From EAD to MARC to Wikidata and Back Again
Tools and Workflows for Archival Metadata at the University of Washington Libraries

15th Semantic Web in Libraries Conference
Berlin, Germany, September 13, 2023

Crystal Yragui, Science Cataloger
Adam L. Schiff, Principal Cataloger
ACKNOWLEDGMENTS

Benjamin Riesenberg
Conor Casey
Elaine Speer
Eulalie Mathieu
Junghae Lee
Mark Carlson
Melissa Morgan
Thad Guidry
Theodore Gerontakos
UW Libraries Administration
UNIVERSITY LIBRARIES
UNIVERSITY of WASHINGTON
Labor Archives of Washington

- Founded in 2010
- Collaboration between the UW Libraries, Harry Bridges Center for Labor Studies, and the labor community
- Over 300 collections, one of the largest repositories for labor materials in the United States
- Mission to ensure that “current and future generations understand the struggles and accomplishments of the working people that built this region”
- Access and preservation of records, records management advice, educational and outreach projects with unions, schools, and the public
Collection Highlights

Topics:
- Political history
- Ethnic and immigrant history
- Legal history
- Civil rights
- Urban history
- Women’s history/gender studies
- Industrial relations
- Social history
- Economic history

Events:
- 1916 Everett Massacre
- 1919 Seattle General Strike
- 1934 Waterfront Strike
- UFW Grape Boycott
- 1999 Seattle WTO protests
LAW Metadata Before Wikidata Enhancements

- Finding aids published by Archives West using Encoded Archival Description (EAD)
- Bibliographic records published in OCLC Worldcat as MARC21
- Name authority records published by the Library of Congress/Name Authority Cooperative Program (NACO) of the Program for Cooperative Cataloging (PCC)
- Wikipedia articles published as text, with knowledge boxes as linked data
- Entities for agents published as linked data by Social Networks and Archival Context (SNAC)
- Various websites and online news outlets publish text
Cannery Workers and Farm Laborers Union, Local 7 records, 1915-2000

Overview of the Collection

Creator: Cannery Workers and Farm Laborers Union, Local 7 (Seattle, Wash.)

Title: Cannery Workers and Farm Laborers Union, Local 7 records

Dates: 1915-2000 (inclusive)

Quantity: 58.77 cubic feet (75 boxes plus 1 oversize folder)

Collection Number: 3927

Summary: Labor union of Seattle, Washington

Table of Contents

Overview of the Collection
Historical Note

Content Description

Use of the Collection
Administrative Information

Detailed Description of the Collection

Accession No. 3927-001: Cannery Workers and Farm Laborers Union, Local 7 records, 1915-1985 (bulk 1933-1982)

Accession No. 3927-002: Cannery Workers and Farm Laborers Union, Local 7 records, 1915-2000

Guide to the Cannery Workers and Farm Laborers Union Local 7 Records

1915-2000

Cannery Workers and Farm Laborers Union Local 7 Records

Special Collections, University of Washington Libraries

Seattle, WA 98195

Finding aid written in English

Finding aid based on DACS
**MARC 21**

![MARC 21 Image]
Founded in 1933 as the Cannery Workers and Farm Laborers Union, Local 18257 of the American Federation of Labor (AFL), it represented Alaska salmon cannery workers and farm workers. In 1937, the union became Cannery Workers and Farm Laborers Union, Local 7 of the United Cannery, Agricultural, Packing, and Allied Workers of America of the Congress of Industrial Organizations (CIO). In 1945, Local 7 became affiliated with the Food, Tobacoo, and Agricultural Workers of America (CIO). In 1951 the union became Local 37 of International Longshoremen’s and Warehousemen’s Union, and ca. 1987 it became Region 37 of IBU/ILWU. The membership historically was Filipino American canner workers.

The Cannery Workers’ and Farm Laborers’ Union was organized June 19, 1933 in Seattle to represent the primarily Filipino-American laborers who worked in the Alaska salmon canneries. Filipino Alaskeros first appeared in the canneries around 1911. In the 1920s as exclusionary immigration laws went into effect, they replaced the Japanese, who had replaced the Chinese in the canneries. Workers were recruited through labor contractors who were paid to provide a work crew for the summer canning season. The contractor paid wages and other expenses. This system led to many abuses and harsh working conditions from which grew the movement toward unionization.

