

6-25-2022

## Enhance the Discovery and Interoperability of Culturally Rich Information: the Chinese Women Poets WikiProject

Sai Deng

*University of Central Florida, [sai.deng@ucf.edu](mailto:sai.deng@ucf.edu)*

Greta Heng

Amanda Xu

Lihong Zhu

Xiaoli Li



Part of the [Archival Science Commons](#), and the [Cataloging and Metadata Commons](#)

Find similar works at: <https://stars.library.ucf.edu/ucfscholar>

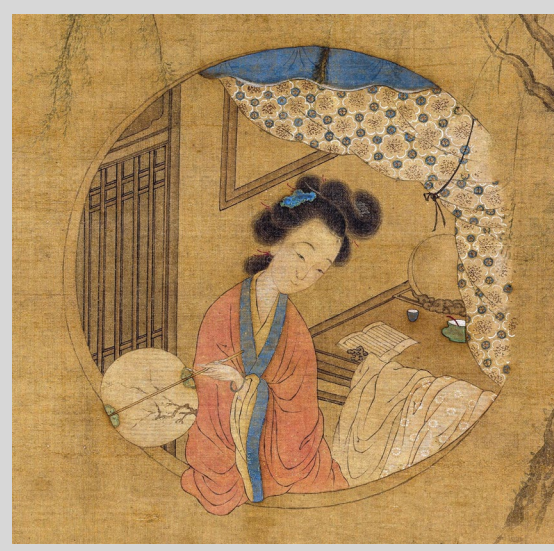
University of Central Florida Libraries <http://library.ucf.edu>

This Poster is brought to you for free and open access by STARS. It has been accepted for inclusion in Faculty Scholarship and Creative Works by an authorized administrator of STARS. For more information, please contact [STARS@ucf.edu](mailto:STARS@ucf.edu).

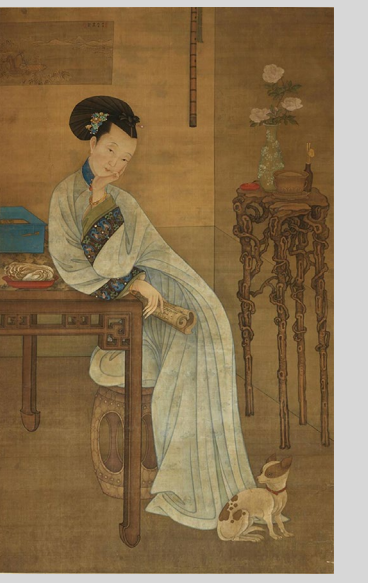
---

### Original Citation

Heng, G., Xu, A., Deng, S., Zhu, L. & Li, X. (2022). Enhance the discovery and interoperability of culturally rich information: the Chinese Women Poets WikiProject. The Chinese American Librarians Association (CALA) Annual Conference 2022 poster session. Washington DC, USA. June 25, 2022.



# Enhance the Discovery and Interoperability of Culturally Rich Information: the Chinese Women Poets WikiProject



Greta Heng<sup>1</sup>, Amanda Z. Xu<sup>2</sup>, Sai Deng<sup>3</sup>, Lihong Zhu<sup>4</sup>, Xiaoli Li<sup>5</sup>

San Diego State University<sup>1</sup>, National Agricultural Library<sup>2</sup>, University of Central Florida<sup>3</sup>, Washington State University<sup>4</sup>, University of California, Davis<sup>5</sup>

## Introduction

The library community has for a number of years questioned how to increase the accessibility and visibility of cultural heritage information stored in centralized databases. Wikidata, a global and open knowledge repository of structured data serving as a hub for linking resources, has emerged as one of possible solutions to library data silo issues.

This research aims to answer the following questions:

- What are the challenges of reconciling names from different sources against Wikidata name entities?
- How to create entries for Chinese women poets and historical places in Wikidata and what are the challenges?
- What are the benefits of using Wikidata to describe cultural heritage related collections and name entities?

This research is important because:

- It enriches Chinese women poets' entries and Chinese historical places in Wikidata and increases the discoverability of that data.
- It contributes to the diversity of data in Wikidata.
- It shows the benefits of Linked Open Data (LOD) in the description and discovery of cultural heritage collections.
- It uses a semi-automated workflow that might be of interest to other librarians doing similar work or Wikidata research.

