Gateway into Linked Data: Breaking Silos with Wikidata

PCC Wikidata Pilot Project @ Texas State University

Mary Aycock, Database and Metadata Management Librarian
Nicole Critchley, Assistant Archivist, University Archives
Amanda Scott, Wittliff Collections Cataloging Assistant

https://digital.library.txstate.edu/handle/10877/13529

ALKEK LIBRARY

It started about 10 years ago...

- Linked data started exploding as a topic at library conferences
- Data that was also machine-readable
- Excitement was in the air about the possibility of breaking our databases (catalog, repositories) out of their library-based silos

TEXAS★STATE
UNIVERSITY LIBRARIES

If you learn it, will they build it?

- Libraries bought ILS vendors' (often pricey) packages offering to convert MARC records to linked data, but results seemed underwhelming
- Many of us channeled our motivation into learning linked data concepts: RDF, Triples, SPARQL
- However, technical infrastructure and abundance of human resources not available to "use it" in most libraries

TEXAS ★ STATE
UNIVERSITY LIBRARIES

Aimed at the Perplexed Librarian

"Yet despite this activity, linked data has become something of a punchline in the GLAM community. For some, linked data is one of this era's hottest technology buzzwords; but others see it as vaporware, a much-hyped project that will ultimately never come to fruition. The former may be true, but the latter certainly is not."

Carlson, S., Lampert, C., Melvin, D., & Washington, A. (2020). Linked Data for the Perplexed Librarian. ALA Editions.

TEXAS★STATE
UNIVERSITY LIBRARIES ®

And then a few years ago, the word became…

What does it have to offer?

- Stability: Created in 2012 as part of the Wikimedia Foundation
- 2014: Google retired Freebase in favor of Wikidata
- Ready-made infrastructure (Wikibase) & ontology
- Shared values of knowledge creation
- Easy to use graphical interface
- Hub of identifiers

If they build it, you can learn...

- Is not a matter of building something and hoping the search engines will find it: Wikidata IS used by Google and other agents (Siri, etc.)
- You can create infoboxes on Wikipedia that pulls from Wikidata

Wikimedia Commons has media related to *Ursula Franklin*.

Archives at

Location	University of Toronto Archives & Records Management Services
Identifiers	B1996-0004, B2015-0005 ✏
Source	Ursula Franklin fonds ⊠

How to use archival material

TEXAS ★ STATE ®
UNIVERSITY LIBRARIES

Not that there aren't challenges!

- Still a learning curve, particularly the deeper you dive
- Difficulty in measuring impact
- Cultural differences between two communities
  - Records/items can be proposed for deletion
  - Anyone can edit (including bots)...

Program for Cooperative Cataloging Wikidata Pilot

- ## Goals of PCC
  - ○ Comparing ease of use and benefits of Wikidata to other registries (LCNAF, ISNI)
  - ○ Assessing the productivity and quality assurance tools that exist (or should exist)
  - ○ Learning about the culture of the Wikidata community
- ## Benefits to our library
  - ○ Gain familiarity of working within linked data platform (Wikidata)
  - ○ Apply linked data knowledge gained over the past years
  - ○ Bring more visibility to affiliated persons, organizations, collections, and projects within our university

Initial Setup

- Recruited people to work on new project (or people proposed new projects!)
- Gave introductory presentation about Wikidata
  - Also demonstrated the creation of a Wikidata record during presentation
- Setup WikiProjects page: https://www.wikidata.org/wiki/Wikidata:WikiProject_PCC_Wikidata_Pilot/Texas_State_University_Libraries

TEXAS ★ STATE
UNIVERSITY LIBRARIES

Four projects at Texas State University

- Wikidata pages for current faculty
- University Archives Oral History Collections
- Wittliff Archives Collections
- Digital Collections Journal Project

TEXAS★STATE ®
UNIVERSITY LIBRARIES

Faculty authority work

- Moving much of traditional catalog authority work from a MARC silo to Wikidata
- Creating/enhancing Wikidata pages for faculty at Texas State University
- Embracing identity management principles
  - Empowering communities to create own items
  - Moving toward stable identifiers (rather than relying on unique headings)

