University Archives in a Web 2.0 World

Schubert Foo

Division of Information Studies
Wee Kim Wee School of Communication & Information

Email: sfoo@pmail.ntu.edu.sg
URL: http://www.ntu.edu.sg/home/assfoo/
Outline

- Sources of University Information
- Information Sources Profile of 51 Korean University Archives
- Library 2.0 applications (sources) in Academic Libraries
- Adoption of specific Web 2.0 Applications in Libraries
- Value of University Information Sources
- Value Proposition for University Archives
- Current Situation
- Goals of University Archives
- Components of University Archives
- University Archives Procedures and Processes
- Some Key Considerations for Success
- Web Archiving
- Lessons Learnt/Discussions/Recommendations
- Conclusions
Sources of University Information

University Administration
- Account resources
- Meeting records
- Official Documents

Education and Research
- Curricula
- Syllabi
- Report cards (Assessment records)
- Thesis/Dissertations
- Research reports

University History
- University publications
- White papers
- Evaluation reports
University History (cont)

University directories
Journals
Photos
Images
Souvenirs
Audiovisual materials
Resources related to retired professors
Resources related to visitors
University maps
Architecture materials
List of graduates

Alumni
Materials related to alumni (chapters, directories, publications, etc.)
Others (Not necessarily less important)
Donations (Donor collections)
Community resources
Others.. (can you think of others that is missing?)
Information Sources Profile of 51 Korean University Archives (Kim & Lee, 2009)

A. University administration
   - A-1 Accounts resources
   - A-2 Meeting records
   - A-3 Official documents

B. Education & research
   - B-1 Curricula
   - B-2 Syllabi
   - B-3 Report cards
   - B-4 Theses
   - B-5 Research reports

C. University history
   - C-1 University publications
   - C-2 White papers
   - C-3 Evaluation reports
   - C-4 University directories
   - C-5 Journals
   - C-6 Photos
   - C-7 Images
   - C-8 Souvenirs
   - C-9 Audiovisual materials
   - C-10 Resources related to retired professors
   - C-11 Resources related to visitors
   - C-12 University maps
   - C-13 Architecture materials
   - C-14 Lists of graduates

D. Alumni
   - D-1 Materials related to alumni
   - D-2 University websites
   - D-3 Donations
   - D-4 Community resources
   - D-5 Other miscellanea

E. Etc.
### From Library 1.0 to Library 2.0: some applications

<table>
<thead>
<tr>
<th>Library 1.0</th>
<th>Library 2.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>References with traditional means</td>
<td>References with Blogs, IM,</td>
</tr>
<tr>
<td></td>
<td>RSS, Tagging, Wikis</td>
</tr>
<tr>
<td>Cataloging</td>
<td>Tagging in OPACs</td>
</tr>
<tr>
<td>Online communities via mailing lists</td>
<td>Online Communities via Social Networks</td>
</tr>
<tr>
<td>Text-based tutorials</td>
<td>Podcast-based Tutorials</td>
</tr>
</tbody>
</table>

**Library 2.0: Four Features.**

- User originated
- Socially rich
- Multimedia enabled
- Communally innovative
The Academic Library 2.0 Model.

Information 2.0
- Open
- Convergent

Web 2.0
- Roles
- Qualifications

Librarian 2.0
- Interactive
- Collaborative

User 2.0
- N-way information flow
- Timely
- Nonlinear

Participatory

(Xu, Ouyang & Chu, 2009)
Adoption of Specific Web 2.0 Applications in Libraries (Xu, Ouyang & Chu, 2009)

Based on review of 81 academic libraries in US.
**Instant Messaging (IM)** – implemented for reference service to replace/complement email and telephone inquiries. Also include librarian-user, user-user.

**Blogs** – librarians’ blogs (mainly news bulletin). Can be combined with RSS to filter right information to users.

**RSS** – feeds up-to-date news or current information to user via their defined preferences.

**Tagging** – allows users own indexing/bookmarking (keyword indexing by users)
Wikis – supports participatory librarianship. Can be combined with blogs to support a learning platform for students.

