Cataloguing for the world: motivation, method and madness

Peter Sidorko, Connie Lam
University of Hong Kong Libraries

OCLC Asia Pacific Regional Council Membership Conference
8 October 2013, Bangkok, Thailand
Outline

- HKUL
- Cataloging in HKUL
- Chinese e-resources cataloguing
- Challenges and solutions
- Cataloguing outputs
HKUL

- Established in 1912
- Comprises 1 main library + 6 branches
- Collection

<table>
<thead>
<tr>
<th></th>
<th>Print</th>
<th>Electronic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Books</td>
<td>2.96 million</td>
<td>3.58 million</td>
</tr>
<tr>
<td>Journals</td>
<td>51,513</td>
<td>81,886</td>
</tr>
</tbody>
</table>

- Staff
  - Professionals: 32
  - Support staff: 197

figures as of June 2013
Objectives of HKUL Cataloguing

- Provide our users quality bibliographic records in timely manner.
  - Catalog individual titles in databases to easy the search for our users.
  - Employ all kinds of technology to speed up the workflow to allow our users have promptly access.

- We focus on our users ➔ Cataloguing for our users.
Challenges we encountered

- Rapid increases of e-resources

**ebooks**

<table>
<thead>
<tr>
<th>Year</th>
<th>2003</th>
<th>2008</th>
<th>2013</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value</td>
<td></td>
<td>1500000</td>
<td>3500000</td>
</tr>
</tbody>
</table>

**ejournals**

<table>
<thead>
<tr>
<th>Year</th>
<th>2003</th>
<th>2008</th>
<th>2013</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value</td>
<td>5000</td>
<td>80000</td>
<td>90000</td>
</tr>
</tbody>
</table>
Challenges we encountered

- Decreases of cataloguing staff

![Chart showing decreases of cataloguing staff over years]

- Support staff
- Professional

2002: 25
2004: 20
2013: 15
What can we do?

- Collection ↑
- Staffing ↓

Vendor supplied records

- For Western e-resources, especially the large package, we rely on vendor records. The quality of the records normally are acceptable.
- But for Chinese e-resources …..
The e-Chinese Collection in HKUL

- **Major acquisitions (to 2005):**
  - 1998: Sikuquanshu (四庫全書)
  - 1999: China Journal Net (China Academic Journal) 5,000+ titles
  - 2000: China Core Newspapers Database
  - China Doctors & Masters Dissertations
  - 2002: Sibucongkan (四部叢刊)
  - 2003: Apabi e-book around 40,000 titles
  - 2005: SuperStar Digital Library around 400,000 titles
Our journey in cataloguing Chinese ER

- **Chinese Journal Net (CAJ)**
  - With the limited information provided by vendor, we started our cataloguing project in 2000.
  - Only a title list in print format. No ISSN, no information on title relationship, no coverage details.
  - We requested vendor to provide more detail information in excel file in later stage, which helped to speed up the cataloging work.
  - The project completed in 2004. We spent 4 years for around 5000 titles.
Our journey in cataloguing Chinese ER

- With the large amount of Chinese ebooks purchased, we asked for Marc records.

- Vendor provided Marc records
  - Problems with the Chinese scripts.
  - Not follow AACR2.
  - Improper tag used, incorrect indicators used.
  - Not follow LC Pinyin guidelines.
  - Multi-records for multi-volume set
  - Multi-records for serial titles.
Our journey in cataloguing Chinese ER

- Communicated with vendor for desired standard.
- Provided advice to enhance the records
- Provided example records to vendor.
- Spent months in communication ➔ no good news ➔ Dead end??
- Why vendor unable to fulfill our requirements
  - Records were generated by metadata in database
  - Problems cannot be tackled by program
In the meantime...

**Message from the Librarian**

**HKU Libraries is Second in WorldCat’s Top Ten**

Anthony Ferguson

We were pleasantly surprised to learn last month that HKU Libraries is the 2nd largest contributor of original cataloging records to the database of the Online Computer Library Center, a membership organization made up of 50,000 libraries in 84 countries. Our 45,750 records submitted to this database provided a large revenue to the Libraries, and also placed us ahead of such famous and large libraries as the Library of Congress, the British Library, etc. The University of Oxford Library came in 2nd with 30,306 records, HKU Libraries second, and the Center for Research Libraries (USA) placed first with 83,549 records. For the second year in a row, our contribution has exceeded 40,000. This is a reflection of the massive amount of work that our Technical Services staff members are able to do year after year.

