

Acquiring Copyright Permission to Digitize and Provide Open Access to Books

by Denise Troll Covey

Digital Library Federation
Council on Library and Information Resources
Washington, D.C.
October 2005

First Digital Library Federation electronic edition, September 2008

Some rights reserved. Published by the Digital Library Federation

Originally published in trade paperback in the United States by the
Digital Library Federation and the Council on Library and Information Resources, Washington, D.C., 2005

This edition licensed under the Creative Commons Attribution-Noncommercial 3.0 Unported License
<<http://creativecommons.org/licenses/by-nc/3.0/>>

The moral rights of the author have been asserted

Digital Library Federation ISBN-13: 978-1-933645-30-8

www.diglib.org

Contents

About the Author	iv
Acknowledgments	v
Foreword	vi
Introduction	1
A Brief History of Law and Practice.....	2
The Implications.....	7
The Response	9
The Random Sample Feasibility Study	11
Overall Results	13
Analysis of Restrictions	15
Analysis of Foreign and Domestic Publications	16
Analysis by Publisher Type.....	17
Analysis by Publication Type.....	17
Analysis by Print Status and Publication Date.....	18
Analysis of Transaction Costs	20
Conclusions and Lessons Learned	21
The Fine and Rare Book Study	21
Overall Results	26
Analysis of Restrictions	28
Analysis of Foreign and Domestic Publications	29
Analysis by Publisher Type.....	30
Analysis by Publication Type.....	31
Analysis by Print Status and Publication Date.....	31
Analysis of Transaction Costs	34
Conclusions and Lessons Learned	35
The Million Book Project Study	38
Tracking the Data	42
Overall Results	43
Analysis by Publisher Type.....	45
Requests that We Provide Lists of Titles.....	48
Request that We Check Copyright-Renewal Records	51
Experiments to Contain Costs	52
Analysis of Transaction Costs	54
Conclusions and Lessons Learned	54
Looking Ahead	58
References.....	60

About the Author

Denise Troll Covey, principal librarian for special projects at Carnegie Mellon University, is responsible for conducting research to inform library administration and strategic planning. She manages Carnegie Mellon University Libraries' performance measures and keeps abreast of technological developments and their social implications and the laws, policies, practices, and standards relevant to digital libraries. Her current projects are engaging Carnegie Mellon faculty members in developing an institutional repository for their scholarly work and conducting an analysis of the public comments and public hearing transcripts regarding the U.S. Copyright Office's investigation of orphan works. Ms. Covey serves on the National Information Standards Organization Standards Development Committee, where she is leading an initiative to develop rights expression and management for scholarly information. She is also secretary of the Measurement, Assessment, and Evaluation Section of the Library Administration and Management Association. Ms. Covey was a Distinguished Fellow at the Digital Library Federation in 2000–2001.

Acknowledgments

Many people were involved in the copyright-permission work reported here. I thank Lily Waters and Leigh Caskey Schenk of the U.S. Army for doing the groundwork for the feasibility study, Tracey Connelly for continuing the work, and Carole George for seeing the study through to completion and conducting the preliminary data analysis. George also created the database for tracking the data in the Posner project and contributed to the design of the publisher database for the Million Book Project. I thank Ruth Ann Schmidt for her help with the Posner permissions and Cynthia Brown for her work on the Thousand Book Project, her assistance in designing the publisher database, and her help finding publisher addresses for the Million Book Project. The time and effort of the librarians and students who helped find publisher addresses are also much appreciated, as are the reading and editing suggestions for this report provided by Cindy Carroll.

Special thanks go to Erin Rhodes, who did the bulk of the permissions work on the Posner and Million Book projects. Without her efforts, diligence, and persistence, this report would not exist. Though her task was sometimes tedious, she persevered. Though she often felt confused and frustrated, she persevered. Despite the inadequacy of our mechanisms for tracking the data, she persevered. And she never complained. I could have had no better assistant.

Special thanks are also extended to Kathlin Smith of the Council on Library and Information Resources for her careful reading and editing suggestions and to the copyright attorney that she recruited to ensure the accuracy of my overview of copyright law.

Those who funded this work must also be thanked: Henry Posner, Jr., and his wife Helen for funding the Posner copyright-permission work and Bruce Miller at the University of California Libraries at Merced for funding the Million Book Project copyright-permission work.

Finally, I thank Gloriana St. Clair, dean of University Libraries at Carnegie Mellon and one of the directors of the Universal Library Project. Her vision and substantial allocation of my time were essential to what we have accomplished.

Foreword

The contemporary academic library and its users have an appetite for digital copies of books that far outstrips the willingness and ability of publishers to provide such access. In the science disciplines, contemporary and historical journal literature is becoming widely available in digital format, albeit at considerable cost. Access to the scholarly record in digital form is already transforming the manner in which science disciplines communicate, publish, research, and review excellence.

This widespread access is not the case for the mass of works in the humanities, arts, and social sciences. Yet it is in these disciplines that the utility of older scholarly books and journal articles tends to be the greatest. Scholars have great interest in digital access to even the very earliest primary works of literature, history, philosophy, religion, and culture that have appeared in print.

While some of this primary material is available in commercial databases, much of it is not. As a result, libraries are increasingly seeking to negotiate noncommercial, free, public, digital access—open access—to copyrighted and noncopyrighted materials that are not available from scholarly publishers. These materials are typically out of print and have little promise for commercial exploitation, yet they are very much alive to scholarly inquiry. Compounding the problem is that nineteenth- and twentieth-century materials are often in a state of physical decay. This only adds urgency to the library's desire to save these materials for current and future scholarship.

What are the stumbling blocks to digitization? Is copyright law a major barrier? Is it easier to negotiate with some types of publishers than with others? To what extent does the age of the material influence permission decisions? This report, by Denise Troll Covey, principal librarian for special projects at Carnegie Mellon University, responds to many of these questions. It begins with a brief, cogent overview of U.S. copyright laws, licensing practices, and technological developments in publishing that serve as the backdrop for the current environment. It then recounts in detail three efforts undertaken at Carnegie-Mellon University to secure copyright permission to digitize and provide open access to books with scholarly content.

The results of this well-documented, meticulous survey are illuminating. The responses to the author's carefully designed inquiries reveal a picture of confusion and chaos in the face of a significant opportunity and growing need. The range of publisher responses and their requests for fees, restrictions, and caveats show a publishing industry that has in no way reached a consensus on how to respond to libraries' growing desire to provide digital access to scholarly materials. Indeed, some publishers are not even aware of what rights they actually own.

From the expense and difficulty of determining copyright status and locating the owner to the struggle to get a response from a publisher when seeking permission to digitize for scholarly use, this timely report provides a detailed account of the challenges facing libraries today. It should be of practical use to publishers and librarians alike as we try to navigate the current situation and work to improve it, through such innovations as the “orphaned works” legislation that is currently under discussion. The lessons learned and reported will inform and aid the rest of us as we wrestle with the same problems.

David Seaman
Executive Director
Digital Library Federation

Introduction

Information users increasingly look to find materials on the Web. Many scholars and librarians dream of creating a “universal digital library,” where high-quality resources are accessible from their desktops. Realizing this dream—creating a digital library that is comparable to an excellent traditional library and providing open access to it—requires negotiating copyright permission.

This report focuses on three efforts at Carnegie Mellon University to acquire copyright permission to digitize and provide *open access* to books—that is, to make books freely available on the Internet for public use.¹ To provide a context for the studies that form the basis of this report, the report begins with an overview of copyright laws, licensing practices, and technological developments that have brought about dramatic changes in the cost and dissemination of scholarly information. This section also describes the impact that these changes have had on research, learning, and libraries. The three studies, including data analyses that explore the response and success rates with different types of publishers and publications and transaction costs, are then presented in detail. Anecdotes illuminate the effort required and problems encountered in trying to acquire copyright permission for open access, from the difficulty of determining copyright status and ownership and locating copyright owners to the questions, concerns, record-keeping methods, and changing contractual practices that constrain publishers’ embrace of open access. The report describes how lessons learned in each study were applied in the next study and the benefits of flexible and innovative approaches to acquiring copyright permission.

¹ SPARC (Scholarly Publishing and Academic Resources Coalition) provides the following definition of open access: “By ‘open access’ to this literature, we mean its free availability on the public internet, permitting any users to read, download, copy, distribute, print, search, or link to the full texts of these articles, crawl them for indexing, pass them as data to software, or use them for any other lawful purpose, without financial, legal, or technical barriers other than those inseparable from gaining access to the internet itself. The only constraint on reproduction and distribution, and the only role for copyright in this domain, should be to give authors control over the integrity of their work and the right to be properly acknowledged and cited” (*SPARC Open Access Newsletter* 2004).

A Brief History of Law and Practice

In the late eighteenth century, James Madison wanted the newly formed United States to offer temporary monopolies to creators as incentives to continue to create, after which their works would become common property—part of what came to be known as the *public domain*—to foster creativity in others. Thomas Jefferson had reservations about such monopolies based on the history of copyright as an instrument of censorship in England (Vaidhyanathan 2002; Thibadeau 2004).² Nevertheless, our founding fathers gave Congress the power “To promote the Progress of Science and useful Arts, by securing for limited Times to Authors and Inventors the exclusive Right to their respective Writings and Discoveries” (United States Constitution 1789, I, 8, 8). Soon thereafter, the first Congress passed this country’s first copyright law as a bargain between creators and users of intellectual property designed to balance the private interest of creators with the public good of others (Copyright Act of 1790).

The initial term of U.S. copyright, legislated in 1790, was 14 years, with the right to renew copyright for another 14 years during the last year of the initial term if the author was still living.³ Federal copyright protection initially applied only to maps, charts, and books. It granted authors or those to whom they transferred their copyrights the sole right to print, reprint, publish or sell these works (Hirtle 2004).⁴ Over the course of the next two centuries, however, the duration and scope of copyright protection were extended, the requirements for acquiring it were changed, and the rights associated with it were redefined. More recently, new technologies evolved that changed scholarly communication and raised questions about the interpretation and application of copyright.

Table 1 shows significant changes in the copyright term.⁵ The Copyright Act of 1870 doubled the duration of the initial copyright term. The Act of 1909 doubled the duration of the renewal period. It required that works be marked with a standard copyright notice to acquire copyright protection and be deposited and promptly registered with the Copyright Office. The 1909 act recognized the right of owners to reproduce, distribute, perform, or make derivatives of intellectual property and acknowledged *works for hire* as a category of works able to acquire copyright protection. It also codified the doc-

² Some of Jefferson’s letters to Madison are available in the Thomas Jefferson Papers Series 1, General Correspondence, 1651–1827, in the American Memory Collection at the Library of Congress Web site. A detailed analysis with links to the correspondence is available at <http://rack1.ul.cs.cmu.edu/jefferson/>. An alternative account is Thomas Nachbar, “Monopoly, Mercantilism, and Intellectual Property,” to be published in the *University of Virginia Law Review* in October 2005.

³ This was changed in 1831 to enable the author’s widows and children to renew the copyright if the author were dead.

⁴ Claims to copyrights and deposits of copyrighted works were made in U.S. district courts until 1870, when these activities were centralized in the Library of Congress.

⁵ There have been 10 retroactive extensions of the copyright-monopoly term, ranging from 1 to 20 years, since 1962 (Brito 2002, Moglen 2002, 12, 15). The current copyright laws of the United States of America are contained in Title 17 of the U.S. Code of Law.

trine of *first sale*, which allows the owner of a lawful copy of a copyrighted work to sell or dispose possession of that copy.

Table 1. Overview of selected extensions of the copyright term

Year of Copyright Act	1790	1870	1909	1976			1998	
Works	All copyrighted works			Works copyrighted prior to 1978	Works copyrighted 1978 or after		All copyrighted works	
Author	All	All	All	All	Personal	Corporate	Personal	Corporate
Initial term (years)	14	28	28	28	Life + 50	Publication + 75 or creation + 100, whichever is shorter	Life + 70	Publication + 95 or creation + 120, whichever is shorter
Renewal term (years)	14	14	28	47	—	—	—	—
Total years	28	42	56	75	varies	varies	varies	varies

Among the many changes to U.S. copyright law, the Copyright Act of 1976 stands out as one of the most dramatic. That act set the duration of copyright for all works created on or after January 1, 1978, to 50 years following the death of the author⁶ or, for works for hire, 75 years after publication or 100 years after creation, whichever expired first. For works copyrighted prior to 1978, the renewal period was extended by 19 years (the initial 28-year term plus the 28-year renewal period plus 19 years, for a total of 75 years). The 1976 Copyright Act clarified or modified the definition of the rights to reproduce, distribute, perform, or make derivatives of intellectual property, and recognized the right of public display. It preempted state copyright laws, which in some cases had provided copyright protection for unpublished works in perpetuity,⁷ distinguished two types of work for hire, and implemented a variety of compulsory licenses. With a few specified exceptions, the 1976 act required works to be marked with a standard copyright notice to acquire copyright protection—a requirement eliminated in the Berne Convention Implementation Act of 1988. The 1976 act eliminated the requirement of “prompt” registration with the Copyright Office, but provided incentives for doing so. Despite these incentives, many works are not registered today.

The 1976 Copyright Act also defined copyright infringement, its defenses and remedies, and exemptions from liability. Section 107 of the act codified for the first time the doctrine of *fair use* of copyrighted works, wherein use “for purposes of criticism, comment, news

⁶ Changing the copyright term to the life of the author plus 50 years was done to comply with the Berne Convention, which then had about 80 members and now has 159.

⁷ The 1976 Copyright Act gave all unpublished works a copyright term of the life of the author plus 50 years or until the end of 2002, whichever was longer. If, during that period, an unpublished work was published, the copyright term was extended until 2027.

reporting, teaching (including multiple copies for classroom use), scholarship or research, is not an infringement of copyright.” In determining whether a use is fair, a court considers the purpose of the use, the nature of the work, the amount and substantiality of the use in relation to the entire work, and the effect of the use on the market for or value of the work. Section 108 includes limited privileges that allow libraries under certain circumstances to make copies for preservation, replacement, or distribution directly to patrons or through interlibrary loan. Section 109 confirms the doctrine of first sale.

Once the 1976 Copyright Act went into effect, copyright protection of both published and unpublished work began the moment that an original work was rendered or fixed in tangible form. Over the next two decades, additional laws were enacted to confirm or extend the scope of copyright, for example, to confirm copyright protection for software (1980), to include the moral rights of creators of selected visual arts (1990), and to protect constructed architectural works (1990) (Copyright Law of the United States of America 2003, iii–viii). Copyright protection now applies to any work “fixed by any method now known or later developed, and from which the work can be perceived, reproduced, or otherwise communicated, either directly or with the aid of a machine or device” (Copyright Law of the United States of America 2003, 2). It does not apply to ideas, facts, titles, names, slogans, procedures, processes, methods, concepts, principles, blank forms, or works produced by the U.S. government.

A subsequent law in 1992, the Copyright Renewal Act, automatically renewed all copyrights secured between 1964 and 1977 and not renewed by the copyright owner, the rationale being that inadvertent failure to comply with formalities such as renewal could result in loss of copyright (Copyright Renewal Act 1992). Six years later, the Sonny Bono Copyright Term Extension Act (CTEA) extended the copyright to the life of the creator plus 70 years or, for works for hire, 95 years from the date of publication or 120 years from the date of creation, whichever expires first (Copyright Term Extension Act 1998). The CTEA extension applied to all works that were copyright protected at the time the law went into effect. Critics of the CTEA dubbed it the “Mickey Mouse Act” because of the Walt Disney Corporation’s active support of the legislation, which prevented Mickey Mouse from entering the public domain for another 20 years (see, for example, Ellis 1999).

In 2002, attempts were made in the case of *Eldred versus Ashcroft* to have the CTEA declared unconstitutional (Downes 2002). The amicus brief argued that repeated retroactive extensions of the copyright term threatened to enact perpetuity by means of installments and to undermine the system of free expression protected by the First Amendment to the Constitution (Brief for Petitioners 2002, 6).⁸ In 2003, the U.S. Supreme Court upheld the CTEA by a vote of

⁸ The charge that proponents of copyright term extensions aim to restrict access to intellectual property in perpetuity is not unfounded. Jack Valenti, former head of the Motion Picture Association of America, suggested that copyright should endure “forever minus a day” (quoted in Boynton 2005). Mary Bono has stated that her late husband Sonny Bono, champion of the CTEA, “wanted the term of copyright protection to last forever” (144 Congressional Record 1998, H9951).

seven to two, noting that although the extension was perhaps unwise on policy grounds, it was nevertheless constitutional (537 U.S. 2003, 17).

Before proceeding further with this discussion of copyright, it is necessary to interject a discussion of two other significant phenomena that have had a strong impact on copyright legislation and practice. First, in the 1960s and 1970s, commercial publishers began acquiring copyright ownership of more and more scholarly work (Pew Higher Education Roundtable 1998). Second, in the 1980s and 1990s, new technologies precipitated dramatic changes in how people create, access, and use intellectual property. The collision of these phenomena has had a profound impact on scholarly communication.

Information technologies enabled scholarly resources to be distributed on the Internet. With the invention of the World Wide Web and the provision of full-text resources online, academic users quickly came to prefer the convenience of Web access to going to the library. The shift to online distribution of scholarly information was accompanied by a shift in library acquisitions, from purchased ownership to licensed access. Many publishers charge a significantly higher price for online access than for traditional access to their content. As a result of their desire to satisfy increasing user demand for online access, libraries now spend more money for materials, but acquire fewer materials, than they did previously. In response to extraordinary increases in the prices of scholarly journals, many libraries have canceled subscriptions—to print, and even to some online journals. As the number of subscriptions decreases, some publishers raise the prices of their journals, which only leads to more cancellations. The spiraling cycle of decreased subscriptions and increased prices is untenable over the long haul.

Thus, we face a paradox. On the one hand, the Web offers easy, speedy, convenient access to abundant content, more content than was ever readily available before. On the other hand, canceled subscriptions and the acquisition of fewer materials by libraries suggest a decline in scholarly resources available to a particular community. This affects not only the research conducted but also the impact of the published research results, as fewer libraries can afford to license the journals or purchase the books. The trend to restrict access, enabled by copyright and contract law, has been referred to as “the progressive commoditization of knowledge” (De Rosa, Dempsey, and Wilson 2003).

Licenses restrict access to members of the licensing community. In terms of the Web, commercially licensed materials reside in the *deep Web*, inaccessible using popular Internet search engines such as Google, which index only materials on the *surface Web*. Furthermore, licenses are covered by contract law: In practice, licenses need not grant public rights such as fair use or interlibrary loan. In conjunction with digital rights management (DRM) technologies, commercially licensed systems control who can access a resource and what they can do with it. When libraries license access to a resource, they agree to the terms of the license and the restrictions of DRM implemented in the delivery system. DRM systems cannot recognize pub-

lic rights such as fair use. Furthermore, they can take the approach that if a right is not explicitly granted, it is denied, thereby prohibiting any innovative use in the future.

The online distribution of commercial information precipitated new laws. Perhaps the most striking is the 1998 Digital Millennium Copyright Act (DMCA), which made it illegal to circumvent, remove, impair, or deactivate technological protections against unlawful access, and illegal to manufacture, sell, or distribute code-cracking devices that would enable unauthorized access or copying (Digital Millennium Copyright Act 1998).⁹ Critics argue that, in effect, the DMCA legalized whatever rights or restrictions copyright holders implemented in computer code (Electronic Privacy Information Center 2002; Lohmann 2002).

Current licensing practices and technological protections, in conjunction with the anticircumvention law, in many cases make it impossible to exercise the first-sale doctrine in the digital environment. In a required follow-up study of the DMCA in 2001, the U.S. Copyright Office concluded that the doctrine of first sale does not apply to online resources. The explanation given was that the doctrine was designed as a distribution right applicable to tangible works where distribution is limited by geography and the natural degradation of the physical work. Digital works are intangible and their distribution infringes the reproduction right of the copyright holder.¹⁰

Fair use is also at risk in the digital realm. Efforts to develop guidelines for fair use of digital works in education and libraries, initiated at the Conference on Fair Use in 1994, failed for the most part (Conference on Fair Use 1998). Little progress has been made in this arena, with the exception of the Technology, Education, and Copyright Harmonization (TEACH) Act passed in June 2001. The TEACH Act legalized the temporary storage and transmission of limited portions of a performance or display, comparable to what could be done in the timeframe of a live classroom session, by educational institutions without their having to acquire permission from the copyright holder (S. 487 2001).