Social Networks – Facebook, Myspace, etc, to create and support virtual communities within the library environment.

Podcasts – recordings of library related tutorials, lecture materials.

So what do we do with all these information generated from various university sources, including those generated in Library 2.0? Is there value in them for us to invest in managing them?
Value of University Information Sources

- Cultural and educational heritage
- Historical/Key records
- Intellectual property (not necessarily restricted to academic community but depends on how ‘open’ are the sources to public)
- Vital and dynamic civic engagement to society at large to make difference to social justice and equality through access of resources
Examples of dynamic civic engagement to society

A learning commons in public libraries contributed by public/academic librarians, library reference/subject guides, library research publications – building sustainable information resources around important events (elections, public education, economic developments))

→ Need to manage them for long term storage, archive and access
Value Proposition for Having a University Archive
(Adapted from Shiozaki & Eisenschitz, 2009)

1. Comprehensive preservation of these sources (Preservation is a public good)
2. Making collected materials available to the library members (and public)
3. Continuity of access to cited materials
4. Evidence showing certain contents once existed
5. Selective preservation of valuable resources
6. Creation of new knowledge (but require advanced access functions)
7. Encouragement of university pride
Current Situation

“In fact, many within the BL (British Library) are starting to consider that the print version is a surrogate for the digital original, rather than the other way round”

Steve Green, The British Library, Boston Spa, UK

Digital versus Print?

Research still shows that it is important to keep both the original and digitized or processed copy – to serve the needs of different user groups.

“Digital transcriptions .. hold little interest for researchers because the information in mediated and unreliable” (Maxwell (a historian), 2010)
More and more publishers are digitizing their back catalogues and making content available online, for either free or fee paying access.

Common scenario – Different parts of the university (or organization) go about creating their own digital repositories in the absence of an organization-wide digital library (archive) plan. This fragmentation results in ‘technology islands’ being created.
Stakeholders involved with managing digital repositories or collections are subject to different organization structures with no clear defined roles across structures. Problems in maintenance, data redundancy and integrity, incompatible formats and configurations, preventing interoperability and exchange of data effectively.

Even with a university archive (e.g. Korea), many lack a legislative system of university archives to ensure that proper deposit of information is carried out by the affected information generating entities in the university.
Goals of University Archives

1. Receive, process, store and preserve any defined type of digital material in perpetuity.
2. Provide **access** to these materials with proper permissions,
3. Ensure that materials are **easy to find** and access.
4. Ensure **authenticity** of the materials.
5. Ensure users can **view** the materials with contemporary applications
6. Ensure that users can, where possible, **experience** the material with original look-and-feel (to maintain the same experience as before)

(Green, 2009)
Components of a University Archive

Software components (through acquiring, subcontracting or developing the system)
Hardware infrastructure (buying, installing and maintaining the servers, networks, links etc. to create a scalable robust environment)

Defining procedures and process (to manage the digital content and ensure staff are proficiently trained)
Work with others (libraries, international institutions, associations), employ international standards (such as preservation standards and processes).

Collectively, promote the central importance of libraries and archives in safeguarding our digital heritage.

(Green, 2009)
University Archive Procedures and Processes
(Adapted from Kim & Lee, 2009)

Foundation principles
• Presence of internal principles as foundation for university archives; Observation of national archiving guidelines
• Collection in various forms and formats – PDFs, complete website, blogs, podcasts, videos, images, etc.)
Functions

1. **Identity** – collecting, managing and preserving resources pertaining to university history
2. **Education** – usage of archival materials for teaching, classroom lectures
3. **Research** – usage of archival materials for research by faculty, staff, students and external communities.
4. **Administration** – records management of university
5. **Services** – provision of archival materials online and offline.
Process Guidelines
Guidelines covering collection, preservation, arrangement, use and disposition. This includes classification schemes for the archived material.

