



CARNEGIE MELLON UNIVERSITY

REPORT TO THE DIGITAL LIBRARY FEDERATION FALL 2003

Table of Contents

- I. Collections, services, and systems
- II. Projects and programs
- III. Specific digital library challenges
- IV. Digital library publications, policies, working papers, and other documents

I. Collections, services, and systems

A. Collections

[Charette Digital Project](#)

The *Charette* Digital Project provides web-based public access to a regional architectural journal. *Charette* (vols. 1-54, 1920-1974) was the journal of the Pittsburgh Architectural Club; and was later co-sponsored by the Pittsburgh Chapter of the American Institute of Architects and the Pennsylvania Society of Architects. Over the years, its coverage of architecture extended to the whole of Pennsylvania, eastern Ohio, and West Virginia. Carnegie Mellon University Libraries holds an extensive run of *Charette*, and borrowed missing issues from other institutions for scanning. Copyright status was investigated and permissions were acquired as needed. The *Charette* Digital Project created digital images of each page of *Charette* including text, photographs, graphics, advertisements, etc. To date, a web site has been created and a test version of the browsing interface is operative; a full-text searching interface is in development. The *Charette* Digital Project is a project of the Carnegie Mellon University Architecture Archives. [contact: Martin Aurand, mal_f@andrew.cmu.edu]

- Aurand, Martin. [Charette Digital Project: An Historic Pittsburgh Architectural Journal Goes Online](#). Presented at the Association of Architecture School Librarians Conference, Louisville, KY, March 2003.

Carnegie Mellon Technical Reports (2002-)

The University Libraries plan to digitize all of the Carnegie Mellon Technical Report series. Permission to digitize the reports will be acquired from the copyright holders. To date, Robotics Institute and Mathematics technical reports have been digitized (nearly 700 reports). The digitized reports are available via links in the Carnegie Mellon library catalog and WorldCat, and in the Million Book Project collection. [contact: Ann Marie Mesco, mesco@andrew.cmu.edu]

Costume Books Project (planning phase)

For the last decade, the library has housed a selection of 157 costume books in non-circulating reserves at faculty request. There is extraordinary contention for these books, which are critical to students preparing for class assignments and productions. The library is seeking funding to make these items accessible via the Internet through a non-destructive digital scanning process. Scanning will eliminate the problem of contention and provide access to the images and text of the books 24/7. Scanning will also extend the life of those items that have become worn or have loose plates. Books whose copyright has expired will be made available on the web for a worldwide audience, while attempts will be made to obtain permissions to make copyrighted material more widely available - otherwise, Internet Protocol authentication will be used so that only Carnegie Mellon users will have access to the items. [contact Bella Gerlich, bg2r@andrew.cmu.edu]

Digital Image Database

Currently in the user-testing phase, the University Libraries Image Database will be introduced to the campus community in the spring of 2004. Based on the library's Slide Collection requests and usage, the images will be searchable and available via the web for viewing and using in a classroom setting. Digitization expands authorized access and eliminates contention for individual slides. For each image there will be three JPEG files accessible in various sizes (constraining tool will automatically adjust height and width): thumbnail (128 x 99 dpi), working (640 x 480 dpi) and presentation (1024 x 768 dpi). In addition, an archival copy of each image will be preserved using the highest possible resolution TIFF format and kept on tape backups. Image resolutions were decided based on recommendations/standards from the Library of Congress (American Memory Collection) website, ARLIS conference presentations, various listservs and professional communiqués. Copyright and/or source information will be provided in for each record. The intellectual content provided to catalog the images will be generated by library staff or, in the case of purchased collections, by the vendor. The Getty Research Institute's Vocabulary Databases, made available via the web to support limited research and cataloging efforts, may be consulted to standardize creator names, confirm dates, etc. The Image Database will be available to Carnegie Mellon users only through Internet Protocol authentication. [contact Bella Gerlich, bg2r@andrew.cmu.edu]

Digital Information Versatile Archive (DIVA)

DIVA allows students and researchers to search, browse, view and print digital images of books, technical reports and archival documents. With specifications developed by Carnegie Mellon librarians and archivists, DIVA provides conventional access to library and archival materials, and adds powerful new functions for searching and retrieving documents, supporting multimedia, and customizing the structure and presentation of collections. At present, DIVA is used for more than one million pages of archival materials (Carnegie Mellon's H. John Heinz III Archives, Allen Newell Collection, and Herbert Simon Collection; and, in cooperation with The Carnegie Museum, the Diplodocus and Douglass Archive). [contact: Chris Kellen, ck05@andrew.cmu.edu]

- [DIVA](#) (project summary)

H&SS Undergraduate Honors Theses

Fulfilling a commitment to digitize undergraduate honors theses from Carnegie Mellon's College of Humanities & Social Science, the University Libraries conducted a pilot project (2002-2003), in which guidelines were tested and 34 theses were digitized and made available via links in the library catalog. New H&SS undergraduate honors theses will be added annually, effective 2004. [contact: Gabrielle Michalek, gm11@andrew.cmu.edu]