TEXAS ★ STATE
UNIVERSITY LIBRARIES

Reassessing effort put into siloed records
Experimentation with creating basic MARC authority records (NACO-lite)

Add Personal Name Attributes ✕

- Birth Date and Death Date (046)
- Other Attributes of Person (368)
- Place of Birth and Place of Death (370)
- Associated Country (370)
- Place of Residence/Headquarters (370)
- Other Associated Place (370)
- Electronic Mail Address (371)
- Field of Activity (372)
- Associated Group (373)
- Occupation (374)
- Gender (375)
- Associated Language (377)
- Fuller Form of Personal Name (378)

010			no2021005485
040			TxSmTSU ‡b eng ‡e rda ‡c TxSmTSU
024	7		Q56892582 ‡2 wikidata ‡1 https://www.wikidata.org/entity/Q56892582
100	1		Castro-Arellano, Ivan
373			Texas State University ‡2 naf
400	1		Arellano, Ivan Castro-
667			Record created as part of PCC Wikidata Pilot. Please do not edit or remove the Wikidata information.
670			Castro-Arellano, Ivan. Ecological patterns of the small mammal communities at El Cielo Biosphere Reserve, Tamaulipas, Mexico, 2007: ‡b title pa...
670			Texas State University Department of Biology faculty directory webpages, viewed January 16, 2021 ‡b (Ivan Castro-Arellano, Ph.D., Associate Pr... (Zoonotic diseases, Hantavirus ecology), Community Ecology (Temporal Niche Overlap, Metacommunity Theory), Mammal Ecology, Invasive S... ecologist to address questions to understand the ecology of zoonotic diseases and invasive species; my goal is to integrate the study of commu... understand the dynamics of pathogen transmission) ‡u https://www.bio.txstate.edu/about/Faculty---Staff/faculty/Ivan-Castro-Arellano.html

TEXAS ★ STATE
UNIVERSITY LIBRARIES ®

Determined data model

Models for Current Faculty [edit]

- These tables cover faculty who currently work at Texas State University
- Include references for every statement unless otherwise specified
- Examples: Todd M. Ahlman (Q99209876), Melinda M Villagran (Q56929652)
- The tables below are modified from Stanford Libraries' Excellent Wikiproject Wikidata:WikiProject_Stanford_Libraries/Data_models#Academic_faculty

Basic statements   [edit]

Property ⬍	Value ⬍	Usage note ⬍
Label	Person's name as given on university website	Do not include titles or degrees
Alias	Other form of name in use or previously used	
Description	based on field of work	See description guidelines (concise, lower case)
instance of (P31)	human (Q5)	no reference needed
occupation (P106)	academic (Q3400985)	
occupation (P106)	university teacher (Q1622272)	also include appropriate roles, such as researcher (Q1650915), professor (Q121594), dean (Q723682), director (Q1162163)
occupation (P106)	based on field of work	for example, anthropologist (Q4773904), musician (Q639669)
field of work (P101)	academic discipline	usually correct value appears with lower capitalization in Wikidata search result

Preparatory steps for Faculty Authors

- Created spreadsheet with faculty that already:
    - Had authority records and/or
    - Had Wikidata records
- Setup structure and Wikidata items for all campus departments so it would be available for various projects through batch tools

Doing our part to increase visibility

Researchers outline lessons to be learned from ongoing COVID-19 pandemic

RESEARCH & INNOVATION   Jayme Blaschke | January 5, 2021

The severity of the ongoing worldwide pandemic has prompted international researchers to warn that COVID-19 could be the first of many novel diseases infecting the human population and call for nations to shift from a reactionary strategy to one of prevention before the inevitable emergence of other zoonotic diseases.

