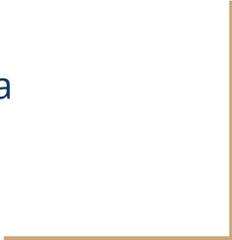




'archives at':

An Opportunity to Leverage MARC to Create
Linked Open Data

Daniela Roviida & Jennifer Brcka
March 11, 2022



UNIVERSITY OF
NOTRE DAME

HESBURGH LIBRARIES

Introduction

- Our work began with local participation in [PCC Wikidata Pilot](#)
- Created a sub-group to explore linked data for archives and rare materials
- Focus on adding “archives at” (P485) statements to Wikidata entries



Background on project

- Modeled and created entries for our repositories
- Chose 12 archival collection test cases
- Among these test cases we saw an opportunity to look at linked data regarding a single creator's dispersed archive
- Case study: Ed Dorn Papers

Wikidata items

Label

'Q' identifier

Property, represented by 'P' value
P31

Value, another 'Q' value, Q5

Ed Dorn (Q5334756) [edit](#)

American writer

Edward Merton Dorn

↳ Reconcil: Most relevant properties which are absent

↳ In more languages

Configure

Language	Label	Description	Also known as
English	Ed Dorn	American writer	Edward Merton Dorn
Spanish	Ed Dorn	escritor estadounidense	
Traditional Chinese	No label defined	No description defined	
Chinese	No label defined	No description defined	

All entered languages

Statements

instance of	human Q5	edit
	↳ 2 references	
+ add value		
sex or gender	male Q5	edit
	↳ 0 references	
+ add reference		
+ add value		

Motivation: Discovery for diasporic archives

Discovery System	Pros	Cons
Archives' Online Finding Aids	Thorough description of materials' scope, content, organization, location, and rules for use & access, typically include inventories	Various formats: EAD, HTML, PDF. Rely on annotation of related or separated materials to discover dispersed content
WorldCat	Allows for discovery of content aggregated from multiple repositories, robust search indexing	Only includes items or collections from member institutions, and only finding aids encoded in MARC
ArchiveGrid	Allows for broad discovery of content aggregated from multiple repositories in EAD, MARC, HTML, and PDF. 'Summary View' allows for advanced filtering	Only includes items or collections from member or contributing institutions. Indexing, authority control not uniform among record types

Motivation, search results, agents & topics

ARCHIVEGRID title:"Edward Dorn" OR title:"Ed Dorn" Search

44 Edward Dorn" OR title

LIST VIEW SUMMARY VIEW

PEOPLE

Name ↓↑	Count ↓↑
Dillof, Dobro Dick	1
Dorn, Edward	41
Dorn, Edward Archives	2
Dorn, Edward Correspondence	2
Duncan, Robert, 1919-1988	2

GROUP

Name ↓↑	Count ↓↑
Mountain, N.C.	1
Dept. of Special Collections & University Archives	1
Edward Dorn fonds	1
University of California, Berkeley Language-Behavior Research Laboratory	1

PLACES

Name ↓↑	Count ↓↑
Black Mountain (N.C.)	1
North Carolina	1

ARCHIVES

Name ↓↑	Count ↓↑
University of California, San Diego	24
Harvard University - Houghton Library	5
Stanford University	3
University of California, Santa Cruz	3
University of Delaware - Hugh M.	2

ARCHIVE LOCATIONS

Name ↓↑	Count ↓↑
United States - California - La Jolla	24
United States - Massachusetts - Cambridge	5
United States - California - Stanford	4
United States - California - Santa Cruz	3

TOPICS

Name ↓↑	Count ↓↑
American poetry - 20th century	1
American poetry -- 20th century	2
Authors, American	1
Black Mountain school (Group of poets)	1

ARCHIVEGRID "Beats (Persons)" Search

277 "Beats (Persons)"

LIST VIEW SUMMARY VIEW

PEOPLE

Name ↓↑	Count ↓↑
Dorfman, Elsa	2
Dorn, Edward	4
Dorn, Edward Portraits	1
Dorn, Edward, 1838-1923 Correspondence	1
Dostoyevsky, Fyodor, 1821-1881	2