The flurry of proposed legislation and current litigation pertaining to copyright law and related public policies is beyond the scope of this report. Interested readers are encouraged to visit the Web sites of the U.S. Copyright Office, the Electronic Frontier Foundation, and Public Knowledge as starting points for keeping informed.¹¹

⁹ The DMCA (Title 17 Section 1201) can be construed as an attempt to promote innovative distribution methods and broader access to copyrighted works by protecting technological measures that safeguard the rights of copyright owners (Copyright Law of the United States of America 2003). However, many view DRM less charitably: "DRM is fundamentally about enforcement and prohibition: 'do not inspect (it's illegal even to *try* to inspect!), do not repurpose, do not copy or play more than the allowable amount, do not expect to be able to read/listen/play once your rights have expired, unless you send us more money'" (Open Digital Rights Language Initiative 2005).

¹⁰ The Copyright Office ruled that the concerns raised by librarians were contractual issues beyond the scope of their investigation (U.S. Copyright Office 2001).

¹¹ See <http://www.copyright.gov/>, <http://www.eff.org/> and <http://www.publicknowledge.org/>.

The Implications

All the copyright term extensions described in the previous section of this report diminish the rate at which creative works enter the public domain. Under current copyright law, if a work is in the public domain, anyone can reproduce, distribute, make derivative works of, or perform or display the work publicly without permission or payment. Legal allowances for use of copyrighted works without permission, such as the doctrine of fair use, the TEACH Act, and library copying privileges, are limited, and the circumstances of their application are sufficiently ambiguous to deter their use.¹² While a work is copyright protected, people more often than not must request permission and often pay the copyright holder a fee for the right to reproduce, distribute, make derivative works of, or perform or display a work. Copyright holders can grant one or more of these rights, and they may do so exclusively or nonexclusively. The many retroactive extensions of the copyright term since 1962 keep “substantially all works with otherwise-expiring copyrights out of the public domain for a generation” (Moglen 2002, 12).

Given the mushrooming volume of publishing over the past century and the current duration of the copyright term, we can assume that the number of books currently in the public domain is relatively small in comparison with the number of books still protected by copyright.¹³ We can also safely assume that most of those books are no longer in print.¹⁴ The commercial marketplace offers limited access to out-of-print books. Libraries supposedly provide access to these books (Lessig 2004).

So what is happening to these millions of out-of-print books presumably residing on shelves in a library or offsite storage facility? If not weeded from the collection, books printed on acidic paper are slowly turning to dust. As fewer copies remain and as they become more brittle, these books cease to circulate or to be available for inter-library loan, making them virtually inaccessible to potential readers. Copyright law allows libraries to make up to three physical copies of a deteriorating book if it is not otherwise available. Given user preferences for online access, however, libraries are not likely to invest

¹² Ample evidence that ambiguity in the law and fear of litigation result in self-censorship and gatekeeping, rather than the exercise of fair use rights, is provided in the responses to the U.S. Copyright Office Notice of Inquiry regarding orphan works (see <http://www.copyright.gov/orphan/>). Similar difficulties have resulted in failure to take advantage of the TEACH Act (see Hutchinson 2003).

¹³ To get some sense of the size of the public domain, in July 2004 Michael Lesk analyzed 36 million catalog records for books in the RLG database. According to his analysis, 6.5 million (18 percent) of these books, published in English, French, German, Italian, and Spanish, are in the public domain (Lesk 2004a). Although there are books in the public domain published in other languages and Lesk’s study was preliminary, the data provide some sense of the relative size of the public domain of books.

¹⁴ According to Lawrence Lessig, of the 10,027 books published in the United States in the 1930s, less than 2 percent are still in print, and the number of out-of-print books far exceeds the number of in-print books through subsequent decades (Lessig 2004). Additional data on the print status of books published in the United States are provided on pages 18–19 of this report.

their limited resources in making and storing physical copies. Copyright law also allows digitization for preservation purposes in certain circumstances, but access to the online copy must be restricted to users physically in the library that created the digital copy. To provide open access, or even authenticated remote access, to these digitized works requires permission from the copyright owner of each title.¹⁵ It is no wonder that according to a recent survey, 89 percent of librarians agree or strongly agree with the statement: “Copyright issues are one of the major challenges to the building of the digital library” (Carroll 2004, 9).

In 2004, Brewster Kahle and Richard Prelinger challenged the constitutionality of existing copyright laws on grounds that the copyright system denies public access to works protected by copyright but no longer available in print without benefiting the creator or the public. The argument raised questions about the constitutional bargain between private interest and public good and focused on the fact that “the copyright system contains no mechanisms to create and maintain useful records of copyright ownership” (Stanford Law School Center for Internet and Society 2004). In the absence of such records, “people who would like to distribute or use orphaned works—digital libraries, or creators who would like to include the work in their own creative expression—often are unable to clear rights” (Stanford Law School Center for Internet and Society 2004). A federal district court in California dismissed the case, but it is currently on appeal to the Ninth Circuit Court of Appeals.¹⁶

However, on January 26, 2005, the U.S. Copyright Office issued a notice of inquiry regarding *orphan works*, tentatively defined as “copyrighted works whose owners are difficult or even impossible to locate.” Prompted by the Senate Judiciary Committee and with support from the House Judiciary Committee, the inquiry is part of an investigation to determine “whether orphaned works are being needlessly removed from public access and their dissemination inhibited” (U.S. Copyright Office 2005). The Copyright Office received 721 initial comments and 146 reply comments in response to its notice, many of which provided detailed answers to the specific questions posed in regard to the age, identification, and designation of orphan works, and the nature of the problems faced by people who want to use them. Many responses also elaborated what remedies should be available to copyright owners who later come forward to challenge the orphan status of their work and the infringement by users.

Recalling the era when U.S. copyright law required renewal to retain or extend copyright for a longer term, one might think that

¹⁵ Ironically, in the appeal of the 1999 *Eldred v. Reno* Supreme Court decision, the U.S. Court of Appeals for the District of Columbia acknowledged the rational basis of the Congressional decision that certain classes of works, such as films, would not be preserved if copyright was *not* extended (U.S. Court of Appeals 2001).

¹⁶ *Kahle v. Ashcroft*, 2004 U.S. Dist. Lexis 24090 (N.D. Cal. Nov 19, 2004), on appeal sub nom. *Kahle v. Gonzales*.

data on copyright renewals could shed light on the rate at which copyrighted works were abandoned by their owners. Research conducted by the Copyright Office in 1961 revealed that less than 15 percent of all registered copyrights were renewed, and that the renewal rate for books was only 7 percent (Ringer 1961, 220). Michael Lesk's recent analysis of two million books published in the United States from 1923 through 1963 revealed that less than 10 percent had their copyrights renewed (Lesk 2004b). The unanswered and unanswerable question is whether the low rate of renewal was inadvertent or intentional. Were these books abandoned because the copyright owners no longer wanted to exercise their rights or because they failed to comply with the formality of copyright renewal in the requisite timeframe? Another compelling and unanswerable question is whether past practice (i.e., the low rate of copyright renewal 40 or more years ago) is necessarily predictive of current or future behavior in a radically different technological environment for the creation and dissemination of copyrighted work.

The Response

Although capitalism has historically trusted the marketplace to be self-correcting over time, by the mid-1990s there were serious problems in the market for scholarly communication. This had several significant results.

Faculty members began putting their work on the surface Web, where access is free, scholarly or educational use is unrestricted, and their work can easily be found using popular Internet search engines. Over time, this grassroots phenomenon became known as the *open-access movement*. In 1997, the Association of Research Libraries initiated the Scholarly Publishing and Academic Resources Coalition (SPARC), which aims to lower the cost and expand the online dissemination and use of peer-reviewed scholarly work by contributing to the development of open-access journals and competitive alternatives to expensive commercial journals, promoting fundamental changes in the system and culture of scholarly communication, and raising awareness of the relevant issues.¹⁷ The movement to provide free online access to scholarly articles was aided significantly by the international Budapest Open Access Initiative in 2002.¹⁸ Since then, substantial research has been conducted to determine the impact

¹⁷ SPARC is developing resources and tools to help authors retain their right to self-archive their work on a personal or an institutional Web site. See <http://www.arl.org/sparc/resources/copyres.html> for information about intellectual property rights and retaining the right to self-archive. The SPARC Open Access Newsletter (SOAN) by Peter Suber, available at <http://www.arl.org/sparc/soa/index.html>, is a rich and timely resource for following the open-access movement. For more information, see <http://www.arl.org/sparc/>.

¹⁸ For information on the BOAI, see <http://www.soros.org/openaccess/read.shtml>.

of open access and to address the concerns of various stakeholders in the scholarly information supply chain.¹⁹ Perhaps the most significant research conducted, in terms of promoting the open-access movement, is the research confirming that open access increases use of material and need not decrease sales when a given work also appears in a commercial publication (see, for example, Pope 1999, Lawrence 2001, Antelman 2004, Harnad and Brody 2004).

The efforts of SPARC and other organizations and individuals engaged in the open-access movement have yielded results. Although definitions of what constitutes open access, in terms of how promptly after publication a work must be made available on the surface Web, vary somewhat among the players, the movement to liberate scholarly work from the deep Web is afoot with intensity. The number of agencies and foundations that require or encourage open access to publications based on research they funded is increasing.²⁰ The number of peer-reviewed, open-access journals is increasing. Some prominent commercial journals have started offering authors the option of paying to have their published work available through open access (Gass and Doyle 2005). The number of universities creating institutional repositories to provide open access to their scholarly assets is further evidence of the spread of open-access initiatives. Though there is much debate about who will pay for open access, consensus regarding the benefits makes it unlikely that the movement will halt any time soon (Davis et. al 2004; Gass and Doyle 2005).

Users clearly prefer the ease and convenience of surface Web access to information. Just as clearly, current copyright laws and licensing practices interfere with meeting their needs and expectations. Most students and faculty (50 percent to 90 percent) perceive a significant gap between their high-priority needs and the service their library is providing (LibQual+™ 2002, 2003). Despite the burgeoning success of the open-access movement, a tremendous amount of work remains to be done. To date, the open-access movement has focused on scholarly journals, but libraries contain more than journals. Creating a digital library that is comparable to an excellent traditional library requires negotiating copyright permission to digitize and to

¹⁹ For example, the Rights Metadata for Open Archiving (RoMEO) project in the United Kingdom in 2002–2003 investigated intellectual property rights issues surrounding the self-archiving of academic research and developed metadata elements for rights information that could be harvested with the Open Archives Initiative (OAI) Protocol for Metadata Harvesting. For information about RoMEO, see <http://www.lboro.ac.uk/departments/lis/disresearch/romeo/>. A related project in the United Kingdom, Securing a Hybrid Environment for Research Preservation and Access (SHERPA), created a searchable database of publisher copyright and self-archiving policies to help authors and libraries discover publishers that allow the archiving of pre- or postprint articles on the surface Web—a.k.a. green publishers. The database is available at <http://www.sherpa.ac.uk/romeo.php>. For information about SHERPA (2002–2005), see <http://www.sherpa.ac.uk/>.

²⁰ The widely publicized National Institutes of Health (NIH) initiative to encourage open access to NIH-funded research within 12 months of publication is evidence of this trend.

provide open access to an array of materials. Given the cost of acquiring and storing redundant library collections, it behooves libraries to explore the possibility of acquiring permission to digitize and provide open access to different kinds of materials.

What follows is a detailed look at three studies conducted by Carnegie Mellon University Libraries to acquire copyright permission to digitize and provide open access to books. The first study was conducted to determine the feasibility of acquiring copyright permission for open access to books. The second and third studies, informed by the results of the feasibility study, were conducted as components of real digitization projects. The work illuminates problems and complexities relevant to the designation of orphan books.

The Random Sample Feasibility Study

Between 1999 and 2001, the Carnegie Mellon University Libraries conducted a feasibility study to determine the likelihood of publishers granting nonexclusive permission to digitize and provide surface Web access to their copyrighted books. The primary goal of the project was to develop an understanding of the process, the time it takes, and the problems encountered. We also wanted to ascertain whether different types of publishers responded differently and whether they responded differently on the basis of the type or print status of their publications.

We consulted a statistician on campus to ensure that the random sample of books we selected from our library catalog would yield statistically valid results. The random sample contained 368 titles. We created a database to track the study. Each record in the database contained fields for capturing the bibliographic information about a title, whether it was in or out of copyright, the name and contact information of the publisher, dates for when initial and follow-up letters were sent, details about the publisher's response, and whether permission was granted or denied. Publishers were given the option of providing open access or of restricting access to Carnegie Mellon users. The database had fields to capture this information and was later amended to capture additional restrictions that publishers applied. The database also enabled coding the type of publisher, type of publication, and whether the title was in or out of print.

The study took two years to complete because it was conducted with intermittent labor. Overall, four people worked on the project, including two visiting librarians from the U.S. Army, Lily Waters and Leigh Caskey Schenk. Waters designed the database and helped populate it with the bibliographic information about the books. Two other researchers, Tracey Connelly and Carole George, subsequently worked on the project, with George completing the preliminary data analysis (George 2001). Meanwhile, librarians coded the print status and type of publisher and publication for each title in the sample.

Of the 368 titles in the sample, 351 (95 percent) were copyright protected. Upon initial examination, 10 percent of the copyrighted titles were eliminated from the study because they were technical reports or theses that had been mistakenly cataloged as books. We also eliminated 3 percent of the books when third-party copyright ownership, for example, of charts, illustrations, or photographs, would have complicated the pursuit of copyright permission. As the study proceeded, another 8 percent of the titles were eliminated when publishers introduced complications from third-party ownership. Ultimately, 11 percent of the copyrighted titles were eliminated as too complicated to pursue. The final sample for which we were seeking copyright permission included 277 titles published by 209 publishers.

Our plan was to send letters to the publishers requesting nonexclusive permission to digitize and to provide free-to-read Web access to their copyrighted books in the sample. If we received no response in a month, we would send a follow-up letter. The initial request letter and follow-up letters were somewhat different:

- The initial request letter described Carnegie Mellon University Libraries' collaboration with the School of Computer Science on the Universal Library Project, which aims to digitize the cultural and intellectual history of humankind. The letter referenced the experience of the National Academies Press when it began to provide open access to its books (open access did not decrease sales) and emphasized digitization as a way for our libraries to address the "urgent need for more space to store physical volumes." The letter asked publishers to tell us who owned the copyright to their titles if they no longer did or if they did not own the copyright to a work in its entirety. It also explained that this was a research project and provided a brief overview of what we expected to learn.
- The follow-up letter referenced the date of the initial request letter, summarized its contents, and further explained that we were working from a random sample of books in our collection. It ended with the provocative statement: "If we do not receive a response from you within 60 days of mailing this letter, we will assume that you have granted permission to digitize the book and offer it free to read by anyone on the Internet." Though we had no intention of digitizing books without permission, we included this statement to elicit a response. To our surprise, only one publisher commented on this approach.²¹

We included a contract with both letters. The contract offered options for publishers to deny permission or to grant permission either for open access or for access restricted to Carnegie Mellon users.

²¹ In March 2001, we received a letter from a publisher explaining that the "threat" of proceeding to digitize titles without permission "is most unprofessional and undermines the worthy objectives of the project." He suggested that we revise our approach and our letter. Nevertheless, this publisher did grant permission to digitize and provide Web access to its title in the random sample.

The first lesson learned was that identifying and locating copyright holders is time-consuming and often unsuccessful. Publishers move, merge, or go out of business, or copyright reverts to the author. Resources used to locate addresses included *Global Books in Print*, *Literary Market Place*, and Internet search engines. We failed to find addresses for 7 percent of the publishers. We sent an initial copyright-permission request letter to each publisher that we could locate. Sometimes we sent initial request letters for the same title to different publishers because the first copyright holder contacted no longer owned the rights and responded with a referral—typically without an address, which started the arduous process of locating the copyright owner all over again. Many letters were returned marked “Address (or Addressee) unknown.”

If the initial letter appeared to have been successfully delivered but we got no response, we sent a follow-up letter. More than 60 percent of the publishers contacted required a second or third letter. The average length of time to receive a response from a publisher was 101 days from the date of the initial letter for a response of “Permission granted,” and 124 days for a response of “Permission denied.” The time to respond was probably affected by our use of intermittent labor, which caused delays in sending follow-up letters. We had planned that follow-up letters would be sent one month after the initial request letter, but two months or more often passed between sending the initial and follow-up letters.

We sent a total of 524 letters: 278 initial request letters and 246 follow-up letters. The number of letters was unnecessarily high because we sent separate letters for each title, rather than sending one letter per publisher that bundled all their titles into one request.

Overall Results

Ultimately, 21 percent of the publishers, accounting for 19 percent of the titles in the sample, could not be located. Half of the publishers of books in the final sample responded to our request letters, and more than a fourth of them granted permission, enabling us to digitize and provide Web access to about a fourth of the copyrighted books in the sample (figure 1).

The preceding analysis of the full sample of publishers and titles sheds light on the difficulty of locating publishers, soliciting a response, and securing copyright permission to digitize and provide Web access to books. However, it skews the success rate in the sense that it measures the success of permissions granted in a context that includes publishers that were never contacted.

Another way of viewing the data is to look only at the publishers we located and the titles in the final sample to which they held the copyright. Looking only at these publishers and titles, more than a third of the publishers did not respond to our letters and more than a third of them granted permission. The permissions granted enabled us to digitize and provide Web access to less than a third of the books in the sample issued by the publishers we contacted (figure 2).

Fig. 1. Analysis of the final random sample of 209 publishers and 277 titles

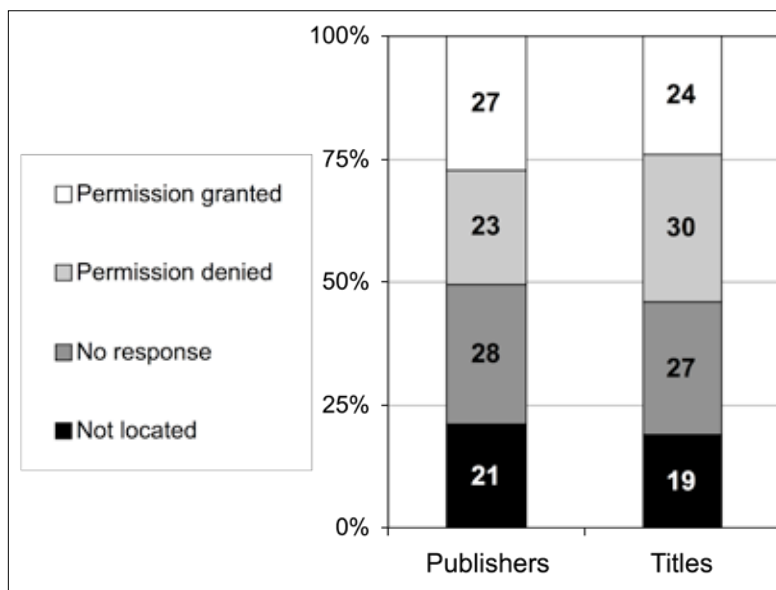
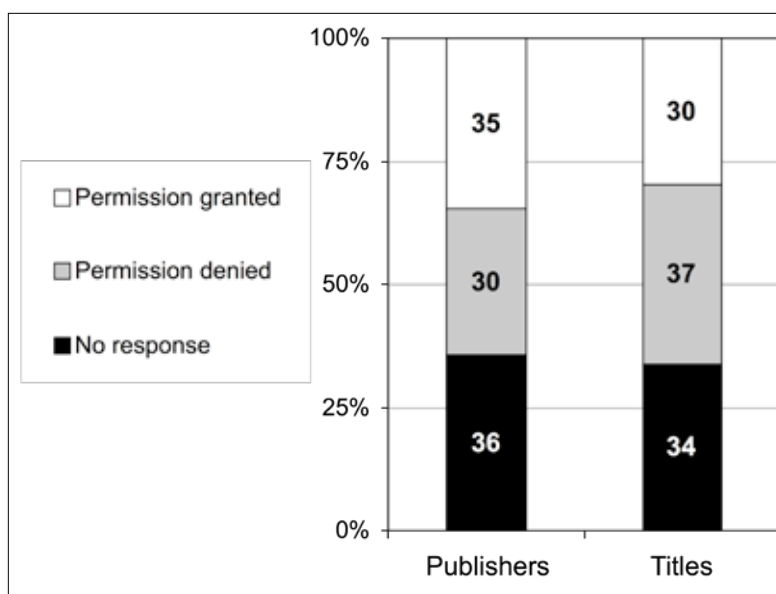


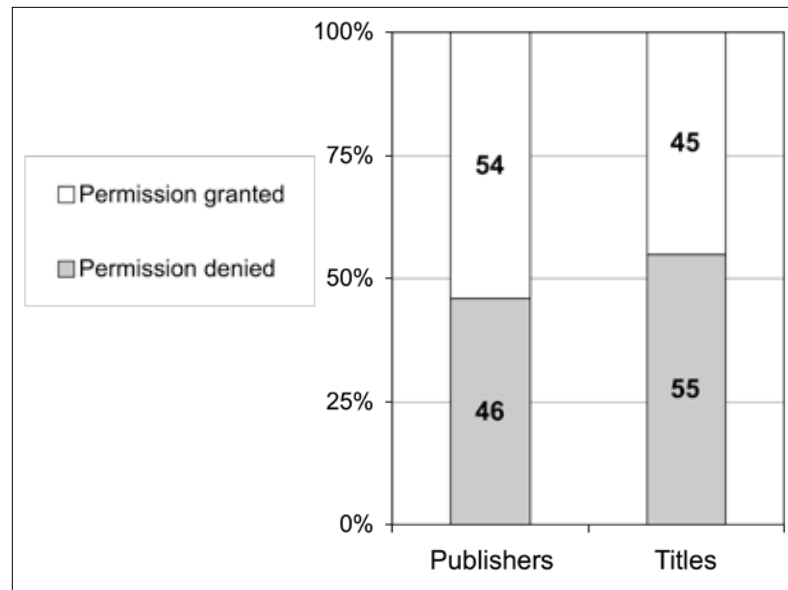
Fig. 2. Analysis of the publishers successfully located



By the time we were analyzing the data from the feasibility study, we had started seeking copyright permission to digitize and provide Web access to books in the Posner Memorial Collection and had revised our process to try to increase the response and success rates. (The Posner study is described later in this report.) We were beginning to believe that increasing the response rate would require one set of strategies and that increasing the success rate among those that did respond would require another. With this in mind and for the purpose of future comparisons, we analyzed the publisher responses in the feasibility study. Looking only at the publishers that

responded and the titles to which they held copyright, more than half of the publishers granted permission for almost half of the titles (figure 3).

