Carnegie Mellon University
Research Showcase @ CMU

University Libraries Research | University Libraries

January 2004

Global Cooperation for Global Access: The Million Book Project

Denise Troll Covey
Carnegie Mellon University, troll@andrew.cmu.edu

Follow this and additional works at: http://repository.cmu.edu/lib_science

Part of the Other Social and Behavioral Sciences Commons

Recommended Citation
Proceedings of the 7th International Conference on Current Research Information Systems: Putting the Sparkle in the Knowledge Society, , , 139- 151

This Book Chapter is brought to you for free and open access by the University Libraries at Research Showcase @ CMU. It has been accepted for inclusion in University Libraries Research by an authorized administrator of Research Showcase @ CMU. For more information, please contact research-showcase@andrew.cmu.edu.

Global Cooperation for Global Access: The Million Book Project

Denise Troll Covey
Associate Dean, Carnegie Mellon University Libraries

Summary

This paper describes the rationale for and international collaborations involved in a project to digitize one million books and offer them free-to-read on the surface web by 2007. Led by computer scientists and librarians at Carnegie Mellon, Million Book Project partners include universities and research institutes; for-profit and not-for-profit organizations; governments and government agencies; librarians and archivists; software developers; and commercial publishers, university presses, and scholarly associations. Areas of cooperation addressed in this paper include collection development, copyright permission, digital registry, book acquisition and shipping, scanning, quality control, sustainability, and added-value services.

1. Introduction

The Million Book Project is an international collaboration to digitize and provide free-to-read access to one million books on the surface web by 2007. The scanning is being done in India and China, with labor funded by the respective governments, following preservation-quality standards developed by the Institute for Museum and Library Services (2001) and endorsed by the Digital Library Federation (2002). The National Science Foundation (NSF) funded the initial collection development meeting in 2001, and in 2002 provided $3.6 million to purchase equipment and support administrative travel. With NSF's approval, some of the travel dollars were allocated to fund the pilot shipment of books to India for scanning.

Carnegie Mellon School of Computer Science and the University Libraries are providing project leadership, with the full support of the University President and Provost. Provost Mark Kamlet led delegations to China in 2001 and India in 2002 to develop and formalize project plans. Raj Reddy, Herbert A. Simon Professor of Computer Science, is responsible for the Project vision and technology. Michael Shamos, Principal Research Scientist and Intellectual Property Attorney, addresses legal issues related to the Project. Gloriana St. Clair, Dean of University Libraries, is responsible for coordinating content selection and acquisition among U.S. library partners. Content is being provided by U.S. libraries and by partner institutions in India and China.

The project in India is being led by the Indian Institute of Science. Participating institutions in India include Anna University; Goa University; Indian Institute of Information Technology; International Institute of Information Technology; Mysore University; Shanmugha Arts, Science, Technology & Research Academy; Tirumala Tirupati Devasthanams; Maharashtra Industrial Development Corporation; and University of Pune.

The project in China is being led by the Chinese Academy of Sciences. Participating institutions in China include Beijing University; Fudan University; Ministry of Education of China; Nanjing University; Peking University; State Planning Commission of China; Tsinghua University; Zheijiang University; and Zhezing University.

Special partners in the Project include the Online Computer Library Center (OCLC) and the Internet Archives. Their roles are described later in this article.

Scanning is well underway in India and China. Approximately 100,000 books should be available by 2004. The books will be replicated on servers around the world, indexed by popular search engines, and freely available on the surface web – accessible anywhere, any time, to anyone with an Internet connection. Any school, public, or academic library will be able to link their library catalog records to the books in the Million Book Collection. The Collection will support education, research, and lifelong learning worldwide. Additional information about the Million Book Project can be found at http://www.library.cmu.edu/Libraries/MBP_FAQ.html.