Million Book Project

Designed to address inequities in the size and accessibility of library collections, to facilitate scholarship and lifelong learning, and provide a large test bed to support digital library research, the international Million Book Project aims to digitize and provide free-to-read access to a million books by 2007. The Million Book Collection will be a multilingual collection of collections selected by librarians, including public domain materials, government documents, and copyrighted books. The National Science Foundation has granted Raj Reddy and Gloriana St. Clair of Carnegie Mellon \$3.6 million for scanning equipment and administrative travel. The governments of India and China are providing the labor to scan the books and capture the metadata. Project partners include the Internet Archive, the Online Computer Library Center (OCLC), libraries in the United States, and universities and research institutes in India and China. One hundred thousand books are targeted to be available by 2004. A pilot test was recently conducted to better understand costs and issues related to shipping books to India. Based on the pilot shipment of six thousand books, strategies have been developed and implemented to reduce shipping costs and turn-around time. [contact: Denise Troll Covey, troll@andrew.cmu.edu]

Copyright permission work for the Million Book Project is discussed later in this report.

- Million Book Project [the collection to date], <http://www.dli.gov.in/home.htm>
REQUIRES INTERNET EXPLORER
- St. Clair, Gloriana. "Knowledge from a Million Books and Palm Leaves," *portal: Libraries and the Academy* 3, 3 (July 2003): vii-ix. Available to Project Muse subscribers at
<http://muse.jhu.edu/journals/portal_libraries_and_the_academy/v003/3.3stclair.html>
- St. Clair, Gloriana. [Million Book Project Today](#). OCLC, Dublin, OH, October 21, 2003.
- St. Clair, Gloriana. [Million Book Project in U.S. and India](#). International Conference on the Future of the Book, Cairns, Australia, April 22, 2003.
- St. Clair, Gloriana. [Million Book Project \(MBP\)](#). Johns Hopkins University, February 5, 2003.
- St. Clair, Gloriana. [Million Book Project \(MBP\)](#). Coalition for Networked Information, San Antonio, Texas, December 5, 2002.
- St. Clair, Gloriana. [Million Book Project: Dreams and Realities](#). Pennsylvania Academic Library Consortium, Inc. (PALCI), June 3 and November 7, 2002.
- Troll Covey, Denise. "Understanding and Assessing the Million Book Project." Presented at the Pennsylvania Library Association Conference, Pittsburgh, PA, October 5, 2003.
- Troll Covey, Denise. 2003. [Copyright Permission: Turning to Dust or Digital](#). *Proceedings of The Future of the Book Conference*, Cairns, Australia. Talk presented at the Future of the Book Conference, April 2003. Paper forthcoming at: <<http://BookConference.Publisher-Site.com>>
- [Million Book Project FAQ](#) (August 2002)
- [Million Book Project](#) (project summary)

One Million Digital Pages

In fall 2001, The University Libraries celebrated a milestone: having digitized more than one million pages of primary source material and making it available to scholars on the web. Accomplished over several years (1993-2001), the achievement includes developing standards for digitization of archival material and for natural language processing, and successfully migrating the digitized collections to new technology (from HELIOS to DIVA) in the course of the work. Three of the university's most important archival collections were digitized: the [H. John Heinz III Archives](#), [Allen Newell Collection](#), and [Herbert Simon Collection](#). [contact: Gabrielle Michalek, gm11@andrew.cmu.edu]

- [University Libraries the First to Digitize One Million Pages](#) (news item)
- [DIVA](#) (project summary)
- [HELIOS](#) (project summary)

Posner Project

A generous gift from Helen and Henry Posner Jr. in honor of Henry Posner Sr. and his wife Ida M. Posner has made possible the digitization of the Posner Memorial Collection. The collection of 622 titles includes landmark titles in the history of western science, beautifully produced books on decorative arts, and fine sets of literature. Examples of volumes of particular importance include works by Copernicus, Kepler, and a Third Folio of Shakespeare. Scanning of the Collection will be completed by 2004. As the digitized books become available, they can be searched and browsed at the Posner Family Collection, <http://posner.library.cmu.edu/Posner/>. Currently only metadata searching is available, but full-text searching will be provided in 2004. The digitized books are also being linked to Carnegie Mellon library catalog records. [contact: Denise Troll Covey, troll@andrew.cmu.edu]

Copyright permission work for the Posner Collection is discussed later in this report.