The CWFLU, under the leadership of its first President, Virgil Duyungan, was chartered as Local 19257 by the American Federation of Labor in 1933. On December 1, 1936 an agent of a labor contractor murdered Duyungan and Secretary Aurelio Simon. Despite this setback, the union was able to win a hiring hall and end the contract labor system in 1937. After Duyungan’s death, Conrad Espe, a Norwegian-American labor organizer, took the leading role in the union. Under the leadership of Duyungan and Espe, the CWFLU made numerous attempts to organize farm workers during the winter months. Farm Division organizers attempted to organize workers in Yakima, Kent, Everett, Bainbridge Island and the White River area, but were often met with harsh opposition from local officials and vigilantes.

Local 18257 came into conflict with the AFL, in 1937 when the parent body, attempting to separate the union along racial lines, recognized a Japanese local organized by Clarence Arai. Local 18257 successfully retained negotiation rights and dispatched its workers in 1937 despite pickets set up by the rival group. Bitterness toward the AFL resulted from the incidents and led to a November 4 vote by the Seattle, Portland, and San Francisco locals to affiliate with the newly formed United Cannery, Agricultural, Packinghouse and Allied Workers of America-CIO (UCAPAWA). In
Cannery Workers and Farm Laborers Union, Local 7

From Wikipedia, the free encyclopedia

The Cannery Workers and Farm Laborers Union, Local 7 was the first Filipino-led union in the United States.\(^1\)

Founded in 1933 as the Cannery Workers and Farm Laborers Union, Local 18257 of the American Federation of Labor (AFL), it represented Alaska salmon cannery workers and farm workers. In 1937, the union became Cannery Workers and Farm Laborers Union, Local 7 of the United Cannery, Agricultural, Packing, and Allied Workers of America. In 1945, Local 7 became affiliated with the Food, Tobacco, Agricultural, and Allied Workers. In 1951 the union became Local 37 of International Longshore and Warehouse Union, and around 1987 it became Region 37 of IBU/ILWU. The membership historically was Filipino American cannery workers.

History [edit source]

Founding [edit source]
Web Resources

Filipino Cannery Unionism Across Three Generations 1930s-1980s

Cannery Workers' and Farm Laborers' Union 1933-39
Their Strength in Unity

by Crystal Fresco

The United States took possession of the Philippines in 1898 and in the decades after that, Filipinos, mostly men, began to make their way to America to seek employment, especially in the fields and canneries. In 1933 some of these men formed the first Filipino-led union ever organized in the United States: the Cannery Workers' and Farm Laborers' Union Local 18257. Based in Seattle, it was organized by "Alaskeros" who worked in the Alaska salmon canneries each summer and in the harvest fields of Washington, Oregon, and California in the other seasons. The union was in its shaky beginnings when two of its founders were murdered. Yet, although its leaders were dead, the union would not die. Instead in the next few years, it grew stronger, becoming effective up and down the West Coast.

This paper investigates the early history of the Cannery Workers' and Farm Laborers' Union and the sense of pride and fraternity within the Alaskero brotherhood that made the union possible. Much of the credit has to be given to the Filipino community. The workers believed in community and unity. The Cannery Workers' and Farmers' Union's
Challenges

- Consistency across platforms
  - Completeness in non-Archives West platforms
  - Updates in Archives West not translating to other platforms
  - Broken or missing links
- Connections between description sets about the same entities
- Connections between interrelated entities (many about the same groups/people/ideas), e.g. Radical Women, IWW, ILWU
- Budgetary and time constraints limit ability to address challenges
What Can Wikidata Offer?

- Identifier hub
- Enhanced discoverability
- Access to more data
  - Archives held at other institutions
  - Connections to people and ideas described outside finding aids
  - SPARQL endpoint
- Community contributions
- Larger pool of volunteer editors
- Open source power editing tools
- No paywalls
- Flexible data structure
Archival Metadata at UW Libraries
Crystal E. Yragui | August 22, 2023

Created using LucidChart. Dashed grey lines are links. Solid lines indicate workflow steps.
Application Profile for Archival Collections

- Developed by creating some Wikidata items by hand and identifying properties that were needed to sufficiently describe a collection
- Helped to identify properties that were lacking
- Includes mandatory and optional elements, including qualifiers that are used
- Includes examples illustrating use of properties
- Living document that can be revised
- Maintained in Wikidata
### Application Profile for Archival Collections