GROUP

Name ↓↑	Count ↓↑
Online Archive of California	11
City Lights Books	10
Naropa Institute	8
Jack Kerouac School of Disembodied Poetics	7
Los Angeles Valley College	5

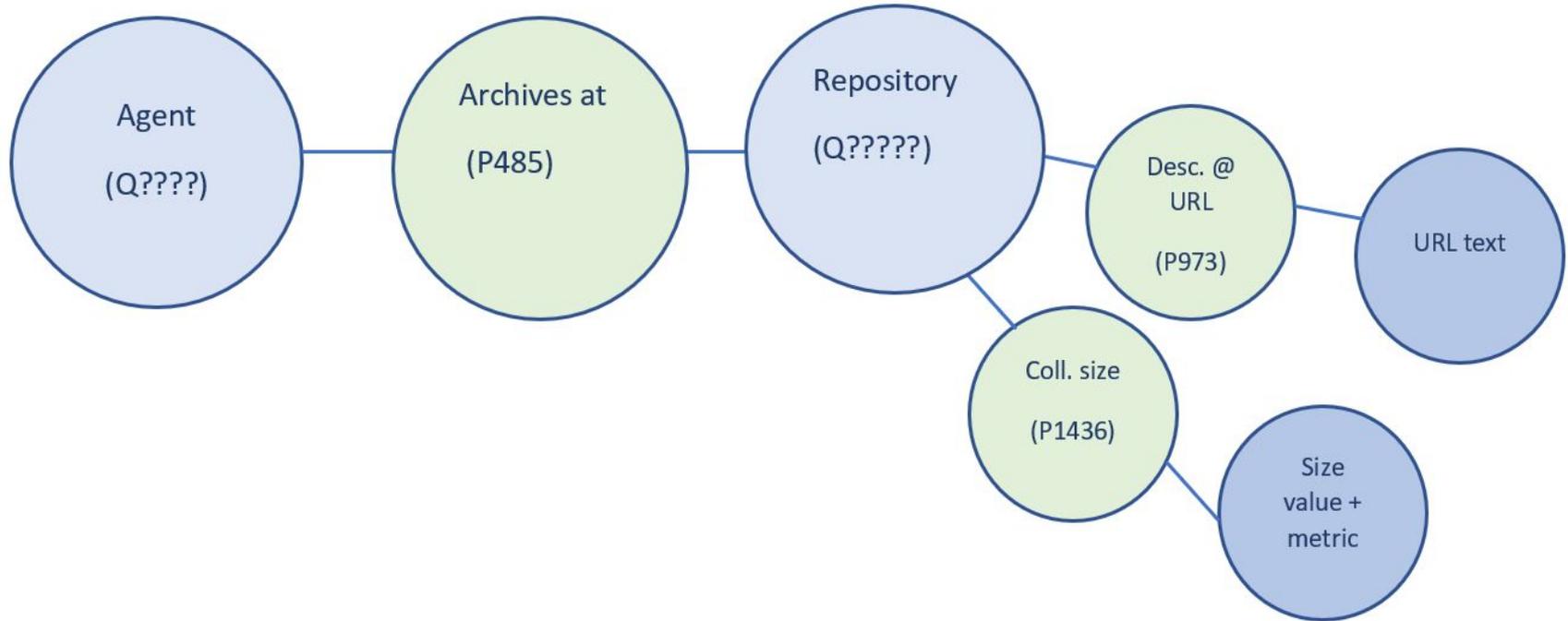
ARCHIVES

Name ↓↑	Count ↓↑
University of California, Berkeley	63
New York Public Library	20
Columbia University	17

ARCHIVE LOCATIONS

Name ↓↑	Count ↓↑
United States - California - Berkeley	63
United States - New York - New York	41
United States - California - Los	23

'archives at' conceptual model



Process - Overview

- Download MARC XML versions of EAD finding aids from ArchivesSpace
- Convert to MARCXML to MARC and join records into one file (MarcEdit)
- Export MARC record elements to tab delimited text file (MarcEdit)
- OpenRefine - Wikidata reconciliation for forms of names
- Build QuickStatements & contribute data to Wikidata

MARC data retrieved

Using MarcEdit's "Export Tab-Delimited Records" tool, 1XX agents were parsed by subfield.