The essay was written in part by Ivan Castro-Arellano, an associate professor in the Department of Biology at Texas State University. Their findings, "Tackling zoonoses in a crowded world: Lessons to be learned from the COVID-19 pandemic," are published in the journal *Acta Tropica* (www.sciencedirect.com/science/article/pii/S0001706X20316934).

Zoonosis describes diseases that can be transmitted to humans from animals.

Dr. Ivan Castro-Arellano

Item   Discussion

Ivan Castro-Arellano (Q56892582)

researcher

▸ Recoin: Most relevant properties which are absent

▸ In more languages

Statements

instance of	human
	▾ 0 references

TEXAS ★ STATE
UNIVERSITY LIBRARIES

SpaceX launch carries Texas State's experiment on biofilm formation to International Space Station

RESEARCH & INNOVATION   Jayme Blaschke | December 4, 2020

Robert J. McLean (Q21520276)

microbiologist

R.J.McLean | Robert McLean | Robert J.C. McLean | Robert Mclean | Robert McLe

▸ Recoin: Most relevant properties which are absent

▸ In more languages

Statements

instance of	human
	▾ 0 references

TEXAS ★ STATE

UNIVERSITY LIBRARIES

Progress in faculty pages

- Met goal of 100 items in April: continuing to add more.
- Credit also to music cataloger who jumped in and created quite a few entries, both manually and through batch editing tools
- Next step: Investigating semi-automatic batch processes?

University Archives Oral History Project

- Why oral histories?
    - Demographic data in several places, spreadsheets (not publicly available), website, or PDF finding aid (not easily searched)
    - Creating linked data to point people to these interviews on our website, helps us connect to our collections and finding aids
- Idea of Notability
    - https://www.wikidata.org/wiki/Wikidata:Notability
    - Has to meet one of three criteria:
        - 2. It refers to an instance of a **clearly identifiable conceptual or material entity**. The entity must be notable, in the sense that it **can be described using serious and publicly available references**.

has works in the collection (P6379)

collection that has works of this person or organisation (use archive location P485 for the archives)

works in collection | has works in the collection(s) | work in collection | has work in the collection | museum housing this person's work

- Using *has works in collection* (P6379) for connection to oral history collection, but added *archives at* (P485), if had full finding aid at Txstate or another institution.
- Ruth Bain example:

archives at	Austin History Center	edit
	▾ 1 reference	
	reference URL https://legacy.lib.utexas.edu/taro/aushc/00136/00136-P.html	
		+ add reference
		+ add value

has works in the collection	Texas 150: Sesquicentennial Oral History Project	edit
	▾ 1 reference	
	reference URL https://www.univarchives.txstate.edu/research-collections/core-collections/070-media-artifacts/media-oral-history/oral-history-name-index/bain-ruth.html	
		+ add reference
		+ add value

TEXAS ★ STATE
UNIVERSITY LIBRARIES

How it started

- Surveyed to see what was already in Wikipedia/data, and other authorities. Also collected links.

Wikipedia page?	Wikidata page?	LCNAF?	UA webpage	Dspace link	SNAC entry	Related collections?
3 (plus one mentioned in related article)	3	16	70	69	14	6

- 3 out of 70 already had Wikidata entries, concentrating on new entries.
- Looked at Stanford Libraries' project to get ideas for properties
  - See Wikidata:WikiProject PCC Wikidata Pilot/Texas State University Libraries for properties included.

TEXAS ★ STATE
UNIVERSITY LIBRARIES ®

The work continues...

- Began to compile data on our participants, starting with our master oral history spreadsheet
  - Biographical sheets interviewees filled out, the interview itself, FindaGrave, obituaries, other institutions finding aids, LC authority records, SNAC records...
- Minimums vs how far down the rabbit hole do you want to go?
  - Example: Piper Professor Award
    - To create one record, had to create two more:
      - Minnie Stevens Piper Foundation
      - Minnie Stevens Piper
    - What information is actually useful in creating links?