**Label**

*Example:* Seattle Minimum Wage History Project Records

**Description**

- archival collection held at the Labor Archives of Washington
- archival collection held at the University of Washington Special Collections
- collection of films shot between 1929 and 1932 by Dr. Joseph F. Rock on his numerous expeditions into the remote border region between China and Tibet, held by the University of Washington Libraries Media Archive
- archival collection of approximately 1500 hours of sound recordings covering World War II era daily news broadcasts, radio dramas, speeches, and interviews, held by the University of Washington Libraries Media Archive

**Also known as**

Variant title(s)

**Statements**

- on focus list of Wikimedia Project (P5008)
  - Q119925630 = Wikidata:WikiProject University of Washington Archival Metadata

- instance of (P31)
  - archival collection (Q0388534) and, if appropriate, other specific values (e.g. manuscript collection (Q42932533), photo archive (Q27032363), sound archive (Q2230431), audio-visual archive (Q2730766))
<table>
<thead>
<tr>
<th>Field</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>title (P1476)</td>
<td>Title proper (MARC 245 a). Use capitalization as found in source metadata (MARC or EAD). Prefer capitalization used in finding aid.</td>
</tr>
<tr>
<td>subtitle (P1680)</td>
<td>Subtitle (MARC 245 b). Use capitalization as found in source metadata (MARC or EAD). Prefer capitalization used in finding aid.</td>
</tr>
<tr>
<td>collection creator (P6241)</td>
<td>Generally will be the entity given in the 1XX field of the MARC record for the collection. Multiple creators would be in 1XX and 7XX.</td>
</tr>
<tr>
<td>country (P17)</td>
<td>United States of America (G30)</td>
</tr>
<tr>
<td>located in the administrative territorial entity (P131)</td>
<td>Seattle (G5983)</td>
</tr>
<tr>
<td>location (P276)</td>
<td>Examples: Labor Archives of Washington (G22096698)</td>
</tr>
<tr>
<td></td>
<td>Washington State Jewish Archives (G109813923)</td>
</tr>
<tr>
<td></td>
<td>University of Washington Libraries Media Archive (G119830690)</td>
</tr>
<tr>
<td></td>
<td>University of Washington Libraries Special Collections Division (G55250361)</td>
</tr>
<tr>
<td>maintained by (P126)</td>
<td>Examples: Labor Archives of Washington (G22096698)</td>
</tr>
<tr>
<td></td>
<td>University of Washington Libraries Media Archive (G119830690)</td>
</tr>
<tr>
<td></td>
<td>University of Washington Libraries Special Collections Division (G55250361)</td>
</tr>
<tr>
<td>part of (P361)</td>
<td>Examples: Labor Archives of Washington (G22096698)</td>
</tr>
<tr>
<td></td>
<td>Washington State Jewish Archives (G109813923)</td>
</tr>
<tr>
<td></td>
<td>University of Washington Libraries Media Archive (G119830690)</td>
</tr>
<tr>
<td></td>
<td>University of Washington Libraries Special Collections Division (G55250361)</td>
</tr>
<tr>
<td>collection (P185)</td>
<td>Examples: Labor Archives of Washington (G22096698)</td>
</tr>
<tr>
<td></td>
<td>Washington State Jewish Archives (G109813923)</td>
</tr>
<tr>
<td></td>
<td>University of Washington Libraries Media Archive (G119830690)</td>
</tr>
<tr>
<td></td>
<td>University of Washington Libraries Special Collections Division (G55250361)</td>
</tr>
<tr>
<td>inventory number (P217)</td>
<td></td>
</tr>
<tr>
<td><strong>inventory number</strong> <em>(P217)</em></td>
<td><strong>collection</strong> <em>(P195)</em></td>
</tr>
<tr>
<td>-------------------------------</td>
<td>------------------------</td>
</tr>
<tr>
<td>start of covered period (MARC 046 $o or 245 date info) <em>(P7103)</em></td>
<td><strong>point in time</strong> <em>(P585)</em></td>
</tr>
<tr>
<td>end of covered period (MARC 046 $p or 245 date info) <em>(P7104)</em></td>
<td><strong>point in time</strong> <em>(P585)</em></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>collection or exhibition size (MARC 300 $a)</strong> <em>(P1436)</em></th>
<th><strong>sourcing circumstance</strong> <em>(P1480)</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>Examples: 0.42 cubic foot 13.8 cubic foot 5 cubic foot sourcing circumstance: approximately 0.72 cubic foot point in time: 3 August 2023 quantity: 2 box 4 videocassette point in time: 3 August 2023</td>
<td><strong>point in time</strong> <em>(P585)</em> <strong>quantity</strong> <em>(P1114)</em></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>scope and content</strong> <em>(P75564)</em></th>
<th><strong>Summary statement providing an overview of the archival collection; equivalent to MARC field 520 first indicator 2</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Examples: Correspondence, minutes, grievance, strike, and subject files, financial records, and ephemera. Correspondents include Forbes Bottomly, J. Bernard Chichester, William E. Morris, and Ross Roeder. Correspondence, organizational and historical materials (1913-1970); minutes (1899-1971); financial records (1913-1968); legal documents (1919-1966); court papers, hearing transcripts, reports, resolutions, speeches, writings, subject files, and other records, relating to the council and affiliated organizations including Labor Temple Corporation; Pierce County Central Labor Council; Politician Education Pierce County Committee; Tacoma Central Labor Council; and its Labor Day and Mother Robinson committees; United Labor Advisory Committee; Washington State Committee on Apprentice Training; Washington State Apprenticeship Council; Women's Union Card and Label League, Tacoma, Wash.; and the newspaper Tacoma Labor Advocate. Major correspondents include Homer Truett Bone and Warren Grant Magnuson. Includes taped interview with council officers Harry S. McVeaugh and Clyde H. Hopp. Photographs of Seattle collected by Jewish civic and business leader Nathan Eckstein. The collection contains views of Seattle, including Bailey Gatzert and the Bailey Gatzert home, Pioneer Square, and Nathan Eckstein marching in a Shriners parade in 1915.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>main subject (MARC 650/610/650) (P921)</strong></th>
<th><strong>Main subjects of this work. May need to create new subject items in order to use this property.</strong></th>
</tr>
</thead>
</table>
Two identifiers are given in every Wikidata collection item:

