Bringing Legacy Digital Collections into a New Digitization Portal: Challenges and Strategies

Jeff LIU, Louisa LAM
University Library System
The Chinese University of Hong Kong
December 11, 2013
Overview

- About The Chinese University of Hong Kong (CUHK)
- Overview of the digital initiatives at the CUHK University Library System (ULS)
- Problems, merits of the current approach
- Challenges of integrating new collections into the legacy collection
- Interim measures to address the challenges
- Limitations and future directions to sustain an on-going digital initiatives that meet the needs of its audience
- Lessons learnt
About CUHK

- Established in 1963 (50th anniversary)
- A comprehensive research university with a 137.3-hectare campus overlooking Tolo Harbor at Shatin, New Territories
- Comprises of 9 colleges, 8 Faculties and Graduate School
- 18,698 undergraduates and postgraduates, 7,157 teaching staff
About CUHK ULS (1)

- Comprises of seven libraries:
  - University Library
  - Lee Quo Wei Law Library
  - Chung Chi College Elizabeth Luce Moore Library
  - Architecture Library
  - New Asia College Ch’ien Mu Library
  - United College Wu Chung Multimedia Library
  - Li Ping Medical Library
### About CUHK ULS (2)

- **Collection size:**

<table>
<thead>
<tr>
<th>Category</th>
<th>Eastern Languages</th>
<th>Western Languages</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Books</td>
<td>1,054,745</td>
<td>1,158,507</td>
<td>2,213,252</td>
</tr>
<tr>
<td>Bound Periodicals</td>
<td>74,262</td>
<td>233,510</td>
<td>307,772</td>
</tr>
<tr>
<td>Current Printed Serials</td>
<td>3,261</td>
<td>6,479</td>
<td>9,740</td>
</tr>
<tr>
<td>Electronic Books</td>
<td>1,882,206</td>
<td>2,674,310</td>
<td>4,556,516</td>
</tr>
<tr>
<td>Electronic Journals</td>
<td>30,589</td>
<td>98,755</td>
<td>129,344</td>
</tr>
<tr>
<td>Electronic Databases</td>
<td>88</td>
<td>544</td>
<td>632</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>3,045,151</strong></td>
<td><strong>4,172,105</strong></td>
<td><strong>7,217,256</strong></td>
</tr>
</tbody>
</table>

- **Special Collections:** range from Shang Dynasty oracle bones to modern Chinese literary archives

- **Patrons**
  - Registered Library users: 58,340
  - No. of users entering the libraries 2012-2013: 2,435,090
Background of Digital Initiatives at ULS (1)

- Why do we need to develop a digital library?
  - To preserve significant academic and intellectual work from ULS and donation, which are not available elsewhere in Hong Kong;
  - To preserve unique, precious but fragile materials for future generation use;
  - To fulfill the wishes of donors;
  - To meet the learning and research needs of the University community;
  - To develop local content to fill the missing gaps in electronic resources e.g. Newspaper database was not available in Hong Kong in the early 90s;
  - To promote online dissemination and sharing of information; and
  - To facilitate scholarly communications
Background of Digital Initiatives at ULS (2)

- What have been accomplished?
  - Started in 1995 with the database Hong Kong NewsBot
  - 23 digitization projects were completed in 18 years: http://udi.lib.cuhk.edu.hk/projects
  - Cover different subject areas ranging from literature, culture, arts, politics, society, and religion and different formats including books, journal articles, manuscripts, newspapers, bamboo stripes, photos, and audio files
  - As at Sep 2013, the total online accessible images has reached a new height of 3,883,694
Long Term Goal of the Library

- Continue to create digital collections in a sustainable and on-going manner
- Current projects include: e-theses and dissertations collections, a translation archive, oral history archive, digitization of more Chinese rare books. There are more unique donation materials need to be preserved and be more widely accessible. More born-digital materials to handle.
- Develop a genuine digital library. A digital library may mean different things to different people.
- In current context, a digital library is conceived as “a focused collection of digital objects, including text, video, and audio, along with access and retrieval, and for selection, organization, and maintenance of the collection.”

