Mapping Library Digital Collections for Chinese Studies

Yan He, Ann James
Sophie Muro and Kahang Ngau
CEAL 2021 Annual Meeting
March 19, 2021

Libraries and Academic Innovation
George Washington University
1. Overview
2. Timeline
3. Accomplishments
   3.1 Dashboard and Widget
   3.2 Website with advanced functionality
   3.3 Metadata
   3.4 Geocoding
   3.5 Quality assurance
4. Potential future plans and concerns
5. Questions
1. Overview

Goals:

- Create an effective tool showing locations of digital collections on Chinese studies in arts, humanities and social sciences
- Provide advance search/browse functionality to accommodate faculty members and students’ need in navigating resources

Significances:

- Discover and promote digital collections
- Fill gap in existing online navigation tools available for researchers using open access sources to support teaching, learning and research
- Help researchers to better plan their field trips through knowing the availability and location of materials
- Provide tools for librarians in library instruction and research consultations
2. Timeline/Workflow

- **06/2020-07/2020**: Evaluate tools/platform; prepare sample dataset; test platform; review literature/existing maps; write student worker job descriptions.
- **08/2020-11/2020**: Data collection and cleaning; geocoding and validating data; mapping data; developing resource description and metadata.
- **11/2020-03/2021**: Designing and implementing online mapping platform and writing Python code for the creation of an auxiliary website.
- **04/2021-06/2021**: Project Phase One wrap up; beta test of online tool; quality assurance checks of data and content; accessibility check. Soliciting feedback and suggestions.
- **06/2021-12/2021 (if continues)**: Project Phase Two: Expand scope from the US and Canada to Europe, East Asia, Oceania and etc.
3. Accomplishments

- 3.1 Dashboard interface customization
- 3.2 Webpage with advanced functionality
- 3.3 Metadata
- 3.4 Geocoding
- 3.5 Quality assurance
3.1 Dashboard Interface Customization

- A Walking Through Tour

Webpage: https://arcg.is/1WbaGO0
A Demo on how to use the Dashboard
Scan the QR code to get access to the app.

Search/Browse App for Cellphone
A Demo on how to use the Browse/Search App
3.2 Web Page with advanced functionality

Multiple Filtering on various categories search on the navigation sidebar.
Sample Python Script

Python Function to check for the validity of the collections’ link.

def check_url(text):
    request = requests.get(text)
    if request.status_code == 200:
        return 'Website exists'
    else:
        return 'Website does not exist'

df['Website_Valid'] = df['URL'].apply(lambda x: check_url(x))
df[df['Website_Valid'] == 'Website does not exist']
### 3.3. Metadata