- Archives West finding aid ID
- OCLC control number
Wikidata Property Proposals

- Archives West is an online aggregator of over 40,000 finding aids for collections in 48 repositories in 6 western U.S. states
- Two property proposals were made for this project:
  - Archives West finding aid ID (P9335)
  - Archives West repository ID (P11366)
- All finding aids have the URL https://archiveswest.orbiscascade.org/ark:80444/ followed by a unique identifier. This stable pattern enables the Archives West finding aid property.
- Archives West repository ID enables retrieval of all finding aids for a particular repository from the Wikidata item for that repository.
Wikidata Property Proposals

- Propose new properties at https://www.wikidata.org/wiki/Wikidata:Property_proposal
- Once a proposal has been made, it is open for community discussion. New proposals are announced in the weekly Status Update https://www.wikidata.org/wiki/Wikidata:Status_updates
- Minimum time for discussion is one week
- If there is consensus after a week (or longer), a property creator will mark the proposal as ready for creation, and eventually a property creator will create the property
- Currently only 48 people have property creation authorization; editors may apply for this status
Wikidata:Property proposal/Archives West finding aid ID

Archives West finding aid ID [edit]

Originally proposed at Wikidata:Property proposal/Generic

- **Done**: Archives West finding aid ID (P9335) (Talk and documentation)

<table>
<thead>
<tr>
<th><strong>Description</strong></th>
<th>identifier of a finding aid in Archives West</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Represents</strong></td>
<td>Archives West (Q7060019)</td>
</tr>
<tr>
<td><strong>Data type</strong></td>
<td>External identifier</td>
</tr>
<tr>
<td><strong>Domain</strong></td>
<td>archival collection (Q9388534)</td>
</tr>
<tr>
<td><strong>Allowed values</strong></td>
<td>xv[0-9][5]</td>
</tr>
<tr>
<td><strong>Example 1</strong></td>
<td>American Federation of Teachers, Local 200 Records (Q165651381) → xv11962</td>
</tr>
<tr>
<td><strong>Example 2</strong></td>
<td>AFL-CIO Region 9 Records (Q105652219) → xv62756</td>
</tr>
<tr>
<td><strong>Example 3</strong></td>
<td>Al Allen Collection on the International Longshore and Warehouse Union (Q105652355) → xv17388</td>
</tr>
<tr>
<td><strong>Example 4</strong></td>
<td>Skowronek Family Papers (Q105653656) → xv50084</td>
</tr>
<tr>
<td><strong>Source</strong></td>
<td><a href="http://archiveswest.orbiscascade.org/">http://archiveswest.orbiscascade.org/</a></td>
</tr>
</tbody>
</table>