Main Features of A Good Digital Library

- Capable of managing digital objects in perpetuity
- Able to support multimedia content along with text
- User-friendly Interface
- Open up access to the content
- Supports a digital library system with a well-defined software architecture and system architecture
- Able to create and handle at least one type of metadata schema
- Interoperable – ability to perform a cross-collection search over diverse sets of metadata records and obtain meaningful results
- Advanced functions to acquire, browse, search, visualize, publish and disseminate digital objects
- Facilitate discoverability of content
- More than a web site – allows easy addition of new materials / collections without any need for manual updating of the structures for access and retrieval
Main Challenges

What approach is to take to achieve the long term goal of the Library?

How to integrate the legacy digital collections with the new ones?
## Digitization Projects (1995-2003 Sep) At A Glance (1)

<table>
<thead>
<tr>
<th>Projects</th>
<th>Year</th>
<th>Source</th>
<th>Digital Content</th>
<th>Full Text</th>
<th>Mode of access</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hong Kong NewsBot (<a href="http://newsbot.lib.cuhk.edu.hk">http://newsbot.lib.cuhk.edu.hk</a>)</td>
<td>1995</td>
<td>local newspapers</td>
<td>Newspaper images in TIFF format</td>
<td>Y</td>
<td>Public access</td>
</tr>
<tr>
<td>Manuscript Collection (various web pages)</td>
<td>1998</td>
<td>CUHK Library collections</td>
<td>Manuscript, images in TIFF format</td>
<td>Y</td>
<td>Public access</td>
</tr>
<tr>
<td>Index to Chinese Periodicals of Hong Kong (HKInChiP) (<a href="http://hkinchippub.lib.cuhk.edu.hk">http://hkinchippub.lib.cuhk.edu.hk</a>)</td>
<td>1999-</td>
<td>Periodicals published in Hong Kong, get permission from publishers</td>
<td>Journal articles, bibliographic records, images in PDF format</td>
<td>Y (21 journals)</td>
<td>Full text restricted to CUHK users only</td>
</tr>
<tr>
<td>Hong Kong Literature Database (<a href="http://hklit.lib.cuhk.edu.hk">http://hklit.lib.cuhk.edu.hk</a>)</td>
<td>2000-</td>
<td>CUHK Library collections and donations, partner with Hong Kong Literature Research Centre</td>
<td>Journals articles, newspaper literary columns, newspaper articles, monographic bibliographic records, theses and dissertations, activities report of Hong Kong literature and literature groups, images in PDF format</td>
<td>Y</td>
<td>Public access</td>
</tr>
</tbody>
</table>
## Digitization Projects (1995-2003 Sep) At A Glance (2)

<table>
<thead>
<tr>
<th>Projects</th>
<th>Year</th>
<th>Source</th>
<th>Digital Content</th>
<th>Full Text</th>
<th>Mode of access</th>
</tr>
</thead>
<tbody>
<tr>
<td>Modern Chinese Drama Database (<a href="http://mcdrama.lib.cuhk.edu.hk">http://mcdrama.lib.cuhk.edu.hk</a>)</td>
<td>2001</td>
<td>based on two Catalogues of Late Chinese Drama, and the catalogue on ULS drama collection</td>
<td>Bibliographic records with links to library OPAC</td>
<td>N (Index only)</td>
<td>Public access</td>
</tr>
<tr>
<td>United College General Education Programme Senior Seminar Papers Database (<a href="http://uugeu.lib.cuhk.edu.hk">http://uugeu.lib.cuhk.edu.hk</a>)</td>
<td>2001-</td>
<td>United College students’ senior seminar paper</td>
<td>Students’ seminar papers, Includes images in PDF Format and slides</td>
<td>Y</td>
<td>CUHK users only</td>
</tr>
<tr>
<td>Zou Ma Lou San Guo Wu Jian, Jia He Li Min Tian Jia Bie Database (<a href="http://rhorse.lib.cuhk.edu.hk">http://rhorse.lib.cuhk.edu.hk</a>)</td>
<td>2002</td>
<td>Donation and partner with Depts. of Chinese Language and Literature, and History</td>
<td>Full text search of the book Changsha Zoumalou San Guo Wu Jian. Jia He Li Min Tian Jia Bie, text with some Chinese characters created as images</td>
<td>Y</td>
<td>Public access</td>
</tr>
</tbody>
</table>
## Digitization Projects (1995-2003 Sep) At A Glance (3)