With a budget that is very high, this high contribution of unique records is one way that HKU Libraries is able to draw back some of the costs, and in some cases, pay completely for some of the many services we subscribe to and offer our users.

With over 12,000 libraries contributing records to WorldCat, and with over 53 million records, this database is the de facto worldwide union catalog. When searching WorldCat, the record is usually already entered, so many libraries buy and catalog the same materials. It is very rare for libraries to enter so many unique records, and WorldCat is a great example of why HKU Libraries has grown and continues to do. We have been able to do this because of the nature of our contribution of records is relatively new compared to other worldwide contributors to WorldCat. We are the largest contributor of Chinese materials in WorldCat. We are the largest of catalogers of Chinese periodicals and ebooks in WorldCat. On the other hand, for most of our English-language materials, instead of contributing records, we download records from WorldCat.

As the largest and premier cataloging database of the world, WorldCat allows contributing members such as HKU Libraries to share resources worldwide with other libraries and universities. In this manner we can increase HKU’s reputation and further fulfill its mission of disseminating knowledge. End-users use WorldCat for information discovery. Upon finding that it resides in HKU Libraries, they can visit the Libraries, request inter-library loan, and now increasingly, simply click to connect to the resource at HKU Libraries. WorldCat is now searchable through Yahoo and Google, which means that HKU Library records will begin to appear in these two giant search engines.

Top 10 original cataloguers

<table>
<thead>
<tr>
<th>Organization</th>
<th>Original Records</th>
<th>Active On</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cornell Research Libraries</td>
<td>43,840</td>
<td>2004</td>
</tr>
<tr>
<td>University of Hong Kong</td>
<td>46,768</td>
<td>2004</td>
</tr>
<tr>
<td>University of Oxford</td>
<td>39,322</td>
<td>2004</td>
</tr>
<tr>
<td>The Ohio State University</td>
<td>20,040</td>
<td>2004</td>
</tr>
<tr>
<td>University of Wisconsin</td>
<td>24,075</td>
<td>2004</td>
</tr>
<tr>
<td>Godfrey Memorial Library</td>
<td>23,047</td>
<td>2004</td>
</tr>
<tr>
<td>University of California, Berkeley</td>
<td>22,822</td>
<td>2004</td>
</tr>
<tr>
<td>Singapore National Library</td>
<td>18,186</td>
<td>2004</td>
</tr>
<tr>
<td>Union Theological Seminary</td>
<td>15,729</td>
<td>2004</td>
</tr>
<tr>
<td>University of Texas at Austin</td>
<td>15,200</td>
<td>2004</td>
</tr>
</tbody>
</table>

**WorldCat**

They can visit the Libraries, request inter-library loan, and now increasingly, simply click to connect to the resources at HKU Libraries. WorldCat is now searchable through Yahoo and Google, which means that HKU Library records will begin to appear in these two giant search engines.
Our journey in cataloguing Chinese ER

With severe budget cuts, this high contribution of unique records is one way that HKU Libraries is able to claw back some of the costs, and in some cases, pay completely for some of the many services we subscribe to and offer our users.

= Way out
Our journey in cataloguing Chinese ER

- We decided:
  - Request vendors to provide title lists with the following data in excel file
    - Author
    - Title
    - Publisher
    - ISBN
    - Direct url
    - Control number
  - Create records ourselves, upload to OCLC to earn more credits.
Our strategy

- Handle titles we have print records first.
- We estimated that 20-30% of the titles have print records in HKUL catalog.
- Based on the title list, we developed scripts
  - To search ISBN or title of print records in HKUL catalog.
  - To generate records based on the print records.
  - To add url and control number to records.
  - To use “Global update function” to convert records from print to electronic.
Try and error process

