WikiProject: Chinese Culture and Heritage

Sai Deng  
*University of Central Florida, sai.deng@ucf.edu*

Lihong Zhu

Amanda Xu

Part of the Archival Science Commons, and the Cataloging and Metadata Commons

Find similar works at: [https://stars.library.ucf.edu/ucfscholar](https://stars.library.ucf.edu/ucfscholar)

University of Central Florida Libraries [http://library.ucf.edu](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.

Original Citation

Introduction and Research Questions

WikiProject: Chinese Culture and Heritage

WikiProject Chinese Culture and Heritage group created the data model for Chinese Female Poets project. The data model is presented at three levels: Concept, Core, Extended. For details, please go to the website of WikiProject Chinese Culture and Heritage.

Challenges in Creating Chinese Related Wikidata Entries

- **Describing content via expressing culture uniqueness, or, using a more universal approach?**
  - Personal Names
    - given name (P1701)
    - family name (P502)
    - periods (P132) vs. (P242)
    - dynasty, time period (P1788)
    - emperor name (P1789)
    - court rank (P1809)
  - Dynasties, Time Period
    - court rank (P1809)
  - Place names
    - place of birth (P190)
    - place of death (P191)

Using External Data to Enhance Existing Wikidata

- **Using Wikidata SPARQL Query for discovery and presentation:**
  - Step 1. Download data from MWDB, make a downloadable version of MWDB (in Access format). The file was downloaded and converted to Excel format. We also exported the post table from MWDB.
  - Step 2. Use OpenRefine to test external data source. Load the Excel format into OpenRefine. In OpenRefine, run the reconciliation process to find out which one in the Excel spreadsheet already has a Wikidata entry. Wikidata reconciliation process found 4283 Wikidata matches with high confidence. For the rest of 3182, Wikidata reconciliation did not find any matches, instead it gave you a list of possibilities. This is the OpenRefine manual for Wikidata users.
  - Step 3. Use information from CGBT. Since there is no guarantee that the reconciliation process in OpenRefine is 100% accurate, we decided to consult this information from China Biographical Database Project (CGBT)–Harvard University.

Opportunities

- **It is important to develop property level (constant) ontology, and consistently markup Wikidata items according to data models and source info, e.g., the birth date of Chia-ying Yeh has two values: January 1, 2024 and July 1, 2024. One person should only have one birth date. Since we are not sure which one is correct, we should indicate which value has higher confidence score.**

Lessons Learned

- **Need data input standard, e.g., *Kang Youwei* is described as researcher (Q11774202) instead of *Kang Youwei*. The publication, he was a Natural Resource Specialist in the U.S. Forest Service. What should the description be?**

Data Model for Chinese Female Poets

- **WikiProject: Chinese Culture and Heritage group created the data model for Chinese Female Poets project.**
- **Data model is presented at three levels: Concept, Core, Extended.**
- **For details, please go to the website of WikiProject Chinese Culture and Heritage.**