## Background

In October 2020, a group of Chinese American librarians from several institutions formed a WikiProject: Chinese Culture and Heritage Group. Since then, we have conducted multiple projects and presented our findings about characteristics of Wikidata on Chinese related topics, challenges in creating Chinese related Wikidata entries, data model for Chinese women poets, and different visualizations using Wikidata SPARQL Query Service. This research is built on works previously reported by Amanda Xu, Lihong Zhu, and Sai Deng with the goals to expand the scale of previous research and reflect on the efficiency of Wikidata batch editing workflows.

WikiProject: Chinese Culture and Heritage Group

Project Coordinator:  
Xiaoli Li (University of California, Davis)

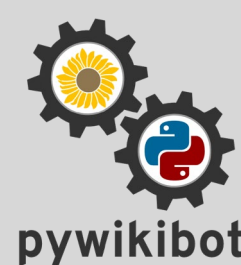
Group Members:  
Sai Deng (University of Central Florida); Jie Huang (Penn State University); Jing Jiang (California Digital Library); Clara Liao (Library of Congress); Hsianghui Liu-Spencer (Carleton College); Greta Heng (San Diego State University); Amanda Xu (National Agricultural Library); Lihong Zhu (Washington State University)

## Data & Methods

### Tool: Pywikibot

Pywikibot is a Python library and collection of tools that automate work on Wikidata. Similar to QuickStatements, it can create new Wikidata items, add new or change statements, and delete statements and items. It also has the following advantages:

- No format restrictions
- Improved efficiency
- More control



We developed several Python scripts using Pywikibot to facilitate the editing process: <https://github.com/sjjiang/ChineseWikiClub>

## Data & Methods

### External Data Sources:

- CBDB: The China Biographical Database
- CHGIS: The China Historical Geographic Information System
- MQWW: Ming Qing Women's Writings