**Planned use**: will add to items for University of Washington finding aids in Archives West

**Expected completeness**: eventually complete (Q21873574)

**Formatter URL**: http://archiveswest.orbiscascade.org/ark:/80444/$1/

**Motivation** [edit]

University of Washington Libraries is creating items for our archival collections and we want to include the identifier for the Archives West finding aid describing each collection.

**Discussion** [edit]

- ✅ Support Pterotamus (talk) 18:15, 24 February 2021 (UTC) [reply]
- ✅ Support I would use this as part of this project to add agents from the Labor Archives of Washington to Wikidata. The project is being done as part of the UW’s PCC Wikidata Pilot with the hope of demonstrating the value of Wikidata to local archivists. --Crystal Clements, University of Washington Libraries (talk) 22:01, 28 February 2021 (UTC) [reply]
- ✅ Support --Jala360 13:58, 2 March 2021 (UTC) [reply]
- ✅ Support I think archives are interested in linked data in general (the SNAC project has been around for awhile) and this could help provide examples they would be very interested in and could use to generate institutional support. --Emwille (talk) 16:10, 3 March 2021 (UTC) [reply]
- ✅ Support I would use this as we begin to contribute data about our archival collections. Lrobare (talk) 19:30, 3 March 2021 (UTC) [reply]

@UWashPrincipalCataloger, Pterotamus, Clements UWL, Jala360, Emwille, Lrobare: ✅ Done Archives West finding aid ID (P9335) Pamputt (A) (talk) 08:13, 20 March 2021 (UTC) [reply]
United Food and Commercial Workers International Union, Local 81 records, 1896-2012

Overview of the Collection

Creator
United Food and Commercial Workers International Union, Local 81 (Seattle, Wash.)

Title
United Food and Commercial Workers International Union, Local 81 records

Dates
1896-2012 (inclusive)

Quantity
96.93 cubic feet (110 boxes)
1 package
1 tube
1 vertical file

Collection Number
5694 (Accession No. 5694-001)

Summary
Records collected and published by United Food and Commercial Workers Union Local 81, including earlier groups which merged into the UFCW Local 81.

Repository
University of Washington Libraries, Special Collections
Special Collections
University of Washington Libraries
Box 352900
Seattle, WA
98195-2900
Telephone: 2065431929
Fax: 2065431931
speccoll@uw.edu

Access Restrictions
Open to all users.

Languages
English

Sponsor
Processed with funds from the Labor Archives Fund, Labor Archives of Washington

Historical Note

The Amalgamated Meat Cutters and Butcher Workmen of North America (AMC & BW of NA), a labor union representing retail butchers and packinghouse workers, was chartered by the American Federation of Labor (AFL) in 1897, consolidating seven local unions in Chicago. The AMC & BW of NA was committed to craft unionism, with 56 departments representing various workers in the meat industry. Workers in each craft within a city had their own council, executive board, business agent and contract. In early 1900, nine Seattle butchers formed the Protective Union of Butchers, Local 81. This local was the first butcher trade union in Washington State, and would officially be chartered Local 81 of the Amalgamated Meat Cutters and Butcher Workmen of North America on April 2, 1900.
United Food and Commercial Workers International Union, Local 81 Records

(Q121301777)•••
ORES predicted quality: D (2.36)

archival collection held at the Labor Archives of Washington
UFCW Local 81 Records

Identifiers

<table>
<thead>
<tr>
<th>Archives West finding aid ID</th>
<th>xv41351</th>
<th>edit</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0 references</td>
<td>+ add reference</td>
</tr>
<tr>
<td>+ add value</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>OCLC control number</th>
<th>864724204</th>
<th>edit</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0 references</td>
<td>+ add reference</td>
</tr>
<tr>
<td>+ add value</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
EAD to Wikidata Workflow for Agents

