DIGITAL LIBRARY
F E D E R A T I O N

The University of Chicago Library

Report to the Digital Library Federation
Spring 2004

Table of Contents

I. Collections, services, and systems

A. Collections

American Environmental Photographs, 1891-1936

A project to digitize more than 4,500 glass lantern slides, glass negatives, and photographic prints created by faculty members and students of the University of Chicago Department of Botany.

These images document natural environments, ecologies, and plant communities in the United States at the end of the nineteenth and the beginning of the twentieth century.

Image capture at 600dpi, 8-bit grayscale of black and white photographic images; uncompressed TIFF files saved from Adobe Photoshop were stored on CD and a tape backup system. JPEGs and GIF thumbnails for each image were derived for viewing on web browsers. Descriptive and technical metadata about each image were entered into a Microsoft Access database. Descriptive metadata were exported as an ASCII file and marked up in SGML according to a custom DTD. This file was sent to LC for processing into the American Memory database. Searching and navigation of this collection is provided by LC's American Memory interface. Work on this project was performed in house.

http://www.lib.uchicago.edu/h/aep

Ancient Near East and the Mediterranean World

Project to digitize approximately 36 volumes from the Library's Ancient Near East and Classics collections.

This digital project was part of a larger project that preserved almost 9,000 volumes in this subject area. Other preservation options in the project included microfilming and physical treatment. In this demonstration portion of the project, the Library explored methods of digitizing and providing access to bound volumes that included both text and illustrations. All digitization work on this project was performed in house. EOS (Electronic Open Stacks) is the Library's home for page-turned image-based texts created for the project. It was primarily built to house the Ancient Near East and Classics titles and includes a small number of other page-turned items. This is no longer seen as the larger home for the Library's digitized books and is not being actively developed. One can navigate through the texts by way of a digital table of contents or by locating a particular page from the book's indices or list of plates and entering the number into the page navigator. EOS books are not full-text searchable.

Pages captured at 300 dpi, 8-bit grayscale; LZW compressed TIFF files were saved from Adobe Photoshop and stored on CD. High and low resolution JPEG files were derived for display on web browsers. Access is provided via digital book software created by The ARTFL Project using a database keyed in by library staff.

http://www.lib.uchicago.edu/e/dl/proj/neh2/

Chopin Early Editions

Digitized versions of all the scores in the University of Chicago Library's Chopin collection. Users can search or browse the Chopin Early Editions collection via a variety of data points, including titles, genres, plate numbers, publishers and place of publication.

Access to individual scores in the digital collection is also available through records in the Library's catalog with links to this Web site. The scanned images of all pages of the scores provide scholars, bibliographers, collectors, performers, and students with new ways to study this collection.

Page images were scanned at 400 dpi, 24-bit color with ruler and color bar. Because this is a Special Collection project, all page edges are displayed and scores were scanned cover to cover (with the exception of library bindings). Uncompressed TIFF images were saved from Adobe Photoshop and stored on CD as well as in a tape backup system. Low and high resolution JPEG files were produced for access via web browsers. Future plans include providing a DjVu file for each score. Intellectual access is supplied from MARC records in the local

catalog and WorldCat as well as through the Greenstone software interface accessible on the Internet.

http://chopin.lib.uchicago.edu/

Electronic Full Text Sources (EFTS)

EFTS provides the primary means of access to, as well as information on, full-text scholarly resources available at the University of Chicago.

Texts are arranged by language, subject, and searching interface.

Nearly all of the Library's searchable, encoded full-text resources are mounted under PhiloLogic, a full-text searching system which The ARTFL Project and the Library have been developing.

http://www.lib.uchicago.edu/h/efts

Electronic Open Stacks (EOS)

EOS is the Library's home for page-turned image-based texts.

One can navigate through these texts by way of the digital table of contents or by locating a particular page from the book's indices or list of plates and entering the number into the page navigator. EOS books are not full-text searchable.

Books are scanned and saved as TIFF files. Access copies are saved as JPEG files and structural metadata are used to provide navigation within each book.

http://www.lib.uchicago.edu/h/eos

Italian Women Writers (IWW)

The Italian Women Writers project (IWW) is a long-term research endeavor to preserve and provide access to an extensive corpus of literature written by Italian women authors.

Its goal is to bring information on and texts by both famous and previously neglected Italian women writers to a wider audience of students, scholars, teachers, and the general public, and to preserve these often fragile texts for generations to come.

IWW combines the use of MySQL for searching metadata and PhiloLogic for full-text searching.

http://www.lib.uchicago.edu/efts/IWW/

The First American West: The Ohio River Valley, 1750-1820

This was a joint project between the University of Chicago Library and the Filson Historical Society of Louisville, Kentucky, to digitize 745 rare books, pamphlets, newspapers, maps, prints, and manuscripts presenting a wide-ranging historical overview of the Ohio River Valley and trans-Appalachian West from the earliest Euro-American settlement to the passing of the frontier beyond the Mississippi River.

