Authority Control for Digital Collections

Jeremy Myntti  
Anna Neatrour  
Digital Library Services

Nate Cothran  
Vice President, Automation Services
Project Overview

KEEP CALM AND MANAGE METADATA
Summary

- Metadata inconsistency
- Overview of authority control process
- Results and problems encountered
- Further work with Open Refine
- Next steps
Inconsistency

Gosiute Indians
Goshute Indians
Navajo Indians
Navaho Indians
Salt Lake
Salt Lake City
Salt Lake City (Utah)
Bishop, Dail Stapley
Bishop, Dale Stapely
Bishop, Dale Stapley
Beckwith, Frank A. (1876-1951)
Beckwith, Frank Asahel (1876-1951)
Beckwith, Frank A.
Beckwith, Frank A. (1876-1951)
Beckwith, Frank Asahel (1876-1951)
Beckwith, Frank Asahel, 1876-1951

Woven basket or jug: http://content.lib.utah.edu/cdm/ref/collection/UU_Photo_Archives/id/13887
Partnership
Methodology

A group of St. George (Sibwit) Paiutes and Wickiups (cedar)

Paiute Indians; Ute Indians--History; Wickiups; Indians of North America--Dwellings

Utah

A group of people sitting and standing in front of a brush shelter

Digitized by: J. Willard Marriott Library, University of Utah

Image; StillImage

image/jpeg
MARC & XML, Normalized

- **MARC**
  - 651 $a United States $x History $y 20th century.
  - 651 $ UNITED STATES $ HISTORY $ 20TH CENTURY

- **XML**
  - United States--History--20th century;
  - 650 $a United States $x History $x 20th century
  - 650 $ UNITED STATES $ HISTORY $ 20TH CENTURY
MARC & XML Examples

- **MARC**
  - 600 $a Smith, John, $d 1947 Apr. 16-
  - 651 $a United States $x History $y 20th century.
  - 650 $a Navajo Indians $x History.
  - 650 $a Religious ceremonies.
  - 650 $a Religion.

- **XML**
  - Smith, John, 1947 Apr. 16-; United States--History--20th century; Navajo Indians--History; Religious ceremonies; Religion;
MARC & XML Conversions

• MARC
  ◦ 600 $a Smith, John, $d 1947 Apr. 16-
  ◦ 651 $a United States $x History $y 20th century.
  ◦ 650 $a Navajo Indians $x History.
  ◦ 650 $a Religious ceremonies.
  ◦ 650 $a Religion.

• XML
  ◦ 600 $a Smith, John, $d 1947 Apr. 16-
  ◦ 650 $a United States $x History $x 20th century
  ◦ 650 $a Navajo Indians $x History
  ◦ 650 $a Religious ceremonies
  ◦ 650 $a Religion

Hoop ceremony, Medicine man is Gray Squirrel (Yucca hoops); http://content.lib.utah.edu/cdm/ref/collection/uaida/id/18210
XML Authority Matching

- **XML**
  - 600 $a Smith, John, $d 1947 Apr. 16-
  - 650 $a United States $x History $x 20th century
  - 650 $a Navajo Indians $x History
  - 650 $a Religious ceremonies
  - 650 $a Religion

- **Authority**
  - 150 $a Rites and ceremonies
  - 450 $a Ceremonies
  - 450 $a Cult
  - 450 $a Cultus
  - 450 $a Ecclesiastical rites and ceremonies
  - 450 $a Religious ceremonies
  - 450 $a Religious rites
  - 450 $a Rites of passage
  - 450 $a Traditions

Navajo man in medicine hogan with large sand painting; http://content.lib.utah.edu/cdm/ref/collection/uida/id/45145
XML Authority Control

• Pre-processing
  • <subject>Smith, John, 1947 Apr. 16-; United States--History--20th century; Navajo Indians--History; Religious ceremonies; Religion; </subject>

• Post-processing
  • <subject>Smith, John, 1947 Apr. 16-; United States--History--20th century; Navajo Indians--History; Rites and ceremonies; Religion;</subject>
  • <keyword>Ceremonies; Cult; Cultus; Ecclesiastical rites and ceremonies; Religious ceremonies; Religious rites; Rites of passage; Traditions;</keyword>
Project Timeline

