

Linking People:

Collaborations Between Metadata Librarians and Programmers

Jeremy Myntti
@jmyntti

Anna Neatrour
@annaneat

Liz Woolcott
@lizwoolcott



Why?

Metadata problems when collections are aggregated:

- Savage, C. R. (Charles Roscoe), 1832-1909
- Savage, C. R.(Charles Roscoe),1832-1909
- Savage, C.R. (Charles Roscoe)
- Savage, Charles R.
- C. R. Savage (Charles Roscoe Savage and George Ottinger), Pioneer Art Gallery, East Temple Street, Salt Lake City, Utah
- Charles R. Savage
- Savage, Charles Roscoe
- C. R. (Charles Roscoe) Savage, photographer
- Savage, C. R.
- Charles R.



Charles R. Savage, perplexed by his many name variants.

Charles R. Savage
Utah State Historical Society Classified Photograph Collection
<https://collections.lib.utah.edu/details?id=432953>

More metadata, more metadata problems

- Vendor based DAMS often not offering good authority control solutions
- Libraries in Utah often consult additional regional names sources
 - Daughters of Utah Pioneers name index:

http://www.dupinternational.org/pioneer_index.php.

Last Name	Given Name	Maiden Name	Birth Date	Death Date
Allred	Ellen	Shepherd	12 May 1845	15 Dec 1928
Ashworth	Mary Elizabeth	Shepherd	24 Mar 1854	11 Apr 1932
Bailey	Charlotte	Shepherd	7 May 1848	7 Jul 1937
Barney	Priscilla Ann	Shepherd	4 May 1842	2 Nov 1916
Blackham	Priscille Ann	Shepherd	4 May 1842	20 Sep 1916
Davidson	Lydia	Shepherd	16 Sep 1836	
Derrick	Mary	Shepherd	12 Aug 1812	17 Feb 1886
Hartley	Ellen	Shepherd	9 Dec 1854	27 Oct 1918
Home	Mary Park	Shepherd	13 Nov 1836	28 Apr 1924
Justesen	Sarah Ann	Shepherd	4 Sep 1842	8 Aug 1933
Miller	Ann	Shepherd	8 May 1825	27 Mar 1910

- Hosting collections for many partners means less control over cataloging practices.
- Use LC Authorities as best we can

Local pilots at Marriott Library

- Pilot project with Backstage Library Works
 - Replicate MARC authority control project in digital collection XML records
- Pilot project using OpenRefine
 - Reconciling data with LCNAF and VIAF



Van de Vanter, Billy's car
<https://collections.lib.utah.edu/details?id=126695>

Overview of grant

1. Investigation
 - a. May 2016 - October 2016
2. Evaluation and Testing
 - a. November 2016 - April 2017
3. Pilot Implementation
 - a. May 2017 - October 2017
4. Assessment and Future Planning
 - a. November 2017 - April 2018



Digital image copyright 2002, Topaz Museum. All rights reserved.

Christmas party at Lafabreques, Topaz Museum
<https://collections.lib.utah.edu/details?id=341268>

Collaborating with programmers

- Digital Library Services (DLS) and Digital Infrastructure Development (DID) are distinct departments at the Marriott Library
- DLS has digitization staff, metadata librarians, digital preservation
- DID has application development, sandbox server support
- PIs Consulted with DID on all phases of grant development - brief proposal, final proposal, budget

Collaborating with other institutions

Project partners include:

- Utah State University
- Brigham Young University
- Utah State Archives
- University of Oregon
- University of Nevada, Reno
- University of Denver

Advisory Board:

- Kevin Ford, Art Institute of Chicago
- Gretchen Gueguen, DPLA
- Eric Miller, Zepheira
- Philip Schreur, Stanford

"Looking Across Donner Lake - To the Summit" - Two people visible in small boat
Utah State University
Western Photographers

<http://digital.lib.usu.edu/cdm/ref/collection/westernphoto/id/779>



Partner Perspectives

DAMs:

- CONTENTdm
- Hydra
- Islandora

Types of collections & content:

- Historical/archival material
- Mostly images and text, increasing audio and video



Partner Perspectives

Outcomes Desired:

- Well organized, well managed regional linked data repository
- Rich documentation of best practices and workflows for creating/sharing local linked-data compliant authorities
- Ability to authorize names through NACO/SACO (in the long run)
- Contributing names to the larger library/archival community

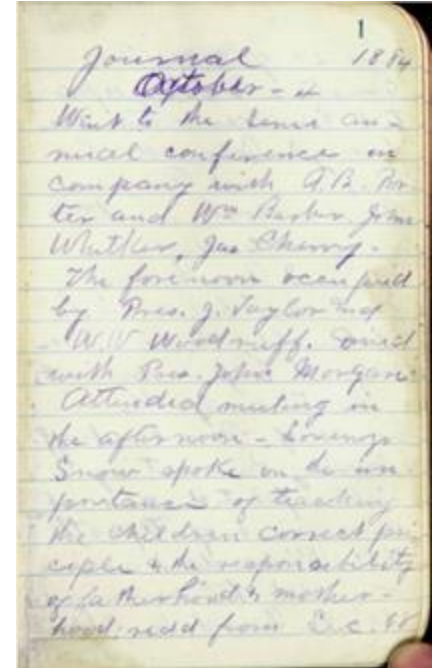
Phase 1: data model review

Data to capture

- Preferred form of name
- Alternate forms of name
- Local authority source
- Institution holdings
- Relationship information

Data Models Explored

- SKOS
- OWL
- BIBFRAME Authorities/Agent/Role
- **EAC-CPF**



Diary of B. H. Roberts, 1884-1885, page 1
<http://content.lib.utah.edu/cdm/ref/collection/uw/id/3069>

Phase 1: data model review

- Discussed possibilities for data modeling with partners
- Based on data samples received and goals for project, thought EAC-CPF best fit our needs
- Archival standard, many digital collections come from Special Collections/Archives
- <http://eac.staatsbibliothek-berlin.de>
- SNAC - <http://socialarchive.iath.virginia.edu/index.html>



Brigham Young portrait
Utah State University

Phase 1: data model review (EAC record)

```
<?xml version="1.0" encoding="UTF-8"?>
<eac-cpf xmlns="urn:isbn:1-931666-33-4" xmlns:xlink="http://www.w3.org/1999/xlink" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="urn:isbn:1-931666-33-4 http://eac.staatsbibliothek-berlin.de/schema/cpf.xsd">
  <control>
    <recordID>http://wnafvocab.org/ark:/12345/67891011/</recordID>
    <maintenanceAgency>
      <agencyName>Western Name Authority File</agencyName>
    </maintenanceAgency>
  </control>
  <cpfDescription>
    <identity>
      <entityType>person</entityType>
      <nameEntry>
        <part>Savage, C. R. (Charles Roscoe), 1832-1909</part>
        <authorizedForm>LC</authorizedForm>
        <authorizedForm>utsu</authorizedForm>
      </nameEntry>
      <nameEntry>
        <part>Savage, Charles Roscoe, 1832-1909</part>
        <alternativeForm>LC</alternativeForm>
      </nameEntry>
    </identity>
    <relations>
      <resourceRelation resourceRelationType="creatorOf" xlink:role="" xlink:actuate="onRequest"
        xlink:href="http://contentdm.lib.byu.edu/cdm/search/collection/Savage/" xlink:show="new"
        xlink:type="simple">
        <relationEntry>Charles R Savage</relationEntry>
      </resourceRelation>
    </relations>
  </cpfDescription>
</eac-cpf>
```

Phase 1: metadata wrangling (data submitted)

Miller, John N.
Miller, P. B.
Mizpah
Mogo, Margaret
Monroe
Montgomery, S. H.
Moody, William
Moore, Eri James
Morgan, David D.
Morgan, John
Morley, Isaac
Morris, R. M.
Morriss, R. M.
Mundock, John