2. Rationale

The Million Book Project is designed to address the following concerns:

- Disparity in the size and accessibility of library collections. In the United States, some single institutions, like Harvard and Yale, have more books in their libraries than some entire states have in all of their libraries combined. The view of the landscape is more grim when expanded to include developing and underdeveloped nations. The five largest libraries in India have only one million books, while Harvard has 12 million and the Library of Congress 30 million. In our rapidly changing world, lifelong learning and access to books are essential to employment, health, peace, and prosperity. Lack of equitable access to information impedes the democratization of knowledge and empowerment of a global citizenry.
- Student use of inappropriate materials to do their assignments. Almost all undergraduate students (96%) believe that information found on the surface web is adequate for doing their homework. Most of them (72%) use popular Internet search engines like Google to find information (Friedlander 2002). Almost half of them (46%-48%) use online resources "all" or "most" of the time, and believe that other sites have better information than the library web site (OCLC 2002). Given that online library resources reside in the deep web, where they are not retrievable by popular Internet search engines, faculty are concerned about the quality of the resources that undergraduate students use to complete their assignments. Only 6% of surface web content is appropriate for student academic work, and no single Internet search engine indexes more than 16% of the surface web (Lawrence and Giles 1999). Furthermore the trend is for the results retrieved by popular search engines to be ranked by fees paid by advertisers or sponsors rather than by relevance to the user's query. Lack of quality resources on the surface web is having a negative impact on the quality of student learning.
- Rapid delivery of research materials. Almost all students and faculty (90%) rate easy, convenient – which means speedy – access to information as a high priority need (Friedlander 2002). To most of them (54%-68%) easy, convenient access also means remote access to full-text electronic information from their homes or offices. Fewer than half of the students and faculty surveyed believe that the library is adequately meeting their needs in this regard; 24% report that they often cannot get information when they need it (Marcum and George 2003). Students perceive vendor licensing restrictions and password requirements as barriers to easy remote access to library resources (OCLC 2002). Despite the fact that students and faculty trust the library more than the Internet, most of them begin their search for information with an Internet search engine because it is easier and faster to find resources

using an Internet search engine than it is using electronic library resources (Marcum and George 2003; OCLC 2002; Jones and Madden 2002; EDNER 2002). The increasing availability and use of online bibliographic databases, the increasing number of scholarly publications, and the increasing cost of library materials have created a situation wherein libraries are spending more money but purchasing fewer materials. Interlibrary loan is increasing, but it is inconvenient and the turn-around time is often inadequate for both the rapid-paced, highly competitive research conducted by faculty and graduate students and the shorter deadlines and last-minute efforts of undergraduate students. Graduate students at Carnegie Mellon report that the limitations of the library collection and need to resort to interlibrary loan – particularly to get older journal articles and out of print books – impact their selection of research topics, the quality of their work, and their grade point average (George, 2004). Lack of speedy access to quality resources on the surface web is having a negative impact on the timeliness and success of academic work.

- Preservation of our cultural and intellectual heritage. Millions of books printed on non-acid-free paper are turning to dust on library shelves. Copyright seriously impedes their preservation through digitization because permission rights must be negotiated for each title. Even when books are digitized, current practice is to restrict their use through licensing provisions or digital rights management technologies that can trump otherwise legal uses of the materials. Lack of commitment to perpetual storage and access threatens digital books with disappearance or darkness in an inaccessible archive.
- Preservation of the core values of librarianship. Librarians traditionally champion equitable access, stewardship of our heritage, service to all of humanity, and individual privacy as handmaidens of intellectual freedom, democracy, and literacy. However in the digital arena, government legislation seems to threaten these values and the once hallowed public domain seems destined to wither from starvation. For example, in the United States the 1998 Copyright Term Extension Act (CTEA) extended copyright to the life of the author plus 70 years or 95 years for a work for hire. Attempts to have the CTEA declared unconstitutional failed in 2002 when the U.S. Supreme Court reviewed the case and ruled that the law appropriately restricted the copyright monopoly to a "limited time."

In addition to addressing the above concerns, the Million Book Collection will provide a large, globally accessible test bed for research in the following areas: search engines; machine translation; intelligent indexing; automatic summarization; information storage, distribution, management, security, and sustainability; accuracy of optical character recognition (OCR); OCR of non-Romanic languages and scripts; open access; copyright; and digital rights management.

