any 3 characters][colon][space]100[space]1[any 1 character]` and replaces it with nothing.

This finds `([all characters up to][comma][all characters up to])[comma or period][all characters]`. The information inside the parentheses will be retained, as noted by “$1”. The last comma or period and all characters after will be deleted.
Build the JSON query in Notepad++:
Option 2: Old name format (cont.)

Find what: (^.*)

Replace with: {
  "target": {
    "record": {
      "type": "bib"
    },
    "field": {
      "marcTag": "100"
    },
    "expr": {
      "op": "starts_with",
      "operands": [$1]
    }
  },
  "or",
}

Replace with: {
  "target": {
    "record": {
      "type": "bib"
    },
    "field": {
      "marcTag": "700"
    },
    "expr": {
      "op": "starts_with",
      "operands": [$1]
    }
  },
  "or",
}

Replace with: {
  "target": {
    "record": {
      "type": "bib"
    },
    "field": {
      "marcTag": "600"
    },
    "expr": {
      "op": "starts_with",
      "operands": [$1]
    }
  },
  "or",
}

This finds everything in that line, or in this case, the name.

This finds bibliographic records with 100 (or 700 or 600) that start with “[this specific name]”. The “or” on the end of the line indicates that there are more searches to perform.
Build the JSON query in Notepad++:
Option 2: Old name format (cont.)

```
{"target": {"record": {"type": "bib"}, "field": {"marcTag": "100"}}, "expr": {"op": "starts_with", "operands": ["Yallob, P. A."], "or", "Yallob, P. S."}, "expr": {"op": "starts_with", "operands": ["Virgilino, Nicholas A."], "or", "Bailyn, John Thomas"}, "expr": {"op": "starts_with", "operands": ["Wheeler-Hilchison, Malcolm"], "or", "Winslow, Donald J."}, "expr": {"op": "starts_with", "operands": ["Ullmann, Elizabeth Xorn"], "or", "Almeida, Fernando de"], "expr": {"op": "starts_with", "operands": ["Averill, Gerald F."], "or", "Bennett, Rodney"}, "expr": {"op": "starts_with", "operands": ["Barczewski, Cezary"], "or", "Berger, Klaus"}, "expr": {"op": "starts_with", "operands": ["Borgo, W. Arthurs"], "or", "Borgo, Don"}, "expr": {"op": "starts_with", "operands": ["Brett, Lionel"]}, "or", "Brett, Lionel"}, "or", 
```
Build the JSON query in Notepad++:
Final step

```json
{
  "queries": [
    "ENTER FORMATTED SEARCH TERMS HERE"
  ]
}
```
Run query through Sierra Create Lists

- Name your file
- For option 1, Store record type = a authority
- For option 2, Store record type = b bibliographic
- Select JSON
Search Results

Option 1:
• 12 results out of 86 record numbers searched

Option 2:
• 100+ results out of 86 old names searched in 258 total searches; 18 were more closely investigated; 9 more names found
Compare Names to Updated List

To Delete Duplicates:

Find:
\^(.\*?)(\s+)?\^\(\?=.*\^\1\)$

Replace with nothing

Check regular expressions
Export updated authority records

Export into a batch file or individually.

Change the headings in the bibliographic record or wait for bibliographic maintenance to make the changes.
Tips

• Google is your friend. If I don't know how to do something in Excel or Regular Expressions, I google it and can usually find a decent formula or two that can meet my needs.

• Create a search in the classic view, then switch over to the JSON view to find how to write the code. Use "INSERT VALUE HERE" as your value and it'll be that much easier to copy your own values.

• Wait until you have several lists to combine and then do the search all at once, instead of setting it up weekly (as would be the case with the names).
Tips

• Try to avoid copying your Regular Expressions from Word because the automatic-formatting might cause problems. For example, " gets changed to “ automatically. This happened to me while developing the slides and it took me a while to figure out the problem.

• Reuse reuse reuse! We figured out how to set up one json query and then reuse the heck out of it. We just change the individual values as needed.
• You will run into times where the code doesn't want to work. Sierra will prevent you from continuing your search with a dialog box. This box shows where the problem is but it's hard to work with it in Sierra. Use your version from Notepad++ to find the line and column and see what might be wrong.

• Experience will make this easier to troubleshoot.
Resources

• Getting Started with Regular Expressions in Marcedit
  • https://www.youtube.com/watch?v=7YXvS4xBEfw

• Regular Expressions in Create Lists

• Regular Expressions Cheat Sheet
  • https://www.rexegg.com/regex-quickstart.html

• Batch Processing in the Client: Overview- OCLC
Links and Questions

• Automating the Authority Control Process – Slides and Procedure
  • https://uknowledge.uky.edu/ovgtsl2019/conf/schedule/17/
• Last year's presentation: Automating Authority Work
  • https://events.library.nd.edu/ovgtsl2018/talk/monaco.shtml
• Contact me:
  • Stacey Wolf
  • stacey.wolf@unt.edu