Bulk entry tools

- Used Quickstatements and OpenRefine to create new entries in Wikidata
  - Manually did edits to existing records even though these bulk entries tools do allow for editing and adding to existing entries
- Quickstatements needed property and Q identifiers to work, time consuming with the varying data I had, was more useful with entries like university departments which only varied a little
- OpenRefine allowed me to use my data collection spreadsheet and 'reconcile' against existing properties

See Resources slide for guides to using bulk entries.

TEXAS ★ STATE
UNIVERSITY LIBRARIES ®

Facet / Filter | Undo / Redo 249 / 249

52 rows | Schema | Issues 230 | Preview

Extensions: Wikidata ▾

Show as: **rows** records Show: 5 **10** 25 50 rows

« first ‹ previous **1 - 10** next › last »

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

All		Name (Label)	Label	Description	Also known as (.	birth name	instance of	sex or gender	country of	given name	family name	date of birth	place of birth	reconciledPlace	date of death	place of deat
												Gustave Xavier Furtado (7)				
												☑ ☑ Vittore Benedetto Antonio Trevisan de Saint-Léon (6)				
												☑ ☑ Create new item				
												Search for match				
☆ ⬜	5.	Hardeman, Sammie Choose new match	Sammie Hardeman	oral history project participant	Sammie L.C. Hardeman		human Choose new match	male Choose new match	United States of America Choose new match	Sammie Choose new match	Hardeman Choose new match	Tue Jul 06 00:00:00 CST 1915 ☑ ☑ Create new item Search for match	Lockhart, Texas	Lockhart Choose new match	Fri Oct 31 00:00:00 CST 2003	
☆ ⬜	6.	Hardesty, Walton D. "Pete" Choose new match	Walton D. "Pete" Hardesty	oral history project participant	Walton D. Hardesty; Walton Duane Hardesty;Pete Hardesty		human Choose new match	male Choose new match	United States of America Choose new match	Walton Choose new match	Hardesty Choose new match	Tue Jul 26 00:00:00 CDT 1921 ☑ ☑ Create new item Search for match	Wichita Falls, Texas	Wichita Falls Choose new match	Thu Jun 27 00:00:00 CDT 2002	
☆ ⬜	7.	Harper, Bruce Choose new match	Bruce Harper	oral history project participant	Bruce F. Harper		human Choose new match	male Choose new match	United States of America Choose new match	Bruce Choose new match	Harper Choose new match	Fri Mar 20 00:00:00 CDT 1942 ☑ ☑ Create new item Search for match	San Marcos, Texas	San Marcos Choose new match	Wed Jul 15 00:00:00 CDT 2020	
☆ ⬜	8.	Haynes, Mary Sue Choose new match	Mary Sue Haynes	oral history project participant			human Choose new match	female Choose new match	United States of America Choose new match	Mary Choose new match	Haynes Choose new match	1898/1900-12-09 ☑ ☑ Create new item	Longview, Texas	Longview Choose new match	1997	
☆ ⬜	9.	Henderson, Richard B. Choose new match	Richard B. Henderson	oral history project participant	Richard Bevier Henderson		human Choose new match	male Choose new match	United States of America Choose new match	Richard Choose new match	Henderson Choose new match	Sun Nov 27 00:00:00 CST 1921 ☑ ☑ Create new item Search for match	Washington, D.C.	Washington, D.C. Choose new match	Wed Nov 19 00:00:00 CST 2003	
☆ ⬜	10.	Holmes, Samuel Choose new match	Samuel Holmes	oral history project participant			human Choose new match	male Choose new match	United States of America Choose new match	Samuel Choose new match	Holmes Choose new match	1924 ☑ ☑ Vladimir Lenin (17) ☑ ☑ George H. W. Bush (17) ☑ ☑ Georges Rouy (17) ☑ ☑ Oriol de Bolòs (15) ☑ ☑ Giacomo Puccini (14) ☑ ☑ Andreas Voss (13) ☑ ☑ Meinhard Michael Moser (13) ☑ ☑ Johannes Eugenius Bülow	Austin, Texas	Austin Choose new match	2011	