National Ballet (Washington, D.C.)
Fauth & Co
Colorado Repertory Dance Company
Colorado School of Mines
Welgos Photos
De Bellis Studios
Colorado State College (Fort Collins, Colo.)

```
{
  "@id": "http://opaquenamespace.org/ns/people/AbdulBassitJalila",
  "@type": "skos:PersonalName",
  "dc:issued": {
    "@value": "2015-07-16",
    "@type": "xsd:date"
  },
  "dc:modified": {
    "@value": "2015-07-16",
    "@type": "xsd:date"
  },
  "rdfs:isDefinedBy": {
    "@id": "http://opaquenamespace.org/VOCAB_PLACEHOLDER.nt"
  },
  "rdfs:label": {
    "@value": "Abdul-Bassit, Jalil A.",
    "@language": "en"
  },
}
```

```
1 Title Description Photographer Image Date Subject
Location Collection Digital Publisher Contributing
Institution Ordering and Permissions Information
Date Digital Donor Image ID Digitization
Specifications Resource Type Format Language OCLC
number Date created Date modified Reference URL
CONTENTdm number CONTENTdm file name CONTENTdm file path
2 Quilici Barn, Reno Photograph of a barn built between
1900 and 1910; photograph taken August 31, 2002;
photographic print, 8 x 10 inches Barnes, Catana L.
2002 Barns -- Nevada Reno (Nev.); Washoe County;
Nevada Sagebrush Vernacular: Rural Architecture in
Nevada Special Collections Department, University of
Nevada, Reno Library University of Nevada, Reno Not to
be reproduced without permission. For permissions
information, please see:
http://knowledgecenter.unr.edu/speccoll/photos.aspx To
order a copy of this image, add it to your favorites, then
visit the shopping cart located here:
http://contentdm.library.unr.edu/shoppingcart/index.php
2003 UNRS-P2003-13-109 These images exist as
archived TIFF images and one or more JPEG versions for
general use. Image image/tiff eng 2003-04-23
2014-10-09
```

Phase 1: metadata wrangling (data compiled)

	A	B	C	D	E
1	Name	Institution	Collection	type	field
2	Savage, Charles R.	USHS	Classified Photos	personal name	Person
3	Savage, Charles R.	USHS	Classified Photos		Photographer
4	Savage, Charles R., 1832-1909 USE Savage, C. R. (Charles Roscoe), 1832-1909	BYU	Savage		photographer
5	Savage, Charles R.;	USHS	Classified Photos		Photographer
6	Savage, Charles R.; Savage, Lenny Louise; Savage, Annie Amelia; Savage, Ray T.	USHS	Classified Photos	personal name	Person
7	Savage, Charles Roscoe	BYU	mormon bib		author
8	Savage, Charles Roscoe, 1832-1909 USE Savage, C. R. (Charles Roscoe), 1832-1909	BYU	DUP	personal name	personal names
9	Savage, Chas. R.	USHS	Classified Photos		Photographer
10	Young, S. A.	Oregon		personal name	
11	Young, Seraph	USHS	Classified Photos	personal name	Person
12	Young, Seymour B.	USHS	Classified Photos	personal name	Person
13	Young, Seymour B. (Seymour Bicknell), 1837-1924; Young, Isaac Bray;	UU	John Mills Whitaker Papers		Creator
14	Young, Shirrel R.	USHS	Classified Photos		Photographer
15	Young, Stuart Michael;	UU	UAIDA		Creator
16	Young, Susan (Susannah) Snively	USHS	Classified Photos	personal name	Person
17	Young, Trevor	DU		personal name	
18	Young, Virginia; Young, Josephine; Lyon, Camilla; Compson, Lucine	USHS	Classified Photos	personal name	Person
19	Young, Whitney	Oregon		personal name	
20	Young, Whitney M.	DU		personal name	
21	Young, Willard	USHS	Classified Photos	personal name	Person

Phase 1: metadata wrangling (data deduplicated)

	A	B	C	D	E	F
	Name	Institution;Collection;field	type	# of collections	# of institutions	# of states
1	Young, Brigham, 1801-1877	UU;EAD;n/a -- BYU;DUP;personal names -- BYU;GEA;personal names -- BYU;DUP;personal names -- BYU;GEA;personal names -- BYU;19th Century Mormon	PN	29	3	1
2	Anderson, John	UU;EAD;n/a -- BYU;Diaries;personal names -- BYU;DUP;personal names -- DU;n/a;n/a -- USHS;Classified Photos;Person -- UU;John Mills	PN	6	4	2
3	Jackson, William Henry, 1843-1942	BYU;Jackson;personal names -- BYU;Jackson;photographer/artist -- BYU;OTImages;photographer/artist/creator --	PN	6	4	2
4	Savage, C. R. (Charles Roscoe), 1832-1909	BYU;DUP;personal names -- BYU;GEA;personal names -- BYU;Savage2.;personal names -- BYU;Savage;personal names -- BYU;DE1;author -- BYU;DE2;author --	PN	25	4	1
5	Utah Fuel Company	UU;EAD;n/a -- BYU;RSMag;company name -- USHS;Classified Photos;Corporate Body -- USU;Lee;creator	CB	4	4	1
6						