Identity Management for Creators during PCC Wikidata Pilot Project

GitHub Repository: https://github.com/uwlibcams/SCArchivesAgents

EAD to Wikidata Workflow for Agents

- Extract selected nodes from EAD XML finding aids using Python
- Perform reconciliation against Wikidata in OpenRefine
- Export Quickstatements batch from OpenRefine for item creation
  - Used “archives at” property with Archives West finding aid ID’s for each entity
- By hand, do more granular identity management in NACO and Wikidata as time allows (still in progress)
- Example:
  [https://www.wikidata.org/wiki/Q16972177](https://www.wikidata.org/wiki/Q16972177)
Region 37 IBU/ILWU (Q16972177)

first Filipino-led union in the United States; labor union local representing cannery workers and farm laborers in Seattle, Washington

Cannery Workers' and Farm Laborers Union Local 18257 | Cannery Workers and Farm Laborers Union Local 7 | CWFLU Local 18257
| CWFLU Local 7 | UCAPAWA Local 7 | Food, Tobacco, Agricultural and Allied Workers Union of America Local 7
| International Longshoremen and Warehousemen Union Local 7-C | ILWU Local 37
| United Cannery, Agricultural, Packinghouse, and Allied Workers of America Local 7
| International Longshoremen's and Warehousemen's Union Local 37 | International Longshore and Warehouse Union Local 37
| Region 37 of the IBU/ILWU

> Reoin. Most relevant properties which are absent

In more languages

<table>
<thead>
<tr>
<th>Language</th>
<th>Label</th>
<th>Description</th>
<th>Also known as</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td>Region 37 IBU/ILWU</td>
<td>first Filipino-led union in the United States; labor union local representing cannery workers and farm laborers in Seattle, Washington</td>
<td>Cannery Workers' and Farm La... Cannery Workers and Farm Lab... CWFLU Local 18257 CWFLU Local 7 UCAPAWA Local 7 Food, Tobacco, Agricultural and ... International Longshoremen an... ILWU Local 37 United Cannery, Agricultural, Pa... International Longshoremen's a... International Longshore and Wa... Region 37 of the IBU/ILWU</td>
</tr>
<tr>
<td>Spanish</td>
<td>No label defined</td>
<td>No description defined</td>
<td></td>
</tr>
</tbody>
</table>

Statements
### Statements

**instance of**

- local union
  - 0 references
  - + add reference
  - + add value

**notable work**

- Cannery Workers and Farm Laborers Union, Local 7 Records
  - 1 reference
  - + add value

**archives at**

- Labor Archives of Washington
  - subject named as: Cannery Workers and Farm Laborers Union, Local 7 photographs
  - inventory number: PH1844
  - Archives West finding aid ID: xv33776
  - 1 reference

- Labor Archives of Washington
  - subject named as: Cannery Workers and Farm Laborers Union, Local 7 records
  - inventory number: 3927
  - Archives West finding aid ID: xv82950
  - 1 reference