### Research Phases:

#### Phase 1: Chinese Women Poets

Creating and enriching Chinese women poets Wikidata entries

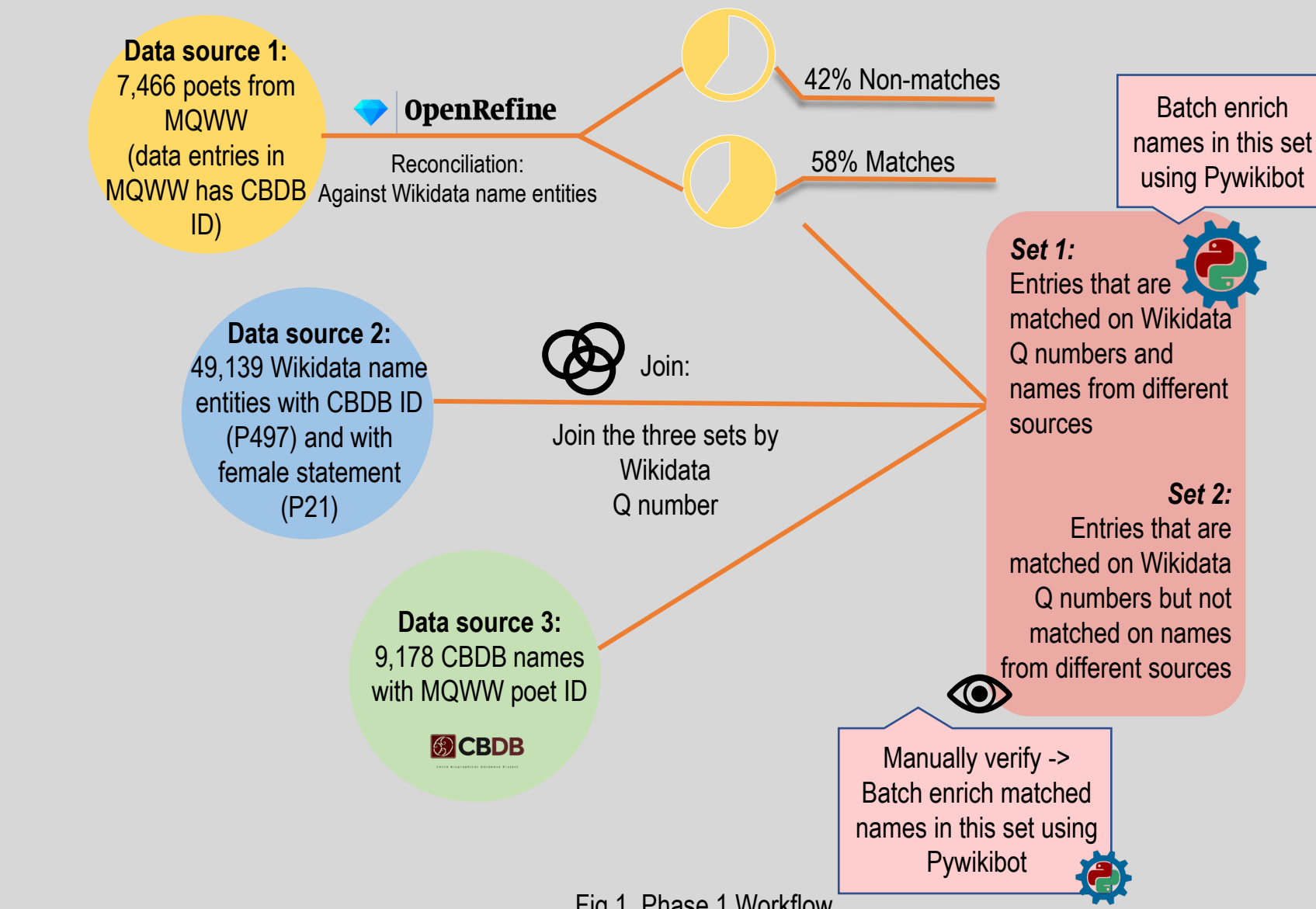


Fig 1. Phase 1 Workflow

More details about the data model of Chinese women poets can be found in the group WikiProject page: [https://www.wikidata.org/wiki/Wikidata:WikiProject\\_Chinese\\_Culture\\_and\\_Heritage](https://www.wikidata.org/wiki/Wikidata:WikiProject_Chinese_Culture_and_Heritage)

#### Phase 2: Chinese Historical Places

Creating Wikidata entries for the birthplace of selected Chinese women poets

Data Model for Historical Places	Value	Note/Example
<b>Basic description</b>		
English label	English Description	For 县, English description should be: ancient county of China
Chinese label	Chinese description	中国历史地名, 今xxxx
Simplified Chinese label	Simplified Chinese description	
Traditional Chinese label	Classical Chinese description	
<b>Core statements</b>		
instance of (P31)	Wikidata item for type of place	ancient county of China (Q28739697)
country (P17)	China (Q29520)	
official name (P1448)	String value	In traditional Chinese
native label (P1705)	String value	In traditional Chinese
coordinate location (P625)	Coordinates	
located in time zone (P421)	UTC+08:00 (Q6905)	
CHGIS ID (P4711)	String value	