- June-Sept. 2012 – Define project
- June 2013 – Contracted with Backstage, ACIG presentation
- June 2013-Feb. 2014 – Continue testing
- Feb.-May 2014 – 17 collections processed
- June-Aug. 2014 – Manual review (intern)
- April-June 2015 – Explore OpenRefine
Results – Statistics

Creators / Contributors

- Photographs
- Manuscripts
- IR

Subjects

- Photographs
- Manuscripts
- IR

[Bar charts showing percentage of changed and matched entries for Creators/Contributors and Subjects categories]
Results – Reports

Unmatched

Changes

Date changes
Results – Standardization

Punctuation

Indians of North America – Education;
Indians of North America -- Education
Indians of North America - Education
Indians of North America--Education;

A group of Navajos at Navajo Mountain government school;
http://content.lib.utah.edu/cdm/ref/collection/uaida/id/43551
Results – Standardization

Capitalization

Coal Mines and Mining; Coal Miners; Coal Trade
coal mines and mining; coal miners; coal trade

Coal mines and mining; Coal miners; Coal trade

Petroglyphs; Moccasin, Arizona [XIII];
http://content.lib.utah.edu/cdm/ref/collection/uuida/id/39218
Results – Standardization

Update Access Points

Liebler, Harold Baxter, 1889-1982
Liebler, H. Baxter (Harold Baxter), 1889-1982

Civil War (U.S.) 1861-1865
United States--History--Civil War, 1861-1865

Father Liebler and a group of Navajos at Hogan Church on Palm Sunday; http://content.lib.utah.edu/cdm/ref/collection/uaida/id/43414
Problems Encountered

Missing MARC tags

- Names treated as topical headings and vice versa
- **Provo** => **Provisional IRA**
- **Cars** => **CARS 2002 (2002: Paris, France)**

Walker War Map 1853-1854; http://content.lib.utah.edu/cdm/ref/collection/uaida/id/15474
Problems Encountered

Local names matched wrong records

<table>
<thead>
<tr>
<th>Local Name</th>
<th>Changed to</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bailey, Ron</td>
<td>Bailey, Ron, 1943-</td>
</tr>
<tr>
<td>Bennett, Myrtle</td>
<td>Bennett, Myrtle, -1991</td>
</tr>
<tr>
<td>Evans, Jack</td>
<td>Evans, Jack, Pulp fiction writer</td>
</tr>
<tr>
<td>Gardner, G.</td>
<td>Gardner, G. (Gary)</td>
</tr>
<tr>
<td>Jeffery, S.</td>
<td>Jeffery, S. (Steve)</td>
</tr>
<tr>
<td>Johnson, Abe</td>
<td>Johnson, F. T.</td>
</tr>
</tbody>
</table>
Problems Encountered

Data in wrong fields

- Date: Price, Hiram, 1814-1901
- Spatial coverage: Thatched roofs

Brush shelter, John Duncan's Camp, White Rocks, Summer '09;
http://content.lib.utah.edu/cdm/ref/collection/uaida/id/14655
Intern review and clean-up
OpenRefine Project

- Used UAIDA as a pilot, since it had the greatest number of unmatched names due to the size of the collection (over 8,000 items)
- 529 unmatched names after Backstage process

Navajo woman weaving, 
http://content.lib.utah.edu/cdm/ref/collection/uaida/id/45379
Open Refine: two approaches

- Reconciliation process developed by Jenn Wright and Matt Carruthers, University of Michigan Library, https://github.com/mcarruthers/LCNAF-Named-Entity-Reconciliation


A group of Navajo children and teenagers, http://content.lib.utah.edu/cdm/ref/collection/uaida/id/43285
Open Refine: differences in results

- Both processes found name matches through searching VIAF.
  - Wright and Carruthers’ process looked for a matching LC authority record in the VIAF cluster
    - 81 records were matched, 132 were false matches, and 312 number had no match
  - Page’s process matched names to authors in a more general fashion
    - 70 records were matched, 37 were false matches, and 449 had no match.
Open Refine: manual work