**country**

- United States of America
  - 0 references
  - + add reference
  - + add value

**located in the administrative territorial entity**

- Seattle
  - 0 references
Identifiers

 VIAF ID

- 12675562
  - 0 references
  
- 129113227
  - 0 references
  
- 153763289
  - 0 references
  
- 127949966
  - 0 references
<table>
<thead>
<tr>
<th>Library of Congress authority ID</th>
<th>Freebase ID</th>
<th>SNAC ARK ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>subject named as Cannery Workers and Farm Laborers Union. Local 7 (Seattle, Wash.)</td>
<td>/m/0rdwotm</td>
<td>v6mb3b4x</td>
</tr>
<tr>
<td>+ 0 references</td>
<td>+ add reference</td>
<td>+ add value</td>
</tr>
<tr>
<td>n88150562 subject named as Cannery Workers and Farm Laborers Union. Local 18257 (Seattle, Wash.)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>+ 0 references</td>
<td>+ add reference</td>
<td>+ add value</td>
</tr>
<tr>
<td>n87890405 subject named as International Longshoremen's and Warehousemen's Union. Local 37 (Seattle, Wash.)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>+ 0 references</td>
<td>+ add reference</td>
<td>+ add value</td>
</tr>
<tr>
<td>n88150565 subject named as Region 37 IBU/LWU</td>
<td></td>
<td></td>
</tr>
<tr>
<td>+ 0 references</td>
<td>+ add reference</td>
<td>+ add value</td>
</tr>
<tr>
<td>Role</td>
<td>Entity</td>
<td></td>
</tr>
<tr>
<td>-------------------------------</td>
<td>---------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>affiliated entity</td>
<td>Seattle Civil Rights Congress</td>
<td></td>
</tr>
<tr>
<td>parent company</td>
<td>International Longshore and Warehouse Union</td>
<td></td>
</tr>
<tr>
<td>has member</td>
<td>Constantine Baruso</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Jimmy Bulosan Ramil</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Richard Gurtiza</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Gene Viernes</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pompeyo Benito Guioy</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Silme Domingo</td>
<td></td>
</tr>
<tr>
<td>occupant of</td>
<td>Old Cannery Building</td>
<td></td>
</tr>
<tr>
<td>creator of collection</td>
<td>Cannery Workers and Farm Laborers Union, Local 7 Records</td>
<td></td>
</tr>
<tr>
<td>main subject of</td>
<td>Michael Withey Papers</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Trinidad Rojo Papers</td>
<td></td>
</tr>
<tr>
<td></td>
<td>C. T. (Barry) Hatten Papers</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Carlos Bulosan Papers</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Chris D. Mensalvas Papers</td>
<td></td>
</tr>
</tbody>
</table>
MARC to Wikidata Workflow for Archival Collections

Adapted from PCC Wikidata Pilot workflow for batches of ETD’s


GitHub Repository: https://github.com/uwlib-cams/EAD2MARC2Wikidata
OpenRefine Cleanup and Reconciliation

- Start with MARC 21 records in OCLC local bibliographic save file
- Open in MARCEdit and convert to a tab-delimited file
- Open as a new project in OpenRefine
- Clean up and reconcile data in OpenRefine
# OpenRefine Cleanup and Reconciliation

<table>
<thead>
<tr>
<th>MARC21 Field</th>
<th>OpenRefine Column(s)</th>
<th>Wikidata Property/ies</th>
<th>Wikidata Qualifier(s)</th>
<th>Instructions</th>
</tr>
</thead>
<tbody>
<tr>
<td>001</td>
<td>OCLC ID</td>
<td>OCLC control number (P243)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>008</td>
<td>language</td>
<td>language of work or name (P407)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>046</td>
<td>start Date, end Date</td>
<td>start of covered period (P7103), end of covered period (P7104)</td>
<td>point in time (P585)</td>
<td>Combine with date information from the 245 field, normalize, and separate into earliest and latest dates. Automation would be helpful here.</td>
</tr>
<tr>
<td>7XX</td>
<td>creator</td>
<td>collection creator (P6241)</td>
<td></td>
<td>facet 7XX with custom text facet for &quot;creator&quot;. Clear everything that doesn't match. For personal names, split columns by &quot;&quot;, and rejoin in direct order. Clear relators and remove subfield values, dates, and qualifiers. Reconcile against Wikidata. Use a simplified schema to create new items for agents that do not have matches.</td>
</tr>
<tr>
<td>245 S$a</td>
<td>title, Label</td>
<td>title (P1476), Label</td>
<td>language of work or name (P407)</td>
<td>Change to title case. Duplicate title column values to make another column for “label”. Reconcile labels against Wikidata, selecting “create new item for each cell” under reconciliation actions if none match existing items.</td>
</tr>
<tr>
<td>246</td>
<td>variantTitle</td>
<td>Also known as</td>
<td></td>
<td>Change to title case</td>
</tr>
<tr>
<td>300 S$a</td>
<td>extent</td>
<td>collection or exhibition size (P1436)</td>
<td>unit sourcing circumstance (P1480) point in time (P585) quantity (P1114)</td>
<td>It can be time consuming to break out different units and sourcing circumstances. Recommendation: remove extents that are not cubic feet, then normalize cubic feet quantities and use those.</td>
</tr>
<tr>
<td>6XX</td>
<td>mainSubject</td>
<td>main subject (P921)</td>
<td></td>
<td>Reconcile against Wikidata. Either create new items or remove values that do not match to Wikidata items.</td>
</tr>
<tr>
<td>856 S$a</td>
<td>Archives West URI</td>
<td>Archives West finding aid ID (P9335)</td>
<td></td>
<td>use substring function to isolate identifiers from the base URIs</td>
</tr>
<tr>
<td>[Alma Holdings] collectionNumber</td>
<td>inventory number (P217)</td>
<td>collection (P195)</td>
<td></td>
<td>separate into multiple columns by separator if there are more than one inventory number attached to each collection (in the case of multiple accessions)</td>
</tr>
<tr>
<td>[Generate] instanceOf</td>
<td>instance of (P31)</td>
<td></td>
<td>&quot;archival collection&quot; (Q29388534)</td>
<td></td>
</tr>
<tr>
<td>[Generate] archive</td>
<td>collection (P195), part of (P381), maintained by (P126), location (P276)</td>
<td>for collection: inventory number (P217)</td>
<td>The repository for an archival collection can be expressed through all of these Wikidata properties. For ideal search results no matter how users look for this information, we decided to use all four. Reconcile against Wikidata and create new items for any that do not match existing items.</td>
<td></td>
</tr>
<tr>
<td>[Generate] country</td>
<td>country (P17)</td>
<td></td>
<td>Reconcile against Wikidata.</td>
<td></td>
</tr>
<tr>
<td>[Generate] locatedIn</td>
<td>located in the administrative territorial entity (P131)</td>
<td></td>
<td>Reconcile against Wikidata.</td>
<td></td>
</tr>
</tbody>
</table>
OpenRefine Wikidata Schema