Kent Finlay (Q105925041)

founder of Cheatham Street Warehouse and music producer; Texas Music Legends Hall of Fame member ✎ edit

James Kent Finlay

▼ In more languages
Configure

Language	Label	Description	Also known as
English	Kent Finlay	founder of Cheatham Street Warehouse and music producer; Texas Music Legends Hall of Fame member	James Kent Finlay
Spanish	No label defined	No description defined	
Traditional Chinese	No label defined	No description defined	
Chinese	No label defined	No description defined	

All entered languages

Statements

instance of	⬦ human	✎ edit
	▼ 0 references	
		+ add reference
		+ add value

sex or gender	⬦ male	✎ edit
	▼ 0 references	
		+ add reference
		+ add value

country of citizenship	⬦ United States of America	✎ edit

Kent Finlay did not have an entry, even though mentioned in the Wikipedia article for the Cheatham Street Warehouse and having an entry in the Handbook of Texas.

TEXAS ★ STATE
UNIVERSITY LIBRARIES

Robert Hardesty

Bob Hardesty (Q15429756)

American educator and speechwriter for President Lyndon B. Johnson   ✎ edit

▾ In more languages
Configure

Language	Label	Description	Also known as
English	Bob Hardesty	American educator and speechwriter for President Lyndon B. Johnson	
Spanish	Bob Hardesty	No description defined	
Traditional Chinese	No label defined	No description defined	
Chinese	No label defined	No description defined	

All entered languages

Robert L. Hardesty (Q15429756)

American educator and speechwriter for President Lyndon B. Johnson; university president   ✎ edit
Bob Hardesty | Robert Lewis Hardesty | Robert Hardesty

▾ In more languages
Configure

Language	Label	Description	Also known as
English	Robert L. Hardesty	American educator and speechwriter for President Lyndon B. Johnson; university president	Bob Hardesty Robert Lewis Hardesty Robert Hardesty
Spanish	Robert L. Hardesty	No description defined	
Traditional Chinese	No label defined	No description defined	
Chinese	No label defined	No description defined	

All entered languages

---

occupation	writer	✎ edit
	▸ 1 reference	
	university president	✎ edit
	▸ 1 reference	
	press secretary	✎ edit
	▸ 1 reference	
	speechwriter	✎ edit
	▸ 1 reference	
	+ add value	

archives at	Lyndon Baines Johnson Library and Museum	✎ edit
	▸ 3 references	
	University Archives, Texas State University	✎ edit
	▾ 0 references	
	+ add reference	
	+ add value	

has works in the collection	Texas 150: Sesquicentennial Oral History Project	✎ edit
	▾ 2 references	
	reference URL — https://www.univarchives.txstate.edu/research-collections/core-collections/070-media-artifacts/media-oral-history/oral-history-name-index/hardesty-robert.html	
	retrieved — 31 August 2020	
	reference URL — https://digital.library.txstate.edu/handle/10877/6657	
	+ add reference	
	+ add value	

Wittliff Archives Collections

- Spreadsheet generated using Wittliff's A-Z Index: https://www.thewittliffcollections.txstate.edu/research/a-z.html
- Project's primary focus is adding an "Archives At" statement for each individual (or creating new Wikidata item if none exists)

Archives: A-Z Index

The Archives: A-Z Index is a complete list of archival holdings at The Wi... and music. For our list of photographers, see our **Southwestern & Mex**

A B C D E F G H I J K L M N O

A

- Susan Wittig Albert
- **Judy Alter**
- **Antone's: Home of the Blues Collection**
- **Bill Arhos: Austin City Limits Collection**
- Brian Atkinson
- **Austin Film Festival / On Story Archive**

B

- **Paul and Kitty Baker**
- **Neal Barrett, Jr.**

TEXAS ★ STATE
UNIVERSITY LIBRARIES

Guidelines for Editing Individual Records

- Collections are currently only being approached as an individual's archives
    - Ex. Antone's: Home of the Blues Collection
  - All edits are manual
  - Adding additional relevant biographical data, such as "Awards received" and other readily available information
  - Accessing additional resources when creating New items
    - Restricted from referencing Wikipedia and personal websites

