

Authority Control *for* Digital Collections



Authority Control Interest Group
June 28, 2015

Jeremy Myntti

Anna Neatrou



J. Willard Marriott Library

THE UNIVERSITY OF UTAH

Digital Library Services

Nate Cothran



Vice President, Automation Services

Project Overview



Summary

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



UTAH AMERICAN INDIAN
DIGITAL ARCHIVE

<http://www.utahindians.org>

Inconsistency

Gosiute Indians

Beckwith, Frank A. (1876-1951)

Goshute Indians

Beckwith, Frank Asahel (1876-1951)

Navajo Indians

Beckwith, Frank A.

Beckwith, Frank A. (1876-1951)

Navaho Indians

Beckwith, Frank Asahel (1876-1951)

Beckwith, Frank Asahel, 1876-1951

Salt Lake

Salt Lake City

Salt Lake City (Utah)

Bishop, Dail Stapley

Bishop, Dale Stapely

Bishop, Dale Stapley



Woven basket or jug;

http://content.lib.utah.edu/cdm/ref/collection/UU_Photo_Archives/id/13887

Partnership



Methodology



<http://content.lib.utah.edu/cdm/ref/collection/uaida/id/14697>

<title>A group of St. George (Sibwit) Paiutes and Wickiups (cedar)**</title>**

<subject>Paiute Indians; Ute Indians--History; Wickiups; Indians of North America--Dwellings;**</subject>**

<covspa>Utah;**</covspa>**

<descri>A group of people sitting and standing in front of a brush shelter;**</descri>**

<publis>Digitized by: J. Willard Marriott Library, University of Utah;**</publis>**

<type>Image;StillImage;**</type>**

<format>image/jpeg;**</format>**

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/uaida/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

Oct. 2012 – May 2013 – Testing

June 2013 – Contracted with Backstage,
ACIG presentation

June 2013-Feb. 2014 – Continue testing

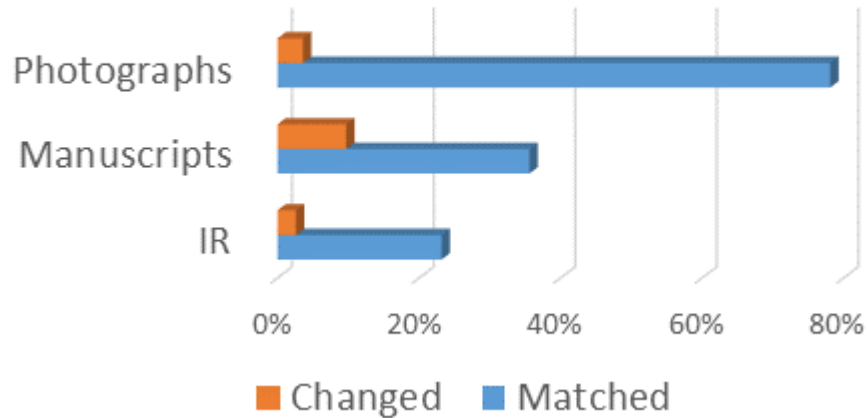
Feb.-May 2014 – 17 collections processed

June-Aug. 2014 – Manual review (intern)