Smart Web Exhibits

Under a grant from the Institute for Museum and Library Services (IMLS), Carnegie Mellon University Libraries, the Carnegie Museum of Natural History, and the School of Computer Science at Carnegie Mellon worked together to solve problems caused by physical space constraints and provide more effective outreach to the public. "Smart Web Exhibits" are designed to deliver information online, on target and on time to a diverse user community. Two exhibits were developed with the grant, based on signature collections in Carnegie Mellon University Archives and the Carnegie Museum of Natural History. [contact: Gabrielle Michalek, gm11@andrew.cmu.edu]

- [Mind Models: Artificial Intelligence Discovery At Carnegie Mellon](#)
- [Diplodocus and Douglass Archives](#)
- [Smart Web Exhibits](#) (project summary)

Swiss Poster Collection

Established in 1985 by Ruedi Ruegg, Swiss graphic designer, and Carnegie Mellon School of Design Professor Daniel Boyarski, The Swiss Poster Collection is a critical selection of more than 300 works representing the Swiss Posters of the Year competition and other Swiss posters from 1971 to the present. The digitization of the posters and the creation and maintenance of the website and electronic database is an example of a successful continuing collaboration effort between the School of Design and the University Libraries. This website has received international attention and has been noted in the design journal *Print*. [contact: Mary Kay Johnsen, mj0g@andrew.cmu.edu]

U. S. Government Documents (2002-)

Many government documents have been targeted for inclusion in the Million Book Project. Contributions from participating Million Book Project members, such as the Carnegie Library of Pittsburgh, will enhance the depth of the collection. [contact: Erika Linke, el08@andrew.cmu.edu]

B. Services

[Ask A Librarian](#) (chat reference service)

Carnegie Mellon University Libraries began its chat reference service in October 2000. The Libraries utilize commercial off-the-shelf software called LivePerson. Service evaluation features of LivePerson provide interesting data for analysis. [contact Paul Neuhaus, neuhaus@andrew.cmu.edu, and Matthew Marsteller, matthewm@andrew.cmu.edu]

- Marsteller, Matthew R. and Paul Neuhaus. "Providing Chat Reference Service: A Survey of Current Practices," in R. David Lankes, Charles R. McClure, Melissa Gross and Jeffrey Pomerantz, eds., *Implementing Digital Reference Services: Setting Standards and Making It Real* (Neal-Schuman, 2002), 61-74.
- Neuhaus, Paul J. and Matthew R. Marsteller. "Chat Reference at Carnegie Mellon University," *Public Services Quarterly* 1, 2 (2002): 29-41.
- [The Chat Reference Experience at Carnegie Mellon](#) (ALA poster session, June 2001)

Electronic Reserves

Carnegie Mellon University Libraries began offering electronic reserve services in fall 2000. Brief item records are created in our online catalog (Sirsi Unicorn) for each title to be put on reserve. Photocopied items submitted for reserves are scanned into PDF format and placed on a server. A program runs hourly on the server to assign each item a URL and then forwards that URL to an email account. The URLs taken from that account are put into the 856 field of the brief record to create the link to the article (or other document).

Users can search the Reserves module in the library catalog by Instructor Name, Course Number, or Course Name. Users call up the record for the item that they need and click the link to view or print the document. Access to electronic reserves is limited to members of the Carnegie Mellon community by IP authentication, and WebISO enables access for remote Carnegie Mellon users. [contact Joan Stein, joan@andrew.cmu.edu]

Patron-Initiated Borrowing (ILLiad and EZBorrow)

Research indicated that users on our campus wanted a system of requesting books and other materials from other libraries that would provide them with speedy delivery, a great deal of personal control, and access to information about the progress of their requests.

We selected ILLiad as our interlibrary loan management system because it provided all of those requirements and more. ILLiad is able to take articles received via Ariel and deliver them to the web, allowing us true desktop delivery for the first time through ILL. It has the added benefit of leveling workload across ILL offices in three library facilities.

We also implemented the Pennsylvania Academic Library Consortium, Inc. (PALCI)'s EZBorrow system for patron-initiated circulation. This allows people to search the catalogs of participating libraries and place requests for books directly (i.e., quickly). Supported by a Pennsylvania-wide delivery service, it provides excellent turnaround time and also allows users to monitor the status of their requests.

A link in the library catalog alerts users to both ILLiad and EZBorrow, and helps users select which system to use based on their needs. [contact Joan Stein, joan@andrew.cmu.edu]

Scanning Services

Carnegie Mellon University Libraries provide public scanning capabilities at all of its reference locations. This service enables students, faculty and staff to create, save and retrieve images files using library materials during the hours of operation at each facility. In addition, a pay-per-print color laser printer was installed in the Arts Reference area to accommodate a growing demand for such a service. [contact Bella Gerlich, bg2r@andrew.cmu.edu]

SFX

Beginning in 2002, the University Libraries have implemented Ex Libris's SFX software, which provides links to journal titles provided by journal publishers and aggregator databases, links to the title record in the library catalog record, links to journal tables of contents and full-text articles, and enables cross-linking between resources (for example, from citations in one resource to full-text articles in another resource). SFX also supports interlibrary loan requests via ILLiad. In 2003, SFX functionality was incorporated into the Automated Reference Assistant (ARA). It was also used to create a web-based list of the University Libraries' electronic journals; adding the Libraries' print journal holdings to this list is being explored. [contact: Erika Linke, el08@andrew.cmu.edu and Alice Bright, ab03@andrew.cmu.edu]