3. Collection Development

Funded by the National Science Foundation, the initial collection development meeting for the Million Book Project was held in November 2001 in Pittsburgh, Pennsylvania. Participants included representatives from the Digital Library Federation (DLF), Center for Research Libraries (CRL), Library of Congress, National Science Foundation (NSF), Online Computer Library Center (OCLC), and librarians from Carnegie Mellon, Haverford College, Indiana University, Pennsylvania State University, Simmons College, Stanford University, University of California Berkeley, University of Chicago, University of Pittsburgh, and University of Washington.

Discussion topics included content focus, copyright considerations, avoiding duplicate scanning, methods of transporting books abroad, and the design of a pilot-study shipment to India.

Participants in the initial collection development meeting swiftly agreed that one million books could not be selected title by title. They also quickly agreed that garnering permission to digitize and provide free-to-read access to copyrighted books would be time consuming and expensive. With these points in mind, the group decided that the Million Book Collection would be a collection of collections, including at least 200,000 indigenous works from India and China, 700,000 public domain works, and a target of 100,000 copyrighted works. The public domain materials would be an assortment of government documents and out-of-copyright (pre-1923) books shipped from the United States. Discussions with university librarians and faculty helped identify subject areas for public domain materials that would be useful to current research.

Efforts to acquire permission to include copyrighted material in the Collection would begin with titles cited in *Books for College Libraries*, which is a five-volume bibliography of books recommended for all academic library collections. The copyright permission work would be considered a separate project requiring separate funding. Everyone agreed that copyright law must be strictly followed for all materials included in the Collection and that letters of assurance must be secured from project partners in India and China. Memorandums of Understanding were completed in 2002.

Carnegie Mellon University Libraries began a pilot study in 2002 to secure non-exclusive copyright permission to include selected titles in the free-to-read Million Book Collection. This work is discussed in a subsequent section of this article. As the pilot evolved, a new strategy for negotiating permission was developed and tested with sufficient success to change the collection development strategy and raise the goal for the Collection to 500,000 copyrighted books.

A second collection development meeting is planned for November 2004 to select additional collections and strategize about how to locate and acquire the books for shipment abroad.

4. Copyright Permission

Approximately 95% of the books ever published are still in copyright, but fewer than 3% of these books are still in print. This means that 92% of the world's books are neither generating revenue for the copyright holder nor easily accessible to potential readers. Furthermore, research shows that during the period when U.S. copyright law required renewal, fewer than 15% of copyrights were renewed (Ringer 1960; Guinan 1961). Clearly most books are abandoned within 28 years of publication. Many academic books appear to be abandoned much more quickly because they frequently go out of print within three years of publication. These findings make U.S. copyright law seem somewhat absurd. Nevertheless the only solution to the current situation is to negotiate permission to digitize and provide open access to copyrighted works.

India and China are responsible for the copyright permission work related to scanning their indigenous materials. Copyright permission work for materials shipped from the United States is being done by Carnegie Mellon University Libraries.

In 2001, Carnegie Mellon University Libraries completed a feasibility study conducted to determine the likelihood of publishers granting permission to digitize copyrighted books and offer them free-to-read on the surface web (George 2001). The study, based on a statistically valid random sample of books in Carnegie Mellon's library catalog, revealed that locating copyright

holders is time-consuming, expensive, and often unsuccessful. Approximately 12% of the publishers could not be located. Half of those located responded to letters of inquiry. Among those who did respond, the success rate for getting permission to digitize their books was 43%. The overall success rate, however, for all the books in the sample was 22%, though the success rate varied with different types of publishers, ranging from 45% for scholarly associations to 12% for commercial publishers. When permission was granted, seldom was it given to offer the digitized book free-to-read on the web. Often restrictions or fees were applied.[1] The feasibility study revealed that it is possible to secure permission to digitize and provide open access to books, but more effective strategies are needed.

