Gateway into Linked Data: Breaking silos with Wikidata

PCC Wikidata Pilot Project @ Texas State University

Mary Aycock
Nicole Critchley
Amanda Scott

Why did we choose to participate in the PCC Wikidata Pilot?

- Well-timed proposal!
- To gain familiarity of working within linked data platform (Wikidata)
- To use linked data knowledge gained over the past years
- To bring more visibility to affiliated persons, organizations, collections, and projects within our university

TEXAS ★ STATE
UNIVERSITY LIBRARIES

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

Projects [edit]

- **Faculty project**
  - Background: In the past, authority records for faculty have been created 1) as need presents and 2) depending on time, priorities, and expertise available. Moreover, some authority records have been created only locally. Through this project, we are working on two fronts: seeking to create basic national authority records for a set of faculty, as well as update or create Wikidata pages for them.
  - Objective: Edit or create faculty Wikidata pages to add identifiers (LCNAF) and other information while also creating brief authority records in NACO. Also create ancillary pages (departments/centers).
- **Wittliff Archives Project**
  - Background: The Wittliff Collections has a number of archives. This project will also work on two fronts: ensuring that a national authority record exists for each of the subjects, as well as editing/creating their Wikidata page with the "Archives At" property.
  - Objective: Add an "Archives at" property tag to relevant Wikidata pages for Wittliff Collections Archives. If pages don't exist, create them and add the LCNAF (or create a basic NACO authority record).
- **University Archives Sesquicentennial Oral History Project**
  - Background: Texas State University Archives has a collection of 70 transcribed oral histories in our Digital Collections platform. In addition to having a relationship to Texas State University (such as a former faculty), many of them are connected to each other.
  - Objective: Create or enhance the Wikidata records for the subjects of the oral histories, expressing relationships through appropriate properties. Add "Works in the collection" (P6379) property tag to relevant pages for University Archives.
- **Digital Collections Journals Project**
  - Background: Five journals are currently published in our institutional repository (Digital Collections).
  - Objective: Create or enhance the Wikidata records for the five journal titles. Use this project to evaluate what additional digital content can be added to Wikidata.

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
  - Moving toward stable identifiers
  - Empowering communities to create own items

Reassessing effort put into siloed records

- Also: experimentation with creating basic MARC authority records (NACO-lite)

010			no2021005485
040			TxSmTSU ǂb eng ǂe rda ǂc TxSmTSU
024	7		Q56892582 ǂ2 wikidata ǂ1 https://www.wikidata.org/entity/Q56892582
100	1		Castro-Arellano, Ivan
373			Texas State University ǂ2 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: ǂb ti
670			Texas State University Department of Biology faculty directory webpages, viewed January 16, 2021 ǂb (Ivan Castro-Arellano, Ph.D., Associa (Zoonotic diseases, Hantavirus ecology), Community Ecology (Temporal Niche Overlap, Metacommunity Theory), Mammal Ecology, Invasi ecologist to address questions to understand the ecology of zoonotic diseases and invasive species; my goal is to integrate the study of co understand the dynamics of pathogen transmission) ǂu https://www.bio.txstate.edu/about/Faculty---Staff/faculty/Ivan-Castro-Arellano.html

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)

TEXAS STATE
UNIVERSITY LIBRARIES

Preparatory steps for Faculty Authors

- Determined Data Model
- Setup WikiProjects page
- 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

TEXAS ★ STATE
UNIVERSITY LIBRARIES ®

Sample QuickStatements for Campus Departments

CREATE						
LAST	Len	"Texas State University Communication Disorders Department"				
LAST	Den	"academic department at Texas State University"				
LAST	Aen	"Texas State University Department of Communication Disorders"				
LAST	P31	Q43229	S854	"https://www.health.txstate.edu/cdis/"	S813	+2020-10-18T00:00:00Z/11
LAST	P31	Q2467461	S854	"https://www.health.txstate.edu/cdis/"	S813	+2020-10-18T00:00:00Z/11
LAST	P361	Q1495387	S854	"https://www.health.txstate.edu/cdis/"	S813	+2020-10-18T00:00:00Z/11
LAST	P361	Q99737527	S854	"https://www.health.txstate.edu/cdis/"	S813	+2020-10-18T00:00:00Z/11
LAST	P101	Q33226	S854	"https://www.health.txstate.edu/cdis/"	S813	+2020-10-18T00:00:00Z/11
LAST	P131	Q128334	S854	"https://www.health.txstate.edu/cdis/"	S813	+2020-10-18T00:00:00Z/11
LAST	P856	"https://www.health.txstate.edu/cdis/"	P407	Q1860	P585	+2020-10-18T00:00:00Z/11
CREATE						
LAST	Len	"Texas State University School of Health Administration"				
LAST	Den	"academic school at Texas State University"				
LAST	Aen	"Texas State University Health Administration School"				
LAST	P31	Q43229	S854	"https://www.health.txstate.edu/ha/"	S813	+2020-10-18T00:00:00Z/11
LAST	P31	Q2467461	S854	"https://www.health.txstate.edu/ha/"	S813	+2020-10-18T00:00:00Z/11
LAST	P361	Q1495387	S854	"https://www.health.txstate.edu/ha/"	S813	+2020-10-18T00:00:00Z/11
LAST	P361	Q99737527	S854	"https://www.health.txstate.edu/ha/"	S813	+2020-10-18T00:00:00Z/11
LAST	P101	Q1137608	S854	"https://www.health.txstate.edu/ha/"	S813	+2020-10-18T00:00:00Z/11
LAST	P131	Q982353	S854	"https://www.health.txstate.edu/ha/"	S813	+2020-10-18T00:00:00Z/11
LAST	P856	"https://www.health.txstate.edu/ha/"	P407	Q1860	P585	+2020-10-18T00:00:00Z/11