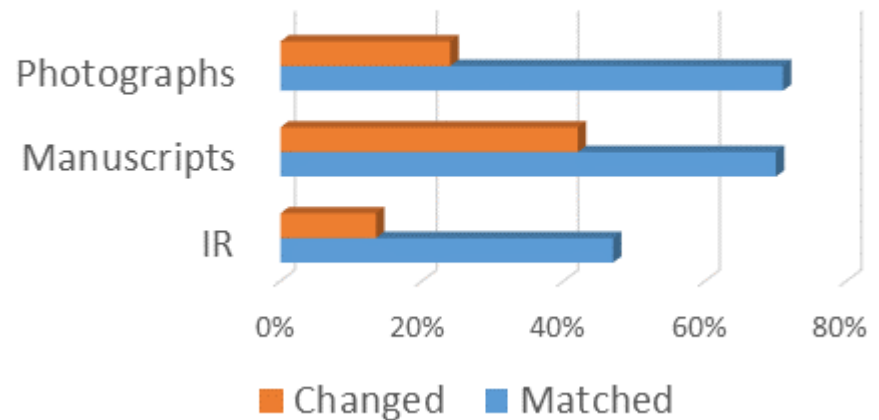
April-June 2015 – Explore OpenRefine

Results – Statistics

Creators / Contributors



Subjects



Results – Reports

Unmatched

7 records	creato	Albin, William M.
1 record	creato	Allen, Rufus C.
1 record	creato	Allred, William J.
8 records	creato	Alter, Cecil J.
2 records	creato	American West Center
15 records	creato	Arentz, Bob
1 record	creato	Arnold Stone, Elizabeth
1 record	creato	Arrowpeen

Changes

2 records	Old	creato	Leupp, Francis Ellington, 1849-1918
	New	creato	Leupp, Francis E. (Francis Ellington), 1849-1918
321 records	Old	creato	Liebler, Harold Baxter, 1889-1982
	New	creato	Liebler, H. Baxter (Harold Baxter), 1889-1982
21 records	Old	creato	Matheson, Alva
	New	creato	Matheson, Alva L.

Date changes

1 record	Old	1907-10-01; 1907-11-06; 1907-11-12; 29 September 1907; 1907-11-20; 1907-11-29; 1907-12-15;
	New	1907-10-01; 1907-11-06; 1907-11-12; 1907-09-29; 1907-11-20; 1907-11-29; 1907-12-15;
1 record	Old	21 April 1908
	New	1908-04-21;
1 record	Old	24 April 1908
	New	1908-04-24;
1 record	Old	1945-1949
	New	1945; 1946; 1947; 1948; 1949;

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](http://content.lib.utah.edu/cdm/ref/collection/uaida/id/43551)



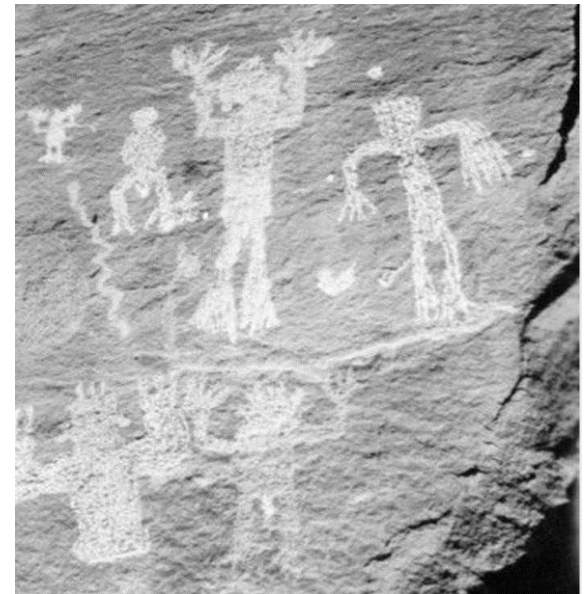
Results – Standardization

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/uaida/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