Phase 1: metadata wrangling (data reconciled)

	A	B	C	D
1	Name	LCNAF Heading	Bad Match	Good Match
2	A. C. C.	Pushkin, Aleksandr Sergeevich, 1799-1837	http://id.loc.gov/authorities/names/n80036579	
3	Aaldeman, Mike			
4	Aas, Morten			
5	Abbot, Edward	Parry, Edward Abbott, Sir, 1863-1943	http://id.loc.gov/authorities/names/n80080252	
6	Abbott, Edith, 1876-1957	Abbott, Edith, 1876-1957		http://id.loc.gov/authorities/names/n50034801
7	Abbott, Ernest Hamlin	Abbott, Lyman, 1835-1922	http://id.loc.gov/authorities/names/n50034805	
8	Abbott, James Smith, 1867-1944	Abbott, James Smith, 1867 or 1868-1944		http://id.loc.gov/authorities/names/no96022313
9	Abbott, John	McBain, Ed, 1926-2005	http://id.loc.gov/authorities/names/n79121690	
10	Abbott, John Austin, 1871-1954	Abbott, John Austin, 1871-1954		http://id.loc.gov/authorities/names/no2013051261
11	Abbott, Mark	Stern, Mark Abbott, 1929-	http://id.loc.gov/authorities/names/n2009070199	
12	Abbott, Myron Alma, 1862-1932	Abbott, Myron Alma, 1862-1932		http://id.loc.gov/authorities/names/no2014098724
13	Abby, Jack	Neidik, Abbey Jack	http://id.loc.gov/authorities/names/no2008053570	
14	Abegg, Isaura Bentley	Abegg, Isaura Bentley		http://id.loc.gov/authorities/names/no2015022649
15	Abel, Albert	Jullien, Louis Antoine, 1812-1860	http://id.loc.gov/authorities/names/n79140958	
16	Abel, Georgie Lee	Abel, Georgie Lee		http://id.loc.gov/authorities/names/no96029282
17	Abel, Mary Hinman, b. 1850	Abel, Mary Hinman, 1850-		http://id.loc.gov/authorities/names/n90652276
18	Abel, Oscar	Ligaluppi, Oscar Abel	http://id.loc.gov/authorities/names/n82156348	
19	Abernathy, Ralph	Abernathy, Ralph, 1926-1990		http://id.loc.gov/authorities/names/n87870916
20	Abernathy, Ralph, 1926-	Abernathy, Ralph, 1926-1990		http://id.loc.gov/authorities/names/n87870916
21	Abernathy, Ralph, 1926-1990	Abernathy, Ralph, 1926-1990		http://id.loc.gov/authorities/names/n87870916
22	Abert, James William, 1820-1897	Abert, J. W. (James William), 1820-1897		http://id.loc.gov/authorities/names/no98079676
23	Abram, James	Garfield, James A. (James Abram), 1831-1881	http://id.loc.gov/authorities/names/n50015182	
24	Abrams, Milton C.	Abrams, Milton C. (Milton Charles), 1918-		http://id.loc.gov/authorities/names/no2008117252

Phase 1: metadata wrangling (data metrics)

- Started with ~500,000 names
- Deduplicated to 76,360
- 7357 -- 2+ collections (9.6%)
- 1484 -- 2+ institutions (1.9%)
- 271 -- 2+ states (0.35%)
- 62,381 personal names
- 10,706 corporate bodies
- 3,273 unknown