Fig. 3. Analysis of completed negotiations



Analysis of Restrictions

The copyright permission request letter offered an option to restrict access to the Carnegie Mellon community, but many publishers mandated other restrictions. Overall, 68 percent of the publishers that granted permission applied some kind of restriction. U.S. publishers were slightly more likely to apply restrictions than foreign publishers were. The most common restriction related to access. Access to more than half of the titles for which permission was granted was restricted to Carnegie Mellon users. Publishers also applied the restrictions or stipulations listed below. The data are based on the number of titles to which the restriction applied, rather than the number of publishers that applied the restriction, because publishers of multiple titles in the sample sometimes applied different restrictions to different titles.

- 23 percent required the display of the full citation.
- 22 percent stipulated that their permission did not apply to components of the work with copyright owned by a third party.
- 8 percent limited the duration of the license to three or four years from the date the contract was signed. After this date, the book would have to be removed from the Web.
- 6 percent prohibited simultaneous use of the digitized book.
- 6 percent required a fee (ranging from \$50 to \$300 per title).
- 3 percent limited permission to digitize the book to a designated date after the contract was signed. Once this date had passed, the book could not be scanned.

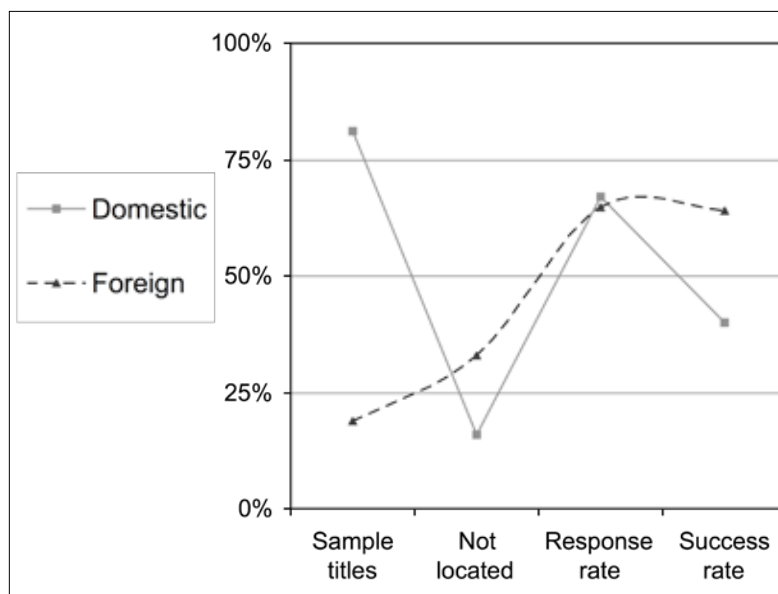
Of the publishers that stipulated that permission did not apply to components of the work with copyright owned by a third party, more than one-third limited the duration of the license to provide Web access to the title, and some also prohibited simultaneous use. Most of the publishers that levied a fee also limited the duration of the license. In addition, 15 percent of the publishers that granted permission requested a copy of their digitized books.

The analyses that follow are based on the number of titles, rather than the number of publishers, in the final sample because publishers with multiple books in the sample sometimes granted permission for some, but not all, of their titles. The *response rate* is based on the number of titles with copyright owned by publishers we successfully contacted. The *success rate* is based on the number of titles with copyright owned by publishers that responded.

Analysis of Foreign and Domestic Publications

Most of the books in the final sample were published in the United States. Foreign publishers were twice as difficult to locate as U.S. publishers. If we located them, the response rates for foreign and domestic publishers were roughly the same. The foreign publishers were more likely to grant permission than U.S. publishers were (figure 4).

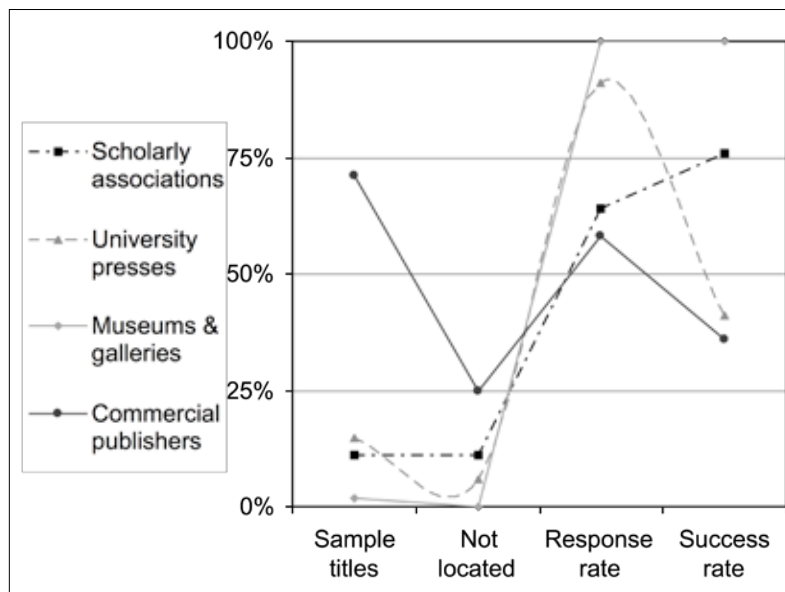
Fig. 4. Analysis of foreign and domestic titles



Analysis by Publisher Type

The response and success rates varied across different types of publishers (figure 5). Although museums and galleries published very little of the content in our sample, they were easy to locate and always responded and granted permission. University presses and scholarly associations also published little of the content in the sample and were relatively easy to locate. University presses were far more likely to respond, but much less likely to grant permission, than scholarly associations were. Most of the books in the sample were published by commercial publishers. They were the most difficult to locate, least likely to respond, and least likely to grant permission. Scholarly associations were slightly more likely to respond than commercial publishers, and university presses were slightly more likely to grant permission than commercial publishers.

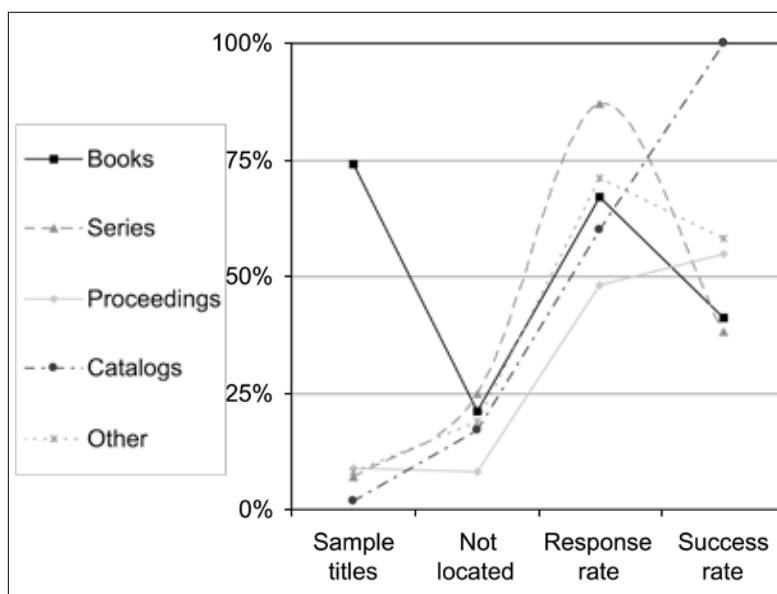
Fig. 5. Analysis by publisher type



Analysis by Publication Type

The response and success rates also varied with different types of publications (figure 6). Most of the sample content was traditional monographs. Monograph publishers were somewhat difficult to locate. Though likely to respond, they were not very likely to grant permission. Publishers of the few series in the sample were the most difficult to locate, the most likely to respond, and the least likely to grant permission. The few publishers of exhibit catalogs were likely to respond and always granted permission. Publishers of the few conference proceedings were the easiest to locate and the least likely to respond; more than half granted permission.

Fig. 6. Analysis by publication type



Analysis by Print Status and Publication Date

Most of the books in the sample were out of print (figure 7). Publishers of out-of-print books were more difficult to locate, less likely to respond, and more likely to grant permission than were publishers of books that were still in print.

Fig. 7. Analysis by print status

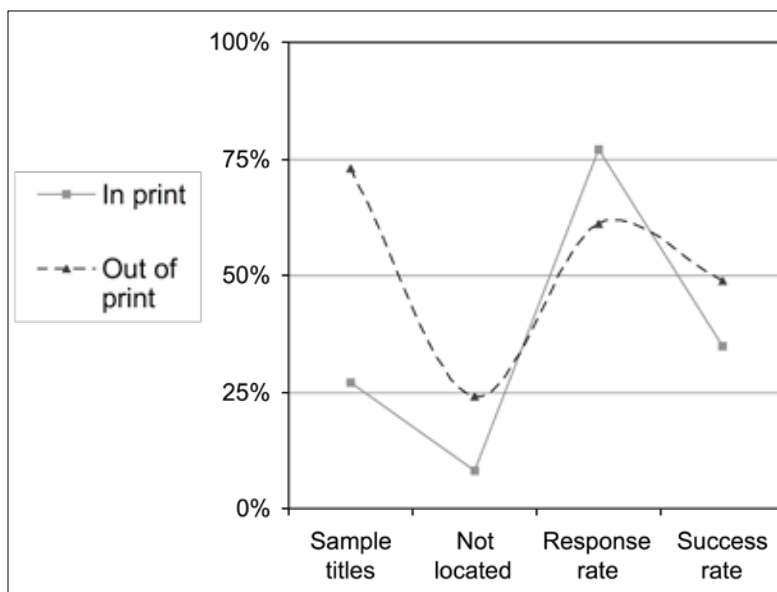


Figure 8 shows the distribution of titles in the sample by publication date and print status. Most of the titles were published relatively recently; only one-third were published before 1970. All the titles published before 1940 and almost all the titles published 1940 to 1960 are out of print. Books in print outnumber books out of print in the sample only in the decade 1990 to 2000.

Fig. 8. Analysis of print status by publication date (number of titles)

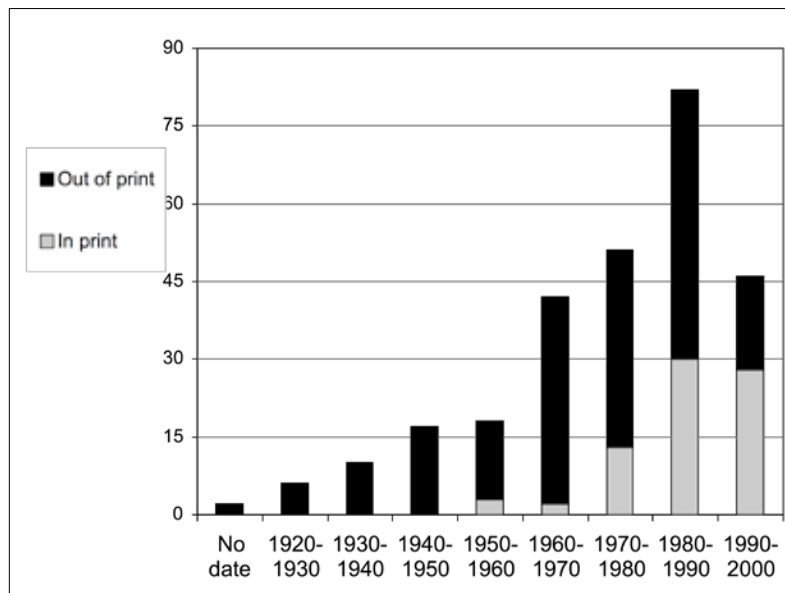


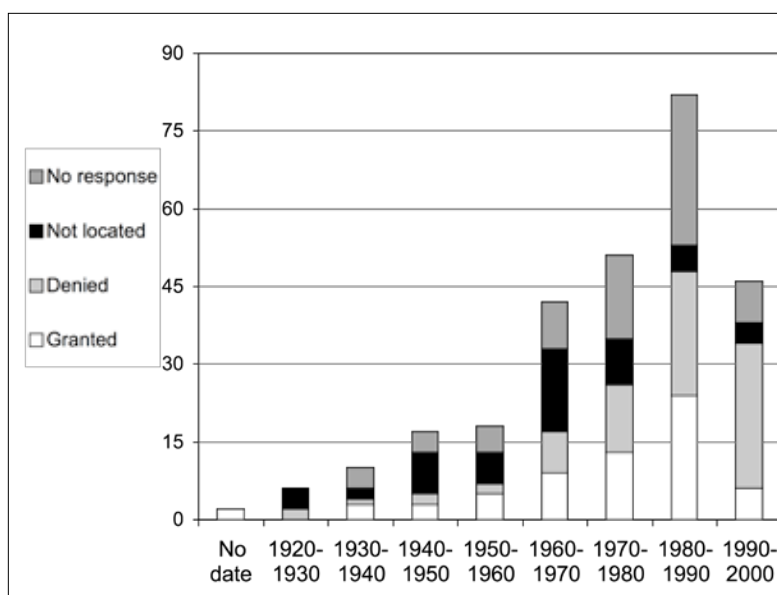
Figure 9 shows the results of our efforts to secure copyright permission by publication date. Because the number of titles in the sample published during each decade varied significantly, the data must be interpreted cautiously. The results suggest that the age of the work did affect the results, but not always in ways we expected.

- With rare exceptions, the older the work, the more difficult it was to locate the publisher. We could not find the publishers of most of the books published between 1920 and 1930 and of almost half of the books published between 1940 and 1950. Publishers of more than a third of the books published from 1950 to 1960 and 1960 to 1970 could not be found. By contrast, few of the publishers of books published 1980 or later could not be found.
- When we could locate the publisher, there did not appear to be a correlation between the date of publication and the response rate. We received no response regarding 30 percent to 40 percent of the 1930–1940, 1970–1980, and 1980–1990 samples, and no response regarding 20 percent to 30 percent of 1940–1950, 1950–1960, and 1960–1970 titles.
- Although permission was sometimes denied for older titles and granted for more recently published titles, the overall trend was as expected: The more recent the date of publication, the more likely that permission would be denied. Permission was denied for more

than half of the titles in the sample published between 1990 and 2000, accounting for 35 percent of the total permissions denied in the study. Only 17 percent of the titles in the sample were published between 1990 and 2000.

- Permission was granted for 20 percent to 30 percent of the titles in the sample published in the 1930s, 1940s, 1950s, 1960s, 1970s, and 1980s. However, with the exception of one decade, the percentage of total titles in the sample published in a given decade was roughly equivalent to that decade's percentage of the total permissions granted in the study. For example, books published between 1960 and 1970 constituted 15 percent of the sample and 14 percent of the total permissions granted. The exception was the decade 1980–1990. Titles published between 1980 and 1990 made up 30 percent of the titles in the sample, but accounted for 37 percent of the total permissions granted in the study. This suggests that 1980–1990 might be a good decade for acquiring copyright permission to digitize and provide open access to books.

Fig. 9. Analysis of results by publication date (number of titles)



Analysis of Transaction Costs

Focused on outcomes, we neglected to track transaction costs in the feasibility study. However, we suspect that the cost per title was high, in part because of the intermittent labor and consequent learning curves. A crude, retrospective speculation about the transaction cost, based on the cost of paper and postage for the letters and a very conservative estimate of labor costs (\$13,000) for Connelly and George, two of the researchers who worked on the project,²² is

²² The two librarians from the U.S. Army who worked on the project, Waters and Schenk, were visiting scholars and were not paid for their time.

roughly \$200 per title for which permission was granted.²³ The speculative cost would be significantly higher if it included my time and the cost of Internet connectivity and database creation.

Conclusions and Lessons Learned

The random sample feasibility study revealed that it is indeed possible to secure permission to digitize and provide open access to books, but the work is tedious and often comes to naught. We learned that even determining the copyright status of a book can be difficult and time-consuming. When we conducted the study, we had a fledgling understanding of U.S. copyright law, but knew very little about foreign copyright law. When in doubt, we assumed that a work was copyright protected and sought permission. In the course of the study, we mistakenly requested permission for four titles that were no longer copyright protected. One publisher denied permission to digitize and provide Web access to three of these titles. Whether this means that the publisher did not know the copyright status of the books, or whether they believed their permission was required regardless of the copyright status of the books is unknown. The feasibility study also demonstrated that identifying and locating current copyright owners, particularly of older books, is a difficult, time-consuming, hit-or-miss, sometimes futile process. We agreed that future studies would track the transaction costs.

The Fine and Rare Book Study

In 2001, the University Libraries at Carnegie Mellon received funding from Henry Posner, Jr., and his wife Helen Posner to digitize and provide Web access to the Posner Memorial Collection of fine and rare books and associated archival material. The collection includes landmark titles of the history of Western science, beautifully produced books on decorative arts, and fine sets of literature. Henry Posner, Sr., formed the collection between 1924 and 1973, starting with literature and decorative arts and, after 1950, focusing on the history of science.²⁴ The funding provided by the Posners was to purchase a high-quality color scanner designed for handling fine and rare books and to pay the scanner operator.

We knew that the collection contained some copyrighted titles and therefore that the project entailed acquiring copyright permission. The Posner project, which took place between 2001 and 2004, became our second copyright-permission study. The library catalog records for each title in the collection were exported and loaded into a database to track the copyright-permission work. Additional data-

²³ This does not include the permission fees that we paid publishers for the right to digitize and provide Web access to their books. We paid fees up to \$100.

²⁴ Henry Posner, Jr., added to the collection a three-part work published in 1998. This title is included in the research and data analysis reported here.

base records were created for copyrighted catalogs and newsletters among the archival material. The database and initial request process were identical to those used in the feasibility study. The request letter offered the option to restrict access to the Carnegie Mellon community. A contract, prepared in consultation with university legal counsel, was included with the letter.

Work began summer 2001 with intermittent labor. The library staff member assigned to the project could dedicate little time to the work, did not consult the copyright-renewal records to determine the copyright status of titles published in the United States from 1923 through 1963, and, as the workers in the feasibility study did, reported having difficulty locating publishers' addresses. As of September 2002, only 75 initial letters and no follow-up letters had been sent. Only a third of the publishers contacted had responded. Of these, 25 percent had granted permission with some kind of restriction or stipulation.

At this point we made several decisions. First, we calculated that at the current rate it would take us four-and-a-half years to complete the copyright-permission work on the Posner titles. We wanted to finish the permission work by the time the books had been digitized, i.e., by the end of 2003. We concluded that we needed to recruit more labor. Second, if a publisher had multiple titles of interest, we decided to list all the titles in a single letter rather than to send one letter per publication. We also decided to call publishers that had not responded to our initial letter rather than to send a second letter. We hoped thereby to increase our success by engaging the publishers in conversation, answering their questions, and addressing their concerns. Follow-up contact was to be initiated several weeks after the initial request letter was sent.