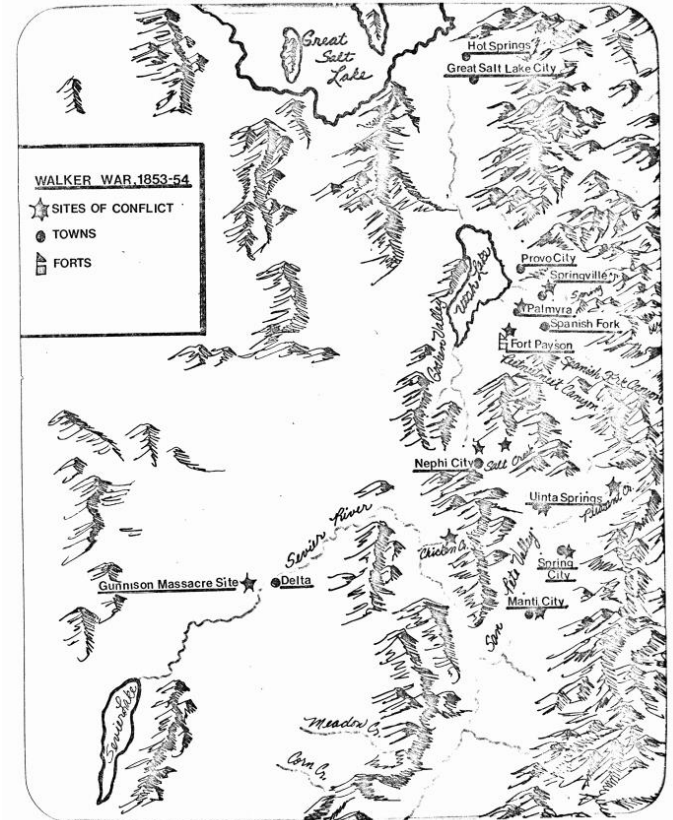
Father Liebler and a group of Navajos at
Hogan Church on Palm Sunday;
[http://content.lib.utah.edu/cdm/ref/
collection/uaida/id/43414](http://content.lib.utah.edu/cdm/ref/collection/uaida/id/43414)



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

Local Name	Changed to
Bailey, Ron	Bailey, Ron, 1943-
Bennett, Myrtle	Bennett, Myrtle, -1991
Evans, Jack	Evans, Jack, Pulp fiction writer
Gardner, G.	Gardner, G. (Gary)
Jeffery, S.	Jeffery, S. (Steve)
Johnson, Abe	Johnson, F. T.

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>
- Reconciliation process developed by Roderic Page, <http://iphylo.blogspot.com/2013/04/reconciling-author-names-using-open.html>



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

<input type="checkbox"/> Heading	<input type="checkbox"/> headings to reconcile
C. Patillo	Santana, Carlos, 1947-.... edit Choose new match
C. S. Photo.	Smith, Ralph C., 1960-.... Choose new match
Cady, A.	Cady, A. <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Cady, Susan A. (1) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Cady, A. Howard (Alice Howard) (1) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Cady, Susan A. (1) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Create new topic
Cady, W.F.	Cady, W.F. <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Create new topic

Open Refine: manual work

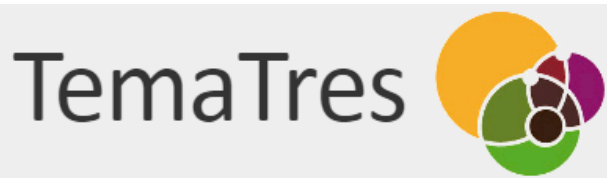
- Wright and Carruthers process had 262 undo/redo actions in the open refine project
- Page's reconciliation process resulted in 424 undo/redo actions
 - Could have refined initial results further



Eight Hopi Baskets,
<http://content.lib.utah.edu/cdm/ref/collection/uaida/id/45009>

