A Batch Solution to the Death Date Problem

A Case Study

Hsianghui Liu-Spencer & Tom Lamb
Carleton College
Northfield, Minn.
IUG Conference, San Francisco
April 25, 2013
The Library of Congress announced a change in policy about adding death dates to personal name headings and a revision to the LCRI for AACR2 rule 22.17 that allows the option for catalogers to add death dates to personal name headings with open dates.
One Bibliographic Record

100 1  Sagan, Françoise, |d1935-
240 10 Garde du coeur. |lEnglish
300 128 p. ; |c21 cm.
600 10 Sagan, Françoise, |d1935-
600 10 Sagan, Françoise, |d1935- |vAnecdotes.
One Bibliographic Record

240 10  Garde du coeur. |l English
245 14  The heart-keeper / |c by |c Françoise Sagan.
300  128 p. ; |c 21 cm.
A Name Change in LC Weekly List 2010-05-09
How Does This Impact a Local Database?

Heading in Bridge Catalog
(Carleton & St. Olaf)

Split headings in other library catalog
A total of 35,208 names over six-year period.

Source: LC Weekly Lists website
A total of 14,143 names over six-year period
What do you do?

Or do you care?

OCLC-CAT Listserv in Feb, 2012
**13 Responses
Bridge Consortium Reacts Over the Years

2005/06
- LC announcement
- **LC weekly lists posted in OCLC site in Feb 2006

2006
- Print lists, check local database. Review heading report.
- **Local editing, coded “a” in AuthoCode2

2010
- Print lists and write notes on library cards
- **Test Kent State DeathFlip Project

2012
The Old Ways We Tried to Keep Up…
Bridge Consortium Reacts Over the Years

2005/06
- LC announcement
- **LC weekly lists posted in OCLC site in Feb 2006

2006
- Print lists, check local database. Review heading report.
- **Local editing, coded “a” in AuthoCode2

2010
- Print lists and write notes on library cards
- **Test Kent State DeathFlip Project
- June~Aug: RDA authority records

2012
- Carleton starts a case study in Feb
- **March: LC’s decision on RDA

Our Goal

- Identify headings in our system that appear in the LC Weekly Closed Dates list and update those headings as needed.

We Need to Be Aware:

1. Not all identified headings require updating.
2. Identified headings may present unique challenges requiring unique approaches:
   - Heading for 1xx and 7xx
   - Heading for 6xx subject
   - Author/Title added entry & Uniform Title (240)
Tools That We Used

http://datatoolbar.com/
1. Batch search in Connexion by LCCN
2. Run macro adding 4xx, 667, 949 ±
3. Export to Millennium
Day One
Observe blind references and let AACP work its magic

Day Two
Check heading report in Millennium
Check *Updated Bib Heading* as well as near-match, duplicate AF, etc.

Day Three
Make sure no more blind references

Blind references – mostly need to create a different 4xx. So, it will flip the following day.

Don’t forget subjects (6xx) and irregular date formats (1xx, 7xx)
Review List for 1XX Name Heading

Search strategy:
contains 1XX but not 110/111 or 100|t

Export 12 data elements to spreadsheet
Review List for 6xx Name Heading as Subject

Search strategy:
6xx but not 610/611; 600|t, |v, or |x

Export 12 data elements to a spreadsheet
... Python Script (a snippet)

This script takes an authority file "fulllist.tab" and a modification file "modifylist.tab" and outputs a csv file containing the necessary changes.

Both input files should be formatted as tab-delimited files "tab"

To specify input files, set the "fulllist" and "modifylist" variables to the desired input file locations. These variables are defined at the end of this script (lines 49-50).

To run the script:
1. Make sure 'authority.py' and both input files are in the same directory.
2. Open Terminal and navigate to this directory.
3. Enter the command "python authority.py".
4. An output file named "authority_changes.csv" will be created in the directory.

```python
def main(fulllist, modifylist):
    fulllist = open(fulllist, 'r')
    modifylist = open(modifylist, 'r')
    changes = open('authority_changes.csv', 'w')
    changes.write(','.join(['id', 'autho code', 'old', 'new', 'author name', 'weekly list']) + '
')
    full = {};
    for line in fulllist.readlines():
        if line in ['\n', '\r']:
            continue
        line = line.split('\\n')
        full[line[0]] = [line[1], line[2], line[3]]
    fulllist.close()
    count = 0
    account = 0
    for line in modifylist.readlines():
        if line in ['\n', '\r']:
            continue
        line = line.split('\\n')
        id = ''.join(line[0].split('\\n'))
        if id in full:
            count = 1
            entry = full[id]
            if line[1] == ':-'
                account = 1
            if entry[1] < line[1]:
                account = 1
        changes.write(','.join(['id', 'autho code', 'old', 'new', 'author name', 'weekly list']) + '
')
    changes.close()
```
Excel Spreadsheets

Formulae:

- Determine final digit  -  =RIGHT(E6,1)
- Remove period  -  =LEFT(E5,LEN(E5)-1)
- Compare cells  -  =(N5=P5)  returns TRUE or FALSE
- Sort spreadsheet by cell color
<table>
<thead>
<tr>
<th>id</th>
<th>author code</th>
<th>author/name</th>
<th>period removed</th>
<th>weekly list</th>
<th>final digit correct heading</th>
<th>STATUS</th>
</tr>
</thead>
<tbody>
<tr>
<td>n 85378827</td>
<td>-</td>
<td>Fessier, Michael, 1907-</td>
<td>Fessier, Michael, 1907-1989</td>
<td>20060301</td>
<td>8</td>
<td>NOT FIXED</td>
</tr>
<tr>
<td>n 86818968</td>
<td>a</td>
<td>Herbert, F. Hugh (Frederick Hugh), 1897-</td>
<td>Herbert, F. Hugh (Frederick Hugh), 1897-1952</td>
<td>20080203</td>
<td>8</td>
<td>NOT FIXED</td>
</tr>
<tr>
<td>n 85017840</td>
<td>a</td>
<td>Ambrose, Thomas, 17/18th cent.</td>
<td>Ambrose, Thomas, 17th/18th cent.</td>
<td>20080706</td>
<td>1</td>
<td>OK</td>
</tr>
<tr>
<td>n 92106624</td>
<td>a</td>
<td>Culp, Robert, 1930-2010</td>
<td>Culp, Robert, 1930-2010</td>
<td>20100328</td>
<td>p</td>
<td>REVIEW</td>
</tr>
<tr>
<td>n 00006682</td>
<td>-</td>
<td>Poulsen, Magnus, 1801-1882</td>
<td>Poulsen, Magnus, 1801-1882</td>
<td>20071107</td>
<td></td>
<td>SUBJECT?</td>
</tr>
<tr>
<td>n 00030853</td>
<td>-</td>
<td>Koppel, Ted, 1940-1940</td>
<td>Koppel, Ted, 1940-Feb. 8</td>
<td>20080525</td>
<td></td>
<td>SUBJECT?</td>
</tr>
<tr>
<td>n 00032732</td>
<td>-</td>
<td>Jackson, David, 1922-2001</td>
<td>Jackson, David, 1922-2001</td>
<td>20100516</td>
<td></td>
<td>SUBJECT?</td>
</tr>
</tbody>
</table>
Connexion macro (a snippet)

400 _9 Last, First, |dYYYY-

667 19 Edited/XXXX. Added 4xx without death date. LC closed date automated process.

949 *atab=anam; ins=XXXX OR

949 *atab=asub; ins=XXXX
A blind reference
## The Numbers

<table>
<thead>
<tr>
<th>Data Sources</th>
<th>Numbers of Headings</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>LC weekly list (02/2006 ~ 02/2012)</td>
<td>35,208</td>
<td></td>
</tr>
<tr>
<td>Millennium AF list up to Feb. 2012 (including everything)</td>
<td>636,954+</td>
<td></td>
</tr>
<tr>
<td>Matches between LC and our database generated by script, March 2012</td>
<td>16,859</td>
<td></td>
</tr>
<tr>
<td>Identify duplicate headings from the spreadsheet (main heading and subject)</td>
<td>2,716</td>
<td></td>
</tr>
<tr>
<td>LC updated headings in our database (16,859-2,716)</td>
<td>14,134</td>
<td>14,134/35,208=40%</td>
</tr>
<tr>
<td>An average cataloger at Carleton would have caught and manually updated,</td>
<td>600</td>
<td>600/14,134=4.3%</td>
</tr>
<tr>
<td>approximately</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Name headings batch-loaded from Connexion to Millennium 60%

- Blind references on Day 2 13% to fix
- Total bib updated through AACP: 42,929+
- Average number of bibs per heading: 4~6 bibs per AF (1~110 bibs per AF)
A Statistics View  (16,859 matching names)

- **1xx/7xx through a batch process** (60%)
- **Fixed prior to 2012-02** (20%)
- **6xx through a semi-manual process** (17%)
- **Date format not in YYYY** (3%)

**Differences – 0.06%, 111 headings**
Are we there yet?

Cycle incomplete. We addressed 1xx/7xx and 6xx. Author/title added entry (240) not yet addressed.
One Bibliographic Record

240 10 Garde du coeur. |l English
300 128 p. ; |c 21 cm.
• Output authority record (240) from Millennium to a spreadsheet
• Sort the data elements by 1) YYYY-YYYY, 2) no date and 3) YYYY-
• Low-hanging fruit – names with the format of YYYY- and batch search in Connexion
• From the result list in Connexion, identify the date YYYY-YYYY (a closed date)
# Successful Searches (5957)