In May 2003, Erin Rhodes was hired as a part-time temporary employee dedicated to the Posner project copyright permission work. Her employment was extended to full-time in September 2003, and the bulk of the permissions work was completed by November of that year. Nevertheless, we were still locating estates and finalizing negotiations for Posner titles through 2004.

The only way to definitively determine the copyright status of a book published between 1923 and 1963 (the period during which copyright renewal had to be formally registered) is to have the Office of Copyright conduct a title search. As an experiment, we asked the Office of Copyright to conduct a title search for seven titles. They immediately charged us \$150 and estimated that it would take four to six weeks to conduct the searches. We received their response 15 weeks later. They found only one of the seven titles. Given the number of titles in the Posner Memorial Collection published between 1923 and 1963, we estimated that it would cost \$6,000 to \$7,000 to have the Office of Copyright conduct title searches. The cost would be closer to \$8,000 if we included the titles with no date of publication. We decided that our time and financial resources were better spent consulting the copyright-renewal records and seeking copyright permission when the copyright status of a work was not clear.

Rhodes consulted the copyright-renewal records for books and serials published in the United States between 1923 and 1963 and coded the records in the database accordingly. As the work progressed, the coded copyright status for items sometimes changed as we learned more about foreign and domestic copyright law.²⁵ In August 2003, we consulted Carnegie Mellon legal counsel to help us determine the copyright status of foreign publications, but it quickly became apparent that the complexity of international copyright law was impeding the project.²⁶ We eventually abandoned efforts to determine the copyright status of many of the foreign books and chose to assume that they were still in copyright and to request permission to digitize them. Later, we consulted university legal counsel about the copyright status of the archival materials associated with the books in the Posner collection. Legal counsel said that we did need permission to digitize and provide Web access to book catalogs, newsletters, broadsides, newspapers, the text of speeches, and correspondence from the book collector or his secretary to book dealers. However, upon examination of sample correspondence from book dealers to the collector, counsel advised us that we did not need permission to digitize and provide Web access to this material because the letters were compilations of facts about the books. The Posner family granted permission to digitize and provide access to personal correspondence from the collector, Henry Posner, Sr., and the work-for-hire correspondence prepared by his secretary. By November 2003, we were still unable to locate some of the publishers of book catalogs, so Rhodes began examining the title pages of book catalogs published in the United States, applying the laws about books published without a copyright notice when notices were required, to determine whether the catalogs were in the public domain.²⁷

²⁵ In the course of the Posner study, we mistakenly requested copyright permission for 74 out-of-copyright books, 54 of which were volumes and supplements to the complete works of William Makepeace Thackeray. All the publishers granted permission. These data are not included in the analyses in this report.

²⁶ Even countries that are World Intellectual Property Organization (WIPO) signatories do not have the same copyright laws. They must meet certain minimal requirements, but how they do that, and what they do in addition to that, can be quite diverse. Additional legislation further complicates the matter. For example, the United Kingdom follows the 2001 European Union Copyright Directive that restored copyright to certain material that had been in the public domain. In the United States, the Uruguay Round Agreements Act (1994) automatically restored copyright to certain foreign works that as of January 1, 1996, were still protected by copyright in their home countries but had fallen into the public domain in the United States because of failure to comply with U.S. formalities or because of a previous lack of copyright relations between the United States and the home country.

²⁷ A "copyright notice" appears in the front matter of a work as "Copyright," "Copr.," or "©," along with the name(s) of the copyright holder(s) and the date of first publication. The law stipulates that if a book is published (a) before 1978 without a copyright notice or (b) between January 1, 1978, and March 1, 1989, without a copyright notice and without being registered with the U.S. Copyright Office within five years of publication, then it is in the public domain. Given the ephemeral nature of book catalogs, for the purposes of the Posner study, we assumed that if a catalog published before 1989 had no copyright notice, then it was in the public domain. We did not consult the Office of Copyright to see whether such a catalog had been registered.

Determining copyright status is one step. Determining copyright ownership is another. Locating the copyright owner is yet another. The three do not necessarily go hand in hand. The publisher or creator cited on the title page of a book is the beginning point for a journey that often resembles traversing a maze. For U.S. works published between 1923 and 1963, renewal records must be consulted to determine the copyright *status*. According to the U.S. Copyright Office, the *claimant* in a copyright-renewal record is the copyright holder at the time of renewal, but not necessarily the current copyright owner. Similar ambiguity applies to the title page of more-recent publications that do not require copyright renewal: The name that appears there might not be the current copyright owner. There is neither a definitive source to identify current copyright holders nor a definitive source for locating those holders once they have been identified. According to copyright attorney Michael Shamos, "If a work is in copyright and the copyright is assigned to a new owner, an assignment document needs to be filed with the Copyright Office. Otherwise, the new owner will not be able to prove his ownership and will not be able to sue anyone for infringement." When asked about publishers we could not locate, he responded, "It is possible that the publishers went defunct and either abandoned their copyrights (not expressly, but by default) or conveyed copyright back to the authors, or sold the copyrights to satisfy creditors in bankruptcy" (e-mail from Michael Shamos to Denise Troll Covey, March 7, 2003). We agreed that the cost of having the Copyright Office conduct a search for each title was prohibitive and that we would consult the Copyright Office renewal records and use our own devices to determine copyright status and try to identify and locate copyright owners. We also agreed, in consultation with university legal counsel, that if we could not locate the copyright owner, we would assume permission was denied and not digitize and provide Web access to the books.

Not counting correspondence or ephemeral material in the archival folders, the Posner Memorial Collection contains 1,106 volumes or cataloged items. We determined that 26 percent (284) were still in copyright or were to be treated as if they were.²⁸ By the conclusion of the study, we determined that these 284 copyrighted works were owned by 104 different copyright holders.

As in the feasibility study, identifying and locating the copyright holders were arduous tasks. There were many publishers that we could not locate using the resources used to find publishers in the feasibility study. An administrative assistant and several librarians were recruited to assist with locating publishers. Again, many letters were returned marked "Address unknown." Letters to foreign publishers were sometimes returned marked simply "Gone away." Publishers often responded by referring us to another publisher, sometimes a foreign publisher,²⁹ the author, or the author's estate.

²⁸ Roughly 70 titles had no publication date, which complicated determining copyright status. The dean of University Libraries examined these works and advised us how to proceed.

²⁹ For example, Farrar, Straus and Giroux in New York referred us to Faber and Faber in the United Kingdom.

The referring publisher seldom provided an address.

To locate authors, we began consulting the Authors Registry,³⁰ Writers, Artists, and Their Copyright Holders (WATCH) File,³¹ the Society of Authors in London,³² and the Authors Licensing and Collecting Society. We had some success consulting these sources, but were still looking for addresses for 13 authors or estates in 2004. Rhodes became quite the detective, making several phone calls to libraries, book dealers, and university professors to discover contact information for the author or estate in question. She also began examining the books themselves, looking for clues. In one case, she discovered that the author had been a professor at City College in New York, so she called a librarian at City College. The librarian helped her locate the author's daughter, who provided her mother's address. The mother, the current copyright holder, granted permission to digitize and provide Web access to the title in the Posner Memorial Collection.

In the course of the Posner study, we encountered many third-party copyright owners. Unlike the feasibility study, we could not eliminate these books from the project. If copyright was held by a publisher, we did not pursue third-party copyright owners. However, when copyright reverted from the publisher to the author, we attempted to contact all the authors and contributors cited in the bibliographic record for the work. For example, the bibliographic record for *The Journal of Christopher Columbus* indicates that the work was translated by one person and revised and annotated by another. Yet another person provided the appendix. Often we were unable to locate all of the third parties.

If a request letter appeared to have been successfully delivered, we conducted a follow-up call or sent an e-mail message a few weeks after the letter was sent. Nevertheless, we frequently sent multiple letters to the same publishers because they had lost or misplaced our letter by the time we spoke to them on the telephone or contacted them by e-mail. In many cases, we also sent multiple letters when the copyright to a title had transferred to another publisher or to the author and we had difficulty locating them. Subsequent letters were frequently sent as attachments in e-mail. By the end of the project, we had sent 174 initial request letters and made 159 follow-up attempts in e-mail or by telephone.

In the discussion that follows, the term *publisher*, unless otherwise distinguished from authors and estates, refers to a unique copyright holder of content in the Posner collection. The term *title* refers

³⁰ The Authors Registry is a New York City-based service that provides contact information for authors. Its staff will try to locate up to 10 authors per week (2 per day) at no charge. See <http://www.authorsregistry.org/welcome.html>.

³¹ WATCH File is an online database maintained by the Harry Ransom Humanities Research Center at the University of Texas at Austin. Available at <http://tyler.hrc.utexas.edu/>.

³² The Society of Authors in London is a literary agency that offers estate information free of charge provided that it represents the estate you wish to locate. Available at <http://www.societyofauthors.org>.

to an item identified in a record in the database created to track the copyright permission work for the Posner project.³³ Because of the way in which the database was constructed, distinguishing *titles* from *volumes* and *parts* would have required manually counting every data point in this report and would have significantly hampered the data analyses. Twelve percent of the copyrighted titles in the collection were multivolume or multipart works. In the analyses below, when not distinguishing titles from volumes or parts made a significant difference in the results, the instance is noted.

Overall Results

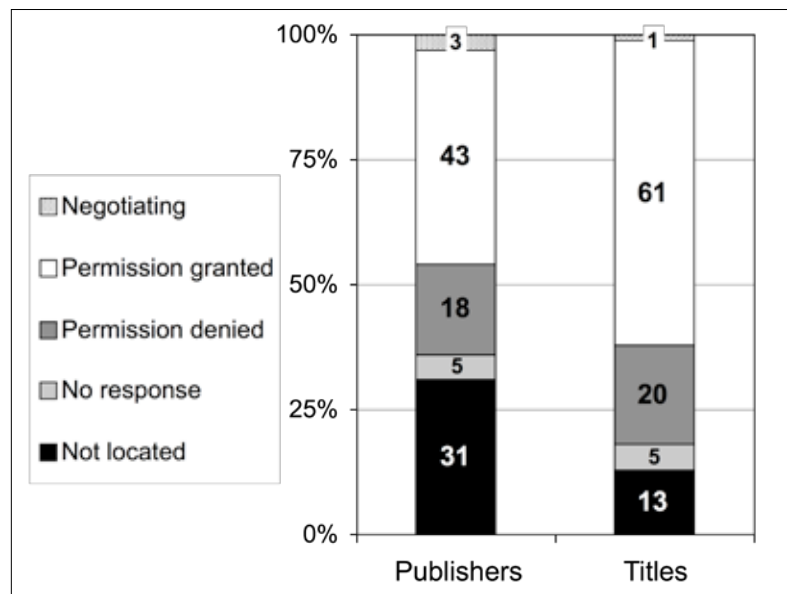
As of November 2004, we were still unable to locate almost a third of the publishers, which meant that we had no opportunity to even try to acquire permission for 13 percent of the copyrighted titles in the Posner Memorial Collection. Almost two-thirds of the publishers responded to our request letter, e-mail, or telephone calls. Almost half of them granted permission to digitize and provide Web access to their works,³⁴ accounting for most of the copyrighted titles in the collection (see figure 10). More than twice as many publishers granted permission as denied permission.

In the context of the Posner study, *permission denied* meant that the publisher either responded “no” to our request or was considered to have denied permission according to the *three-strikes rule*. We established the three-strikes rule in September 2003, in consultation with university legal counsel and the dean of University Libraries, as a way to bring closure to a negotiation if the publisher failed to respond to our initial request letter and two follow-up attempts. For example, three strikes could consist of an initial request letter that was not returned to us marked “address or addressee unknown” and two telephone messages or two e-mail messages that were successfully delivered with no response. According to Carnegie Mellon legal counsel, inability to locate a publisher or lack of response from a publisher, despite due diligence, did not permit us to treat these cases as permission granted. Only two publishers were considered to have denied permission under the three-strikes rule. The few publishers in figure 10 indicated as “No response” are authors and estates that we located in 2004, but that had not yet received two follow-up contacts from us when the data were analyzed for this report.

Of the permissions granted, 12 percent were for multivolume or multipart works with the volumes or parts bound separately:

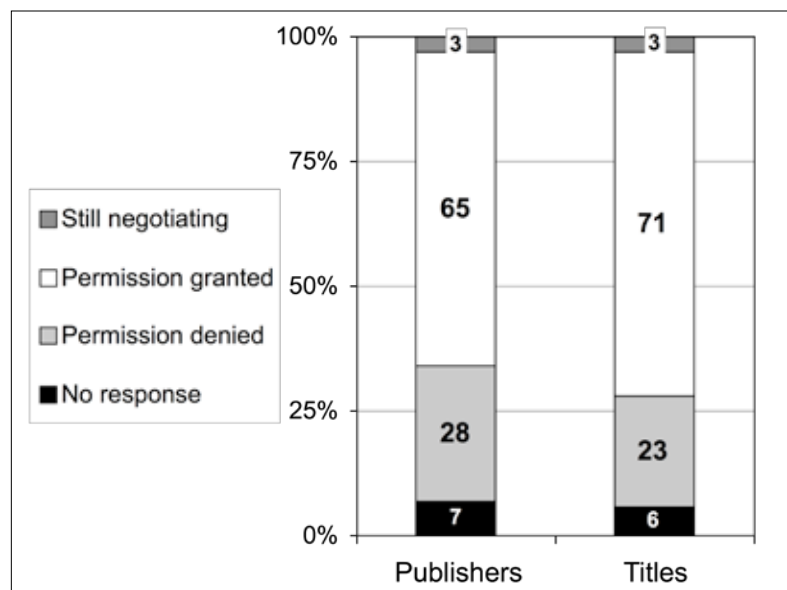
³³ The Posner Memorial Collection contains 633 unique titles. Some titles have many volumes. Some volumes have many parts. Parts and volumes for different titles are sometimes bound differently. Some titles are bound with other titles. Parts, volumes, and titles were not cataloged consistently, so, for example, sometimes separate records existed for each volume in a multivolume work.

³⁴ Frequently publishers, authors, or estates that granted permission did so in e-mail rather than returning a signed contract. University legal counsel confirmed that authenticated e-mail was sufficient for our purposes. We treated the e-mail as a signed contract. We printed and filed it with the other notes and correspondence related to negotiations with that publisher.

Fig. 10. Summary of overall results of the Posner study

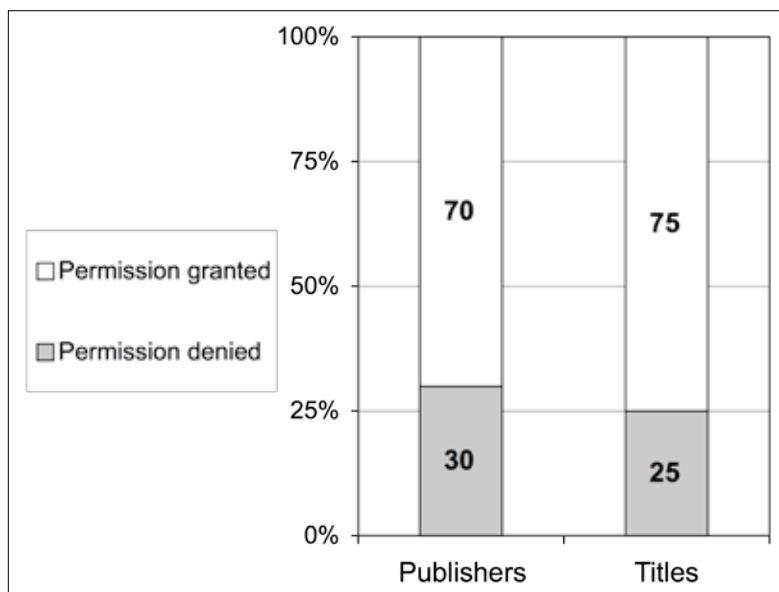
13 titles had 2 volumes, 4 titles had 3 parts or volumes, and 1 had 4 volumes. Of the permissions denied, 13 percent were for multivolume works: 3 titles had 2 volumes, and 1 title had 18 volumes, a supplement, and a catalog. Of the titles for which we could not locate the publishers, 8 percent were for multivolume or multipart works: 2 titles had 3 volumes and one had 3 parts. None of the titles for which we received no response were multivolume or multipart works.

To better understand the outcome of our efforts, we must look strictly at the publishers we located. Of those we contacted, almost all responded and most granted permission. As shown in figure 11, the permissions granted enabled us to digitize and provide Web access to 71 percent of the copyrighted titles published by those we contacted.

Fig. 11. Analysis of the publishers successfully contacted

Looking only at the publishers with which we have completed negotiations and the titles in the Posner collection to which they hold copyright, the overall success rate was 70 percent, granting permission for 75 percent of the titles published by those that responded (figure 12).

Fig. 12. Analysis of completed negotiations



Analysis of Restrictions

As shown in table 2, publishers granting permission to digitize and provide Web access to their books in the Posner collection applied fewer restrictions than did publishers granting permission in the feasibility study.

Table 2. Comparative analysis of restrictions applied

	Feasibility Study	Posner Study
Restrict access to Carnegie Mellon users ³⁵	54%	6%
Display full citation	23%	10%
Permission does not apply to third-party material	22%	5%
License to provide access expires	8%	6%
No simultaneous users	6%	4%
Permission to scan expires	3%	0%

³⁵ One publisher in the Posner study requested that access be restricted to users in the United States. This functionality has not been added to the Posner Memorial Collection online system, so access is restricted to the Carnegie Mellon community.

Of those publishers that stipulated that permission did not apply to components of the work with copyright owned by a third party, all of them limited the duration of the license to provide Web access to the title, and 89 percent prohibited simultaneous use. However, the duration of the license was longer in the Posner study than in the feasibility study. In all but one case, the licenses in the Posner study were six to seven years, rather than the three to four years stipulated in the feasibility study.³⁶ All the publishers that limited the duration of the license in the Posner study were university presses.

Only 1 percent of the publishers that granted permission in the Posner project requested a copy of their digitized books, in comparison with 15 percent in the feasibility study. One publisher made granting permission contingent on our assurance that we would terminate Web access to its four titles in the Posner collection if it gave us 60 days' notice: "A short notice period is essential to allow for the possibility of a reprint license being granted" (e-mail to Erin Rhodes, September 30, 2003). We agreed, and the publisher granted permission. In addition, one current copyright owner, the heir of the author, stipulated that he would grant permission if we would digitize and include his father's notes and updated introduction to the work. We agreed. He denied permission.

Several publishers contacted in the Posner study inquired about royalty fees. We had decided not to pay fees in the Posner project. In one case, the original publisher still owned the copyright to the title, which was published in 1934. Though the publisher was disappointed that we would not pay a royalty, it still granted permission. In another case, the copyright to a title published in 1966 had passed to another publisher that denied permission because we would not pay a royalty fee.

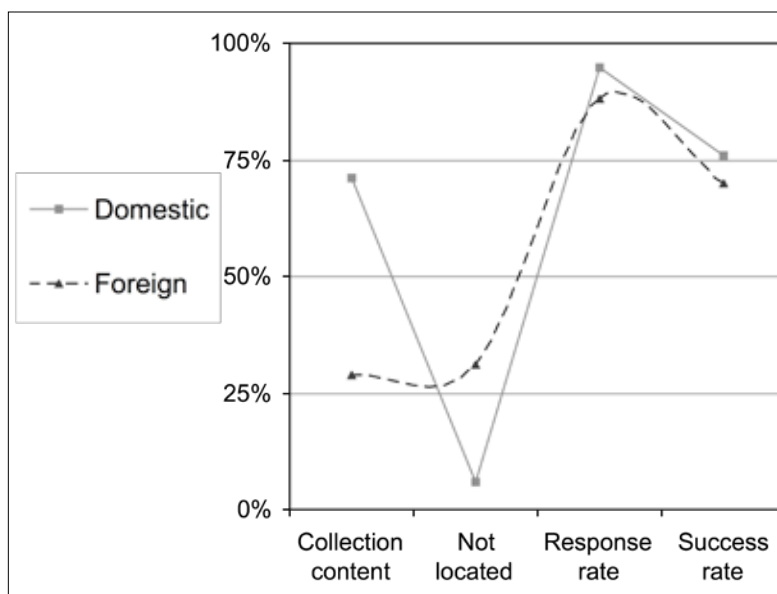
The analyses that follow are based on the number of titles, rather than on the number of publishers, because publishers with multiple books in the Posner collection sometimes granted permission for some titles, but not for others. The *response rate* is based on the number of titles with copyright owned by publishers we successfully contacted. The *success rate* is based on the number of titles with copyright owned by publishers that responded. *Collection content* refers to the copyrighted titles in the Posner Memorial Collection.