Table1 Chinese Historical Places Experimental Stage Data Model

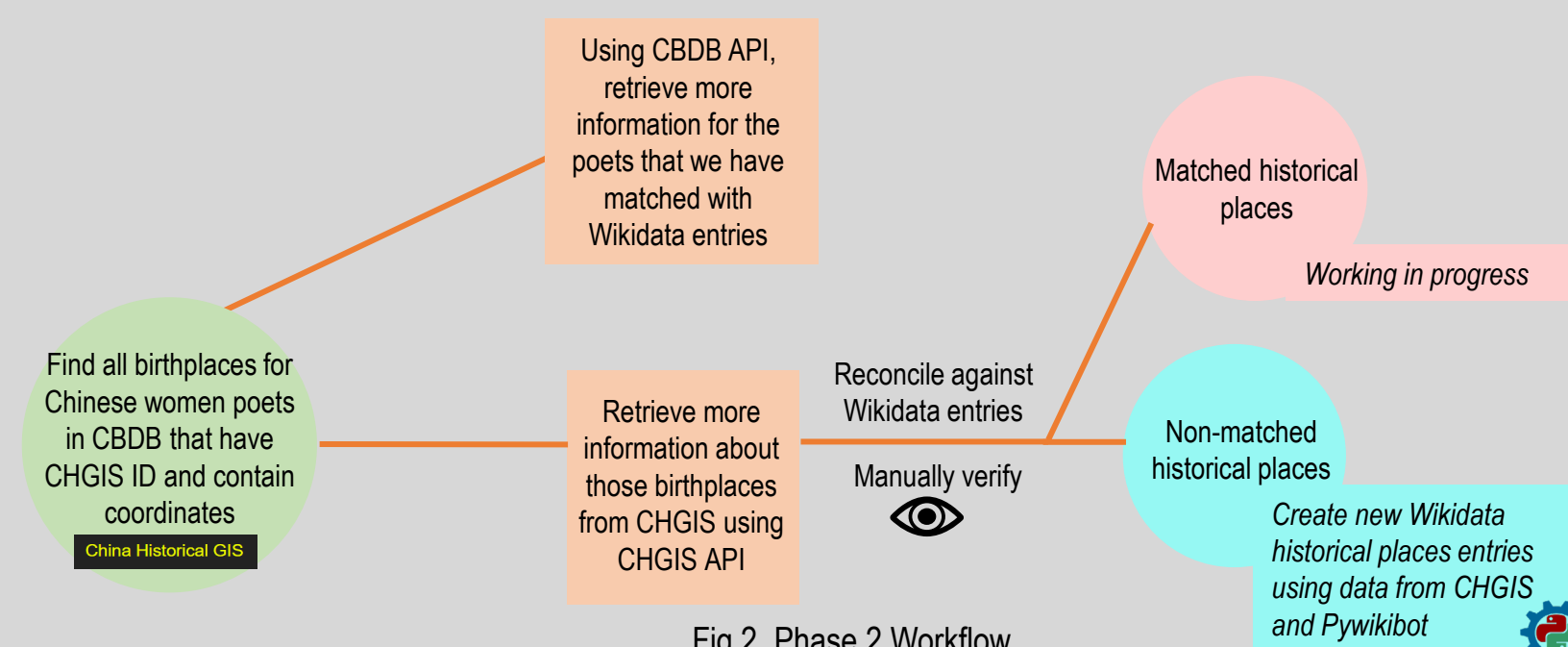


Fig 2. Phase 2 Workflow

## Results

Fig 3. Wikidata Entries Examples (Left: Zhu Gu Nu (a poet), Right: Hedong Xian (a place))