#### Transcription in csv format

<table>
<thead>
<tr>
<th>Collection Title</th>
<th>Institution Name</th>
<th>Library Name</th>
<th>Type of Library</th>
<th>Languages</th>
<th>General Format</th>
<th>Specific Formats</th>
<th>Subject Area</th>
<th>Subjects</th>
<th>Keywords</th>
<th>URL</th>
<th>Start Year</th>
<th>End Year</th>
<th>Year Range</th>
<th>Description</th>
<th>Address</th>
<th>Latitude</th>
<th>Longitude</th>
</tr>
</thead>
<tbody>
<tr>
<td>Li Desheng University &amp; SLAIS stu Academic Chinese</td>
<td>Image</td>
<td>Drawing, Art</td>
<td>History, Cigarette Chhttp://example.com</td>
<td>1885-1950-1885-1950</td>
<td>&quot;360 jobs&quot; is West 16th</td>
<td>49.253</td>
<td>-123.2445</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The Chines Harvard UnFine Arts Academic Chinese, Ja</td>
<td>Image</td>
<td>Rubbing, Art, Religion</td>
<td>Tomb, Ru</td>
<td>Qin Dynastyhttp://example.com</td>
<td>-221-1644-221-1644-1067 full- im Kelley Stre</td>
<td>42.3776</td>
<td>-71.11602</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gaochao B Harvard UnHarvard- Academic Chinese</td>
<td>Text</td>
<td>Rare Book, History</td>
<td>Literature</td>
<td>Scholarshiphttp://example.com</td>
<td>-500-1900-500-1900-1138 public Kelley Stre</td>
<td>42.3776</td>
<td>-71.11602</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chinese Te University &amp; University Academic Chinese</td>
<td>Text</td>
<td>Poem</td>
<td>Literature</td>
<td>Poem</td>
<td>Translationhttp://example.com</td>
<td>476</td>
<td>1868-476-1868</td>
<td>300 Tang Po 155, Rugby</td>
<td>38.0365</td>
<td>-78.52288</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bao Juan C Harvard UnHarvard- Academic Chinese</td>
<td>Text, Im</td>
<td>Rare Book, Literature</td>
<td>Religion</td>
<td>Folk Practihttp://example.com</td>
<td>1368-1644-1368-1644-87 titles of B Kelley Stre</td>
<td>42.3776</td>
<td>-71.11602</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chinese Ra Yale Univer SML East Academic Chinese</td>
<td>Text</td>
<td>Rare Book, Literature</td>
<td>History</td>
<td>Song, Yuan</td>
<td><a href="http://example.com">http://example.com</a></td>
<td>960</td>
<td>1949-960-1949</td>
<td>A collection</td>
<td>41.3153</td>
<td>-72.92604</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cheng Yu T University &amp; East Asia Academic Chinese</td>
<td>Text</td>
<td>Rare Book, Literature, Language</td>
<td>Philosophy</td>
<td><a href="https://example.com">https://example.com</a></td>
<td>1609-1985-1609-1985 A collection</td>
<td>27 King's C</td>
<td>43.6629</td>
<td>-79.39582</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cheng Yu T University &amp; East Asia Academic Chinese, Ja</td>
<td>Text</td>
<td>Book</td>
<td>Literature, Philosophy</td>
<td>Religion, Thhttp://example.com</td>
<td>1786-1985-1786-1985 A collection</td>
<td>27 King's C</td>
<td>43.6629</td>
<td>-79.39582</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Fields include institution name, library name, language, general format, specific format, subject, keywords, URL, start year, end year, year range, description, address, latitude and longitude
- Fields collected but not included: Accessibility, Number of digitized items, Geographic Location
3.3. Metadata: Controlled Vocabularies

○ **Format:**
  - General format: Text, Image, Sound, Video, Dataset
  - Specific format: From collection metadata
  - Refer to Dublin Core & Getty Art & Architecture Thesaurus

**Subject Area/Subjects/Keywords:**
- From collection metadata
- Refer to LCSH & FAST
- Make analysis of collection collection and summarize

**Time Period:**
- Use - (minus sign) to indicate BCE
- Use specific years to transcript dynasties, centuries or decades; e.g. Qing Dynasty (1644 to 1911); 18th century (1700-1799); 1950s-1970s (1950 to 1979)

**Institution Name:**
- Name Authority Records
- Standardized naming convention
3.4. Geocoding Process

- What is geocoding?
- Key challenges and decisions
- The process
- Overcoming limitations
3.5. Quality Assurance Process

- Engage spreadsheet capabilities to
  - Check for duplicates
  - Examine spelling and grammar
  - Spot check for errors
  - Look for nulls (or missing values)
- Annotate and distribute questions to the team
- Mitigate and resolve data issues in underlying data sheet
4. Potential future concerns and plans

1. Expand the geographic scope from US/Canada to the rest of the world
2. Continue cleaning and clarifying metadata
3. Continue seeking enhancements for searching and browsing capabilities
4. Continue to collect recommendations for digital collections on Chinese Studies and gather feedback from users (survey form)

Suggest A Digital Collection or Provide Feedback

https://forms.gle/T6ksEP51hAaYS96X7
Acknowledgement

Externally:

Cathy Yang, Cataloging Librarian, Library of Congress
Shi Deng, Head of CJK Metadata, UC San Diego Libraries
Xiang Li, Chinese & Asian Studies Librarian, University of Colorado Boulder Libraries
Ying Zhang, Research Librarian for Asian Studies, UC Irvine Libraries
Patrick Ryan, Solution Engineer at Esri

Internally:

Geography Department

- Prof. Michael Mann
- Prof. Richard Hinton

Libraries and Academic Innovation:

- Hannah Sommer
- Elizabeth Waraksa
- Peter Cohn
- Matthew Bright
- Cathy Zeljak
- Williams Gillis
- Matthew Mihalik
- Benjamin Horn
- Joshua Gleason
- Laura Wrubel

Confucius Institute
Questions?

Contact Info
Yan (hey@gwu.edu) or Ann (ajames31@gwu.edu)