TEXAS★STATE
UNIVERSITY LIBRARIES ®

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

- Goal: To create/enhance at least 100 records (Have 84 as of 3/9/21)
- Credit also to music cataloger who jumped in and created quite a few entries, both manually and through batch editing tools
- Perhaps a good capstone project for student?

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
- 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 (i.e. tried to look up every name) 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 https://www.wikidata.org/wiki/Wikidata:WikiProject_PCC_Wikidata_Pilot/Texas_State_University_Libraries#Models_for_University_Archives_Oral_History_Collections for properties included.

University Archives Oral History Project

- 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

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
®

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

Statements

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

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

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			
Chinese			

All entered languages

occupation	writer	edit
	▸ 1 reference	
	university president	edit
	▸ 1 reference	
	press secretary	edit
	▸ 1 reference	

archives at	Lyndon Baines Johnson Library and Museum	edit
	▸ 3 references	
	University Archives, Texas State University	edit
	▼ 0 references	

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

Still more to consider

- Too soon to tell what the impact of adding these entries will be
    - Need to establish how to assess these activities
- Time commitment involved and steep learning curve.
- Not wanting to add finding aids links just yet

But…

- Wikidata exposes our data to the public and other can build upon it, breaking down silos. What makes Wikidata a good resource; there is a power and beauty to it.
- What I add just might make a difference.

TEXAS ★ STATE
UNIVERSITY LIBRARIES ®

Wittliff Archives Collections

- Handy alphabetical list from which we created a spreadsheet: https://www.thewittliffcollections.txstate.edu/research/a-z.html
- Project consists of adding "Archives At" statement for each of the subjects (or creating new Wikidata item if none exists)
- Inspired by pioneering work from other libraries (such as Emory)

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.**

Example: Rick Riordan

Rick Riordan (Q212727)

American author most notably known for penning "the Percy Jackson and the Olympians" books

✏ edit

Richard Russell Riordan, Jr. | Richard Russell Riordan Jr

▸ Recoin: Most relevant properties which are absent

▸ In more languages

Statements

instance of	human ✏ edit
	▸ 2 references
	+ add value

| image | Rick Riordan (10439).jpg 3,627 × 2,946; 7.86 MB ▾ 0 references | archives at | Wittliff collections ▾ 1 reference reference URL https://www.thewittliffcollections.txstate.edu/research/a-z/riordan.html retrieved 18 October 2020 |

Wittliff Archives Collections

- Some people already existed in Wikidata but did have to create new items
- Currently researching adding images to WikiCommons to use in Wikidata
- Links may change on finding aids so will have to be updated
- As of mid-March:
  - Nearly 300 edits of 68 articles
- 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 two 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)
- Goal is to add or enhance any existing records in Wikidata
- Status: Done!

Collections in this community

Digital Frontiers

Electronic Journal of Differential Equations

International Research Conference for Graduate Students

Journal of College Academic Support Programs

Journal of Research on Women and Gender

Journal of Texas Music History

Race, Ethnicity, and Place Conference

Texas State Undergraduate Research Journal

Undergraduate Research Conference and Honors Thesis Forum

TEXAS ★ STATE ®
UNIVERSITY LIBRARIES

It Takes a Village (or Several Teams)

Faculty project: Mary Aycock, Sheila Torres-Blank, Amy Eoff, Misty Hopper

Oral History project: Nicole Critchley

Wittliff Archives Collections Project: Amanda Scott, Karen Sigler

Digital Collections Journals: Laura Waugh

TEXAS★STATE
UNIVERSITY LIBRARIES

Further Avenues

- Investigating adding citations to Wikidata
  - Titles from faculty authors
  - Articles from digital journals to Wikidata
- Assessment of Wikidata work
- Developing best practices from pilot
- Editing in Wikipedia
- Opportunity for increase visibility of EDI-related entities?
- Addressing questions of sustainability
  - Can be difficult to find time in existing workflow
  - However, library silo systems aren't sustainable and also relatively invisible. What makes the most sense to put our efforts toward?

TEXAS ★ STATE
UNIVERSITY LIBRARIES ®

Resources

Bulk data entry guides for OpenRefine and QuickStatements:

https://drive.google.com/drive/folders/1y-74zozl23IdxXu6E5GsK0aNBY4jNLzl?usp=sharing

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