A subsequent project (2002-2003) sought permission to digitize and provide open access to the Posner Memorial Collection of fine and rare books, most of them quite old though still in copyright. In this study, 33% of the publishers could not be located. Rather than sending a second letter to publishers who did not respond to the initial letter, follow-up calls were made to the publishers to engage them in conversation, answer questions, and address their concerns. 72% of the publishers contacted responded, and 71% of those responding granted permission. The overall success rate per copyright holder, however, was 34%, accounting for 44% of the copyrighted volumes in the collection. The average transaction cost per volume for which permission was granted was $78. Transaction costs include labor, postage, and telephone charges to identify and locate 114 copyright holders, send 167 letters of request (often only to discover that the address was incorrect or that the copyright had transferred to someone else), and negotiate in 159 phone calls and email messages. Coding of the data has not yet been completed to determine whether the success rate varied with different types of publishers as in the feasibility study, but the coding done to date suggests that authors and estates are as likely to grant permission as university presses (c. 35%).

The two projects described above both took a per-title approach to seeking copyright permission. The Million Book Project initially began with this same approach, requesting permission to digitize and provide open access to the 50,000 titles cited in *Books for College Libraries*. However, as the Posner project unfolded, the per-title approach was deemed too expensive to pursue on a large scale. The approach currently being taken in the Million Book Project is a per-publisher approach based on educating and providing incentives for publishers of quality academic books to grant permission to digitize their out-of-print, in-copyright books and offer them free-to-read on the web.

Letters to publishers briefly introduce the Million Book Project, explicitly state adherence to copyright law, and describe the copyright absurdity wherein out-of-print, in-copyright books are neither generating revenue for the copyright holder nor readily available to potential readers who might be willing to pay for the material. The letters provide an overview of research indicating that users want to find information online, but use it in print (Friedlander 2002); that online access increases use, including use of older materials (Guthrie 2000); and that open access does not decrease revenue and can actually increase sales (Pope 1999). The letters then ask publishers for non-exclusive permission to digitize and offer free-to-read on the web any of the following options:

- All of their out-of-print, in-copyright titles

---

[1] For example, access was to be restricted to the Carnegie Mellon community, a permission fee was levied of $100 to $300 per book, or permission was granted for only a couple years, after which the book would have to be removed from the web.

- All of their titles published prior to a date of their choosing
- All of their titles published *N* or more years ago – they specify *N*
- A list of titles that they specify

The letters explain that the Million Book delivery system will have minimal functionality. Saving and printing will be restricted to one page at a time, like netLibrary books. Carnegie Mellon offers to give participating publishers preservation-quality copies of their digitized books and the associated metadata and OCR text file. To motivate publisher agreement, the letters explain that they can use the electronic files in added-value, fee-based services that they develop or use. For example, Buy buttons and Print-On-Demand service in conjunction with the images could generate revenue for them from the sale of in-print and out-of-print books.

Treating *Books for College Libraries* and other selected collections like an approval plan for publishers obviates the need to check copyright renewal records, reduces the cost of preparing and mailing letters, and with each letter sent potentially secures permission to include more titles in the Million Book Collection than just those in the collection. If only 6% of the 5600 publishers with works cited in *Books for College Libraries* grant permission to digitize 1500 books each, the result will be 504,000 copyrighted works for the Million Book Collection. If negotiations with these publishers achieve the 22% success rate of the feasibility study, the result could be millions of books and the need to negotiate for more labor in India and China.

Results of the 2001 copyright feasibility study revealed that the success rate in seeking copyright permission varies by publisher type. University presses and scholarly associations appear to be three to four times more likely than commercial publishers to grant permission to digitize and provide open access to their copyrighted books on the web. Using this information, copyright permission work for the Million Book Project is focusing on university presses and scholarly associations. Though the initial work was performed intermittently by existing University Libraries staff, a full-time person dedicated to copyright permission work for the Million Book Project was hired in November 2003. A grant proposal is pending with the Institute of Museum and Library Services that would also provide financial support for copyright permission work.