<table>
<thead>
<tr>
<th>Projects</th>
<th>Year</th>
<th>Source</th>
<th>Digital Content</th>
<th>Full Text</th>
<th>Mode of access</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hong Kong Architectural Publications Index</td>
<td>2004</td>
<td>architectural journals published in Hong Kong</td>
<td>Bibliographic records</td>
<td>N (Index only)</td>
<td>Public access</td>
</tr>
<tr>
<td>(<a href="http://hkapi.lib.cuhk.edu.hk">http://hkapi.lib.cuhk.edu.hk</a>)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CUHK E-Journal Initiatives</td>
<td>2004-</td>
<td>CUHK Library collections</td>
<td>Full text from HKInChip, HKLit, and Hong Kong Macau Periodicals Network, images</td>
<td>Y</td>
<td>Full text restricted to CUHK users</td>
</tr>
<tr>
<td>(<a href="http://www.lib.cuhk.edu.hk/DAO/ejournal-project">http://www.lib.cuhk.edu.hk/DAO/ejournal-project</a>)</td>
<td></td>
<td></td>
<td>in PDF format</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chinese Rare Book Digital Collection</td>
<td>2006-</td>
<td>CUHK Library collections</td>
<td>Chinese rare and semi-rare books, e-book reader providing images in Adobe Flash</td>
<td>Y</td>
<td>Public access, 10 pages of full-text</td>
</tr>
<tr>
<td>(<a href="http://chrb.lib.cuhk.edu.hk/">http://chrb.lib.cuhk.edu.hk/</a>)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hong Kong Visual Arts Archive</td>
<td>2006</td>
<td>Citation records of exhibitions and critics of Hong Kong Visual</td>
<td>Bibliographic records</td>
<td>N (Index only)</td>
<td>Public access</td>
</tr>
<tr>
<td>Arts Yearbook</td>
<td></td>
<td>Arts Yearbook</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(<a href="http://hkvaa.lib.cuhk.edu.hk">http://hkvaa.lib.cuhk.edu.hk</a>)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Digitization Projects (1995-2003 Sep) At A Glance (4)

<table>
<thead>
<tr>
<th>Projects</th>
<th>Year</th>
<th>Source</th>
<th>Digital Content</th>
<th>Full Text</th>
<th>Mode of Access</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hong Kong Macau Periodicals Network (<a href="http://hkmpn.lib.cuhk.edu.hk/">http://hkmpn.lib.cuhk.edu.hk/</a>)</td>
<td>2006-</td>
<td>Partner with Macau University Library</td>
<td>Full text from HKInChip and bibliographic records from HKInChip and Macau Periodicals Index, images in PDF formats</td>
<td>Y</td>
<td>Full text restricted to CUHK users only</td>
</tr>
<tr>
<td>Archive of Christian Study Centre on Chinese Religion and Culture (<a href="http://religionpub.lib.cuhk.edu.hk">http://religionpub.lib.cuhk.edu.hk</a>)</td>
<td>2007</td>
<td>Donation</td>
<td>Letters, correspondences, and minutes of church archives; images in PDF Format</td>
<td>Y</td>
<td>Authorized users only to view full text</td>
</tr>
<tr>
<td>Modern Chinese Literature Research Portal ( <a href="http://www.modernchineseliterature.net/">http://www.modernchineseliterature.net/</a>)</td>
<td>2008-</td>
<td>Partner with Dept of Chinese Language and Literature</td>
<td>Bibliographic records (some with full text links to HKLit) providing access to modern Chinese literature including monographs, journal articles, newspaper articles and theses; newsletter and discussion forum, book cover images and TOC characters</td>
<td>N (Index only)</td>
<td>Public access</td>
</tr>
</tbody>
</table>
# Digitization Projects (1995-2003 Sep) At A Glance (5)