VPN IP Address Extension Service

In fall 2002, the University Libraries assisted university Computing Services to test and evaluate the IP Address Extension Service, a Virtual Private Network (VPN) protocol that enables Carnegie Mellon users secure authenticated access to licensed library resources from any platform or location. VPN may be used with dial-up, DSL, or cable connections (available via <http://www.cmu.edu/computing/documentation/VPN/vpn.html>). Users began adopting the new software in December, and are using it as an alternative to the university's proxy server. [contact: Adele Barsh, adele@andrew.cmu.edu]

Web Portal

The Carnegie Mellon Web Portal is an integrated web service that is now available to the students, staff and faculty of Carnegie Mellon University, and will in the future be extended to incoming students and alumni. The portal provides an interface to various services and events that can be customized to include information based on one's interests, activities and organizations. The University Libraries have been active players in the web portal project from the beginning, in 2002. Currently, Carnegie Mellon users can search the library catalog from within the portal, and can access databases and library services. In the future, resources and online services will be tailored and pushed to specific user groups. [contact: Chris Kellen, ck05@andrew.cmu.edu]

C. Systems

Automated Reference Assistant (ARA)

Carnegie Mellon usage statistics indicate that the majority of online catalog and database use occurs by remote access, that is, from outside library facilities. A drawback of remote access is that users do not have traditional reference librarians at their sides to help guide them to relevant and reliable material. The effect is that users often become confused and overwhelmed when searching by remote access. The University Libraries sought funding and developed the Automated Reference Assistant software to help guide remote users to relevant high-quality information online. [contact: Chris Kellen, ck05@andrew.cmu.edu]

- Automated Reference Assistant, <http://www.library.cmu.edu/Research/Ara/>
- George, Carole A. "Heuristic Evaluation: Automated Reference Assistant (ARA)." (Carnegie Mellon University Libraries, 2003).
- Troll Covey, Denise. "Automated Reference Assistance: Reference for a New Generation." Presented at the Coalition for Networked Information, Washington, DC, April 2002.
- [ARA: Automated Reference Assistant](#) (project summary)

Management Information System

The University Libraries continues to develop and maintain a suite of databases to facilitate library operations, for example, databases for monitoring equipment, journal subscriptions, requests for travel reimbursement, capital purchases, technology-related work requests, and student hiring.

A Management Information System Task Force was formed in 2001 to

- Assess current and proposed data gathering practices
- Recommend what data should be gathered and how
- Develop a requirements specification for a management information system (MIS) that will solve existing problems and facilitate data management, analysis, and use
- Oversee the implementation of the new MIS

The data audit was completed in 2001. Recommendations were approved and requirements specified in 2002. Work is now underway to select software and begin implementing the new MIS. The new system, targeted for completion in 2004, will simplify data entry, compilation, and the generation of trend lines and cross correlations. [contact: Denise Troll Covey, troll@andrew.cmu.edu]

- Troll Covey, Denise. "Evaluating Library Services: Using Data to Improve Service to Our Customers." Invited to develop and present full-day workshop at the Pittsburgh District Library Center, Pittsburgh, PA, May 29, 2003.
- Troll Covey, Denise. "What the Assessment Gurus Don't Tell You." Invited keynote speaker at the Palinet Institute, "Evaluating Your Library Services," Harrisburg, PA, October 2002.
- Troll Covey, Denise. [Accountability versus Count-Ability](#). Invited speaker at the NERCOMP Workshop on Statistics and Measurement for Library Services: Why, How, and What to Do with the Numbers, Middletown, CT, April 2002. Encore presentation for the Pennsylvania Library Association (PaLA) College and Research Library Division, Philadelphia, PA, September 2002.

MetaScan (metadata capture software)

The MetaScan software is a data entry tool which allows scanning operators to easily and reliably enter metadata about the objects they are scanning. MetaScan provides the ability to search and extract information from a library catalog for the object being scanned, thus making the metadata entry easier and less error-prone. All information is stored in the industry-standard XML file format. [contact: Chris Kellen, ck05@andrew.cmu.edu]

QuestionPoint Collaborative Reference Service

The Carnegie Mellon University Libraries have been participating in the QuestionPoint reference service since November 2002. QuestionPoint is a collaborative reference service developed by the Library of Congress and OCLC. It allows member libraries to refer questions to expert resources through a global web-based network in which an automated “request manager” routes questions from one library to another, based on metadata about the question and predefined profiles of the collection, subject and staff strengths of members of the network, etc. It also provides a global knowledge base of previously asked and answered reference questions. Between November 2002 and October 14, 2003, Carnegie Mellon has responded to 344 questions from around the world. [contact: Jean Alexander, jeana@andrew.cmu.edu]

Shibboleth Pilot (2002-2003)