University of Chicago Library components of the digital collection are drawn from the rare book and manuscript holdings of the Special Collections Research Center and the general collections of the Joseph Regenstein Library, John Crerar Library, and D'Angelo Law Library.

Materials were scanned in house at each institution with each institution taking responsibility for its master files. Capture specifications were either 400 dpi, 8-bit grayscale or 400 dpi, 24-bit color as appropriate for the original materials. Reference JPEGs and thumbnails were created for browser access. Intellectual access is provided through the American Memory interface using descriptive metadata supplied by the institutions. Descriptive metadata for all items was submitted as a Microsoft Access database; LC was able to do its own extraction once we had mapped our elements to Dublin Core. Approximately 50% of the items have searchable text behind the page images. The transcription and encoding were vended out and produced according to the American Memory DTD. For some portion of the book and journal items MARC records also exist and have been loaded into local catalogs or WorldCat. The items are accessible through the American Memory interface. The transcription of manuscript letters was performed in house. All other transcriptions and all encoding was vended out. All other work was performed in house.

http://www.lib.uchicago.edu/h/faw

The Smart Museum of Art Online Collection

The Smart Museum of Art Online Collection catalog was developed by The ARTFL Project and is being maintained by the Library.

Catalog data are loaded into a MySQL relational database and made searchable on the web using Perl.

http://www.lib.uchicago.edu/efts/smart/

The University of Chicago Centennial Catalogues

The University of Chicago Centennial Exhibition Catalogues presents the text and images from four catalogues published in 1991-92 to accompany a series of exhibitions celebrating the centennial of the University of Chicago.

Drawing on documents, photographs, and other materials in the University of Chicago Archives, the exhibitions examined the University's history through the perspectives of its faculty, student experience, relationships between the University and the City of Chicago, and the University's ten chief executives. The contents of the original catalogues have been preserved but adapted to the textual and graphic features of online presentation. Work on this project was performed in house.

The catalogs were scanned at 400 dpi, 8-bit grayscale and saved as TIFF files and OCR'd using OmniPagePro. Derivative files were made for the images within the catalog pages and the final web catalog was produced from a combination of the OCR'd text and individual images from within the catalog.

http://www.lib.uchicago.edu/staffweb/depts/dldc/projects/centcat/centcats/home.html

B. Services

Society for the Study of Early China Web Site

A new web site for the Society for the Study of Early China which will act as a high-quality scholarly resource for the study of Early China.

The web site will house electronic versions of the print publications of the SSEC, as well as research published exclusively on the web site. The original SSEC web site moved from place to place as editorship of the society's journal changed. This newly-designed web site will find a permanent home at the University of Chicago, thus allowing the development of a more stable and sophisticated research site. Faculty and graduate students from the Department of East Asian Languages and Literature maintain the content on the site in conjunction with the current editor of the society's publications and scholars from around the world.

The web site was designed by Library staff and is housed on Library server computers. The navigational elements of the web site are generated using eauth, a locally developed web management system written in TCL.

http://www.lib.uchicago.edu/earlychina/

C. Systems

Archives and Manuscripts Finding Aids

This system provides web-accessible electronic finding aids for the Library's archives and manuscript materials.

New finding aids are created electronically and paper finding aids are converted into electronic format.

EAD-encoded finding aids are marked up in XML using locally written software, one for the frontmatter and one for the inventory, then uploaded into the Tamino XML Server from Software AG for search and retrieval. Web access is provided via Java servlets and JSP, with server-side transformations of XML to HTML using XSLT.

http://ead.lib.uchicago.edu/

Macroeconomic Time Series Data Source Locator

This resource provides a list of macroeconomic data sources available at the University of Chicago categorized by year, geographic location and subject.

Types of sources include books, journals, yearbooks, CD-ROMs and Internet resources. For each source we determined the oldest date listed for a particular data category (balance of payments, bond market related, etc.).

Bibliographic data are entered in a MySQL relational database and made searchable on the web using PHP.

http://www.lib.uchicago.edu/e/busecon/macroecon/

Mamluk Bibliography Online

The Mamluk Bibliography is an on-going project of the Middle East Documentation Center at the University of Chicago, the aim of which is to compile comprehensive bibliographies of all primary sources relating to the Mamluk sultanate of Egypt and Syria, as well as all research and discussion-- scholarly and popular--germane to the subject.

The project takes the form of two bibliographies: the primary and the secondary. The bibliography is maintained in a ProCite database which is exported as a flat file. The data are made available on the web using NAND: A Non-Relational Database, which provides a search and browse interface to the data.

http://www.lib.uchicago.edu/e/su/mideast/mamluk/

II. Projects and programs

A. Projects

Digital South Asia Library

This pilot project was intended to improve access through development of the infrastructure for inter-continental electronic document delivery to and from selected South Asia libraries and creation of new electronic reference resources, comprised primarily of data entered by project participants in South Asia.