Using facets and filters

Use facets and filters to select subsets of your data to act on. Choose facet and filter methods from the menus at the top of each data column.

Not sure how to get started? Watch these screencasts
OpenRefine Wikidata Extension

- Preview changes
- Use Wikidata extension to “Upload edits to Wikibase”
- Once items were created, we created a new column from the reconciled Label column with all the new Qids to facilitate links back to Wikidata from the finding aids and MARC records
**Wikidata Links**

**MARC 21**
- Use MARCEdit’s OCLC API integration to open MARC 21 records
- Convert to MARC XML
- Take records out of MARC 21 and process using XSLT written by Theo Gerontakos, along with table with Wikidata Qids
- Convert output to .mrc
- Copy and paste output into MARCEdit, replacing initially imported records
- Push changes via OCLC API

<table>
<thead>
<tr>
<th>758</th>
<th>i Wikidata item: a Charles Al Miller Papers 2 wikidata 4 <a href="http://www.wikidata.org/entity/P2888">http://www.wikidata.org/entity/P2888</a> 0 (wikidata)Q121301612</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1 <a href="http://www.wikidata.org/entity/Q121301612">http://www.wikidata.org/entity/Q121301612</a></td>
</tr>
</tbody>
</table>

**EAD**
- Use Python script to insert Wikidata links into copies of finding aids created by Special Collections staff
Challenges

**MARC 21 Data**
- Collections had fallen through the cracks over years of changes to semi-automated cataloging workflows
- Changes to finding aids (as with new accessions) had not carried over to MARC records
- No inventory numbers
- Missing or incorrect Archives West URIs

**EAD Finding Aid Data**
- Changes in MARC descriptions (correcting typos, for instance) had not carried over to finding aids

**Wikidata & OpenRefine:**
- Removing or changing statements in bulk when application profile decisions change
- OpenRefine reconciliation can be a slow business
- Data cleaning processes in OpenRefine could use further automation
Future Work

- MARC 21 pre-processing guidance and tools
  - Putting dates in 046 fields
  - Including inventory numbers in a consistent place
  - Repairing links
- Further integration of MARC 21 and archival finding aid workflows
- Add items for LAW oral histories from Internet Archive
- Extend pilot to larger archival collections at UW
- Automate selected tasks as possible
Thank You!

Questions?

Contact:
Crystal Yragui, cec23@uw.edu
Adam L. Schiff, aschiff@uw.edu
Resources

EAD: Encoded Archival Description

EAD2MARC2Wikidata GitHub Repository

MARC Standards

SNAC

University of Washington EAD to MARC conversion stylesheets (Mark Carlson)

WikiProject University of Washington Archival Metadata

WikiProject PCC Wikidata Pilot University of Washington

SCArchivesAgents GitHub Repository