Facilities
• Staff numbers (including archivist)
• Status of archives (where located, management and control – e.g. library, museum, administration building) and work areas (e.g. for physical preservation), space allocated for archives
Acquisitions and Management
Process for systematic acquisition, collection development.

Access and Services
Public access to records, website management, electronic records management program, provision of search tools, user groups, resources frequently used, user education program, co-operation with other libraries, external institutions (such as associations dealing with records management, archives)
Some Key Considerations for Success (CSF)

**Metadata** (Description and Knowledge Discovery issue)

Need good metadata to describe the digital resources to support searching, knowledge discovery and access, maintenance, etc.

Adopt standard metadata (e.g. Dublin Core, EAD…) as much as possible taking note of the trade-off between the number of elements and the cost to populate them, and ensuring they are endowed with entities to support future trends (e.g. Semantic web).
Person and Corporate Schema Application Profile
(Courtesy of National Library Board, Singapore)

Introduction
The Person and Corporate Schema employs 16 elements. With the exception of the following DC Elements: Identifier, DateBirth, DateDeath, DateStart, DateEnd, and Nationality, all are localised to NLB.

Format of entries:
- Property: A unique token assigned to the term
- Resource (URI): A Uniform Resource Identifier used to identify the term.
- Defined By: An identifier of a namespace, pointer to a schema, or bibliographic reference for a document within which the term is defined.
- Source Definition: The definition of the term in the namespace in which the term was originated.
- Source Comments: Comments on the term from the namespace in which the term originated. Includes authoritative documentation related to the term.
- Local (NLB) Comments: NLB comments about the term.
- Type of Term: A grammatical category of the term (e.g., “Element”, “Element Refinement”, or “Encoding Scheme”).
- Has Encoding Scheme: The described term is qualified by the referenced Encoding Scheme.
- Obligation: Indicates whether the element is required to always or sometimes be present (i.e., contain a value).
- Property value Type: Indicates the type of data that can be present in the value of the element.
- Occurrence: Indicates any limit to the repetitiveness of the element.

Elements

Encoding Schemes (DC namespace):
- Person
- Personal
- PersonalAlternate
- Corporate
- CorporateAlternate
- Identifier
- CorporateMember
- DateBirth
- DateDeath
- DateStart
- DateEnd
- Nationality
- Award
- Biography
- Works
- Affiliations
- Immac

Encoding Schemes (NLB namespace):
- Person
- Classification

Table of Contents

1. Name of Term | Personal
Term URI | http://mims.libraryhub.com/nlb/terms/nlbperson
Label | Personal
Defined By | 
Source Definition | 
DC-Lib Definition | 
Local (NLB) Comments | Name of the person in an internationally recognized format
Definition | 
Source Comments | 
DC-Lib Comments | 
Local (NLB) Comments | 
Type of Term | 
Refined By | 
Has Encoding Scheme | 
Obligation | mandatory if applicable
Datatype | 
Occurrence | not repeatable
### Subject
Audience
Educational Context
Educational Level
Skill Level
Age Range
Literacy Level
Language of Intended User
Prerequisites
Learning Preference
Learning Goals
Achievement Standards
Course
Instructional Method
Level of Difficulty
Skills being Practiced
Typical Learning Time
Origin
Designed for reuse

---

<table>
<thead>
<tr>
<th>Domain Requirement</th>
<th>Existing Properties</th>
<th>Definition</th>
<th>Property Range (V2/V3)</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1: Subject</td>
<td>dcterms:subject</td>
<td>The topic of the resource.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>lom:keyword (1,5)</td>
<td>A keyword or phrase describing the topic of this learning object.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Standards at the Library of Congress (http://www.loc.gov/standards/)

MARC | EAD | METS | PREMIS etc.