The University Libraries, in conjunction with Carnegie Mellon Computing Services, Internet2, database vendors, and other institutions worked to implement a pilot project using new client/server technology. Shibboleth is a potential alternative to IP-address restriction that improves access to e-resources. The software enabled authenticated users to access “shibbolized” library resources (particularly, databases purchased and licensed for campus users) with one log-in. Part of the beta program helped determine the viability of Shibboleth as a tool for libraries and also whether Internet2 and vendors can support the software. The Shibboleth Pilot Project Team developed a web page which offered shibbolized links to two databases, JSTOR and FirstSearch, and which explained the project and the need for it. [contact: Sue Collins, sc24@andrew.cmu.edu]

- <http://www.library.cmu.edu/Services/shib.html>

WolfPack

Creating a digital library requires converting the original scanned images into various formats. WolfPack is a software system that performs these large data conversion tasks in a distributed manner. The WolfPack framework allows the best off-the-shelf conversion programs to be used in an automated system, and it runs the conversions in parallel on a large number of machines. WolfPack is currently used to perform image cropping, deskewing, despeckling and OCR, as well as to create JPEG and Acrobat files from scanned images. [contact: Chris Kellen, ck05@andrew.cmu.edu]

II. Projects and programs

A. Projects

PROJECT UPDATES

Copyright Permissions Projects

Carnegie Mellon University Libraries have undertaken three copyright permission projects beginning in 1999. Lessons learned from one project are applied in the next. [contact: Denise Troll Covey, troll@andrew.cmu.edu]

1. ***Feasibility Study (1999-2000)***. Based on a statistically valid random sample of books in the library catalog, this study was designed to determine the feasibility of getting copyright permission to digitize books and provide them free-to-read on the surface web. The study revealed that
 - Copyright ownership of approximately 3% of the books was too complicated to pursue.
 - Locating publishers is difficult, time consuming, and often unsuccessful. Over 10% of the publishers could not be located.
 - If the publisher could be located, there was only a 50-50 chance of getting a response, even to a second letter of inquiry.
 - Publishers of out-of-print books are more difficult to locate and more likely not to respond than publishers of in print books.
 - If a publisher responded, there was less than a 50-50 chance of getting permission to digitize the book and provide free-to-read access to it, regardless of whether the book was in print or out-of-print.
 - If permission was granted, often restrictions were applied, for example, restrict access to Carnegie Mellon users only, provide access to the book for only two years, or pay \$300 for permission per title.

The overall success rate in securing copyright permission was 22%. However, the likelihood of success varied significantly with different types of publishers, ranging from 45% for scholarly associations to 12% for commercial publishers.

Several different employees conducted this study, none of whom were dedicated to the task. Consequently there were sometimes substantial delays between sending the initial letter of inquiry and sending the follow-up letter. No data were gathered on the transaction costs of pursuing permission to digitize the random sample of books.

2. ***Posner Memorial Collection (2002-2003)***. A full-time temporary employee was hired to do the copyright permission work on the Posner Memorial Collection. A study of the collection and search of the copyright renewal records revealed that

approximately 34% of the books in the collection are copyright protected. The work is ongoing. Letters of inquiry are sent to each publisher of copyrighted titles in the collection, offering them the options of granting open access to their books on the web, restricting access to the Carnegie Mellon community only, or (of course) denying permission to digitize and provide online access to their books. Requests are for non-exclusive permission to digitize and provide online access to the books. Typically the initial letter of inquiry does not get a response from the publisher. However, rather than sending a second letter, the follow-up is done by telephone or email. Often copyright has reverted to the author's estate, so the estate has to be located and contacted.

The copyright permission work for the Posner Memorial Collection will be completed by 2004. As of the end of August 2003, the success rate for copyright permission requests was 56%, with a transaction cost of \$37 per title. Approximately 13% of the copyright holders cannot be located.

3. ***Million Book Project (2003-2007)***. The initial focus of the copyright permission work for the Million Book Project was to acquire permission to include the titles cited in *Books for College Libraries*, which is a five-volume bibliography of books recommended for all academic library collections. *Books for College Libraries* lists approximately 50,000 titles. The University Libraries quickly realized that pursuing permission per title was prohibitively expensive, so a new strategy was needed. The current collection development strategy for copyrighted works targeted for the Million Book Collection is to use librarian-selected bibliographies as approval plans to identify publishers of quality books. Letters are being sent to publishers of works cited in the selected bibliographies, beginning with *Books for College Libraries*. The letters introduce the Million Book Project and educate publishers about user behaviors and preferences and new business models, specifically:

- Users want to find information online, but use it in print.
- Online access increases use of materials, including use of older materials.
- Open access doesn't decrease, and can actually increase sales.
- 95% of the books ever published are still in copyright, but out of print – meaning that they are neither generating revenue for publishers nor easily accessible to potential readers who might be willing to pay for them.

The letters then ask publishers for *non-exclusive* permission to digitize and provide open access to:

- All of their in-copyright, out-of-print books.
- All titles published prior to a date of their choosing.
- All titles published some number of years ago (they specify the number).
- A list of titles that they provide.