Analysis of Foreign and Domestic Publications

As in the feasibility study, most of the books in the Posner collection were published in the United States and foreign publishers were far more difficult to locate than U.S. publishers. However, domestic publishers were more likely to respond and more likely than foreign publishers to grant permission in the Posner study, as compared with the feasibility study (figure 13).

³⁶ Functionality was added to the Posner Memorial Collection online system to automatically remove a book from the Web when its license expires. Tracking when to contact the publishers to request an extension of the license is still done manually. Dates for when to begin contacting the publishers are noted in the author's online calendar.

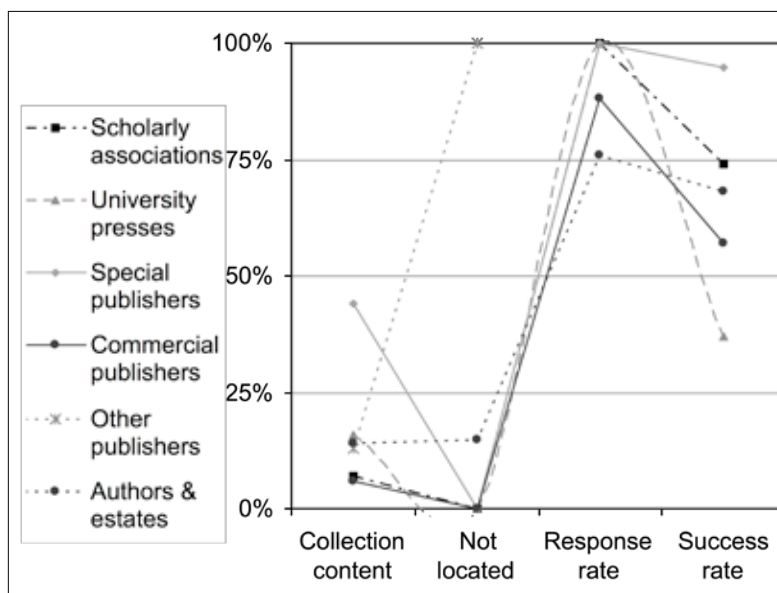
Fig. 13. Analysis of foreign and domestic titles



Analysis by Publisher Type

Again, the response and success rates varied across different types of publishers. We successfully located all the scholarly associations, university presses, and commercial and special publishers of collection content. Special publishers own the copyright to the largest proportion of the content, followed by university presses, and authors and estates. Scholarly associations and commercial publishers own copyright to little of the material. Copyright to 13 percent of the collection content is owned by units that we could neither identify (code by publisher type) nor locate. The response rates of the publishers that we contacted were very good (figure 14). Special publishers almost always granted permission. Scholarly associations,

Fig. 14. Analysis by publisher type

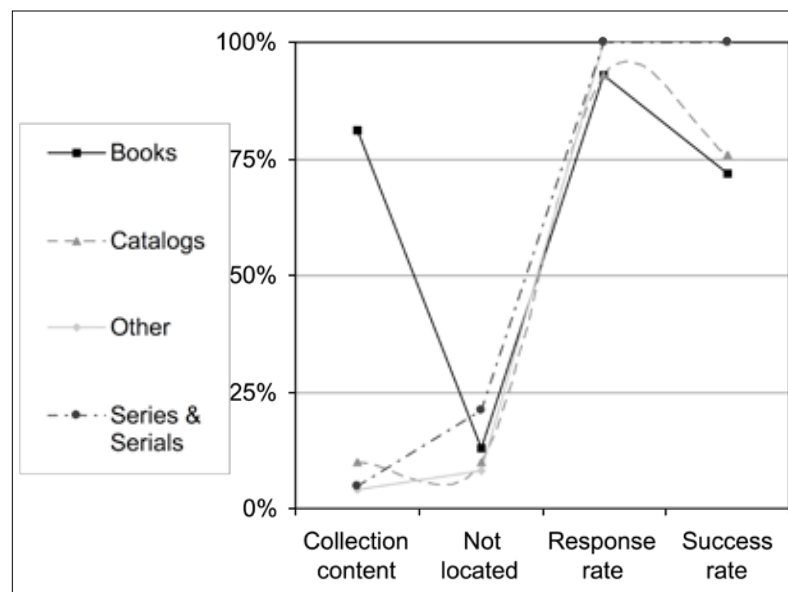


and authors and estates were likely to grant permission, although authors and estates were difficult to locate. More than half of the commercial publishers granted permission. University presses were the least likely to grant permission.

Analysis by Publication Type

The response and success rates also varied with different types of publications. Most of the copyrighted content in the collection is traditional monographs; 10 percent is book catalogs. The book and catalog publishers we located were very likely to respond, and most granted permission (figure 15). Publishers of the few series and serials in the collection were more difficult to locate, but all those that we successfully contacted responded and granted permission.³⁷ The few miscellaneous copyrighted archival materials in the collection, for example, newsletters and newspapers, were coded as “Other” publications. The owners of these materials were relatively easy to locate, and all of them responded and granted permission.

Fig. 15. Analysis by publication type



Analysis by Print Status and Publication Date

Given the age and nature of the Posner Memorial Collection and data on print status by publication date in the feasibility study (see figure 8), we strongly suspected that most of the copyrighted content in the Posner collection is out of print. When we began coding the

³⁷ Though *series* and *serials* are different types of publications, the coding and analysis enabled by our database did not distinguish between them. The database had been designed for the feasibility study, which contained series but no serials. Given the small number of these items in the Posner collection, we did not believe that the distinction warranted reconfiguring the database.

print status of copyrighted books in the collection, we quickly ran into snags. For example:

- If there is a reprint of a title that is in the Posner collection and the reprint is available on demand through Books on Demand, do we code the Posner title as “in print”?
- What if we can find no record for the exact item (i.e., edition, publisher, publication date) that is in the Posner collection?
- If there is a more recent, revised edition of a title in the Posner collection and that edition is still in print, do we code the Posner title as “in print”?

When Rhodes raised these questions, the dean of University Libraries provided an answer, but we were simultaneously discovering in our work on copyright permissions for the Million Book Project (described later) that publishers answer these questions differently. In light of this fact, we came to believe that an analysis of print status as defined by a librarian would be meaningless for our current purposes and chose not to complete this analysis. Details are provided on pages 49–51 of this report.

The copyrighted titles in the Posner collection are significantly older and the distribution of titles published per decade is more even than that of the books in the random sample feasibility study. Roughly 88 percent of the Posner titles were published before 1970, compared with 35 percent of the random sample. Figure 16 shows the comparative distribution by publication date of copyrighted titles in the two studies. Figure 17 shows the results of our efforts in the Posner study to acquire copyright permission by publication date.

Fig. 16. Comparative distribution of project content by publication date (number of titles)

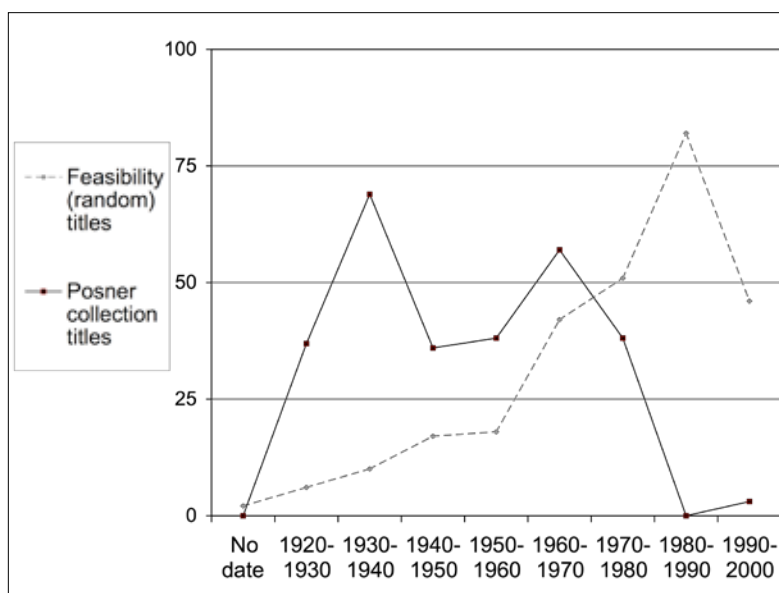
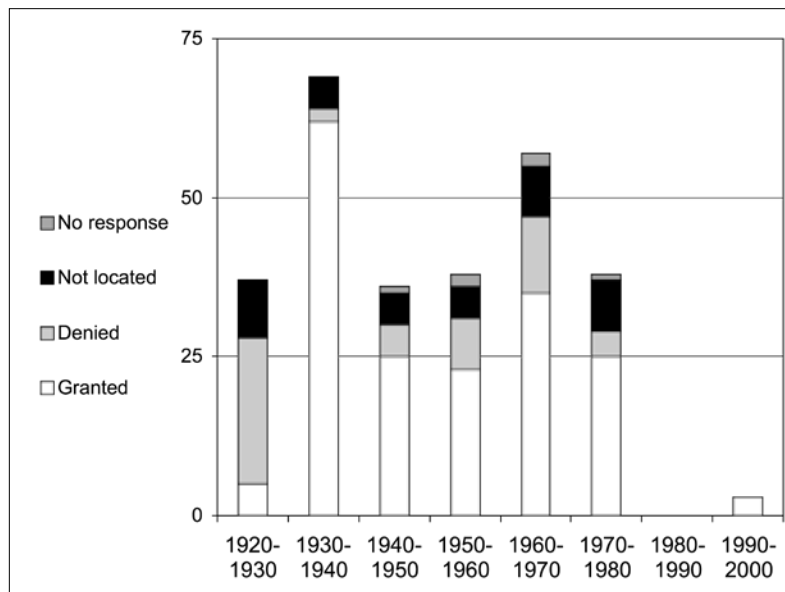


Fig. 17. Analysis of Posner study results by publication date
(number of titles)



The extent to which the age of the work affected the results in the Posner study is unclear.

- Publishers of older material in the Posner collection were not conspicuously more difficult to locate than were publishers of more-recent material. More diligence and persistence were expended on locating and following up with publishers in the Posner study than in the feasibility study; consequently, more publishers were found and more of them responded than in the feasibility study. In the Posner study, there was no striking difference in the ability to find publishers of titles published between 1920 and 1930 and titles published between 1970 and 1980: Almost one-fourth of them could not be found. Although roughly a third of the collection content was published between 1960 and 1980, 40 percent of the publishers we could not locate were publishers of titles published during these two decades.
- It appears as if permission was frequently denied for titles published between 1920 and 1930. However, this is an instance when not distinguishing titles from volumes or parts in the collection skews the data. A closer examination revealed that of the 23 so-called copyrighted “titles” in the collection published in that decade and for which permission had been denied, 20 of them pertain to one actual title (in 18 numbered volumes, a catalog, and a supplement). Two of the remaining three “titles” are a two-volume work. Though there are multivolume and multipart works in the Posner collection published in subsequent decades, none exceeds four parts or volumes, and the total per decade does not dramatically skew the data.³⁸

³⁸ Among all the copyrighted books in the Posner collection, there are 17 two-volume or two-part titles, 6 three-volume or three-part titles, 1 four-volume title, and the 18-volume work with supplement and catalog noted in the text.

- Looking at the data per decade, permission was granted for more than 60 percent of the titles published in the 1930s, 1940s, 1950s, 1960s, and 1970s. With the exception of two decades, the percentage of total copyrighted titles in the collection published in a given decade was roughly equivalent to that decade's percentage of the total permissions granted in the study. For example, books published between 1960 and 1970 constituted 20 percent of the sample and 21 percent of the total permissions granted. The exceptional decades were 1920–1930 and 1930–1940. Titles published 1930–1940 made up 25 percent of the copyrighted titles in the collection. Permission was granted for 90 percent of the titles published during this decade, accounting for 35 percent of the total permissions granted in the study. In contrast, titles published 1920–1930 constituted 13 percent of the copyrighted collection, but accounted for only 3 percent of the total permissions granted in the study (this was, however, the decade where not distinguishing titles from volumes or parts skews the data). Over a third of the total permissions granted were for titles published in 1960 or later.

Analysis of Transaction Costs

We closely monitored the labor costs of copyright-permission assistant Erin Rhodes.³⁹ Rhodes determined the copyright status of the materials in the Posner collection, identified and located the copyright holders, prepared the initial request letters, followed up by e-mail or by telephone, updated the database, and prepared the preliminary statistics. We also monitored the cost of paper and postage for initial request letters and long-distance telephone charges. We did not factor in the cost of Internet connectivity, database creation, consultation with university legal counsel, or administrator time. University legal counsel did not levy a fee for consultations and advice. As project administrator, I answered many questions from the copyright-permission assistant, often in consultation with the dean of University Libraries, and the more-difficult questions from publishers.⁴⁰ I also conducted the data analyses.

On the basis of the costs monitored from May through October 2003, we spent roughly \$10,808 on labor (wages and benefits), \$379 on long-distance phone calls, and \$100 on paper and postage. The average transaction cost per copyrighted title in the Posner collection for which permission was granted was \$78. The cost would be significantly higher if Rhodes's work with authors and estates in 2004, my time, and the cost of Internet connectivity and database creation were included.

³⁹ We did not track the cost of the intermittent labor that worked on the Posner permissions project before Rhodes was hired.

⁴⁰ The number of e-mail messages regarding the copyright permission work provides an estimate of the volume of administrative activity on the project. I sent to or received from Rhodes 364 e-mail messages in the course of the project.

Conclusions and Lessons Learned

Although we located fewer of the publishers of copyrighted content in the Posner project than in the feasibility study, we greatly increased the response and success rates during the Posner study. Of the publishers that we successfully contacted in the latter study, almost all responded to our request, while only two-thirds of those that we contacted in the feasibility study responded. Of the publishers that responded in the Posner study, 75 percent granted permission, in comparison with 45 percent in the feasibility study.

We attribute the increased success in the Posner project to a more informative initial request letter, to prompt follow-up by e-mail or telephone, and to the publishers' ability to see the quality of the digitized books in the Posner collection on the Web.⁴¹ We believe that the age and nature of the Posner Memorial Collection were also significant factors. The Posner collection contains more old books than the random sample did, which probably accounts for the greater difficulty we encountered locating publishers of the Posner works. Special publishers own the copyright to most of the titles in the Posner collection but to very few titles in the random sample feasibility study. Results from the feasibility study suggest that special publishers are more likely to grant permission than traditional publishers are. Furthermore, it is conceivable that publishers of the works in the Posner collection liked the idea of seeing high-quality digital replicas of their books in an online special collection almost a third of which is classic works published from the fifteenth through the nineteenth century.

The Posner project confirmed our belief that it is possible to secure copyright permission to digitize books and to provide open access to them on the Web. It also confirmed what we had learned in the feasibility study about how difficult and time-consuming it is to determine copyright status and to identify and locate copyright holders, particularly authors and estates. However, by dedicating personnel and adjusting our processes, we significantly reduced the cost per title for which permission was granted. Further adjustments to our workflow or refinements to our negotiation strategies could yield even greater cost savings.

The Posner study also made us aware that many publishers do not keep good records. Some do not really know what they have published. On several occasions, we had to photocopy the title page of a book and fax it to the publisher because it claimed it had not published the book. Frequently, publishers reported that they did not

⁴¹ Most of the Posner Memorial Collection is out of copyright. For preservation and security reasons, all the books that could be digitized were digitized in call number order. A small number of volumes (3 percent) could not be scanned because the binding was too tight, the pages were uncut, or the book was too small or too large to manage on the scanner. After the books had been digitized, the archival materials were digitized. Public domain books and archival documents were made available on the Web site at <http://posner.library.cmu.edu/> as they were digitized. Access to copyrighted materials was provided as permissions were granted.

know whether they had the right to grant nonexclusive permission to digitize and provide open access to their books. Some responded that the author had not granted them this right, so they denied permission. Given the age of the books in the Posner collection, it is unlikely that any author explicitly granted electronic rights to the publisher, so we suspect that the publishers that granted permission assumed they had this right because it was not explicitly denied.

As expected, many publishers expressed concern about open access and lost revenue, regardless of the fact that they were not generating revenue from these older, presumably out-of-print, books. The questions they asked related to access restrictions, the quality of the digitized books, and whether the delivery system enabled users to download or print the books.⁴² Negotiating with publishers was frequently confusing, even frustrating, but always enlightening. A few examples will illustrate.

- A bibliophiles group granted permission for two titles (published in 1964 and 1965) without first checking to determine whether it still owned the copyright. We discovered this when we contacted the group again, about a third title that we had overlooked when we sent the initial request letter. When we asked about the third title (published in 1973), the person we spoke with confessed that he had not contacted the authors' estates when he granted permission for the first two titles. He explained that the bibliophiles "controlled" the first two books and that the authors "didn't really write much of anything else," but referred us to the author's heir for the third title. Though we contacted other members of the bibliophiles group in an effort to locate the heir, we were unsuccessful. However, one of the people whom we contacted described the work as a "small, pamphlet-like book." Rhodes checked the actual item in the Posner collection and confirmed that it had no copyright notice, which meant that it was in the public domain (Rhodes's conversation with two members of the bibliophiles group, November 3–4, 2004)
- A commercial publisher of a "little essay," as he called it, sent an e-mail message stating, "Given the slightness of the book in question, I see no reason to write a formal contract. Please consider this e-mail a formal grant of permission to digitize [title] by [author] for use at Carnegie Mellon's University Libraries, subject to the following restrictions." The message then listed six restrictions, including access restriction to the Carnegie Mellon community. However, in the first paragraph of the message he stated that the copyright belonged to the translator of the work, who also wrote the introduction. When we contacted the publisher again to request contact information for the translator, he responded, "This is turning out to be more trouble than it is worth. [Publisher] published the book as a favor to [translator], at our expense. While [translator] (or his heirs) may own the copyright, [publisher], as

⁴² The delivery system for the Posner collection restricts saving and printing to one page at a time, which we believe is sufficient deterrent to prevent users from printing or downloading entire books.

publisher, controls it, and we are within our rights to grant the permission. If you require more than that, you'll have to look elsewhere. We've had no contact with [translator] in years, and can not give you any further information." The essay was published in 1970 (e-mail to Erin Rhodes, October 22 and 27, 2003.)

- After the initial contact, a university press referred us to the author of a work published in 1965. When we inadvertently followed up with the press, it denied permission. We pursued the author and eventually located the author's heir, who granted permission.
- One commercial publisher granted permission for a work published in 1973 to which it confirmed it owned the rights. Yet this publisher also sent us a disclaimer stating that we could digitize the work at our own risk because the author died in 1946 and the 70-year copyright term past the author's death has not expired.
- The son of a defunct publisher granted permission for a title published in 1975, though the publisher no longer owned the copyright and neither he nor the son had any idea who did. (The book had been published with no indication of the author.) He signed and returned a contract noting the caveat about copyright ownership.
- One publisher responded that copyright to a work published in 1950 had not been renewed. He explained that another publisher had issued a reprint without its permission, so he saw no reason why we should not or could not digitize and provide Web access to the book.
- A university press granted permission for a title published in 1940 but restricted the duration of the license to display the book on the Web. We later learned that copyright had reverted to the author and that the author's heir had renewed the copyright. Shortly before the license with the university press expired, we contacted the press to request a renewal of the license. The press renewed the license but suggested that we also contact the author's heir "since she seems to hold the copyright" (e-mail to Erin Rhodes, September 8, 2004).
- A special publisher responded to our initial contact with a request that we contact it again in a few months. We marked our calendar and contacted the publisher again months later. After a couple more follow-up attempts with no response, the publisher was coded "Permission denied."

We agreed that future copyright-permission studies should experiment with ways to reduce the transaction costs and should formulate and test strategies to increase the response and success rates. We also believed that, whenever possible, we should examine physical books published between 1923 and 1989 to see whether they have a copyright notice as part of our effort to determine copyright status.⁴³ We also knew that we needed to develop a better way to

⁴³ Books published from 1923 through 1977 without a copyright notice and books published from 1978 to March 1, 1989, without a copyright notice and without being registered with the U.S. Copyright Office within five years of publication are in the public domain.

manage the data and routinely calculate statistics. Inadequate methods of analyzing the data unnecessarily delayed analyses that might have guided us to change strategies and correct course in a more timely way.

