Testing the Barriers to Digital Libraries: A Study Seeking Copyright Permission to Digitize Published Works

Carole A. George, Carnegie Mellon University

Available at: https://works.bepress.com/carole_george/3/
Testing the Barriers to Digital Libraries: A study seeking copyright permission to digitize published works

Carole A. George

April 12, 2005

Abstract

Purpose: The aim was to explore the issues related to acquiring copyright permission with the goal of determining effectiveness and efficiency using the least complex process.

Methodology: A random sample of books was chosen, relevant information was recorded, request letters were sent and tracked, and results (permission received or denied) were analyzed with respect to publisher, publication data, time required, and issues related to the process.

Findings: About 52 percent responded with a yes or no with 24 percent Yes responses. Nearly 25 percent never responded, addresses were not found for about 16 percent, approximately 7 percent were too complicated to pursue and response time averaged about 3 months.

Implications: The low rate of positive responses indicates the need to focus on publications and publishers most likely to provide permission: older and out-of-print materials, non-commercial publishers, special collections, while using designated staff and personal contact to improve effectiveness.

Value: Few previous studies exist in this area. This study might benefit other libraries with respect to planning, defining procedures, and improving results when seeking permissions for copyrighted works.
Introduction

In the Carnegie Mellon Digital Library Plan 2000 - 2007, the goal was to migrate to a predominantly digital library (St. Clair, Linke, Richards, Troll, 1999). This goal addresses two issues: preference of students and faculty for full-text resources available online and lack of space for the physical collection. This study examined the process of acquiring copyright permission to digitize books for access via the Internet.

The process and methods, the amount of staff time required, and other related issues were explored in the Copyright Permissions project (1999-2001). The project explored the process of acquiring permission to digitize and provide free Internet access to a randomly selected sample of publications from the Libraries’ collection. The motivation was to learn more about issues related to the process and complexity of negotiations for different types of materials. More specifically this study examines the process of negotiating permission to digitize the sample of publications using the least intrusive and least expensive methods in order to address the following questions:

- How effective is the process (used in this study) in acquiring permission to digitize the publications?
- What factors are related to the positive responses i.e. publisher, year of publication, publication type?
- How much time and staff are required in the process?
- What are the difficulties and issues related to the process?
- How can the process be improved and effectiveness increased?

This paper summarizes the results after two years of pursuing copyright permissions.

Methodology

Random Sample

A statistically significant random sample of publications was drawn from the Libraries’ collection. In order to determine issues related to digitization of diverse library collections, we thought a random sample would provide the best representation. After eliminating duplicates, technical reports, dissertations and theses, and missing books, a total of 337 titles remained. The sample varied with respect to publication type, publisher type, publisher’s origin, print status, and publication date. The materials were published as early as 1880 and as late as 1998, with the median year 1979. Books comprised 70 percent of the sample and the majority (56 percent) of the publishers was commercial. From this sample, we derived our working sample, the group still protected by copyright.

Working Sample

At the onset of the study, the initial sample was reduced from 337 titles to 273 titles—for a variety of reasons. Not all publications in the initial sample required permission for use. About 6 percent were in the public domain, including the government materials and those that were out of copyright. Approximately 5 percent had multiple copyright holders or third-
party ownership because of numerous charts, photographs, etc. These were eliminated so as to focus the study on items for which it might be reasonable to be able to pursue copyright permission. Finally, it proved impossible to locate the copyright holders for about 7 percent of the initial sample.

The remaining 273 copyright holders, 81 percent of the eligible sample, became our working sample or permissions request sample. These copyright holders were sent one or more letters requesting permission to provide web-based, digital access to their materials. Because our interest was in recording and learning from results of permission request, the data obtained as a result of these contacts forms the bulk of this report.

The materials in the permission request sample were published as early as 1892 and as late as 1998, with 1981 the median year. Books comprised 73 percent of the sample (Figure 1). The sample also included conference proceedings (8 percent), series (7 percent), exhibit catalogs and annual reports (2 percent), and various other types of publications (9 percent). Based on information we could locate regarding the sample, at least 73 percent of the publications were out-of-print.

When trying to locate the copyright holder, we started with the publisher. Sixty percent (Figure 2) of the publishers in this sample were commercial while associations, university presses, government, museums, art galleries and others made up the remainder. Eighty percent of the publishers sampled were domestic and twenty percent were international.

![Figure 1. Permission request sample - publications](image1.png)

![Figure 2. Permission request sample - publishers](image2.png)
Procedure

To track data and progress, detailed information was collected and stored for each publication. The record included bibliographic citation, contact information; and details about the request letters and responses, such as dates and outcomes. As discussed previously and shown in Figure 3, the first step was to determine the copyright holder or holder’s agent and search for contact information for each item. We used traditional print sources for publishers’ addresses (Books in Print, Literary Market Place) as well as online resources such as search engines and databases. In most cases the initial communication went to the publisher. The mailing included a cover letter and a response form that offered four options:

1. Copyright holder could give full permission for the library to digitize and provide free-to-read Web access to the item. Web users who wanted to buy the publication would be referred to the publisher.