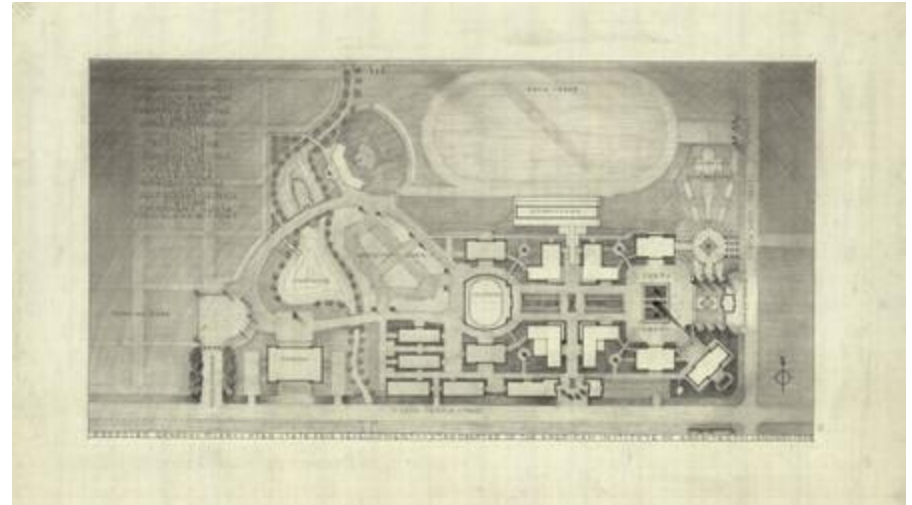
- 1091 are single words
- 2400+ cross references
- 500+ PN are First Last

Total names submitted

- Brigham Young University - 30,535
- University of Utah - 7533
- Utah State University - 2067
- Utah State Historical Society - 12,138
- Utah State Archives - 3657
- University of Nevada, Reno - 1277
- Oregon Digital - 4170
- University of Denver - 16,608

Phase 2: investigating tools

- With help of programmers:
 - Sandbox Environment- Virtual Machines, typically 2 cores, 2gb memory, all using ubuntu 14.04 or 16.04, mysql, php, apache, nginx, tomcat, utilizing the git version control system



Utah State Fair development: Utah Chapter of the American Institute of Architects: sheet 17

George Cannon Young Architecture

<https://collections.lib.utah.edu/details?id=865147>

Phase 2: investigating tools (evaluation matrix)

<https://goo.gl/wMQRbY>

Name	Ready for Testing?	Local URL	Technical Support Requirements	Software type (backend, middleware, framework, complete solution)	Batch Import Support	Batch editing of controlled vocabulary terms	LOD publishing
Skosmos		skosmos.stagin.g.lib.utah.edu	From website: "one or more SKOS vocabularies PHP capable web server SPARQL triple store (we recommend Apache Jena Fuseki with jena-text) HTTP cache (e.g. Varnish) - optional but recommended"	complete solution			
Collective Access	Yes	https://collectiveaccess.stagin.g.lib.utah.edu	Apache, PHP, MySQL http://docs.collectiveaccess.org/wiki/Requirements	complete solution			

Phase 2: investigating tools

What Metadata People care about:

Workflow issues:

- Batch Import Support
- Batch editing of terms
- Collaborative Workflows
- Local URI support

Discovery Issues:

- Data models supported
- Data visualization
- SPARQL Endpoint
- LOD publishing
- Advanced search capabilities
- Browse capabilities

Phase 2: investigating tools

What digital infrastructure/programmers care about:

- Open Source Software Development Support - full community or just one person's project
- Ease of Installation for DID
- Ongoing support considerations from DID
- API availability
- Technical Support Requirements
- Software type (backend, middleware, complete solution)

Phase 2: investigating tools

TemaTres



xEAC



ThManager



protégé

CA COLLECTIVE ACCESS

VocBench 2.0



RAMP

Opaque Namespace
Controlled Vocabulary Manager

redis



Skosmos

Vitro



Phase 2: investigating tools (tools ruled out)



xEAC

RAMP



Phase 2: investigating tools (starting tools)



Phase 2: investigating tools (tools to install)



Phase 2: investigating tools (need to look at)

VocBench 2.0

The logo for iQvoc features a stylized cube with a green top face and an orange front face, followed by the text "iQvoc" in a bold, blue, sans-serif font.