As incentive to participate in the Million Book Project, Carnegie Mellon offers to give the publishers copies of the digitized books and accompanying metadata, which they can use in fee-based, added-value services, for example, a print-on-demand service that would generate revenue from out-of-print books. Introductory letters are followed-up with a phone call or visit to the publisher.

This strategy yields many more books for inclusion in the Million Book Collection for the same transaction costs incurred to negotiate permission per title. Using intermittent labor, as of September 1, 2003, letters had been sent to 184 publishers, but few follow-up calls had been made. The success rate at that time was 4%, with permission granted to include thousands of in-copyright, out-of-print books in the Million Book Collection.

A full time staff member dedicated to copyright permission work for the Million Book Project was hired in October 2003. Future copyright permission work for the Million Book Project will focus on scholarly associations and university presses because the results of the feasibility study indicate that these types of publishers are three to four times more likely to grant permission for open access than commercial publishers.

Digital Audio Reserves Pilot Project (2001-2003)

Planning for Carnegie Mellon's digital audio project began in early 2001 with monitoring the Music Library Association's listserv and a digital audio listserv, gathering correspondence on digital audio projects. This included various opinions on software and legal issues. In June 2001, representatives from Library Instructional Technology and the Arts and Special Collections unit went to Penn State to speak with the people responsible for developing and implementing their digital audio reserves. In fall 2001, preliminary workflows, policies and procedures for creating, mounting and saving audio e-reserves files were established. A front-end for the web site was created, and implementation of an audio e-reserves pilot was ready for the academic year 2002-2003. Two courses were targeted for testing. To measure service capabilities, a survey was created and distributed to the pilot group in spring 2003. Survey results will dictate changes in procedures/website prior to the anticipated release and marketing of service availability in 2004. [contact Bella Gerlich, bg2r@andrew.cmu.edu]

Challenges regarding Digital Audio Reserves are discussed later in this report.

User Studies: Think Aloud Protocols for Mind Models Exhibit (2001-2002)

Funded by a grant from IMLS, and in partnership with the Carnegie Museum of Natural History and the School of Computer Science at Carnegie Mellon, the University Libraries developed a smart web exhibit—an online exhibit designed to deliver archival information via the web to users who might not otherwise use archives. To test site functionality and expose possible problems during development,

think-aloud protocols were conducted with an exhibit prototype. Participants were asked to “think aloud” while performing a set of tasks. As a result of the study, design changes were made to improve navigation, labeling and visual appeal of the [Mind Models: Artificial Intelligence Discovery At Carnegie Mellon](#) exhibit.

- George, Carole A. [Mind Models Exhibit: User Studies](#) [final report]. (Carnegie Mellon University Libraries, 2002).

III. Specific Digital Library Challenges

Digital Audio Reserves Challenges

The Carnegie Mellon University Libraries process uses a freeware program to rip sound files. Files are saved at 96K, as the difference between 124K and 96 is indiscernible, and the smaller files allows for easier streaming. A hyperlink is created using the E-Reserves process currently in place and inserted into a duplicate ‘reserve’ record in the library catalog. The hyperlink launches a web page & Java applet that was developed in-house that requires a client run on the user’s machine allowing the streaming to run locally, as opposed to the server sending the stream.

Web statistics collected during the duration of the pilot study counted the 1,912 successful requests for Audio-E-Reserves pages. There was an average of five successful requests for pages per day. In addition to the web statistics, the respondents to the class survey returned an overall favorable response to the Audio-E-Reserves option and service. Faculty who piloted the service were complimentary, and indicated they would like to use the service for future classes.

Problems were experienced late in the trial when the browsers (Netscape and Explorer) were upgraded system-wide in library facilities. This problem may be eliminated with the addition of a new script that would automatically generate an upgrade as needed, but further research is being done first in order to implement the best delivery system to the users. Currently, staff are contacting other peer institutions and investigating the ROI of continually upgrading the homegrown product as is versus purchasing and servicing existing streaming technologies. [contact Bella Gerlich, bg2r@andrew.cmu.edu]

Copyright Permissions Challenges

Approximately 95% of the books ever published are still in copyright, but out of print. These books are neither generating revenue for publishers nor easily accessible to potential readers who might be willing to pay for them. During the period when U.S. copyright required renewal, research shows that fewer than 15% of copyrights were renewed. Only about 2% of the books published in the 1920s and 1930s have continuing commercial value, yet the Copyright Term Extension Act of 1998 extended their copyright to at least 2023. Copyright law and the cost of seeking copyright permission are among the biggest barriers to creating the digital library.

Primary difficulties in acquiring copyright permission to digitize materials include:

- Learning what, when, and how to apply U.S. and foreign copyright laws.
- Determining copyright status, which requires consulting copyright renewal records (not easily accessible for all media) and knowing the applicable copyright laws.
- Identifying and locating the copyright holders:
 - There could be multiple copyright holders for text, images, and other intellectual assets incorporated in a single work.
 - Copyright often reverts to the author or to the author's estate, but this information is not easily accessible and takes time to uncover and track.
 - Out-of-business publishers, book dealers and estates are difficult to locate.
 - Many publisher addresses in *Global Books in Print* are incorrect.