The Million Book Project Study

The Million Book Project (MBP) is funded by the National Science Foundation (NSF) and the governments of India and China. Its goal is to digitize and provide open access to 1 million books by 2007. With rare exception, the books for the Million Book Collection are being scanned in India and China. The MBP is part of the larger Universal Library Project, which is a partnership of Carnegie Mellon School of Computer Science and the University Libraries. The Universal Library project directors aim to digitize the cultural and intellectual history of humankind. While the vision of the universal library is unlikely to be achieved in our lifetime, the philosophy makes the sequence in which materials are digitized inconsequential.⁴⁴

The initial MBP collection-development meeting was held in November 2001.⁴⁵ Participants swiftly agreed that 1 million books could not be selected title by title. They also quickly agreed that garnering permission to digitize and provide open 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 partner institutions in India and China, 700,000 public domain works, and a target of 100,000 copyrighted works. Efforts to acquire permission to include copyrighted material in the collection would begin with titles cited in *Books for College Libraries (BCL)*, a five-volume bibliography of books compiled by librarians and 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. Partners in India and China would be responsible for securing permission to

⁴⁴ This perspective is not shared by all librarians. Those who lament the approach taken in the Universal Library Project are likely the same librarians objecting to the Google Print Project and to the "blind, wholesale digitization" of books. Some critics of the project forecast disastrous consequences sure to follow from having free online access to older books (see, for example, Tennant 2005).

⁴⁵ Participants included representatives from the Digital Library Federation, Center for Research Libraries, Library of Congress, NSF, and OCLC, and librarians from Carnegie Mellon, Haverford College, Indiana University, Pennsylvania State University, Simmons College, Stanford University, University of California at Berkeley, University of Chicago, University of Pittsburgh, and University of Washington.

include copyrighted books published in India and China.⁴⁶ Carnegie Mellon would be responsible for securing permission to include copyrighted books published in the United States.⁴⁷

Plans were to seek funding to ask publishers for permission to digitize and provide open access to the titles they published that were cited in *BCL*. In the meantime, eager to get started and secure copyrighted content for the Million Book Collection, one of the MBP directors, Raj Reddy, instructed the University Libraries to send letters to significant publishers of scholarly monographs, asking them to participate in the MBP by providing out-of-print books. In June 2002, letters were sent to 32 commercial publishers, 11 university presses, and 1 scholarly association selected by our head of acquisitions, Denise Novak. Using intermittent labor, little follow-up was done and little accomplished. Only seven of the commercial publishers we contacted responded: two granted permission, three denied permission, and two explained that copyright reverted to the author when their books went out of print. Another commercial publisher was considered "Permission denied" under the three-strikes rule. The remaining 24 commercial publishers were abandoned on the basis of preliminary data from the feasibility study that indicated they were the least likely to grant permission (George 2001). The initial 11 university presses were eventually contacted again by copyright-permission assistant Rhodes when she completed the bulk of the permissions work on the Posner project and turned her attention to the MBP in November 2003. Rhodes also followed up with some of the publishers of designated titles that we had contacted in a previous project, the Thousand Book Project, which was folded into the MBP when it began. Eventually, many of these publishers were also abandoned so that we could focus our efforts on publishers of works cited in *BCL*.

As we requested copyright permission for designated titles in the Posner project, we realized that the transaction cost of pursuing copyright permission per title (about \$78 per book) was too high to pursue on a large scale. There are roughly 50,000 titles cited in *BCL*.⁴⁸ Assuming, for the sake of a cursory analysis, that the cited titles were published in the United States:

- 2,200 titles (4 percent) were published before 1923, which means they are in the public domain and can be digitized and included in the Million Book Collection without acquiring copyright permission
- 35,500 titles (71 percent) were published from 1923 through 1963, when U.S. copyright required renewal. Research suggests that

⁴⁶ At the annual meeting of MBP partners in May 2004, Siva Venkamma of the Digital Library of India reported that they had acquired copyright permission to digitize 6,841 books published in India. The copyright-permission work is being done in India by employees of the Registrar of Publications.

⁴⁷ For more information on the MBP, see http://www.library.cmu.edu/Libraries/MBP_FAQ.html.

⁴⁸ We chose to use the 1988 edition of *BCL* to ensure that many of the titles would be out of print.

copyright to at least 90 percent of the books published during this period was not renewed (Ringer 1961, 220; Lesk 2004b). Therefore, 32,000 *BCL* titles could currently be out of copyright. However, the renewal records for all 35,500 books would need to be consulted.

- 12,300 titles (25 percent) were published after 1964, which means that they are copyright protected and require permission to digitize and include in the Million Book Collection.

The 50,000 titles cited in *BCL* were published by about 5,600 publishers. On the basis of the transaction costs from the Posner study, I proposed that we change to a per-publisher approach for the MBP. After discussions with the dean and associate dean, we agreed to treat *BCL* like an approval plan for publishers, assuming that if they had published books cited in *BCL* then they were among the best publishers in the country. Many libraries use publisher-based approval plans to select books for their collections. We subsequently began asking the publishers of books cited in *BCL* for permission to digitize all of their out-of-print, in-copyright books to facilitate collection development for the MBP and to reduce the cost of acquiring copyright permission. Treating *BCL* like an approval plan for publishers substantially reduced the transaction cost by obviating the need to check copyright-renewal records for cited titles, simplifying letter preparation, and reducing the cost of paper and postage. Consider the effort required to prepare letters containing lists of designated titles: about 950 titles cited in *BCL* were published by Harvard University Press; 356 titles were published by Indiana University Press.⁴⁹ Furthermore, using a per-publisher rather than a per-title approach meant that each letter could potentially secure permission to include more titles in the Million Book Collection than just those cited in *BCL*. This was already apparent from two of the publishers that we had initially contacted in June 2002. The National Academies Press, with only 26 titles cited in *BCL*, had granted permission for about 3,400 titles published through 1994. Rand McNally, with two titles cited in *BCL*, had granted permission for all of its out-of-print, in-copyright books except atlases—roughly 900 titles. We calculated that if only 10 percent of the 5,600 publishers with works cited in *BCL* granted permission to digitize 500 books each, the result would be 280,000 copyrighted works for the Million Book Collection.

In August 2003, MBP project partner University of California Libraries at Merced (UC Merced) provided funding for a full-time copyright permission assistant at Carnegie Mellon, Erin Rhodes, and a part-time copyright permission assistant at UC Merced, Sarah Sheets.⁵⁰ With dedicated labor, in November 2003 we began sending letters to publishers of books cited in *BCL*. Letters to publishers briefly introduced the MBP, explicitly stated adherence to copyright

⁴⁹ Project partner OCLC provided a list of publishers and the number of titles they have cited in *BCL*.

⁵⁰ Two proposals to the Institute of Museum and Library Services (2003 and 2004) seeking financial support for copyright-permission work were not funded.

law, and described the copyright absurdity wherein out-of-print, in-copyright books are neither generating revenue for the copyright holder nor readily available to potential readers. The letters provided 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 (Pope 1999). The letters then asked publishers for nonexclusive permission to digitize and offer free-to-read on the Web any of the following options:

- all their out-of-print, in-copyright titles
- all their titles published prior to a date of their choosing
- all their titles published *N* or more years ago (they specify *N*)
- a list of titles that they provide

The letters explained that the Million Book delivery system will have minimal functionality. They closed with an offer to give participating publishers preservation-quality copies of their digitized books and the associated OCR text file, explaining that they could 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. Unlike the feasibility study and Posner project, the MBP offered no option to restrict access to the Carnegie Mellon community.

Over time, we revised the letter to include answers to common questions asked or concerns raised by the publishers. For example, we updated the letter to state that the Million Book delivery system restricts saving and printing to one page at a time, as netLibrary does. Later, we included a sentence indicating that we were seeking a partner to provide print-on-demand service for the Million Book Collection. As more and more publishers indicated that they were not inclined to participate because there was no direct financial reward, we reorganized the letter to foreground our efforts to provide print-on-demand service, highlighting that it would generate revenue for them. For a short time, we included letters of endorsement for the project. These letters articulated some of the work involved in participating, but praised the project and noted that the benefit was worth the cost. However, when new publishers we contacted commented that the work described in the letters was discouraging and a reason not to participate, we discontinued including the endorsement letters.

When Rhodes turned her attention to the MBP, we had not completed the data analyses from the feasibility study or the Posner project. We relied on the preliminary analysis of the data from the feasibility study to guide our copyright-permission work in the MBP. On the basis of the preliminary finding that university presses and scholarly associations were more likely than commercial publishers were to grant permission to digitize and provide open access to their copyrighted books (George 2001), copyright-permission work on the MBP started with university presses and scholarly associations. When we had contacted all the university presses and scholarly asso-

ciations with books cited in *BCL*, we began sending letters to commercial publishers, but soon stopped. Funding for the MBP copyright-permission work was running out, and we decided to dedicate our efforts to closing negotiations with publishers we had already contacted.

As in the Posner project, we often sent multiple letters to the same publisher because they had lost or misplaced the initial letter by the time we spoke to them on the telephone or contacted them by e-mail. We often sent subsequent letters as attachments to an e-mail message. To expedite the process, eventually we began sending even initial request letters as enclosures in e-mail if we could find an e-mail address for the publisher. From August to December 2004, 71 percent of the letters were sent by e-mail.

In the beginning, Rhodes was conducting follow-up calls or sending follow-up e-mails two weeks after we sent the initial letters. We discovered that in almost all cases, the publisher had not had a chance to even look at the letter in that period of time. We extended the period to three weeks, with little change in the results. By May 2004, we had extended the period to four weeks.

As of January 24, 2005, we had sent 665 initial request letters and made 782 follow-up attempts,⁵¹ either by telephone or e-mail, to reach 431 publishers. Over time, we abandoned 67 of the publishers, mostly commercial presses, because they were contacted before we had labor dedicated to the MBP permission work and too much time had passed with no response or follow-up. We had also significantly changed our request letter and strategy. The data analyses in this report are based on the 364 publishers with which we sought to close negotiations.

Tracking the Data

Because the Million Book Project takes a per-publisher, rather than per-title, approach to seeking copyright permission, we had to create a new database to track the work. The publisher database was designed and implemented in 2003. The previous database had a record for each title for which we were seeking permission; the new database has a record for each publisher. The publisher database contains fields for the name and address of the publisher, the name(s), phone number(s), and e-mail address(es) of the person or people we contacted at the publisher, and the dates when letters were sent and follow-up attempts made. Each record also contains buttons for coding the publisher type, indicating whether permission was granted or denied, and if granted, which option was designated in the signed contract (e.g., all out-of-print titles, designated titles). The database also contains a field for entering the date when permission was granted or denied and a field for entering notes about the negotiations.

We envisioned contacting thousands of publishers and, thinking it would facilitate our efforts, imported into the database the names and addresses of the publishers that we had successfully contacted

⁵¹ The number of follow-up contacts does not include the few follow-ups done by the intermittent labor that initially contacted publishers in June 2002.

in the feasibility study and Posner project. In hindsight, this was not a good idea. Many, perhaps most, of these publishers were not university presses or scholarly associations with books cited in *Books for College Libraries*; consequently, many of these publishers have not been contacted for the Million Book Project. The design of the database makes it difficult to do statistical analyses of only the publishers that we have contacted during the project. Furthermore, the report features of the database have not been implemented. Consequently, the statistical data have been tracked in spreadsheets.

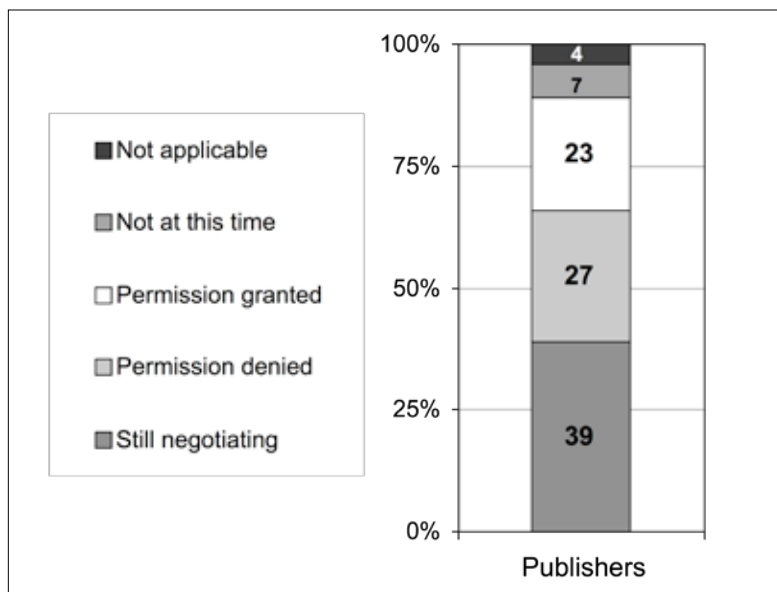
The spreadsheets evolved over time in several ways. First, we revised the basic spreadsheet periodically to facilitate different kinds of statistical analyses. For example, in July 2004 we decided to track the data separately for significantly different versions of the request letter to see whether the revision had any effect on the response or success rate. We also revised the spreadsheet to incorporate new ways to code responses from publishers that we had not expected. Initially, completed MBP copyright negotiations were simply recorded as "Permission granted" or "Permission denied." Over time, we created two new categories to record other types of responses. Publishers that expressed interest in the project but did not have the resources to invest in determining the current print status or copyright ownership of their titles were coded as "Not at this time." Publishers that explained that copyright reverted to the author when their books went out of print were coded as "Not applicable." Sometimes we sent multiple letters or made multiple follow-up contacts only to discover that rights to out-of-print books had reverted to the author.

Overall Results

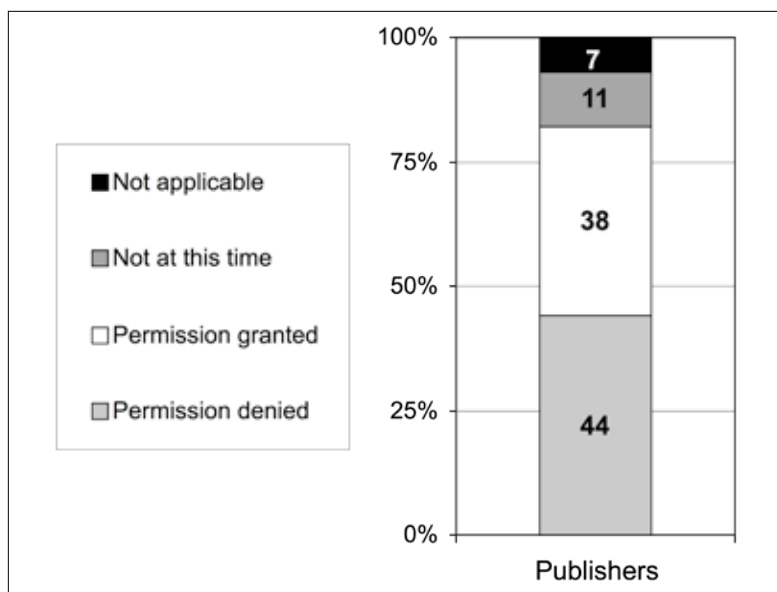
We located all the publishers that we attempted to contact in the MBP. Not counting the 67 publishers that we abandoned, as of February 2005, 61 percent of the negotiations had been completed. The others are under way. Almost one-fourth of the publishers have granted permission to include at least some of their titles in the Million Book Collection; slightly more have denied permission. Some responded "Not at this time," and a few are considered "Not applicable" because copyright for their out-of-print books reverted to the author (figure 18).

Given the difficulty of locating authors and estates (a lesson learned from the Posner study) and the expense of the per-title approach to seeking copyright permission, we did not systematically try to locate authors or their estates in all cases where copyright reverted to the author. We did, however, conduct some experiments to assess the difficulty and potential cost of locating authors or estates that owned the copyright to selected titles cited in *BCL*.⁵² The results of these experiments are discussed on pages 53–54 of this report.

⁵² Letters to authors requested permission for one of the following: (1) titles cited in *BCL* to which they own the copyright (the titles were listed in the letter); (2) all the out-of-print titles to which they own the copyright (they provide the list); and (3) a list of designated titles to which they own copyright (they provide the list).

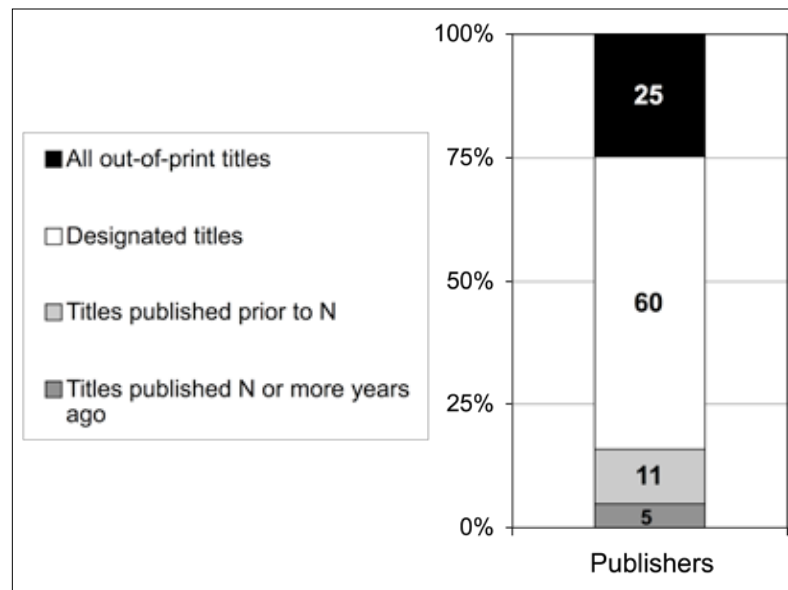
Fig. 18. Analysis of the 364 publishers contacted and not abandoned

Looking at only the completed negotiations (figure 19), more than a third of the publishers granted permission and close to half denied permission. Of those that denied permission, 80 percent denied permission outright. The rest were considered “Permission denied” according to the three-strikes rule.

Fig. 19. Analysis of completed negotiations

Of the publishers that granted permission, one-fourth granted permission for all or most of their out-of-print titles.⁵³ More than half granted permission for titles that they specified. Some designated a date of publication prior to which their books could be scanned. Few specified a number of years from the date of publication prior to which their books could be scanned (figure 20).

Fig. 20. Analysis of the permissions granted



The number of titles for which permission was granted is not yet known. Lists must be compiled of participating publishers' out-of-print, in-copyright titles or the titles they published prior to the designated date or time period. Project partner OCLC is helping us compile these lists, which we will send to the publishers for approval before locating the books and shipping them to India or China for scanning. In other cases, we are waiting for the publisher to provide its list of designated titles. Lists received to date range from half a dozen to several thousand titles. Most lists include several hundred titles. Without making projections for the lists yet to be received, we estimate that as of mid-February 2005 we had been granted permission to digitize at least 52,900 titles from publishers, authors, or estates with works cited in *BCL*.

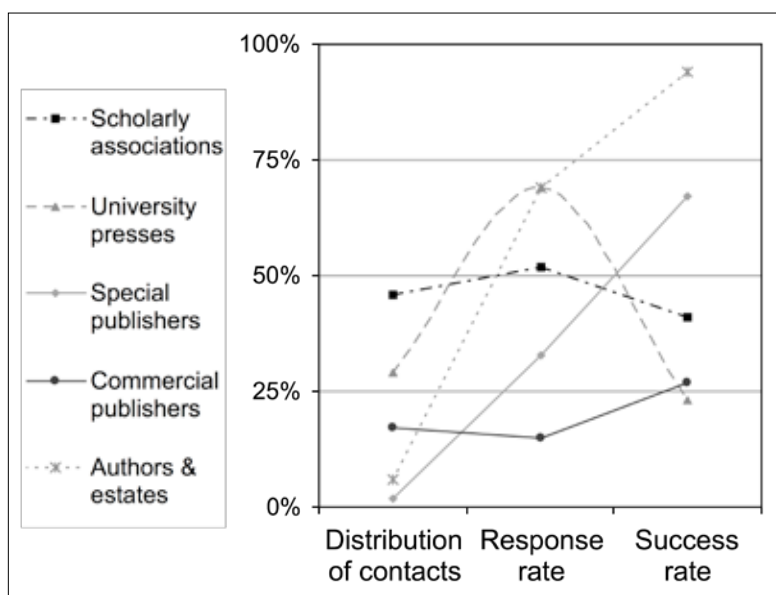
Analysis by Publisher Type

As they had in the previous two studies, the response and success rates for completed negotiations varied across different types of publishers. The response rate for completed negotiations for all publisher types was significantly lower in the MBP than in the Posner

⁵³ A few publishers granted permission for all of their out-of-print titles except those that they specified on the contract.

project. Most of the publishers contacted in the MBP were university presses and scholarly associations. As in the Posner study, special publishers, authors and estates, and scholarly associations were the most likely to grant permission. University presses were the least likely. Although the success rate for scholarly publishers and university presses was comparable in the feasibility and Posner studies, it was significantly lower in the MBP (figure 21).