protégé

ThManager

Opaque Namespace
Controlled Vocabulary Manager

redis

Vitro

Phase 3: pilot implementation

- Perform full evaluation of selected software - harvest, standardize, reconcile, and import controlled vocabulary information into the software of choice and make data available as LOD
- Hire and train student assistant to facilitate data entry, vocabulary reconciliation and enrichment, research, vocabulary maintenance, and assessment tasks.
- Enrich data with relationships and collections holding information.
- Explore the possibility of setting up an OpenRefine reconciliation service for the vocabulary
- Develop and revise collaborative workflows

Phase 4: assessment (workflows)

- Managing vocabulary
- Reconciliation issues
 - Many partners on CONTENTdm
 - Need external system not tied to specific digital library software/infrastructure
- Explore impact of WNAF on metadata creators and users
- Training opportunities



Block Prints of Lennox and Catherine Tierney Private Art Collection [002]: Geisha under Flowering Tree
Lennox and Catherine Tierney Photograph Collection
<https://collections.lib.utah.edu/details?id=329534>

Phase 4: assessment (statistics)

- Capture and assess data on the percentage of names not in a national authority file, the number of names unique to one institution, and number of relationships we are able to express with the vocabulary.

8. Navaho man and child. 1913 Dr. Elliott, photo
Utah American Indian Digital Archive
<https://collections.lib.utah.edu/details?id=390032>



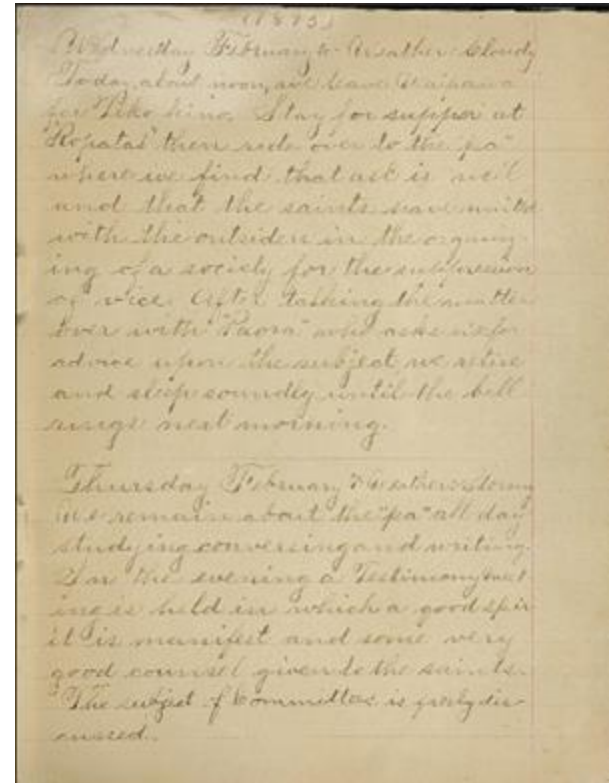
Phase 4: assessment (statistics)

Search results in December 2016	Creator		Subject	
	MWDL	DPLA	MWDL	DPLA
Savage, C. R.	2022	2226	44	50
Savage, C. R. (Charles Roscoe)	1830	2023	43	44
Savage, C. R. (Charles Roscoe), 1832-1909	1708	1901	43	43
Savage, C. R., 1832-1909	1708	1901	43	43
Savage, Charles	1936	2165	60	126
Savage, Charles R.	1932	2128	58	65
Savage, Charles R., 1832-1909	1708	1902	43	43
Savage, Charles Roscoe	1833	2034	44	53
Savage, Charles Roscoe, 1832-1909	1708	1904	43	43
Savage, Charles, 1832-1909	1708	1905	43	43
Savage, Chas. R.	1	1	0	1
Savage, Chas. R., 1832-1909	0	0	0	0

Phase 4: assessment (expanding use)

- Develop a plan for expanding the controlled vocabulary to more institutions.
- Consider what would be needed to move into full production for the project.

Thomas Jefferson O'Brien journal commencing Feb. 6, 1895
https://collections.lib.utah.edu/details?id=1042220#t_1042220



Further reading and acknowledgements

Full Grant Narrative

<https://www.imls.gov/grants/awarded/LG-72-16-0002-16>

Project Webpage

<https://sites.google.com/site/westernnameauthorityfile/home>

This project was made possible in part by the
Institute of Museum and Library Services LG-72-
16-0002-16.