- Search titles by program → Fail
  - Simply rely on title is not reliable.
  - A lot of titles in the file have errors.
  - Same title sometimes has duplicate entries due to
    - Different punctuation used
    - Simplified/Traditional characters
    - Different space used
  - We tried one batch and found that too much effort required for the clean up.
Try and error process

- Search ISBN by program → Fail
  - The matching rate is much lower than our estimation.
  - Some publishers use same ISBN.
    - For different titles.
    - For same title, but different editions.
  - We used ISBN program to process 3 batches and have to re-check every title in these batches to ensure the records are correct.
Our solution

- The title list provided has a lot of incorrect data and duplicated entries.
- Manually search for matching records is the only solution.
- We expand the search to other libraries via Z39.5 and also OCLC.
- We set up a small project team for cataloguing of Chinese e-books.
Our tactics

- Outsourcing cataloguing of Western monographs to free some manpower.
- Load vendor supplied records to our catalog for patron driven acquisition package without editing.
  - Not to waste time to handle titles we don’t buy
  - Upgrade those titles we purchased
- Minimize local practice
  - Give up all local practices, except Hong Kong subject headings for Hong Kong materials.
Our tactics

- Plan more before action.
  - Analyze title list to locate serial titles, collectanea and have them excluded from searching list.
- Members of the newly set up team are mainly responsible for searching.
  - Minimal training
  - High productivity
- Gradually develop their cataloging skills by assigning different tasks based on performance.
Our tactics

- Task breakdown
  - Searching
  - Descriptive cataloging
  - Authority checking
  - Call number and subject assigning
  - Review and quality control

- Matching work with staff skills
  - Tasks are allocated to staff according to experience and knowledge
  - Staff only handle tasks they are able to handle
Our outcomes

- Titles in collectanea which are not searchable in the databases, can be located by searching HKUL catalog.

- More precise title information in HKUL catalog than the databases which help users to locate the resources easily.

- By uploading records to OCLC and with Open WorldCat, our users can identify the e-books by searching on the web. Our collection becomes more visible.
<table>
<thead>
<tr>
<th>Title</th>
<th>Limit results to available items</th>
<th>Search</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>EBOOK</td>
<td>1995</td>
</tr>
<tr>
<td>2</td>
<td>EBOOK</td>
<td>1995</td>
</tr>
</tbody>
</table>

**Titles (1-2 of 2)**

1. **電子商務: electronic resource**
   - Title: 电子商贸: 电子资源
   - Author: China: Qing 1995.
   - Call Number: 614.3159
   - Click to access

2. **電子商務: electronic resource**
   - Title: 电子商贸: 电子资源
   - Author: China: Qing 1995.
   - Call Number: 614.3159
   - Click to access
AUTHOR
Cai, Qiong. 1453-1509.

TITLE
Ai an mi zhen [electronic resource]; Yi Juan; He Luo si jian; Yi Juan; Tai Ji tu shuo; Yi Juan / Cao Qing zhuang.

IMPRINT

EDITION
D1 ban.

NOTE
"Zhongguo ke xue yuan tu shu guan cang Qing Qianlong qian nian (1742) Cai Tingkui ke ben" ying yin.

LC SUBJECT
XJ.jpg

ALT AUTHOR

Show similar items
Divination.
Divination -- China.

Show similar items

Show similar items
### TITLES (1-50 of 2502)

<table>
<thead>
<tr>
<th>Title</th>
<th>Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>发挥主要电子资源</td>
<td>1995</td>
</tr>
<tr>
<td>孔子圣像电子资源</td>
<td>1995</td>
</tr>
<tr>
<td>孔子圣像 电子资源</td>
<td>1995</td>
</tr>
<tr>
<td>千秋圣像 电子资源</td>
<td>1995</td>
</tr>
<tr>
<td>本端圣像 电子资源</td>
<td>1995</td>
</tr>
<tr>
<td>本端圣像 电子资源</td>
<td>1995</td>
</tr>
<tr>
<td>本端圣像 电子资源</td>
<td>1995</td>
</tr>
<tr>
<td>正堂大 电子资源</td>
<td>1995</td>
</tr>
</tbody>
</table>