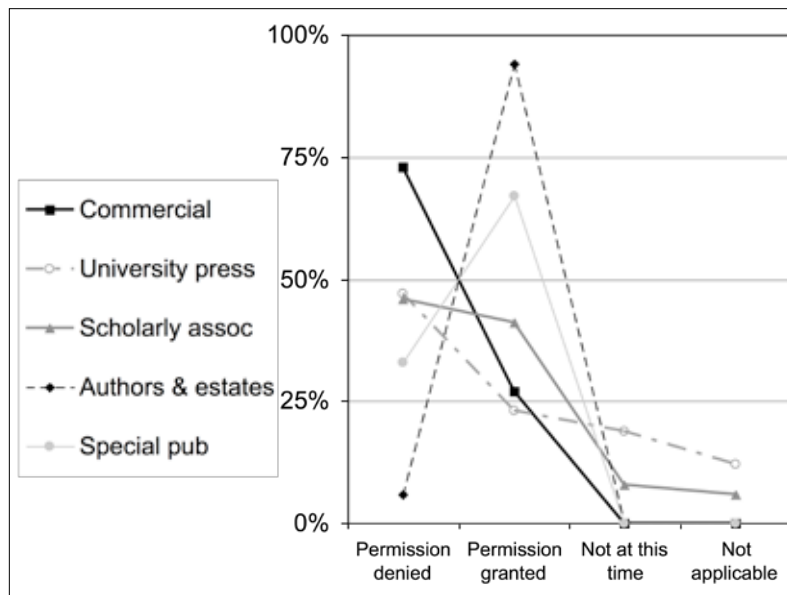
Fig. 21. Analysis by publisher type



In response to feedback from publishers indicating that their participation in the MBP was contingent on generating revenue, we revised our initial request letter in July 2004 to foreground plans to provide print-on-demand service for the Million Book Collection. Print-on-demand would generate new revenue for participating publishers. The only noteworthy change in response to the new letter was that the success rate of completed negotiations with scholarly associations increased from 39 percent to 48 percent. We do not know whether this increase was directly related to our emphasis on the potential for revenue.

Figure 22 provides a more-detailed analysis of completed negotiations by publisher type. Few (9 percent) of the permissions denied were based on the three-strikes rule, and all of them were from university presses and scholarly associations. University presses were the most likely to respond "Not at this time" or "Not applicable."

Figure 23 provides an analysis of permissions granted by publisher type. Data on authors and estates and on special publishers are combined as "Other" copyright holders. Only four scholarly associations chose the "moving-wall" model of JSTOR, granting us permission to digitize and provide Web access to books published 1, 2, 7, and 15 or more years ago. One university press and eight scholarly associations granted permission for books published prior to a

Fig. 22. Analysis of completed negotiations by publisher type

designated date. The university press allowed us to digitize books published prior to 1990. The scholarly associations gave permission for us to digitize books published prior to 1951, 1980, 1991, 1992, 1995, 2000, 2001, and prior to "the present." Not included in this analysis is the publisher that granted permission for all its titles published prior to 1923, i.e., all its out-of-copyright titles, for which its permission is not required. The two participating commercial publishers granted permission for all of their out-of-print titles, although one designated some exceptions. Most university presses and scholarly associations chose to provide a list of designated titles. This was also the most popular choice for special publishers and authors and estates.

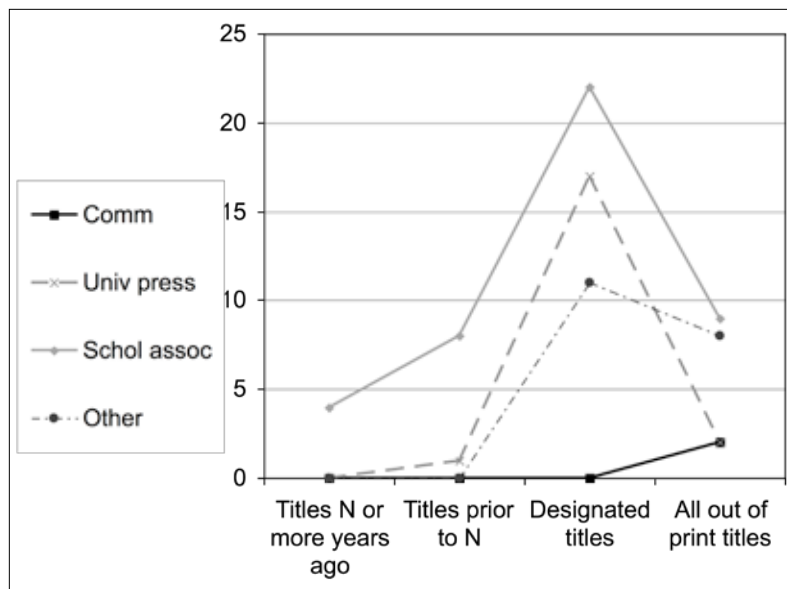
Fig. 23. Analysis of permissions granted by publisher type (number of publishers)

Table 3 lists the publishers that had granted permission as of February 2005. Seventeen authors and estates have also agreed to participate in the project. Some university presses and scholarly associations have agreed to consider participating if we send them a list of their titles to review and approve for inclusion in the Million Book Collection.

Requests that We Provide Lists of Titles

Twelve university presses and two scholarly associations responded they would consider participating in the MBP if we provided a list of their out-of-print titles of interest to us. We compiled lists for six of the presses. In one case, at the request of the press, we compiled and provided a second list. We experimented with how to reduce the time and labor cost of producing lists of titles. The experiment is described on pages 52–53 of this report.

One university press responded that the list of titles we provided was “inappropriate.” The press no longer owned the copyright to some of the titles and some of the listed books had third-party copyrights. We coded this publisher as “Permission denied” according to the three-strikes rule. Another press denied permission. The others are still reviewing the lists we provided. Recent correspondence from one of the presses indicates the work involved for the publisher even when we provide the list of titles:

My sincere apologies for the length of time it is taking us to research these copyrights. It gets a little difficult with the older titles, as the records are often incomplete and correspondence with our authors to verify rights can be slow. . . . It’s also, unfortunately, just me doing this research around the schedule of my main Acquisitions job. I am currently trying to slog through in-print backlist titles to check our electronic rights, and hopefully this includes some of the books on your list. But to answer your question, yes, we are definitely still interested in participating (e-mail from a university press to Erin Rhodes, February 23, 2005).

We began compiling lists of books the university presses had published that were cited in *BCL*. We checked the copyright-renewal records for cited U.S. titles published between 1923 and 1963 and included a title on the list only if the copyright had been renewed and the claimant was the press.⁵⁴ The lists we prepared indicated that these titles were our top priority. If the press had fewer than 100 titles cited in *BCL*, we added titles until we had a list of about 100 titles. When we needed to add titles not cited in *BCL*, we started with older titles, moving forward in time until we reached 100 titles. We soon changed our strategy to begin adding titles published in 1964 or later

⁵⁴ The U.S. Copyright Office confirmed that the “claimant” is the copyright holder at the time of renewal. We assumed that if the claimant were the author, that copyright had reverted from the publisher to the author.

Table 3. Publishers granting permission in the MBP as of February 2005

American Antiquarian Society	Law and Justice Foundation of New South Wales
American Enterprise Institute	Louisiana State University Press
American Folklore Society	Marquette University Press
American Geographical Society	Medical Library Association
American Institute of Biological Sciences	Mercer University Press
American Library Association	Michigan State University Press
American Meteorological Society	National Academies Press
American Musicological Society	National Association for the Education of Young Children
American Society of Agricultural Engineers	National Association of Counties
Arctic Institute of North America	Pacific Research Institute
Barnes Foundation Press	Polynesian Society
Brookings Institution Press	Rand McNally
Carl Vinson Institute of Government	Robert Schalkenbach Foundation
Catholic University of America Press	Rockefeller University Press
Cato Institute	Royal Institute of International Affairs
Colonial Williamsburg Foundation	Southern Illinois University Press
Dante Society	Special Libraries Association
Educational Products Information Exchange	Summer Institute of Linguistics
Employee Benefit Research Institute	Totem Design Press
Forging Industry Association	Texas A&M University Press
Fraser Institute	Texas Christian University Press
Geological Society of America	Texas Tech University Press
Hansard Society	Urban Institute
Haynes Foundation	University of Alabama Press
Hebrew Union College Press	University of British Columbia Press
Institute of Carmelite Studies	University of Hawaii Press
Institute of Economic Affairs	University of New Mexico Press
Institute of Industrial Relations	University of North Carolina Press
Institute of Intergovernmental Relations	University of Texas Press
Institute of Labor and Industrial Relations	Wenner-Gren Foundation for Anthropological Research
International Institute of Refrigeration	W. E. Upjohn Institute for Employment Research
Johns Hopkins University	Wilfrid Laurier University Press

to avoid incurring the cost of checking the copyright-renewal records for titles not cited in *BCL*. We discovered additional titles by searching *Global Books in Print (GBP)*, specifying a date range, and limiting the query to out-of-print books. At the time, we believed that compiling lists of older, out-of-print titles would increase our success rate. With no way to determine whether the publisher was still the copyright owner of titles published in 1964 or later, we simply assumed that they were.

We soon learned that *GBP* does not include many older titles and that publishers define *out of print* differently from librarians. Presses responded that titles on our list were not out of print, regardless of whether *GBP* designated the item on our list (author, title, edition, publication date, ISBN) to be so.

An example of negotiations with one university press conveys the complexity of the situation. We checked the copyright-renewal records for its titles cited in *BCL* and for additional titles found in *GBP*. In the process, we discovered that the author had apparently renewed copyright to 26 of the titles. We then asked the press whether copyright to its books reverts to the author when a book goes out of print. The press's response sheds light on the different or changing definitions of *out of print* and how changing contractual practices

make it difficult even for publishers to determine copyright ownership:

Not to parse words too carefully, but the answer here at [university press] is not all that straightforward. First, there is the issue of "out of print." We seem to have had several differing views of that concept over the years. My director and I feel that to the extent we have any copies at all on our warehouse shelf, the book is not out of print. Second, with the print-on-demand possibilities these days, there's an argument that a book is never out of print if you control the rights. Others, of course, look at things a little differently.

Second, there are the contractual issues involved. In other words, the vast majority of our author contracts do not call for any automatic reversion of rights. If we run out of copies to sell, the author can ask us to declare the book out of print and ask that we revert the rights. If they don't ask, then we don't have to do anything at all. That phraseology has not always existed in our contracts and so, on a number of older books, we apparently did declare them out of print and we formally transferred the rights back to the author. I'd bet that many of the renewals you have seen for our materials have been through this mechanism, with the author renewing copyright in his or her own right after we reverted (e-mail from a university press to Erin Rhodes, August 2, 2004).

An example from another university press further illustrates the current state of affairs:

Copyright can be renewed by our authors, but only if they request the rights (including copyright) reversion from the Press. We do not allow copyright to revert automatically. What I assume happened with many of these titles is that they went out of print and the authors requested the rights revert to them. After the Press processed the necessary paperwork, the authors registered the renewals in their names. This would be standard procedure now, of course. We do not have records detailing the progression of copyright reversion and/or renewal for our older titles, such as those on the Million Books Project list (e-mail from a university press to Erin Rhodes August 18, 2004).

The bottom line is twofold. First, publishers need to consult the contract for each title to determine who owns the copyright. Second, we need to understand their definition of *out of print*. We did not know how the presses that asked us for a list of their out-of-print titles defined *out of print* and even if we did, we could not know, for example, how many copies of a title they had in their warehouses. Consequently, we stopped checking the print status of books as we compiled lists of titles. When we no longer needed to limit our query to out-of-print books, we switched from searching *GBP* to searching *WorldCat* to discover titles because *WorldCat* cataloged more older

books. Ironically, by August 2004 we started compiling lists of more-recent titles because in our follow-up negotiations with university presses they frequently responded that copyright to older books had reverted to the author.

Request that We Check Copyright-Renewal Records

One university press offered to give us a list of all of its out-of-print books if we would check the copyright-renewal records and report back on the titles to which it still held the copyright. They explained that although they “technically” know which copyrights have reverted to authors, the information is not in a database. They would have to go through all of their paper files “to see if there is a reversion-of-rights contract.” When we told our contact at the publisher about the copyright-renewal records, she responded, “I didn’t realize there was a simple way for you to find out which books are not in [university press] copyright. It’s great to hear there is” (e-mail from university press to Erin Rhodes, August 2, 2004). If we would identify all the out-of-print books to which this press owns the copyright, it would give us permission to digitize and provide Web access to them.

There are several problems with this approach. First, the MBP has limited resources to do the copyright-permission work for the project. Second, even if we had the resources to check the copyright-renewal records for publishers, the proposal is seriously flawed. It reveals how little some publishers know about copyright.⁵⁵ The copyright-renewal records apply only to books published between 1923 and 1963, and even if the press had renewed a copyright, there is no guarantee that copyright did not revert to the author sometime after it was renewed. Furthermore, it is highly likely that the press published books after 1963 that are now out of print. We have no way of verifying who owns the copyright to any book. Only the publisher can verify whether it still owns the copyright and, if not, who owned it after them. However, time and time again throughout the MBP, publishers responded that they have no easy way to determine either copyright ownership or the print status of the books they have published. The copyright data exist in paper files that must be consulted one by one. Nineteenth-century record-keeping methods seriously impede acquiring copyright permission by any means other than a title-by-title request.

We did not agree to consult the copyright-renewal records for this press. They do, however, plan to participate in the MBP and are slowly working their way through their paper records to identify appropriate titles.

⁵⁵ Further evidence that publishers do not understand copyright is apparent in the response of one scholarly association, which returned a signed contract granting us permission to digitize all its titles published prior to 1923—titles that are out of copyright and therefore did not require their permission to digitize. The contract is dated July 26, 2004.

Experiments to Contain Costs

Compiling Lists of Titles. By February 2004, several university presses had asked us to provide lists of their out-of-print titles that we wanted to include in the Million Book Collection. Our top priority was the titles cited in *BCL*.

The first press that asked us for a list had almost 1,000 titles cited in *BCL*. We decided to send them a sample list to test their response, rather than invest resources in compiling a comprehensive list. Three steps were involved in compiling the test list of 76 titles. We quickly discovered that many of the press's citations in *BCL* were incomplete, so our first step was to verify the citations. If the book was published between 1923 and 1963, we then checked the copyright-renewal records to see whether the copyright had been renewed and, if so, who owned the copyright at the time of renewal. If the title was still in copyright and the copyright was owned by the publisher at the time of renewal, we then consulted *Global Books in Print* to determine whether the title was out of print. If it was out of print, we put it on the list. This was a tedious process. We devised some experiments to see how long each of these steps was taking and what we could do to reduce the time.

Working from a spreadsheet of problematic citation data derived from attempts to parse the OCR text of a digitized copy of *BCL*, Rhodes and Brown compared the time it took to verify a citation by searching *WorldCat* with the time it took to verify a citation by searching the OCR from the digitized *BCL*. Verifying a citation using *WorldCat* took substantially longer because the search often yielded multiple results, which meant that time had to be spent examining the full bibliographic records to determine which item was the exact book cited in *BCL*.

Rhodes and Brown also compared the time it took to consult the copyright-renewal records using

- a digitized version of the U.S. Copyright Office Catalog of Copyright Entries, which contains renewal records for books published between 1923 and 1949 and renewed before 1978⁵⁶
- the Office of Copyright's online Copyright Catalog, which contains renewal records for books published 1950 or later and renewed after 1978⁵⁷

⁵⁶ Available at <http://digital.library.upenn.edu/books/cce/>. This site provides access to digitized copies of the renewal records for books and serials published between 1923 and 1950. The records were digitized at Carnegie Mellon. John Ockerbloom, formerly of Carnegie Mellon and now at the University of Pennsylvania, organized and maintains the Web site. Hypertext links are provided to navigate the volumes and pages. To mid-1973, when the U.S. Copyright Office stopped alphabetizing renewal records, each page link includes a summary of the alphabetical range of copyright-holder names covered by the page. Copyrights renewed from mid-1973 to 1977 must be located by registration number. Links are provided to searchable text transcriptions of copyright renewals from these years. The transcriptions were prepared by Project Gutenberg. Registration numbers are also available in the multivolume *Catalog of Copyright Entries*, which is not available online.

⁵⁷ Available at <http://www.copyright.gov/records/cohm.html>. The U.S. Copyright Office provides access to a database of renewal records for books, performing and visual arts, sound recordings, and other registered works

- the rekeyed, ASCII version of all the copyright-renewal records prepared by Distributed Proofreaders, with a search interface provided by Rutgers University professor and MBP partner Michael Lesk⁵⁸

The “Lesk database,” as we affectionately call it, proved to be the fastest, most flexible, and most effective means to check copyright-renewal records.

As a result of these time trials and feedback from university presses, we began using the online text of *BCL* to verify citations (resorting to WorldCat only when the *BCL* citation was incomplete), consistently used the Lesk database to check renewal records, and stopped checking the print status of books. Ultimately, we reduced from nine to six minutes the time it took to put one title on the list. It took about three minutes to verify the citation of a title that did not require us to consult the renewal records. Even with the new, streamlined approach, the cost of preparing a list of 100 titles was \$80 to \$160, depending on whether the renewal records needed to be consulted.

We prepared lists of titles using Microsoft Word because it was easy to copy, paste, and reformat the citations using this program. Our plan was to subsequently create a list in Excel only for titles that the publisher granted permission to digitize and include in the Million Book Collection. Excel would be a better vehicle for tracking when the books had been located, shipped abroad for scanning, and returned to the lending library.

Locating Authors and Estates. University presses often told us that copyright to older, out-of-print books had reverted to the author. Our experience in the Posner project had taught us that authors and estates are difficult to locate, but very likely to grant copyright permission to digitize and provide Web access to their books. We managed to locate most of the authors and estates that owned the copyright to books in the Posner collection, but it took a considerable investment of time, which is money. Although the approach to seeking copyright permission in the MBP is per publisher, rather than per title, we decided to experiment with ways to reduce the cost of locating authors in an effort to secure copyright permission for titles cited in *BCL* for which copyright had reverted to the author.

We contacted the Authors Registry. During the Posner project, we had learned that they would try to locate 10 authors per week (2 per day) at no charge. For the MBP, we asked the registry to locate 25 authors or estates. They charged us \$2.50 per author or estate for which they could locate an address, and they responded the same day we

(published between 1950 and 1977), with the exception of serials. Fielded searching is supported, e.g., author, title, claimant (copyright holder at the time of renewal), and registration number. The U.S. Copyright Office renewal records for serials are available at <http://www.copyright.gov/records/cohs.html>.

⁵⁸ Available at <http://www.scils.rutgers.edu/~lesk/copyrenew.html>. Keyword searching is supported.

submitted our request. They found addresses for only about half of the authors and estates for which we requested information, but almost all the contact information they provided was correct. Of the authors who responded to our query, 94 percent granted permission.

While the experiment demonstrated that the Authors Registry is a cost-effective way to locate authors, the benefit of contacting authors or their estates for the MBP was quite small. A successful negotiation with an author or estate typically yields permission for only one or two books, while a successful negotiation with a publisher typically yields hundreds of books. The transaction cost of the negotiations per title is much higher with authors than with publishers. The MBP copyright-permission work will continue to focus on publishers.

Analysis of Transaction Costs

We closely monitored the labor costs of the full-time copyright-permission assistant at Carnegie Mellon and the part-time assistant at UC Merced. We also tracked the cost of paper and postage at both universities and long-distance telephone charges at Carnegie Mellon. We were unable to track long-distance charges at UC Merced. We did not factor in the cost of the intermittent labor that began the project, Internet connectivity, database creation, consultation with university legal counsel, or administrator time. As in the Posner project, university legal counsel did not levy a fee for consultations and advice. As project administrator, I answered many questions from the copyright-permission assistants and publishers, and conducted the data analyses.⁵⁹

From November 2003 through January 2005 we spent \$35,876 on labor (wages and benefits), \$615 on long-distance phone calls, and \$217 on paper and postage. On the basis of the conservative estimate of the number of titles for which permission was granted (52,900), the average transaction cost for the Million Book Collection was \$0.69 cents per title. The cost would have been significantly higher if it had included my time and the cost of Internet connectivity and database creation.

Conclusions and Lessons Learned

We encountered significantly less difficulty locating publishers of works cited in the 1988 edition of *BCL* than we did in the feasibility and Posner studies. Perhaps there are fewer older titles cited in *BCL* than in the random sample or the Posner collection. Not counting the publishers we abandoned early in the project, almost all the publishers we contacted in the MBP responded to our request. Only 9 per-

⁵⁹ The number of e-mail messages regarding the copyright-permission work provides an estimate of the volume of administrator activity on the project. I sent to or received from the copyright-permission assistants 629 e-mail messages from 2001 through 2004. Another 591 messages were sent to or received from publishers.

cent of the completed negotiations were coded “Permission denied” through lack of response.