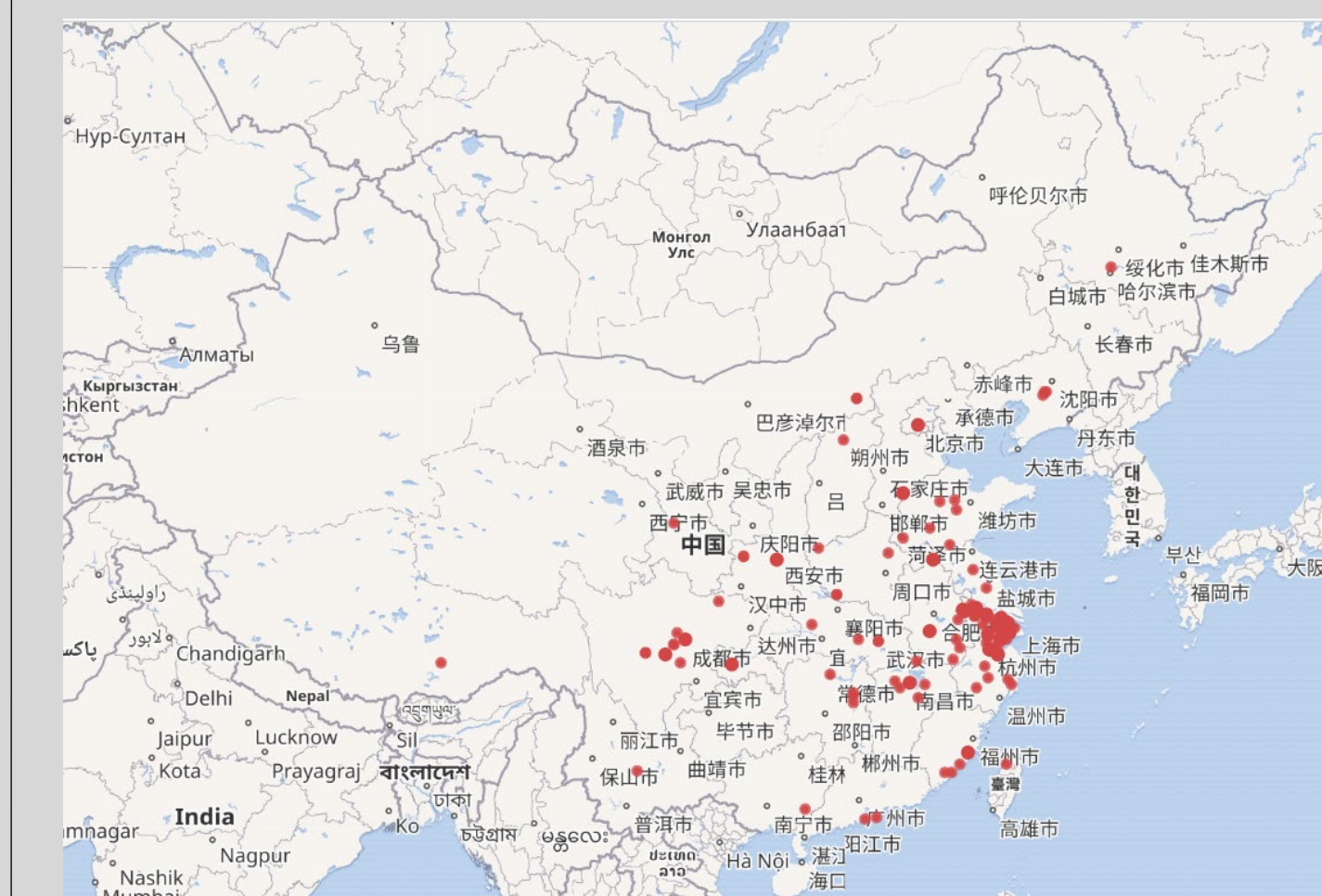


Fig 4. 152 Birthplaces Map View

## Discussion

### 1 Challenges of describing Chinese women poets and reconciling names in different sources against Wikidata name entities:

#### 1.1 Accuracy concern of information in Wikidata

- Wrong gender information
- Information conflict
  - Peishu (Q45625354) is described as a Tang dynasty person but has both Tang and Qing dynasty values of country of citizenship.

Fig 5. Information Conflict Example

## Discussion

### 1.2 Some Wikidata constraints cannot be applied to women in ancient China

- Wikidata constraint: if a family name (P734) is added for a name entity, a given name (Q202444) has to be added as well.
- While women in ancient China had personal names, they were not always recorded. A married woman was referred to by either both her surname and her husband's surname or just her surname.

Fig 6. Example of the Name of a Women in Ancient China

### 1.3 Minimal information for Some Poets in Wikidata

- Some poets' Wikidata entries have minimal information and no CBDB ID
- Manual verification is required

### 2 Challenges of reconciling & describing Chinese historical places

#### 2.1 Different administrative units between ancient China and the Western World

- Different dynasties in China may have different level of administrative units; hard to find the exact match term in English.
- The group has agreed to use county to describe Xian (县) and use prefecture to describe Fu (府). However, we haven't found an equivalent word for Zhou (州) yet..

#### 2.2 Matching criteria for historical places

- Varied data quality and granularity for Chinese historical places in Wikidata
- The name of the same place may change in different dynasties
- Debates about the key elements used in historical places reconciliation

### 3 Potentials of using Wikidata for libraries & cultural heritage collections

- Lowers the barriers for a library to adopt LOD
- Encourages contributions from different institutions or different departments of same institution; improves the interdisciplinary engagement
- Provides data visualization tools and multilingual supports
- Improves the discoverability of cultural heritage collections

## Future Work

- Enrich Chinese women poets Wikidata entries
  - Now: either created or enriched Chinese women poets Wikidata entries with basic information
  - Future: provide more extended properties for those names, such as literary name (Hao, 号) and courtesy name (Zi, 字)
- Find the best practice for describing Chinese historical places
  - Now: 23 historical places were created that had no match in Wikidata
  - Future: keep working on historical places that have Wikidata entries; find the best practices for describing Chinese historical places
- Apply for a property number for MQWW
  - Now: no MQWW property number in Wikidata
  - Future: apply for a P number for MQWW and link the poets' Wikidata entries to their poet ID in MQWW
- Improve Pywikibot speed.