Material for retrospectively digitization is being drawn from four other projects of the Southern Asia Department, specifically: the Roja Muthiah Research Library, the collection of the Urdu Research Library Consortium, the Official Publications of India collection on deposit at the University of Chicago, and publications addressed under the South Asia Reference Tools Program. The project has received subsequent funding.

http://dsal.uchicago.edu/

Kanji Alive

Kanji alive is an interactive web-based tool which allows students to study and look up Kanji by their meaning, pronunciation, stroke order, or radical.

Students can watch a Quick Time Movie that displays the correct stroke order for each Kanji in handwritten style and see examples of frequently used compounds in which the Kanji appear.

Kanji alive is cross-platform and does not require any Japanese fonts. It was developed by an independent team at the University of Chicago with support from the Library.

http://kanjialive.lib.uchicago.edu/

eCUIP Digital Library Project

An ongoing collaboration with the Chicago Public Schools/University of Chicago Internet Project (CUIP) to build a K-12 digital library of materials for use by both teachers and students in the Chicago Public School System.

The web site is housed on the Library's web server and is maintained using Dreamweaver templates.

http://www.lib.uchicago.edu/ecuip/

B. Programs

Digital Reformatting Program

As part of its broader mission to preserve and provide long-term access to Library's collections, the Preservation Department is engaged in an ongoing program to digitize Library materials. The focus of the Digital Reformatting Program is on materials from the general and special collections throughout the Library. Individual items, groups of items, or larger collections for inclusion in the Program are materials that may be fragile or so deteriorated that they are at risk of being damaged with handling or use. Materials are identified by bibliographers reviewing collections, by staff as items circulate, and by faculty for research and teaching. The Preservation Department is responsible for managing work in house in an equipped and staffed Digital Reformatting Unit and for developing and monitoring contracts with external vendors. Digitally reformatted titles and collections are freely available through the Library's online catalog, through the OCLC database, and through individually designed sites on the Library's Web. The Digital Reformatting Program is part of a broader digital library initiative at the Library in support of research and instruction at the University. The Preservation Department works collaboratively with other partners in the Library, including the Digital Library Development Center and the Special Collections Research Center to accomplish this work.

Electronic Reserves

Program to convert reserve material to electronic format for delivery over the web.

Materials include journal articles, book chapters, problem sets, and sample exams.

Materials are scanned and converted into PDF using Adobe Acrobat.

http://www.lib.uchicago.edu/h/2

III. Specific digital library challenges

Digital Archiving

As custodians of cultural heritage, libraries have served the role of repositories of traditional research resources and have been entrusted with their accessibility and long-term care. Extending from traditional resources to also include resources in electronic form, and assuming responsibility for creating, converting, and acquiring such resources, libraries have de facto also assumed the responsibility for providing access to and ensuring maintenance of these resources over time.

An important current focus of the University of Chicago Library is on long-term maintenance of digital resources. In this context, we define digital archiving as the managed activities necessary for ensuring both the long-term maintenance of a bit stream and continued accessibility of its contents.

The inherent fragility of digital materials requires more attention, often much sooner than resources on paper. In addition to descriptive metadata, other kinds of metadata, including preservation, technical, administrative, structural, and rights metadata, are required for the maintenance and long-term usability of electronic resources. Decisions regarding preservation metadata and the creation thereof must take place up front, as early on in the process as possible, but standards for preservation metadata are still evolving, and the relation between digital preservation and property rights is not yet clear. However, understanding, documenting, and adhering to relevant standards is crucial for future interoperability.

Digital archiving is expensive, and it is not yet clear how it will be funded, which technologies will prove to be dominant, cost-effective, or most suitable for leveraging in support of other digital initiatives being undertaken. Yet there is pressure to archive digital materials today despite these uncertainties. As a short-term strategy to relieve some of this pressure, the University of Chicago Library is adopting GNU EPrints as a low-cost solution for allowing the deposit of some digital materials into a central location and providing information about them via the Open Archives Initiative (OAI). Though the software does not support all digital archiving functions, it is mature, and is being viewed as a way of gaining some experience in what is required from a long-term solution, both in terms of functional requirements, and also in terms of how those requirements might be met cost-effectively.

IV. Digital library publications, policies, working papers, and other documents

Building Collections with Greenstone Digital Library Software. Tod A. Olson, Programmer/Analyst Digital Library Development Center, University of Chicago Library

NAND: A New Tool for an Old Problem. Charles Blair, Elisabeth Long, and Keith Waclena, The Digital Library Development Center, The University of Chicago Library