**Resource Description Formats**
- MARC 21 formats - Representation and communication of descriptive metadata about information items
- MARCXML - MARC 21 data in an XML structure
- MODS (Metadata Object Description Standard) - XML markup for selected metadata from existing MARC 21 records as well as original resource description
- MADS (Metadata Authority Description Standard) - XML markup for selected authority data from MARC21 records as well as original authority data
- EAD (Encoded Archival Description) - XML markup designed for encoding finding aids

**Digital Library Standards**
- METS (Metadata Encoding & Transmission Standard) - Structure for encoding descriptive, administrative, and structural metadata
- MIX (NISO Metadata for Images in XML) - XML schema for encoding technical data elements required to manage digital image collections
- PREMIS (Preservation Metadata) - A data dictionary and supporting XML schemas for core preservation metadata needed to support the long-term preservation of digital materials
- TextMD (Technical Metadata for Text) - XML schema that details technical metadata for text-based digital objects.
- ISO/DIS 26577 - Information and documentation - MarcXchange.
- ISO 20775 - Schema for Holdings Information.
- ALTO - Technical Metadata for Optical Character Recognition

**Information Resource Retrieval Protocols**
- Z39.50 - Supports information retrieval among different information systems

**Current Highlights**
- TextMD schema now maintained by the Library
- SRU and Open Search Initiative begun
- PREMIS report on implementation approaches issued
- MIX Version 2.0 issued
- Resource papers on URIs released
Crosswalk: MARC to DC – to support mapping between metadata systems
(Courtesy of National Library Board, Singapore)

<table>
<thead>
<tr>
<th>DC Element</th>
<th>DC Qualifier</th>
<th>Encoding Scheme</th>
<th>MARC Tag</th>
<th>IND 1</th>
<th>IND 2</th>
<th>Subfields to apply</th>
<th>MARC Tag Name</th>
<th>Remarks on Changes made</th>
<th>Instruction for Vendor</th>
<th>Instruction to note for stylesheet</th>
<th>Updated DC XML Snippet</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title</td>
<td></td>
<td></td>
<td>245</td>
<td>0/1</td>
<td></td>
<td>0/1</td>
<td>Title</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alternate</td>
<td></td>
<td></td>
<td>246</td>
<td>0/3</td>
<td></td>
<td>#/0-3</td>
<td>$a $b $c</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Exception 2 - multiple marc tags mapped to a single dc field should be crosswalked as individual line entries, the dc field can be repeatable viz; all marc tags mapped to dc_title would have individual line entries after being crosswalked.

Exception 4 - multiple subfields in a single marc tag which is mapped to a single dc field should be crosswalked as a single line entry, with the exception of $6 and $8 which is to be crosswalked as a corporate line entry for the dc field it is mapped to.

$6 - Linkage is used to link field data to other fields. $6 is for data linking fields that are different script representations of each other. Vendor to ensure $6 is to be interpreted as linkage and not as a value. To apply Instruction to Vendor (4) under Chinese & Tamil records tab.

$8 - Field link and sequence number. Identifies linked fields and may also propose a sequence for the linked fields. The structure and syntax are $8 [linking number] [sequence number] [field link type]. $8 is to be interpreted as linkage and not as a value.

In the absence of punctuation between subfields, to insert full stop followed by space.

Concatenate each subfield value in the order they appear in the marc tag.
Note - do not concatenate subfields alphabetically.
Some Key Considerations for Success (cont)

Planning and Managing Costs (Cost management issue)

Selecting, seeking permissions, quality assurance process, cataloguing, storage and maintenance, processes for long term preservation, staff costs, salaries, daily operational costs, etc.)
Some Key Considerations for Success (cont)

Seeking and Growing Collection to Maximize Its Value (Collection issue)

Need to constantly identify, gather and create content and not waiting it for it to be deposited in the archive.
Some Key Considerations for Success (cont)

Collaboration with others (Sharing and learning issue)

Consortia of collaborating institutions, sharing one system with other university archives and institutions) to allow expansion of scope of collection and services through cost-sharing and optimization.
Web Archiving

Means to counter the changing and fragile characteristics of the Web by archiving websites to ensure access by future generations on the accumulated cultural, education and institutional heritage of the organization.