<table>
<thead>
<tr>
<th>Projects</th>
<th>Year</th>
<th>Source</th>
<th>Digital Content</th>
<th>Full Text</th>
<th>Mode of access</th>
</tr>
</thead>
<tbody>
<tr>
<td>Renbai Collection</td>
<td>2008</td>
<td>ULS provide only the storage facilities</td>
<td>Cantonese opera materials including photos, libretti, operatic song scripts,</td>
<td>Y</td>
<td>Physical access only</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>program notes, posters, publications</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rulan Chao Pian Music Collection</td>
<td>2008</td>
<td>donation</td>
<td>web pages for her donation, with some photos in jpeg format and audio-clips;</td>
<td>Y</td>
<td>Public access</td>
</tr>
<tr>
<td>(<a href="http://rulanchaopian.lib.cuhk.edu.hk/">http://rulanchaopian.lib.cuhk.edu.hk/</a>)</td>
<td></td>
<td></td>
<td>open-reel tapes in CD-ROM</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ULS E-Book Initiative (<a href="http://ebook.lib.cuhk.edu.hk/">http://ebook.lib.cuhk.edu.hk/</a>)</td>
<td>2010</td>
<td>CUHK Library collections</td>
<td>Rare books, semi-rare books, Minguo (1911-1949) books and out-of-print books</td>
<td>Y</td>
<td>Library’s dedicated PCs only</td>
</tr>
</tbody>
</table>
How about following the current approach? Yes, but…
Problems with the current legacy digital collections (1)

- Different web sites with different user interface were created for different projects

<table>
<thead>
<tr>
<th>User</th>
<th>OPAC / Primo</th>
</tr>
</thead>
<tbody>
<tr>
<td>Different search interface for different in-house developed DBs</td>
<td></td>
</tr>
<tr>
<td>Tamino XML Database</td>
<td></td>
</tr>
<tr>
<td>Search result displayed separately in the web site for each collection</td>
<td></td>
</tr>
</tbody>
</table>

Rare Book DB support direct link from OPAC/Primo

Rare Book DB support direct link from OPAC/Primo
Problems with the current legacy digital collections (2)

- In 2002, Digital Access Online was launched as a web gateway to access different digitization projects.
DAO is far from the standard of a digital repository equipped with powerful search tool and interface!
Problems with the current legacy digital collections (3)

At the frontend:

- Different web interfaces for browsing and searching collections may cause confusion to users (No single theme and not under CUHK ULS branding)
- Inconsistent look and feel, and different language interface (3 projects are trilingual, 7 are bilingual and 11 in single language) - create undesirable user experience
- Most interfaces have not been revamped for a long time and updated with advanced features such as RSS, social media tools
- Unable to perform one-stop search across all collections
- Some do not have links to library OPAC and none to discovery platform
- Not all provides full text access and images, some are indexes only
Problems with the current legacy digital collections (4)

- Various level of accessibility ranging from local library computers access only to the general public, but none is open access.
  - Because of copyright issues /donor agreement / confidentiality / other reasons, and none is harvestable
  - ULS eBook system provides more than 18,000 volumes of E-books, due to copyright concern, these e-books can only be accessible at dedicated PCs in the Library.
  - The materials from Renbai Collections could not be available online according to the agreement
Problems with the current legacy digital collections (5)

At the backend:

- **Software and System Architecture**
  - Tamino XML Database offers:
    - Good support of CJK handling including:
      - Character/phrase search
      - Simplified / Traditional Chinese conversion indexing and searching
    - But it is not very popular
      - No application frontend
      - Costly to maintain
  - No standardized schema for existing dataset and the data fields varies a lot
  - None is compliant with OAI-PMH protocol for harvesting and interoperability
  - Difficult to maintain and manage each separate web site in case of any breakdown
Problems with the current legacy digital collections (6)

- Digitization Strategy & Policy
  - No digitization strategy & policy
    - Stop-Go manner – not proactive to secure funding and support to digital collections
    - No policy on what kind of materials need digitization
  - In the old days, digitization is done mainly for preservation and archival purposes – accessibility is given a lower priority
  - Not all signed agreements grant open access rights to ULS
  - Some collections do not have signed agreements
How to preserve the existing merits of the legal digital collections in the development of a new digital library system?
Merits of the current digital collections (1)

- Reach a critical mass of digitized material

<table>
<thead>
<tr>
<th></th>
<th>Printed Materials</th>
<th>Photo</th>
<th>Multimedia</th>
<th>Book Cover &amp; TOC</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Vol</td>
<td>Image</td>
<td>Image</td>
<td>Audio / minute</td>
</tr>
<tr>
<td>2005/06</td>
<td>128</td>
<td>43,386</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2006/07</td>
<td>405</td>
<td>222,235</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2007/08</td>
<td>2,676</td>
<td>560,168</td>
<td>0</td>
<td>13,576</td>
</tr>
<tr>
<td>2008/09</td>
<td>13,660</td>
<td>2,337,778</td>
<td>6,374</td>
<td>20,979</td>
</tr>
<tr>
<td>2009/10</td>
<td>1,826</td>
<td>614,113</td>
<td>2</td>
<td>21,437</td>
</tr>
<tr>
<td>2010/11</td>
<td>300</td>
<td>60,749</td>
<td>1,409</td>
<td>3,711</td>
</tr>
<tr>
<td>2011/12</td>
<td>86</td>
<td>20,359</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2012/13</td>
<td>4152</td>
<td>596,704</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Total:</td>
<td>22,847</td>
<td>4,455,492</td>
<td>7,785</td>
<td>59,703</td>
</tr>
</tbody>
</table>