Furthermore, publishers' permission departments are very slow to respond, publishers sometimes cannot find contracts for titles they published, and publishers are sometimes uncertain whether their contracts grant them electronic rights.

Along with the above difficulties, publishers have inclinations that impede acquiring non-exclusive permission to digitize and provide open access to their works. For example, they are afraid that their books will lose value or be commercially distributed by someone other than themselves if they are digitized and made available on the web. They often want to be paid copyright permission fees or royalties.
[contact: Denise Troll Covey, troll@andrew.cmu.edu]

Challenges of Collaboration

As we undertake collaborative digital projects with colleagues in academia and the private sector, on campus and around the world, we have discovered that although we share a vision we also have many differences that need to be communicated and addressed along the way. Differences in approach, expectations, understanding and philosophy must be negotiated among disciplines, and between the public and private sectors. Differences in bandwidth, in the broadest sense—including hardware, software, and the learning curve for people from various backgrounds who must understand and use new technologies—is another constant that must be anticipated. Handing off projects from one group to another may result in discomfort, discontinuity and delay. Far from discouraging collaboration, dealing with the difficult issues of group work heightens awareness, encourages creativity and produces better, stronger projects.

DOI (Persistent URLs)

The challenge of creating persistent URLs to our digital items was met through the adoption the Digital Object Identifier (DOI) standard. Through the use of DOI and the creation of software that allows for simple entry and maintenance of our DOI data, we have created a system which solves the persistent URL challenge. [contact: Chris Kellen, ck05@andrew.cmu.edu]

- [Carnegie Mellon University Libraries Digital Object Proxy Server](#)

Challenges to Workflow and Workload

In an environment in which building the digital library is everyone's job, developing procedures and workflow for the digital library pose organizational, resource and social challenges. Efforts that center on essential workflow tasks alone may not thoroughly address organizational and social issues. Several means of addressing challenges to workflow and workload include the following:

- Communicate the digital library vision effectively.
- Develop a mutual understanding of how individuals and departments contribute to the effort.
- Develop and flowchart work in an inclusive manner. Engaging staff responsible for the tasks in the process ensures better insight into the workflow and better workflow development.
- Recognize that digital library work can and will disrupt what staff may consider core responsibilities.
- Recognize that some core library activities are paramount to digital library tasks.
- Thoughtfully consider and analyze workflow and processes.
- Evaluate how staff are deployed and add or reassign staff as necessary.
- Manage change with intention, care and training.

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

MOST RECENT PUBLICATIONS & TALKS

Aurand, Martin. [Charette Digital Project: An Historic Pittsburgh Architectural Journal Goes Online](#). Presented at the Association of Architecture School Librarians Conference, Louisville, KY, March 2003.

Birdie, Christina and Erika Linke. "Identifying and Selecting Content for the Million Book Project." Libraries and Education in the Networked Environment (24th IATUL Conference). June 2-5, 2003. Forthcoming in the conference proceedings.

Collins, Susan L. "Developing Web Resource Guides for Local History: A Pittsburgh Example," *Public Services Quarterly* 1, 2 (2002): 67-78.

Collins, Susan L. [In the Aggregate: A Quantitative Study of the Impact of Interdisciplinary Full Text Databases on Historical Research](#), *Journal of the Association for History and Computing* 6, 1 (April 2003).

Collins, Susan L. [That was Then and This is Now: Technological Change and Library Services](#). Presented at American Association for History and Computing, DePauw University, Greencastle, IN, 2003.

George, Carole A. "Heuristic Evaluation: Automated Reference Assistant (ARA)." (Carnegie Mellon University Libraries, 2003).

George, Carole A. [Mind Models Exhibit: User Studies](#) [final report]. (Carnegie Mellon University Libraries, 2002).

Gerlich, Bella Karr and Amy Perrier. "Arts Instruction in the Age of Technology: Providing Library Services to Support Studio and Survey Faculty Who Use Technology for Instruction," *Information Technology and Libraries* 22, 2 (June 2003): 79-83

Harvey, Melissa J. "Opening Doors: Libraries and Campus Portals." Invited speaker at the Pennsylvania Library Association Annual Conference, October 4, 2003.

Linke, Erika C. "Million Book Project" in Miriam Drake (ed.), *Encyclopedia of Library and Information Science*, 2nd edition (New York, Marcel Dekker, 2003).

Linke, Erika C. "Creating a Free-to-Read International Digital Library." In *Proceedings of Digital Library 2002: IT Opportunities and Challenges in the New Millennium* (Beijing: Beijing Library Press, 2002), 213-219.

Linke, Erika C. "Identifying and Selecting Content for the Million Book Project." Invited speaker at Libraries and Education in the Networked Environment (24th IATUL Conference), June 2-5, 2003.

Linke, Erika C. "Creating a Free-to-Read International Digital Library." Invited speaker at Digital Library 2002: IT Opportunities and Challenges in the New Millennium, Beijing, China, July 8-12, 2002.