100\$a	100\$b	100\$c	100\$q	100\$d	100\$e	110\$a	110\$b
Ginzburg, Evgeniia.							
Brown, Mike,					creator.		
McCabe, Pat,				1955-	creator.		
Dorn, Edward.							
Roskina, Nataliai.							
Creeley, Robert,				1926-2005.			
Kompaneyets, Katya,					creator.		
Chesterton, G. K.			(Gilbert Keith),	1874-1936.			
Gomberg-Verzhbinskaia, Eleonora Petrovna.							
Fleischer, Nat,				1887-1972,	creator.		
Wetmore, William S.			(William Shepard),	1801-1862.			
Tunney, Gene,				1897-1978,	creator.		
Magalotti, Lorenzo,		Cardinal,		1584-1637,	creator.		
Nanda, Edward				1800-1882	creator		

MARC data retrieved

The following fields were also included:

300: Extent (extents given in cubic feet manually converted to linear feet)

506: Access note (Is content available to users?)

520: Scope and content (Was content generated by agent)

856: Electronic access (Location of the finding aid)

506	520	856
	Correspondence concerning the Vietnam War, Lucey's trip to South Vietnam as an observer of the 1967 election, his friend's trip to South Vietnam, and his trip to South Vietnam as an observer of the 1967 election, his friend's trip to South Vietnam, and his trip to South Vietnam as an observer of the 1967 election.	http://archives.nd.edu/findaids/ead/xml/luc.xml
	Five reels of letters from the United States to the Ludwig Mission Society and two reels of the society's magazine, <i>Annales de la Mission de Saint-Louis</i> .	http://archives.nd.edu/findaids/ead/xml/lud.xml
	General office files, director's files, and journal files.	http://archives.nd.edu/findaids/ead/xml/lwc.xml
Restricted by policy.	Annual reports of the Law Librarian, other reports and general correspondence, letters sent by the Law Librarian, budgets, and correspondence.	http://archives.nd.edu/findaids/ead/xml/LWL.xml
Contractual restrictions may apply.	Correspondence and papers of Marcus Lyon; notes on protozoology, 1916; manuscripts of Marcus and Martha Lyon; medical notes and papers.	http://archives.nd.edu/findaids/ead/xml/lyf.xml
Restricted by contract.	Class notes and papers written at University of Notre Dame, University of Pittsburgh, and Bowling Green State University.	http://archives.nd.edu/findaids/ead/xml/lyo.xml
	Clippings of literary events and reviews of plays at Notre Dame, 1858-1885; manuscripts of works by Lyons; ribbons, awards, and other items.	http://archives.nd.edu/findaids/ead/xml/lys.xml
Restricted by policy.	Ernst and Young YMP Program MS in Accountancy student files, classes of 2000-2005; also Orientation Files, Survey Files, and other files.	http://archives.nd.edu/findaids/ead/xml/mac.xml
Contractual restrictions may apply.	Minutes, reports, and correspondence for the Notre Dame history department, 1958-1966, and the Catholic Commission on the History of the United States.	http://archives.nd.edu/findaids/ead/xml/maf.xml
	Correspondence, 1925-1957; writings, speech notes, and speeches by Mahony, 1920s-1950; diary of Mahony, 1928; Mahony's papers.	http://archives.nd.edu/findaids/ead/xml/mah.xml
	Correspondence (1930s-1983) with Catherine de Hueck Doherty, Paul Hanley Furfey, Reynold Hillenbrand, Bernard J. Sheehy, and others.	http://archives.nd.edu/findaids/ead/xml/mak.xml
	Video tapes concerning population and environment, cultural pluralism, ethnic identity, the melting pot myth, and William F. Buckley, Jr.	http://archives.nd.edu/findaids/ead/xml/man.xml
	Correspondence with Lewis Gray, Secretary of War, Alexis Grawford, Auditor of South Bend, Edward Savin, CSC, and others.	http://archives.nd.edu/findaids/ead/xml/mas.xml

OpenRefine

Names were reconciled against Wikidata via *Column 1* > *Reconcile* > *Start reconciling*.