Images available at in ULS Network or Standalone PCs: 3,117,504

Images available on the web: 903,612

Last updated: June, 2013

PNC Annual Conference and Joint Meetings 2013
Merits of the current digital collections (2)

- **High usage:**
  
  ULS Usage Statistics from July 1, 2012 to June 30, 2013

<table>
<thead>
<tr>
<th></th>
<th>Bamboo Database</th>
<th>HKAPI</th>
<th>HKInChip</th>
<th>HK Lit</th>
<th>HK Macau Periodicals Networks</th>
<th>HK Newsbot</th>
<th>Modern Chinese Drama Database</th>
<th>Modern Chinese Literature</th>
<th>UC Senior Seminar Database</th>
<th>Zoumalou San Guo Wu Jian Database</th>
</tr>
</thead>
<tbody>
<tr>
<td>(No. of requests)</td>
<td>52,400</td>
<td>67,699</td>
<td>596,664</td>
<td>5,114,613</td>
<td>311,583</td>
<td>91,417</td>
<td>60,461</td>
<td>515,511</td>
<td>463</td>
<td>3,516</td>
</tr>
<tr>
<td>(No. of hits selecting “Browse by year”)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>52,400</td>
<td>67,699</td>
<td>596,664</td>
<td>5,114,613</td>
<td>311,583</td>
<td>91,417</td>
<td>60,461</td>
<td>515,511</td>
<td>463</td>
<td>3,516</td>
</tr>
</tbody>
</table>

- **Good support of CJK handling including:**
  - Character/phrase search
  - Simplified / Traditional Chinese conversion indexing and searching

- **Specific metadata fields for each project**
- **Thematic design and presentation of each project – respect the wish of the collaborative parties and donors**
- **Staff already well acquainted to current design and structure of database - no learning curve for further development**
Are these approaches possible?
Yes, but
Pros:
- Standardize and map existing data into DC or other standard schema
- Can have one stop search over existing data (interface can be S. Chi, T. Chi and English
- Shorter time for implementation, flat learning curve
- Good CJK support includes word/phrase search and Inter search of Simplified/Traditional Chinese offered by Tamino database
- After data mapping, enable OAI for system interoperability
Local data characteristics can still be kept in the result page
Records linked by permanent link/identity

Cons:
- The one stop search interface may not have facet or other advanced function
- Not a total solution for standard schema for existing dataset (e.g. METS, MODS, etc)
- No modern features such as RSS, social media tools, OAI-PHM Harvester
- Need a longer time for development and implementation

PNC Annual Conference and Joint Meetings 2013
Use a OAI harvester to harvest data, make it searchable at Primo/vufind

Pros:
- Standardize and map existing data into DC or other standard schema that could be easier when migrating to other standard system
- Can have one stop search over existing data (interface can be S. Chi, T. Chi and English
- The one stop search interface have facet or other advanced function
- Can have modern features such as RSS, social media tools
- Records linked by permanent link/identity
- Good CJK support includes word/phrase search and Inter search of Simplified/Traditional Chinese

Cons:
- Localized display and result may be lost
- Need some time for implementation, flat learning curve
- Not a real standardized digitized repository software
- Not a total solution for standard schema for existing dataset (e.g. METS, MODS, etc)
Use a OAI harvester to harvest data, make it searchable at Primo/vufind

**Pros:**
- Standardize and map existing data into DC or other standard schema that could be easier when migrating to other standard system
- Can have one stop search over existing data (interface can be S. Chi, T. Chi and English
- The one stop search interface have facet or other advanced function
- Can have modern features such as RSS, social media tools
- Records linked by permanent link/identity
- Good CJK support includes word/phrase search and Inter search of Simplified/Traditional Chinese