<table>
<thead>
<tr>
<th>Search Key</th>
<th># Records</th>
<th>Save #</th>
<th>Control #</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>In:no2001040027</td>
<td>1</td>
<td>1</td>
<td>5489849</td>
<td>Tudor, David, †d 1926-1996. †t Rain</td>
</tr>
<tr>
<td>In:n 2001038289</td>
<td>1</td>
<td>2</td>
<td>5529617</td>
<td>Janson, Drude Krog, †d 1846- †t Sad</td>
</tr>
<tr>
<td>In:n 83800735</td>
<td>1</td>
<td>3</td>
<td>1054308</td>
<td>Pringsheim, Alfred, †d 1850-1941. †t Vorlesungen über Zahlen- und Funktionenlehre</td>
</tr>
<tr>
<td>In:n 2001069451</td>
<td>1</td>
<td>4</td>
<td>5646871</td>
<td>Jakobowski, Edward, †d 1858-1927. †t Erminie. †p Lullaby</td>
</tr>
<tr>
<td>In:n 82042472</td>
<td>1</td>
<td>5</td>
<td>739480</td>
<td>Jackson, Emily, †d 1861- †t Silhouette</td>
</tr>
<tr>
<td>In:n 83185023</td>
<td>1</td>
<td>6</td>
<td>1001752</td>
<td>Radford, William A., †d 1865- †t Portfolio of details of building construction</td>
</tr>
<tr>
<td>In:n 85180165</td>
<td>1</td>
<td>7</td>
<td>1750929</td>
<td>Boer, T. J. de †q (Tjitze J.), †d 1866- †t Geschichte der Philosophie im Islam. †t English</td>
</tr>
<tr>
<td>In:n 81128919</td>
<td>1</td>
<td>8</td>
<td>673814</td>
<td>Ramsey, A. S. †q (Arthur Stanley), †d 1867-1954. †t Introduction to the theory of Newtonian attraction</td>
</tr>
</tbody>
</table>
Keeping up – the basic steps

1. Scrape data from LC site (spreadsheet vs. RSS)
2. Output auth records from Millennium
3. Generate a list of matching LCCN
4. Review list to sort headings into categories
5. Batch export from Connexion into Millennium
6. Attend to next day Headings Reports
7. Address headings requiring manual attention
8. Address author/title headings
Bridge Consortium Reacts Over the Years

2005/06
LC announcement
**LC weekly lists posted in OCLC site in Feb 2006

2006
Print lists, check local database. Review heading report.
**local editing, coded “a” in AuthoCode2

2010
Print lists and write notes on library cards
**Test Kent State DeathFlip Project

2012
Carleton starts a case study in Feb
**March: LC’s decision on RDA
**June-Aug: RDA authority records

2013
Catch up!
Clean up one full year of LC weekly list
(dated 2013-02-29)
Data Comparison with Access
Numbers for the 7th year 03/2012 to 12/2013

<table>
<thead>
<tr>
<th>Data Sources</th>
<th>Numbers of Headings</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>LC weekly list (03/2012 ~ 02/2013)</td>
<td>4,503</td>
<td>**note: LC didn’t publish any lists in August 2012</td>
</tr>
<tr>
<td>Millennium AF list up to Feb. 2013 (only contains 1xx and 6xx with no subfields x,v,t)</td>
<td>656,898+</td>
<td></td>
</tr>
<tr>
<td>Matches between LC and our database generated by MS Access, March 2013</td>
<td>1,571</td>
<td></td>
</tr>
<tr>
<td>LC updated headings in our database</td>
<td>1,571+</td>
<td>1,571/4,503=<strong>34.88%</strong></td>
</tr>
</tbody>
</table>
Lessons Learned

• Saving time?

• Taking advantage of OCLC batch searching and exporting

• Understanding the process of AACP in Millennium

• Subject template in Millennium load table: 949 *atab=asub

• Global update in 400 19 |wnnea Name

• Making sure to keep up with heading reports and clear up the space for loading more records

• Check the setting for AACP: Millennium III manual: #107824
  ➢ Set this option to YES to enable the application of name authority record updates to name-title bibliographic headings
Our Goal

- Identify headings in our system that appear in the LC Weekly Closed Dates list and update those headings as needed.

We Need to Be Aware:

1. Not all identified headings require updating.
2. Identified headings may present unique challenges requiring unique approaches:
   - Heading for 1xx and 7xx
   - Heading for 6xx subject
   - Author/Title added entry & Uniform Title (240)
The Plan to Move Forward

- Keep up with the RDA changes to personal name headings
- Develop our own monthly schedule of Closed Date updating
Thanks Go To

Kathy Blough
St. Olaf Music Cataloger

Jason Cohn
Student Worker in Carleton Archives

Nat Wilson
Digital Archivist & Technology Coordinator, Carleton College

Mark Ehlert
Coordinator of Digitization, Cataloging & Metadata Education, Minitex
Credit Also Goes To

Susanne Nevin  
St. Olaf Cataloger

Sue Ims  
Carleton College Cataloger

JoEllen LaPrade  
St. Olaf Cataloger
Q & A
Thank you!

Tom Lamb tlamb@carleton.edu
Hsianghui Liu-Spencer hliuspen@carleton.edu