**Note:** The text is not in English and cannot be accurately translated.
<table>
<thead>
<tr>
<th>序号</th>
<th>类型</th>
<th>标题</th>
<th>作者</th>
<th>出版社</th>
<th>出版日期</th>
<th>价格</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>EBOOK</td>
<td>20世纪中国散文史 [电子资源] / 哈九唯, 沙小倪</td>
<td>上海: 上海教育出版社</td>
<td>1988</td>
<td>电子资源</td>
<td>虚拟货币</td>
</tr>
<tr>
<td>2</td>
<td>EBOOK</td>
<td>20世纪中国散文史 [电子资源] / 沙小倪</td>
<td>上海: 上海教育出版社</td>
<td>1988</td>
<td>电子资源</td>
<td>虚拟货币</td>
</tr>
<tr>
<td>3</td>
<td>EBOOK</td>
<td>20世纪中国散文史 [电子资源] / 哈九唯</td>
<td>上海: 上海教育出版社</td>
<td>1988</td>
<td>电子资源</td>
<td>虚拟货币</td>
</tr>
<tr>
<td>4</td>
<td>EBOOK</td>
<td>20世纪中国散文史 [电子资源] / 沙小倪</td>
<td>上海: 上海教育出版社</td>
<td>1988</td>
<td>电子资源</td>
<td>虚拟货币</td>
</tr>
<tr>
<td>5</td>
<td>EBOOK</td>
<td>20世纪中国散文史 [电子资源] / 哈九唯</td>
<td>上海: 上海教育出版社</td>
<td>1988</td>
<td>电子资源</td>
<td>虚拟货币</td>
</tr>
</tbody>
</table>

注意：上述信息可能为虚构或示例数据，实际内容需根据具体情况进行调整。
Our outcomes

- We developed our skills and knowledge in
  - Planning the work
  - Monitoring the workflow
  - Employing different automated tools
The e-Chinese Collection in HKUL

- Major acquisitions (from 2005)
  - 2008:
    - China yearbook full-text database
    - China reference works online
  - 2009:
    - Database of Chinese classic ancient books (中國基本古籍庫)
    - Taiwan electronic periodical services
    - Apabi e-references
  - 2010: Chinamaxx digital library
  - 2011: China online journals. [Medical journals]
  - 2012: Wanfang video. Medicine
Chinese e-resources we cataloged

- Chinese e-journals: 18,273 titles
- Chinese e-books: 447,051 titles
- Chinese e-news: 947 titles

We manage to have records for all our e-resources, except:

- Chinamaxx digital library (50% cataloged)
- China Doctors & Masters Dissertations (no title list provided)
Collection sets in OCLC

- Colleagues in US learnt that we have all China academic journals (CAJ) cataloged and urged OCLC to approach us for creating a collection set.
- We created our first set for CAJ in 2006
- Added another set for China Yearbook Full-Text Database in 2009
- We maintain the updates of these two sets
Cataloging outputs

- Top 3 original catalogers in OCLC since 2004

<table>
<thead>
<tr>
<th>Year</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>2004</td>
<td>2</td>
</tr>
<tr>
<td>2005</td>
<td>3</td>
</tr>
<tr>
<td>2006</td>
<td>1</td>
</tr>
<tr>
<td>2007</td>
<td>2</td>
</tr>
<tr>
<td>2008</td>
<td>2</td>
</tr>
<tr>
<td>2009</td>
<td>1</td>
</tr>
<tr>
<td>2010</td>
<td>1</td>
</tr>
<tr>
<td>2011</td>
<td>2</td>
</tr>
<tr>
<td>2012</td>
<td>1</td>
</tr>
</tbody>
</table>
Are we mad?

- While others are accepting vendor records even they are poor in quality, we decide to create records by our own.

- While others are waiting for records in OCLC for downloading, we contribute huge amount of records to OCLC.

- NO!
  - It is because we understand both ways are unable to provide our users quality records in timely manner.
Conclusion

- Cataloging for our users is our objective.
- Cataloging for the world is the side effect.
Thank You!

Peter Sidorko
peters@hku.hk

Connie Lam
csllam@hku.hk