- Check matches against collection and discard false matches

<table>
<thead>
<tr>
<th>Heading</th>
<th>headings to reconcile</th>
</tr>
</thead>
<tbody>
<tr>
<td>C. Patillo</td>
<td>Santana, Carlos, 1947-....</td>
</tr>
<tr>
<td>C. S. Photo.</td>
<td>Smith, Ralph C., 1960-....</td>
</tr>
<tr>
<td>Cady, A.</td>
<td>Cady, A.</td>
</tr>
<tr>
<td></td>
<td>✓ ✓ Cady, Susan A. (1)</td>
</tr>
<tr>
<td></td>
<td>✓ ✓ Cady, A. Howard (Alice Howard) (1)</td>
</tr>
<tr>
<td></td>
<td>✓ ✓ Cady, Susan A. (1)</td>
</tr>
<tr>
<td></td>
<td>✓ Create new topic</td>
</tr>
<tr>
<td>Cady, W.F.</td>
<td>Cady, W.F.</td>
</tr>
<tr>
<td></td>
<td>✓ ✓ Create new topic</td>
</tr>
</tbody>
</table>
Open Refine: manual work

- Wright and Carruthers process had 262 undo/redo actions in the open refine project
- Page’s reconciliation process resulted in 424 undo/redo actions
  - Could have refined initial results further

Next Steps

Create local controlled vocabularies
Next Steps

Convert LCSH spatial data to Geonames

**LCSH:**
Salt Lake City (Utah)

**Geonames:**
Salt Lake City, Salt Lake County, Utah, United States,
http://sws.geonames.org/5780993/

Map of Northern Utah, Early Settlements;
http://content.lib.utah.edu/cdm/ref/collection/uida/id/14541
Next Steps
Next Steps

URIs in CONTENTdm, MWDL (Primo), and DPLA

<table>
<thead>
<tr>
<th>Title</th>
<th>Three people outside a teepee;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject</td>
<td>St. Christopher’s Mission to the Navajo (Bluff, Utah), <a href="http://id.loc.gov/authorities/names/nr91019909">http://id.loc.gov/authorities/names/nr91019909</a>; Liebler, Harold Baxter (Harold Baxter), 1889-1982, <a href="http://id.loc.gov/authorities/names/n94048511">http://id.loc.gov/authorities/names/n94048511</a>; Tipis, <a href="http://id.loc.gov/authorities/subjects/sh94002186">http://id.loc.gov/authorities/subjects/sh94002186</a>;</td>
</tr>
<tr>
<td>Creator</td>
<td>Liebler, H. Baxter (Harold Baxter), 1889-1982, <a href="http://id.loc.gov/authorities/names/n94048511">http://id.loc.gov/authorities/names/n94048511</a>;</td>
</tr>
<tr>
<td>Type</td>
<td>Image:StillImage, <a href="http://purl.org/dc/dcmitype/StillImage">http://purl.org/dc/dcmitype/StillImage</a>;</td>
</tr>
<tr>
<td>Language</td>
<td>eng, <a href="http://id.loc.gov/vocabulary/languages/eng">http://id.loc.gov/vocabulary/languages/eng</a>;</td>
</tr>
<tr>
<td>Coverage</td>
<td>Bluff, San Juan County, Utah, United States, <a href="http://www.geonames.org/5535638/">http://www.geonames.org/5535638/</a>;</td>
</tr>
</tbody>
</table>
Next Steps

Convert data to RDF

@prefix dc: <http://purl.org/dc/elements/1.1/> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .

<http://lib.utah.edu/ark:/87278/s6pr8bn0> a rdfs:Resource ;
    dc:title "Cherry Blossom, a Ute Indian Maiden from U and O;" ;
    dc:publisher http://id.loc.gov/authorities/names/no2014076689 ;
    dc:date "2010-05-05" ;
    dc:subject http://id.loc.gov/authorities/subjects/sh2010117611,
                 http://id.loc.gov/authorities/subjects/sh97003343 ;
    dc:spatial http://sws.geonames.org/5783932/ ;
    dc:description "Photograph of a woman standing among brush;" ;
    dc:type http://purl.org/dc/dcmitype/StillImage ;
    dc:creator http://id.loc.gov/authorities/names/n82112911 ;
    dc:format "image/jp2" .

http://content.lib.utah.edu/cdm/ref/collection/uaida/id/14757
Questions

Jeremy Myntti | jeremy.myntti@utah.edu
Interim Head, Digital Library Services

Anna Neatrour | anna.neatrour@utah.edu
Metadata Librarian

Nate Cothran | nate@bslw.com
Vice President, Automation Services