Marsteller, Matthew R. [Birds of a Feather: Subject-Based Consortia for VR](#). Invited panel speaker at the Virtual Reference Desk Conference (VRD 2002), Chicago, IL, November 11, 2002.

Marsteller, Matthew R. "Virtual Reference Services in PALCI Member Libraries." Panel speaker at the Fall Membership Meeting, University of Pittsburgh, November 8, 2002.

Marsteller, Matthew R. [We Deliver: Bringing Live Reference to Our Users](#). Panel speaker at the Discussion Forum sponsored by the Products and Services Committee, RUSA/MARS, American Library Association Midwinter Conference, January 20, 2002.

Marsteller, Matthew R. and Dianne Mizzy. "Exploring the Synchronous Digital Reference Interaction for Query Types, Question Negotiation, and Patron Response," *Internet Reference Services Quarterly* 8, 1/2 (2003): 149-165.

Marsteller, Matthew R. and Paul Neuhaus. "Providing Chat Reference Service: A Survey of Current Practices," in R. David Lankes, Charles R. McClure, Melissa Gross and Jeffrey Pomerantz (eds.), *Implementing Digital Reference Services: Setting Standards and Making It Real* (Neal-Schuman, 2002), 61-74.

Marsteller, Matthew R. and Kate Thomes. "What Librarians Want and Need From Journals." Invited speaker at the Council of Science Editors Annual Meeting, May 6, 2003.

Neuhaus, Paul J. "Privacy & Confidentiality in Digital Reference," *Reference & User Services Quarterly* 43, 1 (Fall 2003). Forthcoming.

Neuhaus, Paul J. and Matthew R. Marsteller. "Chat Reference at Carnegie Mellon University," *Public Services Quarterly* 1, 2 (2002): 29-41.

St. Clair, Gloriana. "Dysfunction in Scholarly Communications," Carnegie Mellon University, March 10, 2003.

St. Clair, Gloriana. "Knowledge from a Million Books and Palm Leaves," editorial statement in *portal: Libraries and the Academy* 3, 3 (July 2003). Available to subscribers at http://muse.jhu.edu/journals/portal_libraries_and_the_academy/v003/3.3stclair.html.

St. Clair, Gloriana. "Managing Information for Learning" in Sigrún Klara Hannesdóttir (ed.), *Global Issues in 21st Century Research Librarianship* (Helsinki, Finland: NORDINFO, 2002).

St. Clair, Gloriana. Recent talks about the Million Book Project include:

- [Million Book Project Today](#). OCLC, Dublin, OH, October 21, 2003.
- [Million Book Project in U.S. and India](#). International Conference on the Future of the Book, Cairns, Australia, April 22, 2003.
- [Million Book Project \(MBP\)](#). Johns Hopkins University, February 5, 2003.

- [Million Book Project \(MBP\)](#). Coalition for Networked Information, San Antonio, Texas, December 5, 2002.
- [Million Book Project: Dreams and Realities](#). Pennsylvania Academic Library Consortium, Inc. (PALCI), June 3 and November 7, 2002.

St. Clair, Gloriana and Erika Linke. "The Library of the Future: The Future of the Library" in Daniel P. Resnick (ed.), *The Innovative University* (Pittsburgh, PA: Carnegie Mellon Press, 2003).

St. Clair, Gloriana and Susan Ware and Sara Lou Whildin. "The Disquieting Dilemmas of Digital Libraries" in Paul S. Goodman (ed.), *Technology Enhanced Learning: Opportunities for Change* (Mahwah, NJ: L. Erlbaum, 2002).

Troll Covey, Denise. [Accountability versus Count-Ability](#). Invited speaker at the NERCOMP Workshop on Statistics and Measurement for Library Services: Why, How, and What to Do with the Numbers, Middletown, CT, April 2002. Encore presentation for the Pennsylvania Library Association (PaLA) College and Research Library Division, Philadelphia, PA, September 2002.

Troll Covey, Denise. "Automated Reference Assistance: Reference for a New Generation." Presented at the Coalition for Networked Information, Washington, DC, April 2002.

Troll Covey, Denise. [Copyright Permission: Turning to Dust or Digital](#). Presented at the Future of the Book Conference, Cairns, Australia, April 2003. Paper forthcoming in *Proceedings of The Future of the Book Conference*, at <http://BookConference.Publisher-Site.com>.

Troll Covey, Denise. "Evaluating Library Services: Using Data to Improve Service to Our Customers." Invited to develop and present full-day workshop at the Pittsburgh District Library Center, Pittsburgh, PA, May 29, 2003.

Troll Covey, Denise. "Understanding and Assessing the Million Book Project." Presented at the Pennsylvania Library Association Conference, Pittsburgh, PA, October 5, 2003.

Troll Covey, Denise. "What the Assessment Gurus Don't Tell You." Invited keynote speaker at the Palinet Institute, "Evaluating Your Library Services," Harrisburg, PA, October 2002.