**Cons:**
Need some time for implementation, steep learning curve
Not a real standardized digitized repository software
Not a total solution for standard schema for existing dataset (e.g. METS, MODS, etc)
Tamino XML Database

Ingest Data to market/open source professional Digital Repository Tools

Fedora Commons/VUDL/Digitool/eprints, etc

Integrated search and result retrieval in one modern repository system

Pros:
- Standardize and map existing data into DC or other standard schema that could be easier when migrating to other standard system
- Can have one stop search over existing data (but the interface may not be shown in all S. Chi, T. Chi and English
- The one stop search interface have facet or other advanced function
- Can have modern features such as RSS, social media tools, OAI-PMH Harvester
- A real implementation of standardized digitized repository software that is a total solution for standard schema for existing dataset (e.g. METS, MODS, etc)
- Web interface for staff to handle data import/export
  A longer term and sustainable solution

Cons:
- Need long time (at least 1 year) for system comparison, evaluation, implementation, steep learning curve
- Need to configure in order to have CJK support includes word/phrase search
- Cross search of simplified/traditional Chinese most probably not available
- The access control of full text materials (e.g. PDF) may not be control as free as in-house development
Interim Measures to address the existing problems

- Shorter time to develop as the interface for the existing collections need to be updated
- Retain the strength of the current approach
- For those collections that allow open access, a standard metadata schema compliant with OAI-PMH or other open access protocols should be established so as to enhance their utilization in the region and worldwide
- Buy time to implement a long-term solution
Interim Plan for A New Digital Repository

Pros:
- Standardize and map existing data into DC or other standard schema
- A one-stop search over existing data
- Take a shorter time for implementation
- Flat learning curve
- Good CJK support includes word/phrase search and Inter search of Simplified/Traditional Chinese offered by Tamino database
- After data mapping, can enable OAI-PMH for system interoperability
- Still allow search in specific fields required by some collections

Cons:
- This one-stop search interface may not have facet or other advanced functions as offered by a standard digital repository
- Not as easy as a digital repository to implement standard metadata schema (e.g. METS, MODS, etc)
- No modern features offered by a standard digital repository such as RSS, social media tools, OAI-PHM Harvester

Develop a single search interface for all digital collections

Tamino XML Database

Search result displayed separately on each DBs in their original layout
Brand New Web Presence: A new portal (1)

- An unified web interface for all digital collections to enhance user experience
- URL: http://udi.lib.cuhk.edu.hk
Brand New Web Presence: A new portal (2)

- To save time and effort to maintain so many different websites in different web folders/servers, a modern Content Management System (CMS) Drupal was selected.
Brand New Web Presence: A new portal (3)

- Drupal was chosen due to:
  - Ease to use, implement and support
  - On popular and standard LAMP platform
  - Scalable with manageable templates
  - Modularized function (e.g. Internationalization to support trilingual, Google Analytics, Apache Solr, OAI-PMH etc...)
  - Modern design and architecture
  - Open source with large user community, documentation and forums
  - Some Libraries are already using Drupal for their website and repository platform

Meet our needs!
Brand New Web Presence: A new portal (4)

Highlights:

- Migrated these collections totally into the new portal including search and display
  - Chinese Rare Book Digital Collection
  - Manuscript Collection
  - Oracle Bones
The revamped Chinese Rare Book Digital Collection is compliant with OAI-PMH protocol so as to enable interoperability of records, the collection was registered in several Open Access Repository Directories.
Brand New Web Presence: A new portal (7)

- To further promote the website and revitalize the projects, a thematic Exhibition of the launch of new Portal with Open Access was held in the Main Library.
Brand New Web Presence: A new portal (8)

- Press Release and newsletter for promotion
Brand New Web Presence: A new portal (9)

- Search Engine Optimization (SEO)
  - To optimize UDI website to appear in higher ranking in Search Engine (e.g. Google)
Brand New Web Presence: A new portal

- Again, are we done?

- No!
Brand New Web Presence: A new portal

- **Google Analytics**
  - We can analysis more about our users in terms of geographic locations
  - We can know more about our users interest, how they know our pages by referral
  - Able to identify which pages and links your visitors click the most
  - Fine tune the website to better fit their needs
Add More Collections for Open Access

- **Chinese Rare Book Digital Collection (Already launched)**
  - The digital collection has been enabled full text view instead of 10-pages preview
  - 3,000 more volumes (around 350,000 pages) of Chinese rare books from Ming Dynasty (Year 1368-1644) will be added to the digital collection in 2014.