Added a column based on column 1, populated with corresponding Wikidata Q numbers (GREL: cell.recon.match.id)

OpenRefine names test 2 txt [Permalink](#)

Facet / Filter <

Undo / Redo 15 / 15

15 rows

Show as: **rows** records Show: 5 10 25 50 100 500 1000 rows

<input type="checkbox"/>	<input type="checkbox"/>	1.	All-American Girls Professional Baseball League All-Star Team Choose new match	Q4727791
<input type="checkbox"/>	<input type="checkbox"/>	2.	Ed Dorn Choose new match	Q5334756
<input type="checkbox"/>	<input type="checkbox"/>	3.	Cristina Peri Rossi Choose new match	Q290502
<input type="checkbox"/>	<input type="checkbox"/>	4.	Jack Pfefer Choose new match	Q16007691
<input type="checkbox"/>	<input type="checkbox"/>	5.	Éilís Ní Dhuibhne Choose new match	Q8078168
<input type="checkbox"/>	<input type="checkbox"/>	6.	Stanley Weston Choose new match	Q55287093

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](#)

QuickStatements

'Archives at' statements were added to Wikidata in a batch.

Data in green cells was pasted into the QuickStatements tool at:

<https://quickstatements.toolforge.org/#/batch>

QUICKSTATEMENTS MODEL TO ADD ARCHIVES@ PROPERTIES							linear feet = .844			
(Load green cells)										
Name	Q number for entity with archives	archives at	Archive: RBSC	described at URL	Finding Aid URL (place in quotes)	collection or exhibition size	number plus measurement constraint unit, (linear feet) Q changed to U	number	constraint units, linear feet	
Ed Dorn	Q5334756	P485	Q108653830	P973	"https://archivesspace.library.nd.edu/repositor	P1436	22.5U97958628	22.5	U97958628	
All-American Girls Profession:	Q1052634	P485	Q108653830	P973	"https://archivesspace.library.nd.edu/repositor	P1436	7.5U97958628	7.5	U97958628	
Peri Rossi, Cristina, 1941-	Q290502	P485	Q108653830	P973	"https://archivesspace.library.nd.edu/repositor	P1436	5U97958628	5	U97958628	
Pfefer, Jack (Jacob), 1894-197	Q16007691	P485	Q108653830	P973	"https://archivesspace.library.nd.edu/repositor	P1436	104U97958628	104	U97958628	
Ní Dhuibhne, Éilís, 1954-	Q8078168	P485	Q108653830	P973	"https://archivesspace.library.nd.edu/repositor	P1436	20U97958628	20	U97958628	
Weston, Stanley	Q55287093	P485	Q108653830	P973	"https://archivesspace.library.nd.edu/repositor	P1436	27U97958628	27	U97958628	

Results

Overview: Query for both repositories: <https://w.wiki/4uYk>

Case Study: Allows user to readily locate individual elements of a dispersed archive, and identify where material is concentrated: <https://www.wikidata.org/wiki/Q5334756>

Semantic structure allows uniform topical discovery regardless of subject analysis present within finding aids: <https://w.wiki/4wSn>

Data and relationships created are impactful, lowered threshold to participate

Can serve as powerful complement to discovery systems already in place

Acknowledgements

Thanks to our colleagues for your project contributions and support:

Peggy Griesinger, Rachel Boyd, Lisa Stienbarger, Eric Lease Morgan

Sources:

Aberdeen Tower Block Archives. "How to upload a spreadsheet of data to Wikidata using Quickstatements".

https://www.ed.ac.uk/files/atoms/files/tower_block_archive_documentation.pdf

Bolliger, S., Brüderlin, B., Gasser, M., Lyskawa, J., Maier, P., Schmitt, L. (2020). How to Link Your Institution's Collections to Wikidata? : a short manual to a semi-automatic way of using the "archives at" property (P485). <https://doi.org/10.3929/ethz-b-000393724>

Wikidata: WikiProject University of Toronto Libraries: [Data Model for Libraries and Archives](#)

Questions?

Thanks for your attention and please feel free to contact:

Daniela Roviada: drovida@nd.edu

Jennifer Brcka: jbrcka@nd.edu

If you have any questions about our project.

Our Wikidata project description:

https://www.wikidata.org/wiki/Wikidata:WikiProject_PCC_Wikidata_Pilot/University_Notre_Dame#Discover_Archives_Project