

Western Name Authority File: Creating a Collaborative Regional Vocabulary

Anna Neatrour (@annaneat)

Jeremy Myntti (@jmyntti)



J. Willard Marriott Library

THE UNIVERSITY OF UTAH

DPLA Fest, April 2017

Why?

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
BYU Digital Collections

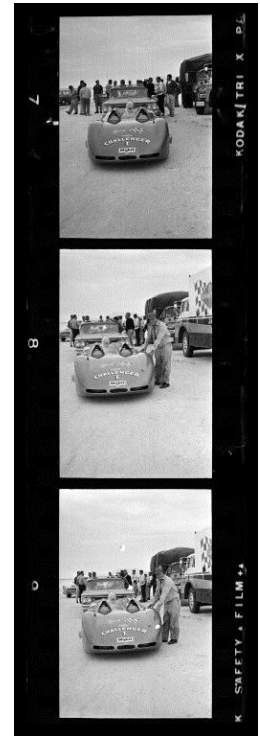
<http://contentdm.lib.byu.edu/cdm/ref/collection/Savage2/id/1749>

More metadata, more
metadata problems...

IMLS Grant for regional authority control

4 phases:

- 1) investigating data models to express local/regional name authority data using linked data standards
- 2) evaluation of tools used for creating, maintaining, and making this data available
- 3) pilot implementation using the tools investigated in the second phase
- 4) assessment of how this type of authority data can improve digital collection metadata on a local, regional, and national level.



What will save us?

Libraries Working Together

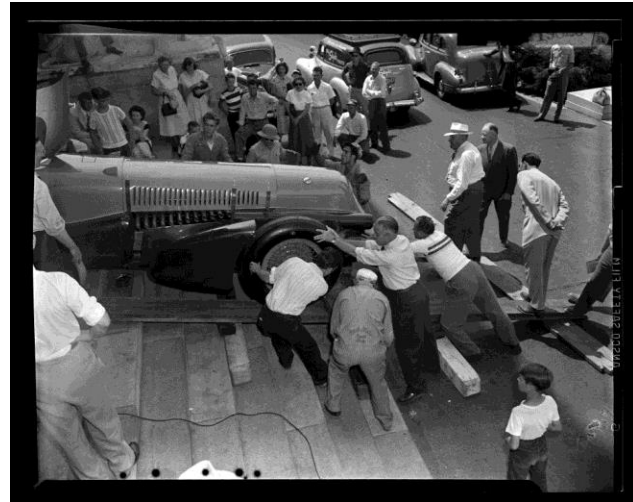
Metadata Review

Received data from 8 institutions

- Brigham Young University
- Oregon Digital (Univ. of Oregon and Oregon State Univ.)
- University of Nevada, Reno
- University of Utah
- University of Denver
- Utah Department of Heritage and Arts
- Utah State Archives
- Utah State University

Data gathered

- Name (including associated dates)
- Institution
- Collection
- Field(s)
- Type (personal name or corporate body)
- Cross references



How Will It Happen?

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

(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.staging.lib.usf.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.staging.lib.usf.edu	Apache, PHP, MySQL http://docs.collectiveaccess.org/wiki/Requirements	complete solution			

Phase 2: investigating tools/frameworks

TemaTres



ThManager

VocBench 2.0

CA COLLECTIVE
ACCESS

Opaque Namespace
Controlled Vocabulary
Manager

redis



Refine OPEN



Skosmos

Vitro

DATA MODELS!!!!!!

Data Models: Collaborative Investigation

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** – look to SNAC as a model
<http://socialarchive.iath.virginia.edu/>



Campbell, Donald; Salt Flats – Shot 2
<https://collections.lib.utah.edu/details?id=691545>

Documenting Change

Stats script leveraging DPLA API

SQLite database, use python Requests library to query DPLA API - get stats on coverage for name variants in MWDL and DPLA.

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

Workflow considerations - 2nd year

- 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



Thompson, Mickey; Salt Flats -Shot 11
Utah Department of Heritage and Arts, Salt Lake Tribune
Negative Collection
<https://collections.lib.utah.edu/details?id=694077>

Further reading and acknowledgements

Full Grant Narrative

<https://www.ims.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.