Web archiving currently done by national libraries around the world, Internet Archives, and others.

Some national libraries have passed legislation and/or regulations for collecting web content, others have archiving based on contracts with the institutions.
IIPC (International Internet Preservation Consortium)  [http://www.netpreserve.org/](http://www.netpreserve.org/)

Aims to acquire, preserve and make accessible knowledge and information from the Internet for future generations everywhere, promoting global exchange and international relations.

Members include the national libraries of Australia, Canada, Denmark, Finland, France, Iceland, Italy, Norway, Sweden, The British Library (UK), The Library of Congress (USA) and the Internet Archive (USA).
Early Members

Internet Archives (1996)
PANDORA Australia’s Web Archive (1996)
The UK Government Web Archive (1997)
New Zealand Web Archive (1999)
OASIS (Korea – 2005)
Web Archive Singapore (ard 2006)

(Full list available from http://www.netpreserve.org/about/archiveList.php)
Internet Archives (http://www.archive.org/index.php)

The Internet Archive continues to build a digital library of Internet sites and other cultural artifacts in digital form. Like a paper library, it provides free access to researchers, historians, scholars, and the general public through the use of the Wayback Machine and other means to archived sites.
Wayback Machine
Browse through over 150 billion web pages archived from 1996 to a few months ago. To start surfing the Wayback, type in the web address of a site or page where you would like to start, and press enter. Then select from the archived dates available. The resulting pages point to other archived pages at as close a date as possible. Keyword searching is not currently supported.

City University of Hong Kong
Run Run Shaw Library
http://www.cityu.edu.hk/lib/
### Search Results for Jan 01, 1996 - Oct 15, 2009

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: some duplicates are not shown. See FAQ.

* Denotes when site was updated.

Material typically becomes available here 6 months after collection. See FAQ.
February 7, 2003
June 24, 2008 (last available)
Now (April 2010) but not available in Internet Archives
Web Archive by Specific Events/Themes

Example: Asian Tsunami (http://tsunami.archive.org/)
Archived by NTU SCI – The Singapore Internet Research Centre
Archive-IT – Subscription Based Web Archiving

Allows institutions to build and preserve their own web archive of born digital content, through a user friendly web application, without requiring any technical expertise or hosting facilities. Subscribers can harvest, catalog, and archive their collections, and then search and browse the collections when complete. Collections are hosted at the Internet Archive data center, and accessible to the public with full text search. Current it contains over 125 partners including state archives, university libraries, federal institutions, state libraries, non government non profits, museums, historians, and independent researchers.
Lessons Learnt & Recommendations

1. Consider setting up a proper University Archives to achieve the various benefits.

2. Learn from others – adopt best practices. Form alliances with them whenever possible.

3. Consider legislation – or set procedures for compulsory deposit (ie. Legal deposit)
4. Join IIPC – allow Internet Archives/Archive-IT/National Libraries to do archiving of your website. Preservation is at the centre in the culture of such institutions. Among these options, national libraries have the best long term survival potential as it is a national institution supported by government.

5. Deposit key materials with the national e-legal deposit. As noted, there is longevity in such national institutions. Redundancy is fine.
Conclusions

1. Potential for the university library plus other co-operating departments (IT, museum, administration) to design and implement a University Archive. The value proposition of such an archive is sound and appealing. In reality, it is not so easy to convince university top management due to costs and budget challenges.

2. Long term preservation of archival university resources is difficult and costly but important and essential.
3. Resources will be a key impediment to implement a good sustainable university archive. Some form of university legislation can help to ensure the archive’s viability and long term sustainability and well as ensure the quality of the archive in the long term.
References


-- Thank you --

For more information, please contact

Schubert Foo
sfoo@pmail.ntu.edu.sg