- **Minguo e-books (Published in year 1911-1949)**
  - It is a hidden treasure collection that can only be accessible within library network before.
  - It is proposed to make public open access of all 18,000 volumes of e-books with a take-down policy if it is agreed by the senior management.
Add More Collections for Open Access

- CUHK Electronic Theses and Dissertation Collection to be available in 2014
  - Retrospective conversion of around 14,000 vols. of Masters’ theses and PhD Dissertations from 1963-2011 with more than 1 million images, plus
  - Annual additions of about 600 vols. of born-digital theses and dissertations submitted to the Graduate School

- Various manuscript collections and donations received by ULS
  - Depends on the agreement and donor’s wish
  - Digitization priority will be placed for those that could be open access
Are these sufficient enough, any limitations for now? What else to be done in future?
Limitations and Future Direction (1)

- **The New UDI Platform**
  - The new platform only migrate the static pages of the existing projects.
  - Currently only index and content of Chinese Rare Book Digital Collection, Oracle Bones and Manuscript Collection were fully migrated to the new system.
  - Index and full text content of other collections are not migrated at the moment due to technical reason and inconsistent metadata elements (such as XML, DC, etc).
Limitations and Future Direction (2)

- **Unified Digital Repository Platform**
  - In order to conduct cross search for all digital collections, implementing an unified Digital Repository Platform is inevitable.

- **System Requirements for Digital Repository**
  - Powerful CJK support
  - Well structured databases for easy management
  - Easy to set up and maintain
  - Large user group (preferably in the region) for easier support
  - Sustainability for future development
Limitations and Future Direction (3)

- Unified Digital Repository Platform
  - **Pros:**
    - A total solution for standard schema for existing dataset (e.g. METS, MODS, etc.)
    - Standardize schema allow easy migration to other repository tools in future
    - Offers one-stop search over existing data
    - Provides user-friendly functions such as facets
    - Social networking capability such as RSS, social media tools
    - OAI-PMH compliant for discovery
    - Web interface for staff to handle data import/export
    - More sustainable for future development
  - **Cons:**
    - Takes longer time (at least 1 year) for system comparison, evaluation, implementation
    - Steep learning curve for technical staff
    - Need to configure to support CJK word / phrase search
Limitations and Future Direction (4)

Examples:
• Fedora Commons: http://lcdl.library.cofc.edu/
• VuDL: http://digital.library.villanova.edu/

PNC Annual Conference and Joint Meetings 2013
Limitations and Future Direction (5)

- Integration with other systems that are under preparation in the ULS
  - ArchiveSpace
    - It would be beneficial to both systems as it can provide cataloguing and finding aids for digital objects
  - Digital Humanities Tools (Planning underway)
    - In order to extend the sustainability, Digital Humanities tools and specialist will be sought to analyze our digital collections and ULS’s role as well as add value to our collections by GIS data, linked data model, etc
Limitations and Future Direction (6)

- Establish ULS Digitization Strategy and Policy
  - Should we digitize materials that do not allow open access?
  - The Digitization is for preservation, access or for both?
  - What are the strategies for collaborative projects with internal departments/colleges and external parties?
  - Should indexes be kept?
  - What is the appropriate metadata schema for ULS digital collection? Dublin Core? Not all fields in DC meet all our needs.
  - What is the priority for collections to be digitized?
Lessons Learnt (1)

- **Change Management**
  - Not easy to make a tiny change on the structure in a large digital collection
- **Personnel**
  - The core members of the digitization team have been changed
  - Some new viewpoints and ideas were brought into consideration during the planning process
  - Green light was given from senior management in terms of decision and funding
- **Technical**
  - Existing users and staff may be reluctant to change
  - The change of the new UDI platform paves the way for further changes
  - The team need to learn from scratch the new platform and technology

Source: http://www.novosolutions.com/
Lessons Learnt (2)

- More planning and analysis are required to conduct before the implementation especially working with stakeholders other than the team
The Team Still Have A Long Way To Go!
The End
Thank You!

jeffliu@lib.cuhk.edu.hk
louisalam@lib.cuhk.edu.hk

Please visit us at:
http://udi.lib.cuhk.edu.hk