Next Steps

Create local controlled vocabularies



Next Steps

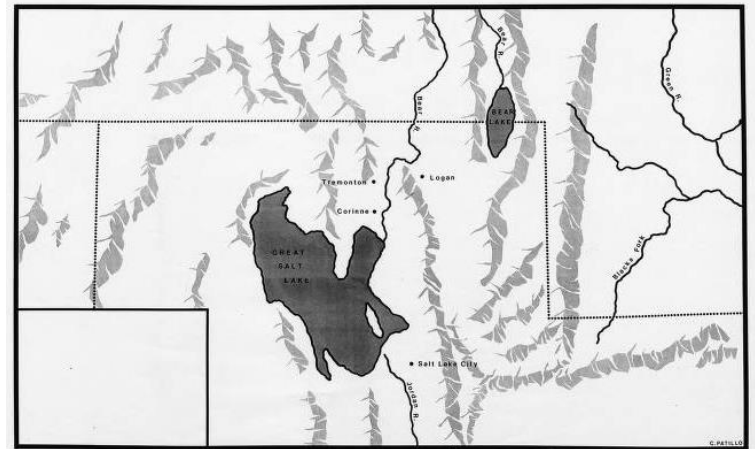
Convert LCSH spatial data to Geonames

LCSH:

Salt Lake City (Utah)

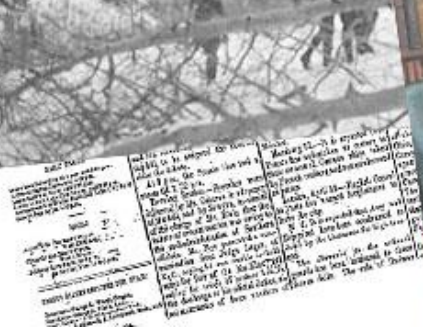
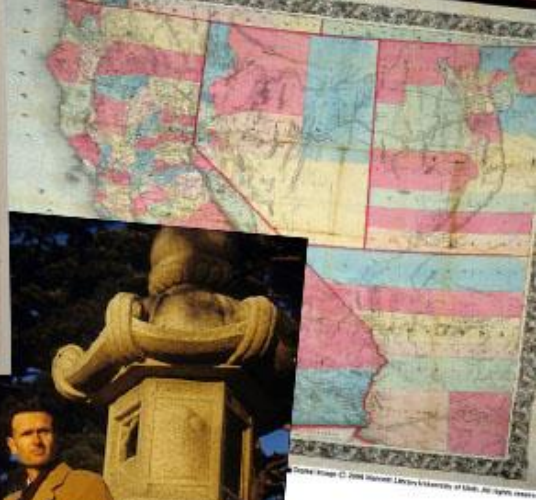
Geonames:

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/uaida/id/14541>

Next Steps



Next Steps

URIs in CONTENTdm, MWDL (Primo), and DPLA

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

<http://content.lib.utah.edu/cdm/ref/collection/uaida/id/43183>



Next Steps

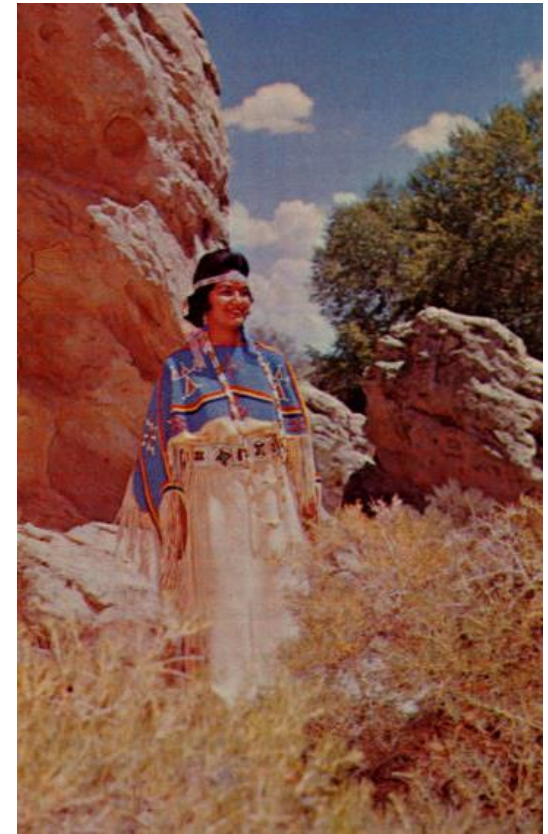
Convert data to RDF

@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/jpeg" .
```



<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