To date the following publishers have granted permission: National Academy Press, Brookings Institution Press, Rand McNally, Louisiana State University Press, Mercer University Press, Southern Illinois University Press, Texas Christian University Press, University of Alabama Press, University of Hawaii Press, University of North Carolina Press, and Wayne State University Press. Ten authors and estates have also agreed to participate in the Project, and a dozen other university presses and scholarly associations have agreed to consider a list of titles provided by the Project. Experiments are underway to find the most cost-effective way to generate for these publishers lists of their out-of-print, in-copyright titles cited in *Books for College Libraries*. The hypothesis of the new copyright permission strategy has been confirmed. For example, though *Books for College Libraries* lists only 26 titles published by the National Academy Press, the publisher granted permission to digitize 3,400 titles. As more publishers agree to participate, the Project might already have permission to digitize many of the copyrighted books in the additional collections to be selected in 2004.

To date, 267 letters have been sent in an attempt to reach 157 copyright holders, and 93 follow-up have been made. Of those with whom negotiations have been completed, 61% have granted permission to digitize and provide open access to many of their copyrighted titles. The number of titles for which permission has been granted is not yet known. Lists must be compiled, for example, of all of a publisher's out-of-print, in-copyright titles or all of the titles they published

prior to a specified date. Project partner OCLC is helping to compile these lists. In other cases, we are waiting for the publisher to provide the list.

5. Digital Registry

Participants at the initial collection development meeting in November 2001 expressed concern about duplicate shipping and scanning of materials. OCLC agreed to explore the development of a three-tiered digital registry that would enable registering books that have been scanned, books selected or queued for scanning, and books that would be desirable to scan. They also agreed to consider allowing books to be registered by organizations that are not OCLC members.

Since the initial meeting, OCLC has invested considerable resources in developing a Digital Registry that will soon be ready for testing. Based on functional requirements produced by the Digital Library Federation (Digital Library Federation 2001) and collaboratively designed by OCLC and the Digital Library Federation Digital Registry Working Group, the pilot enables the registry of born-digital materials and digital reproductions of paper-based monographs, including works intended to be scanned and works already scanned. Carnegie Mellon University Libraries is a member of this Working Group, and in October 2003 visited OCLC to begin discussion of how best to integrate the Digital Registry into the workflow for the Million Book Project.

The Registry captures item identification, access restriction, preservation elements and actions, production or reproduction notes, and system information. The technology enables the conversion of MARC records to MARCXML, adding mandatory and optional registry elements in a batch, and converting the records back to MARC.

Registry signals that the materials will be preserved and accessibility maintained; that digitization complies with established standards and best practices; that the materials are stored in professionally managed systems; and that a use copy is available online for public access (though perhaps not for free).

6. Acquisition and Shipping

The initial discussion in November 2001 of transporting books to India and China acknowledged that sending books by air would be faster but more expensive, while sending books by sea would be slower and less expensive. Participants discussed whether shipments should be coordinated among participating libraries, centralized, or individualized, but no decisions were made. Meeting participants expressed concern about whether the books would be returned and in good condition. All agreed that additional information was needed to make sound decisions. Several attendees at the meeting expressed reservations about participating in the Million Book Project until they knew more about what was entailed in pulling books from their collection, packing and tracking the materials, the time away and safe return of their books, and the quality of the scanning being done in the Project. A pilot shipment was needed to gather data.

In August 2002, the pilot shipment of approximately 6000 books – 243 boxes weighing 11,298 pounds – left New York for Chennai on nine palettes in a twenty-foot ocean container at a cost of $2.00 per book round trip. The trip to India took twenty-five days. From Chennai, the books went to the central distribution center at Deemed University in Tanjore, from where they were

distributed to the scanning centers. Two-thirds of the books did not need to be returned to the United States. Most of the books that were to be returned were returned in good condition in August 2003. The missing books still have not been located.

A year is entirely too long for books to be gone from their home library. Analysis revealed that significant delays were created at customs, because project partners were unfamiliar with the procedures, and in the redistribution of the books from Tanjore to the scanning centers. The pilot provided the requisite education to streamline moving future shipments through customs. To solve the problem of redistribution, several international centers were inaugurated in India in January 2004. In the future, all U.S. shipments will go directly from seaport to these centers, where they will be scanned and boxed for return (if necessary). The second important lesson learned from the pilot was how to reduce the cost to $1.00 per book round trip ($0.50 one way) – by packing the books in crates instead of on palettes, 1500 pounds per crate.