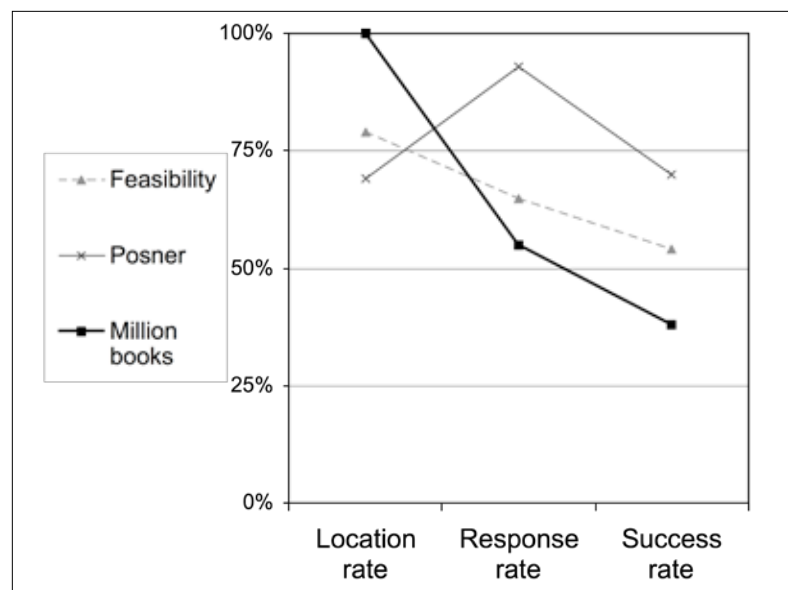
Table 4 and figure 24 provide comparisons of the results of the random sample feasibility study, the Posner study, and the MBP copyright-permission work as of February 2005. The total number of publishers that responded in the MBP is the number of negotiations that were closed by a response from the publisher. It does not include the 19 publishers coded as “Permission denied” according to the three-strikes rule. In figure 24, the *location rate* is based on the number of publishers we tried to contact, the *response rate* is based on the number of publishers located, and the *success rate* is based on the number of responses or, in the case of the MBP, the number of completed negotiations.

Table 4. Comparison of the results of the three studies

	Publishers				Titles	
	Total overall	Total located	Total responded	Granted permission	Permission granted	Transaction cost per title granted
Feasibility	209	165	106	57	66	\$200.00
Posner	104	72	67	45	178	\$78.00
Million Books	364	364	202	84	52,900	\$0.69

Figure 24 suggests that the Posner study is our most successful project to date in terms of response and success rates. However, the transaction cost per title for permissions granted in the Posner study, though significantly better than in the feasibility study, is far too high to pursue on a large scale. Despite the lower overall success rate, the per-publisher approach taken in the MBP garnered permission for

Fig. 24. Comparison of location, response, and success rates



significantly more titles at less cost than the per-title approach of the previous projects.

Publishers contacted in the MBP expressed the same reservations as publishers in the earlier projects did. Their reasons for not participating in the project were fear of open access and lost revenue. Scholarly associations and, to a lesser extent, university presses are more receptive to open access to older works than are commercial publishers. Many publishers, particularly university presses, indicated that they wanted to participate, but could not because copyright reverts to the author when their books go out of print. Many publishers also noted that older contracts did not grant them electronic rights to the books or that they were uncertain of their rights in this regard. Copyright reversion and contractual rights turned out to be significant barriers to our per-publisher approach. As in the previous studies, we found authors to be much more helpful, willing, and receptive than publishers were. The authors' overall perceptions of open access differ greatly from those of publishers. Rather than seeing open access as a channel for lost revenue, many authors see it as a way to preserve their work online.

In addition to the concerns and constraints noted above, some publishers were hesitant to participate in the MBP because they were already involved in other electronic projects. Few publishers named these projects. Those that did mentioned the Bibliovault. Even though the MBP requests nonexclusive permission, publishers simply did not want to attach themselves to more than one or two online projects at a time. Their comments suggest that they were uncertain whether they had granted exclusive rights to these other projects.

The most common response from publishers that chose not to participate in the MBP was they did not have the time and staffing necessary to participate. Their preference was to grant permission for a designated list of titles, but they did not have the resources to compile the lists. Given their methods of record keeping, preparing a list of titles would require examining their paper files title by title to determine copyright status and ownership, what rights they had, and, depending on their definition of *out of print*, checking their warehouses to determine print status. University presses, in particular, mentioned tight budgets prohibiting them from engaging in any activities other than their necessary daily functions. Some publishers said that they could not allocate the necessary resources to the MBP because it would not generate revenue.

As in the Posner study, publishers were concerned about the quality of the scanning being done in the MBP and the functionality of the delivery system. In response to these concerns, we provided our digitization and functional specifications, and referred them to the MBP project Web sites, where they could see public domain titles scanned in the project.⁶⁰ A few publishers introduced a new concern

⁶⁰ The Million Book Collection is not yet integrated in its entirety on a single Web site. The books that have been integrated are available at the Universal Library site in the United States at <http://www.ulib.org>. Books scanned in China are available at <http://www.ulib.org.cn>. Books scanned in India are available at <http://www.dli.ernet.in/>. Use Internet Explorer to access the collection.

that we found rather odd. They were concerned that their titles were “not appropriate” for the Million Book Collection. One scholarly association commented that its out-of-print books were “too specific.” We reassured all the publishers that we were interested in including monographs, serials, and bibliographies in the collection. One scholarly association explained that it denied permission because all its out-of-print books were available in a commercial database, “therefore [the association] does not wish to participate in the Million Book Project as librarians consistently tell us that they do not wish to have the same electronic content available from multiple sources” (Letter from publisher to Dean of University Libraries Gloriana St. Clair, July 7, 2004).

Unlike the publishers in the previous studies, those in the MBP sometimes changed their responses. For example:

- A university press initially granted permission, but later informed us that permission was denied. Apparently a new director had just been hired, and this person chose not to participate. We changed the response from “Permission granted” to “Permission denied.”
- One university press verbally granted permission and sent us a list of titles to review and indicate which ones we wanted to include in the Million Book Collection. Librarians marked the titles of interest, and we returned the list to the press. The head of the press later responded that rights to its out-of-print books revert to the author. We changed their response from “Permission granted” to “Not applicable.”
- Another university press planned to participate, then discovered that the copyright to its out-of-print books reverted to the author. This press now plans to inform authors about the MBP when copyright reverts to them.
- A scholarly association granted permission for designated titles. Later, when they discovered how long it would take to figure out whether they still owned the copyright to a title or whether the book was out of print, they reported that they could not participate at this time. We changed this association’s response from “Permission granted” to “Not at this time.”
- Another scholarly association responded that copyright to its out-of-print titles reverted to the author. Intrigued by the MBP and confident that authors would grant permission, they offered to contact the authors and negotiate permission to include their works in the Million Book Collection. When they discovered that most of their titles had multiple authors, they got discouraged at the effort required and changed their minds.
- One university press initially granted permission for a list of designated titles but later decided to grant permission for all its out-of-print books.

The MBP confirmed that dedicated personnel, experimentation, and flexibility are critical to success in acquiring copyright permission to digitize and provide open access to books. Adapting strate-

gies and adjusting processes to accommodate what we learn day-to-day could further improve the results of our efforts. Again, we need to develop a better way to manage the data and routinely calculate statistics. More-sophisticated, ongoing analyses might expose trends that could be leveraged during the project to reduce the cost and increase the success of seeking copyright permission for open access.

Looking Ahead

Initial meetings with Carnegie Mellon legal counsel in October 2002 led to the preparation of multiple drafts of a “reasonable effort” document intended to detail the steps and diligence required to determine the copyright status and identify and locate the copyright owners of books. The understanding was that if we designed and followed a rigorous workflow approved by legal counsel and documented our efforts, then we could digitize and provide Web access to books without permission under certain conditions (for example, if the publisher had gone out of business or we could not ascertain who owned the copyright to a work). We agreed that if we digitized a book and made it Web accessible without permission and the copyright owner then contacted us, we would remove that book from the Web at the owner’s request. However, in May 2003, university legal counsel changed their minds and took a more conservative approach: no permission, no digitization and access. They are now reconsidering this decision.

At the request of the American Library Association Office of Information Technology Policy (ALA OITP), I revised the “reasonable effort” document in August 2004 and submitted it to the office as the basis for beginning development of a best practice for pursuing copyright permission to digitize and provide open access to books.

Invited by the OITP, I presented the results of Carnegie Mellon’s copyright-permission research to the ALA congressional lobbyists in November 2004. The lobbyists responded that the per-publisher approach used in the MBP, which reduced the transaction cost to \$0.69 per title, would not persuade Congress that acquiring copyright permission is prohibitively expensive under the current copyright regime. The transaction cost of the per-title approach taken in the Posner project, \$78 per book, is more likely to be persuasive and yield changes in public policy. We all agreed that though the per-publisher approach of the MBP is consistent with the vision of the Universal Library Project, the approach is artificial in terms of what libraries typically do in regard to digitizing collections. Standard library practice is to target designated collections—as we did with the Posner Memorial Collection—and to seek copyright permission as appropriate for titles in those collections. Efforts to inform public policy should be based on the more practical and typical per-title approach to acquiring copyright permission, where the transaction cost per title will be substantially higher.

Future copyright-permission research conducted by Carnegie Mellon University Libraries will be shaped to inform public policy. Although MBP work will continue to use a per-publisher approach, our other digitization projects will take the per-title approach. All future efforts to acquire copyright permission for open access will track labor costs at a finer grain of detail to better understand how the time is spent. In addition, several ideas have been proposed to a private foundation to conduct research aimed at informing public policy, including convening meetings of experts and stakeholders to

- draft legislation and promotional materials for worldwide copyright exemptions and safe harbors for library, archival, educational, and research uses of orphan works; and
- design a study of copyright misuse (i.e., misuse of copyright law by copyright holders) in licenses and DRM technologies for online resources.

The doctrine of *copyright misuse* provides a mechanism for users of copyrighted works who have been charged with copyright infringement to hold copyright owners accountable when such owners make improperly broad claims to their rights. Grounded in case law beginning in 1990, the doctrine forbids copyright owners from attempting to secure exclusive rights, for example, through restrictive licensing practices or DRM technologies that are contrary to public policy or not granted by copyright law. The penalty for copyright misuse is unenforceability of the copyright in court until the misuse has been purged and its effects no longer exist. A finding of copyright misuse is “tantamount to losing the copyright” temporarily (Hollaar 2002; see also Carney 2003).

I encourage librarians to adopt or adapt the workflows and strategies described in this report in their copyright-permission efforts if they appear to be helpful. I encourage librarians to continue advocating for open or affordable access to scholarly information. And I urge them to lobby for the development of laws, licenses, and technologies that do not sacrifice public rights.

References

*All URLs were valid as of
October 15, 2005.*

537 U.S. January 15, 2003. Opinion of the Supreme Court of the United States, *Eldred v. Ashcroft*, No. 01-618. Available at <http://straylight.law.cornell.edu/supct/pdf/01-618P.ZO>.

Antelman, Kristin. 2004. Do Open-Access Articles Have a Greater Research Impact? *College and Research Libraries* 65(5): 372–382.

Boynton, Robert S. 2005. Righting Copyright: Fair Use and “Digital Environmentalism.” *Bookforum* (February-March). Available at <http://www.bookforum.com/boynton.html>.

Brief for Petitioners. May 20, 2002. Supreme Court of the United States, *Eldred v. Ashcroft*, No. 01-618. Available at <http://cyber.law.harvard.edu/openlaw/eldredvashcroft/supct/opening-brief.pdf>.

Brito, Jerry. 2002. Copyright and the Court’s Right. *Brainwash* (Oct. 5). Available at <http://www.affbrainwash.com/archives/007526.php>.

Carney, David. 2003. 3rd Circuit Breaks New Ground on Copyright Misuse. *Tech Law Journal* (August 26). Available at <http://www.techlawjournal.com/topstories/2003/20030826.asp>.

Carroll, Steve. December 2004. Librarians and Copyright. *Library Connect newsletter* 2(4): 9.

Conference on Fair Use. 1998. *Final Report to the Commissioner on the Conclusion of the Conference on Fair Use*. Available through the Internet Archive at <http://Web.archive.org/Web/20021103235358/www.uspto.gov/Web/offices/dcom/olia/confu/confurep.pdf>.

Congressional Record 144 (House 9946-9954). October 7, 1998. Document ID cr07oc98-78.

Copyright Act of 1790. Available at <http://earlyamerica.com/earlyamerica/firsts/copyright/>.

Copyright Law of the United States of America. 2003. *Circular 92*. Available at <http://www.copyright.gov/title17/circ92.pdf>.

Copyright Renewal Act. 1992. Available at http://thomas.loc.gov/cgi-bin/cpquery/?&db_id=cp104&r_n=sr315.104&sel=TOC_133055&.

Copyright Term Extension Act. 1998. Available at <http://www.copyright.gov/legislation/s505.pdf>.

Davis, Phil, Terry Ehling, Oliver Habicht, Sarah How, John M. Saylor, and Kizer Walker. 2004. *Report of the Cornell University Library’s Task Force on Open Access*. Available at <http://techreports.library.cornell.edu:8081/Dienst/UI/1.0/Display/cul.lib/2004-3>.

De Rosa, Cathy, Lorcan Dempsey, and Alane Wilson. 2003. Lifelong Learning in the Community. *Environmental Scan*. Dublin, Ohio: Online Computer Library Center (OCLC). Available at <http://www.oclc.org/membership/escan/research/learning.htm>.

Digital Millennium Copyright Act. 1998. Available at <http://www.loc.gov/copyright/legislation/hr2281.pdf>.

Downes, Larry. 2002. "Free the Mouse" for Creativity's Sake. *USA Today* (October 7). Available at http://www.usatoday.com/news/opinion/editorials/2002-10-07-oplede_x.htm.

Electronic Privacy Information Center. 2002. Before the Technology Administration, Department of Commerce, In the Matter of Digital Entertainment and Rights Management. Available at <http://www.epic.org/privacy/drm/tadrmcomments7.17.02.html>.

Ellis, David Roy. 1999. Copyright Term Extension: Congress Enacts a Mickey Mouse Law and Legislation for the New Millennium. *Clearwater Bar Association Newsletter* (January). Available at <http://www.clwbar.org/resipsa/jan99/copyright.html>.

Evaluation of the Distributed National Electronic Resource (EDNER) Project. 2002. How Students Search: Information Seeking and Electronic Resource Use. Issues Paper 8.

Friedlander, Amy. October 2002. *Dimensions and Use of the Scholarly Information Environment: Introduction to a Data Set*. Washington D.C.: Council on Library and Information Resources. Available at www.clir.org/pubs/reports/pub110/contents.html.

Gass, Andy, and Helen Doyle. 2005. The Reality of Open-Access Journal Articles. *The Chronicle of Higher Education* (February 18): B13.

George, Carole A. 2004. Analysis of the raw data from a survey of how graduate students discover and obtain information. Unpublished. Contact cgeorge@andrew.cmu.edu.

George, Carole A. 2001. *Exploring the Feasibility of Seeking Copyright Permission*. Available at <http://www.library.cmu.edu/Libraries/FeasibilityStudyFinalReport.pdf>.

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

Harnad, Stevan, and Tim Brody. 2004. Comparing the Impact of Open Access (OA) vs. Non-OA Articles in the Same Journals. *D-Lib Magazine* 10(6). Available at <http://www.dlib.org/dlib/june04/harnad/06harnad.html>.

Hirtle, Peter B. 2004. *Copyright Term and the Public Domain in the United States 1 January 2005*. Available at http://www.copyright.cornell.edu/training/Hirtle_Public_Domain.htm.

Hollaar, Lee A. 2002. Misuse of Copyright. Chapter I, Section IIK in *Legal Protection of Digital Information*. Available at <http://digital-law-online.info/lpdi1.0/treatise15.html#secII.K>.

Hutchinson, Kristine H. 2003. The TEACH Act: Copyright Law and Online Education. *New York University Law Review* 78(6): 2204.

Lawrence, Steve. 2001. Online or Invisible. Preprint available at <http://Web.archive.org/Web/20040202012156/http://www.neci.nec.com/~lawrence/papers/online-nature01/>. An edited version appears in *Nature* 411 (May 2001): 521.

Lesk, Michael. 2004a. E-mail messages to Dean of Carnegie Mellon University Libraries Gloriana St. Clair, July 15–16, 2004.

Lesk, Michael. 2004b. E-mail message to Dean of Carnegie Mellon University Libraries Gloriana St. Clair, July 15, 2004.

Lessig, Lawrence. 2004. Open Access, Open Minds. Presentation at the American Library Association Annual Conference, Orlando, Fla., June 27, 2004.

LibQual+™. 2002. Survey Results. Association of Research Libraries. Texas A&M University. Available at <http://www.libqual.org/>.

LibQual+™. 2003. Survey Results. Association of Research Libraries. Texas A&M University. Available at <http://www.libqual.org/>.

Lohmann, Fred von. 2002. Fair Use and Digital Rights Management: Preliminary Thoughts on the (Irreconcilable?) Tension Between Them. *Computers, Freedom, and Privacy*. Available at http://www.eff.org/IP/DRM/cfp_fair_use_and_drm.pdf.

Moglen, Eben. 2002. Brief *Amicus Curiae* of the Free Software Foundation in Support of Petitioners Eric Eldred, et al., v. John D. Ashcroft On Writ of Certiorari to the United States Court of Appeals for the District of Columbia Circuit, No. 01-618. Available at <http://www.mirror5.com/philosophy/eldred-amicus.pdf>.

Open Digital Rights Language Initiative. 2005. ODRL Initiative Releases Draft Specification of ODRL Creative Commons Profile. 2005. *Cover Pages* (March 7). Available at <http://xml.coverpages.org/ni2005-03-07-a.html>.

Pew Higher Education Roundtable (PHER), Association of Research Libraries (ARL), and Association of American Universities (AAU).

1998. To Publish or Perish. *Policy Perspectives* 7(4). Available at <http://www.arl.org/scomm/pew/pewrept.html>.

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

Ringer, Barbara. A. 1961. Study No. 31: Renewal of Copyright. *Copyright Law Revision: Studies Prepared for the Subcommittee on Patents, Trademarks, and Copyrights of the Committee on the Judiciary, United States Senate, Eighty-sixth Congress, first [-second] session*. Washington, D.C.: U.S. Government Printing Office. First published 1960 in *1 Studies on Copyright*, Arthur Fisher Memorial Ed.: 513–514.

S. 487 June 7, 2001. *An Act*. Available at http://frWebgate.access.gpo.gov/cgi-bin/getdoc.cgi?dbname=107_cong_bills&docid=f:s487es.txt.pdf.

SPARC Open Access Newsletter. 2004. Issue #77. Available at <http://www.earlham.edu/~peters/fos/newsletter/09-02-04.htm>.

Stanford Law School Center for Internet and Society. 2004. District Court Dismisses Internet Archive's Challenge to Copyright Regime. *Packets, Cyberlaw Newsletter* 2(4). Available at http://cyberlaw.stanford.edu/packets/vol_2_no_4/002915.shtml.

Tennant, Roy. 2005. Google Out of Print. *Library Journal* (February 15). Available at <http://libraryjournal.com/index.asp?layout=articlePrint&articleID=CA502014>.

Thibadeau, Robert. 2004. *Thomas Jefferson and Intellectual Property including Copyrights and Patents* (August 28). Available at <http://rack1.ul.cs.cmu.edu/jefferson/>.

United States Constitution. 1789. Available at <http://www.house.gov/Constitution/Constitution.html>.

U.S. Copyright Office. August 2001. *DMCA Section 104 Report: 78–105*. Available at <http://www.copyright.gov/reports/studies/dmca/sec-104-report-vol-1.pdf>.

U.S. Copyright Office. January 26, 2005. Orphan Works. *Federal Register* 70(16): 3739–3743. Available at <http://www.copyright.gov/fedreg/2005/70fr3739.html>.

U.S. Court of Appeals for the District of Columbia. 2001. *Eldred v. Reno*. No. 99–5430. Available at <http://caselaw.lp.findlaw.com/scripts/getcase.pl?court=dc&navby=case&no=995430B&exact=1>.

Vaidhyanathan, Siva. 2002. Copyright as Cudgel. *The Chronicle of Higher Education* (August 2): B7.