Progress in Wittliff Archives Project

- Of the 152 listed Archives, 70 have been edited and 13 created
  - Leaving 69 to be reviewed
- Currently researching adding images to WikiCommons for use in Wikidata
- Google search of name with "archives" consistently brings Wittliff archives as top hits (Rick Riordan was #3, Sandra Cisneros was #1)

About 769,000 results (0.53 seconds)

www.thewittliffcollections.txstate.edu › research › cisne...

Sandra Cisneros : The Wittliff Collections

Abstract: The **Sandra Cisneros** Papers span 1954-2014 and are divided into twe that document her entire life and literary career up until the ...

TEXAS ★ STATE ®

UNIVERSITY LIBRARIES

Digital Collections Journals

- Five journals currently published in our institutional repository (Digital Collections)
  - Electronic Journal of Differential Equations
  - Journal of College Academic Support Programs
  - Journal of Research on Women and Gender
  - Journal of Texas Music History
  - Texas State Undergraduate Research Journal
  - World History Review Journal
- Goal is to add or enhance any existing records in Wikidata
- Status: Done!

It Takes a Village (or Several Teams)

- <u>Faculty project</u>: Mary Aycock, Sheila Torres-Blank, Amy Eoff, Misty Hopper

- <u>Oral History project</u>: Nicole Critchley

- <u>Wittliff Archives Collections Project</u>: Amanda Scott, Karen Sigler

- <u>Digital Collections Journals</u>: Laura Waugh

TEXAS ★ STATE
UNIVERSITY LIBRARIES ®

Further Considerations & Avenues

- Investigating adding citations to Wikidata
  - Book titles from faculty authors
  - Articles from digital journals to Wikidata
- Assessment of Wikidata work
- Developing best practices from pilot
- Link changes for finding aids
- Editing in Wikipedia
- Opportunity for increased visibility of EDI-related entities?

TEXAS ★ STATE
UNIVERSITY LIBRARIES ®

Addressing questions of sustainability

- Can be difficult to find time in existing workflow
  - Side project status
- Data maintenance - keeping up with changes
- However, library silo systems aren't sustainable and also relatively invisible. What makes the most sense to put our efforts toward?
- Wikidata exposes our data to the public and other can build upon it, breaking down silos. It's what makes Wikidata a good resource.
  - What we add just might make a difference.

TEXAS ★ STATE
UNIVERSITY LIBRARIES ®

Resources

Carlson, S., Lampert, C., Melvin, D. & Washington, A., (2020). *Linked Data for the Perplexed Librarian.* ALA Editions.

Introduction to Wikidata for Librarians (OCLC webinar): https://www.oclc.org/research/events/2018/06-12.html

ARL Whitepaper on Wikidata: https://www.arl.org/wp-content/uploads/2019/04/2019.04.18-ARL-white-paper-on-Wikidata.pdf

UT Linked Data Study Group: https://wikis.utexas.edu/display/UTLinkedDataLearning/UT+Linked+Data+Learning+Group+Home

TEXAS ★ STATE
UNIVERSITY LIBRARIES ®

More Resources

Bulk data entry guides for OpenRefine and QuickStatements:
https://digital.library.txstate.edu/handle/10877/13529

PCC Wikidata Pilot meetings: Meeting Notes, Slides & Recordings

LD4 Affinity Group page: Wikidata:WikiProject LD4 Wikidata Affinity Group/Affinity Group Calls

Texas State Wikidata PCC pilot page with data models: Wikidata:WikiProject PCC Wikidata Pilot/Texas State University Libraries

TEXAS★STATE
UNIVERSITY LIBRARIES

Questions?

Mary Aycock, mba18@txstate.edu
Nicole Critchley, ncritchley@txstate.edu
Amanda Scott, als3@txstate.edu