Plans to ship books to China were postponed when the Project encountered content restrictions and customs regulations that would drive up the cost and cause significant delays in the Project. Rather than ship books to China, the initial approach was to send Chinese partners lists of titles for which copyright permission had been granted, have them scan books on the lists that they could locate in their library collections, then update and return the lists. In 2004 the scanning centers in China were declared "free enterprise zones," which means that the Project books will not have to go through customs. Books can now be shipped to China for scanning.

To reduce shipping costs and obviate concerns about the safe return and length of time that books will be away from the home library, efforts are being made to acquire duplicate copies or weeded collections so that the books need only be transported one way. Project partner OCLC is helping locate books among partner libraries in the U.S. The Internet Archives acquired a weeded collection of approximately 100,000 books and shipped them to India.

7. Scanning, Quality Control, and Sustainability

Using equipment funded by the National Science Foundation and labor funded by the respective governments, scanning is underway in India and China. Scanning follows established standards for bibliographic metadata and file formats to ensure interoperability with existing systems and migration to new technologies in the future. Pages are scanned at 600 DPI, the images post-processed to de-skew, de-speckle, and crop using the ScanFix software, and OCR'd using Abby Fine Reader to provide text files that will support full-text searching of the Collection.

Many scanning centers are operational and more are planned. A typical center has 6 to 8 Minolta scanners operating two eight-hour shifts per day. The productivity rate is 16 books per day per scanner or roughly 4000 books per year per scanner. Color scanners and microfilm scanners are being purchased. The goal is to have 100 scanners in operation, digitizing 400,000 books per year. Even if the pace significantly deteriorates, the Project can be completed in five years.

Librarians are responsible for metadata capture. When a MARC record is available for a title, librarians download it from OCLC WorldCat. OCLC provided Guest IDs for non-member libraries. When no MARC record is available, the librarians create a Dublin Core record. Some of the initial work in India did not adhere to these standards, so plans are for OCLC to provide additional training in the near future. The earlier metadata records will be corrected.

The workflow was developed, tested and documented by Carnegie Mellon University Libraries,

which also provided the initial training in India and China. The manual is revised as problems are encountered and solutions developed. For example, a procedure was devised to sort weeded collections, purchased sight unseen and shipped abroad, to ensure that copyrighted books are not scanned without permission. In the near future, registering the Million Book Collection in OCLC's Digital Registry must be integrated into the workflow.

Currently the scanned materials are only available on servers in the country where they were scanned and the user interfaces are not easy to use. Integrating and mirroring the Collection will be topics on the agenda when Indian and Chinese partners convene at Carnegie Mellon in May 2004. The user interface is currently being redesigned by a human-computer interaction class at Carnegie Mellon.

Plans are to mirror the Million Book Collection at sites around the world. Organizations currently committed to host the Collection in perpetuity include Carnegie Mellon, the Internet Archive, and the recently announced Digital Library of India. OCLC might also host the collection and link the books to WorldCat. The goal is to have ten copies of the Collection located around the globe. Estimated hardware, infrastructure, and connectivity cost to host the Collection is one million dollars. Estimated size of the Collection is 20 terabytes.

8. Added-Value Services

Since much of the Million Book Collection will be out-of-print books, the most critically needed added-value service is print on demand. The service has been tested in the United States with the Internet Archive's Internet Bookmobile (Internet Archive 2002), and in India with facilities in Hyderabad and more recently with a Digital e-Library Bookmobile (C-DAC Noida 2003).

The goal of Million Book Project leaders is to have a printed book from the Collection cost whatever a cup of coffee costs in the country in which the print-on-demand book is requested. UMI/ProQuest approached Carnegie Mellon in the fall 2003 with interest in providing print-on-demand service for the Project, but they could not support a diversified global business model. Amazon.com has expressed interest in the Project.

References

C-DAC Noida (2003): *Providing Books at Your Doorsteps: Education for All*. Department of Information Technology, MCIT, Government of India. Available: <http://mobilelibrary.cdacnoida.com/>

Digital Library Federation (2002): *DLF Endorsement for Framework of Guidance for Building Good Digital Collections*. Available: <http://www.diglib.org/standards/imlsframe.htm>

Digital Library Federation (2001): *Registry of Digital Reproductions of Paper-based Monographs and Serials*. Available: <http://www.diglib.org/collections/reg/regpapfunc.htm>

Formative Evaluation of the Distributed National Electronic Resource (EDNER) Project (2002): "How students search: Information seeking and electronic resource use." Issues Paper 8. Available: < http://www.cerlim.ac.uk/edner/ip/ip08.rtf >

Friedlander, A. (October 2002): *Dimensions and Use of the Scholarly Information Environment: Introduction to a Data Set.* Washington DC: Council on Library and Information Resources publication 100. Available: <www.clir.org/pubs/reports/pub110/contents.html>

George, C. (2001): *Exploring the Feasibility of Seeking Copyright Permission*. Available: <http://zeeb.library.cmu.edu/Libraries/FeasibilityStudyFinalReport.pdf>

George, C. (2004): Analysis of raw data from survey of how graduate students discover and obtain information. Report not yet available. Contact Denise Troll Covey (troll@andrew.cmu.edu) or Carole A. George (cgeorge@andrew.cmu.edu).

Guinan, J. J. Jr. (1961): Duration of Copyright. *Copyright Law Revision*. Prepared for the Subcommittee on Patents, Trademarks, and Copyrights of the Committee of the Judiciary, U.S. Senate, 86th Congress, 8.

Guthrie, K. M. (March 23, 2000): *Revitalizing Older Published Literature: Preliminary Lessons from the Use of JSTOR*. Available: <http://www.jstor.org/about/preliminarylessons.html>

Institute for Museum and Library Services (November 2001): *A Framework of Guidance for Building Good Digital Collections*. Available: <http://www.imls.gov/pubs/forumframework.htm>

Internet Archive (2002): *Internet Archive Bookmobile*. Available: <http://www.archive.org/texts/bookmobile.php>

Jones, S.; Madden, M. (September 15, 2002): "The Internet Goes to College: How Students Are Living in the Future with Today's Technology." Pew Internet & American Life Project. Available: < http://www.pewinternet.org/reports/toc.asp?Report=71 >

Lawrence, S.; Giles, L. (1999): Accessibility and Distribution of Information on the Web. *Nature* 400, p. 107-109. Summary of findings available at http://www.wwwmetrics.com

Marcum, D.B.; George, G. (October 2003): "Who Uses What? Report on a National Survey of Information Users in Colleges and Universities." *D-Lib Magazine*, Vol. 9, No. 10. Available: < http://www.dlib.org/october03/george/10george.html >

Online Computer Library Center (OCLC) (June 2002): *How Academic Librarians Can Influence Students' Web-Based Information Choices*. Dublin, Ohio: OCLC White Paper on Information Habits of College Students. Available: <http://www2.oclc.org/oclc/pdf/printondemand/informationhabits.pdf >

Pope, B. K. (June, 1999): National Academy Press: A Case Study. *The Journal of Electronic Publishing* 4, 4 Available: <http://www.press.umich.edu/jep/04-04/pope.html>

Quint, B. (March 3, 2003): Overture Acquires Two Major Web Search Engines. *Information Today, Inc. Online*. Available: <http://www.infotoday.com/newsbreaks/nb030303-1.shtml>

Ringer, B. A. (June 1960): Study No. 31: Renewal of Copyright. In *1 Studies on Copyright*, Arthur Fisher Memorial Ed., p. 513-514.

Contact Information

Denise Troll Covey
Associate Dean, University Libraries
Carnegie Mellon
4909 Frew Street, Hunt Library
Pittsburgh, PA 15213
USA

troll@andrew.cmu.edu