2. Copyright holders could give permission for the library to digitize and provide free-to-read Web access to the item to Carnegie Mellon University users only. Carnegie Mellon web users who wanted to buy the publication would be referred to the publisher.

3. The person/entity contacted could respond that they did not hold copyright for the item.

4. The copyright holder could deny permission to digitize the item.

If the original letter had no response within four to six weeks, we planned to send a follow-up letter. If three attempts to contact the copyright holder got no response, no further attempts were made. When the response included a referral, we repeated the process with the new address. When a request was successful and permission granted, the publication was

Figure 3. The procedure for copyright permission
Results

Difficulties

The first step in the process, determining copyright holders and finding contact information, took much longer than originally anticipated. Many publishing companies that appeared in our sample no longer existed. Some companies had been acquired or merged with other companies. Others were little known or foreign publishers. In some cases, request letters were returned to us with referrals to other copyright holders, usually an author, and often no referral address was provided. When this happened, we repeated the lengthy search for contact information.

About 16 percent resulted in the inability to determine accurate contact information; either the initial letter returned a referral without contact data or the letter was simply returned due to incorrect address. Often problems occurred when the initial contact referred us to the author or another copyright holder but did not include contact information.

After the initial contact, other titles were later found to be too complicated or too difficult to pursue. In all, about five percent of the titles were withdrawn from the study after initial contact, including some sampled titles that subsequently were removed or lost from the library collection.

Though full-time staff worked on the project, they also had other library-related responsibilities and were not able to devote their full-time. Therefore, follow-up to non-responses after the initial contact was often six to ten weeks or more and was limited to an additional written request.

Responses

Though 77 percent of those contacted responded to our request, only fifty-two percent of the 273 resulted in a Yes or No response, granting or denying permission (see Figure 4). Permission to digitize was granted for only 24 percent of the items. Responses also included requests for additional information, statements that they no longer owned the copyright, referral information, or further details i.e. the letter is forwarded to another agent.
Yes/No Responses

One hundred forty-three (fifty-two percent) copyright holders answered with permission either granted or denied. As shown in Table 1, approximately 28 percent of copyright holders denied permission. Less than one-fourth, 24 percent, granted permission to digitize materials. Of these, 41 copyright holders (15 percent) placed specific restrictions, sometimes multiple, on the use of digitized materials. Most of the restrictions were minor and included such limits as:

- Restricted access to Carnegie Mellon University users only (15).
- Restricted the amount of time the electronic version could be made available (2).
- Requested fees ranged from $40-$300, $104 average (8)

Others restrictions included requests for a digital copy, no offer for sale of publication, inclusion of source data, and no third party rights.

Table 1.

<table>
<thead>
<tr>
<th>Yes/No Replies to Letters of Request</th>
</tr>
</thead>
<tbody>
<tr>
<td>52 percent (143) of initial request letters responded with a Yes or No</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Permission</th>
<th>Number</th>
<th>Percent of Requests</th>
<th>Percent of Yes/no responses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Denied</td>
<td>77</td>
<td>28%*</td>
<td>46%</td>
</tr>
<tr>
<td>Granted</td>
<td>66</td>
<td>24%*</td>
<td>54%</td>
</tr>
</tbody>
</table>

Total number of permission request letters sent—273

Responses by Publication Date and Publisher

The summary of responses by publisher category is shown in Table 2. The median year of publication for Yes responses was 1979, slightly less that the median for the sample (1981) indicating a greater willingness to share publication rights for holders of older publications. The Associations category returned the greatest proportion of Yes responses, 49 percent of 37 requests with permission to digitize. This category included such publishers as Materials Research Society, Council of Planning Librarians, and New York Academy of Sciences. Coming in second, the University Press category returned 38 percent of 48 requests with permission to digitize.

Table 2

<table>
<thead>
<tr>
<th>Results by Publisher</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percent of total request letters in publisher category</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Publisher Category</th>
<th>Permission Granted %</th>
<th>Permission Denied %</th>
<th>No Response Percent category %</th>
</tr>
</thead>
<tbody>
<tr>
<td>About 49 percent of associations gave permission.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Associations (37)</td>
<td>49 (18)</td>
<td>11 (4)</td>
<td>30 (11)</td>
</tr>
<tr>
<td>University Press (48)</td>
<td>38 (18)</td>
<td>50 (24)</td>
<td>8 (4)</td>
</tr>
<tr>
<td>Museums, galleries, misc. (30)</td>
<td>33 (10)</td>
<td>10 (3)</td>
<td>20 (6)</td>
</tr>
<tr>
<td>About 13 percent of commercial publishers gave permission.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Commercial publishers (158)</td>
<td>13 (20)</td>
<td>29 (46)</td>
<td>30 (47)</td>
</tr>
<tr>
<td>Total requests (273)</td>
<td>66</td>
<td>77</td>
<td></td>
</tr>
</tbody>
</table>

C. A. George, Testing the barriers…, New Library World
The museums, art galleries, and miscellaneous publisher category included art galleries, museums, book companies, and miscellaneous nonprofit groups. This category returned 33 percent of 30 requests with permission to digitize. Twenty percent of requests in this category resulted in no response, primarily due to our difficulty in locating the holder of the copyright.

Commercial publishers returned the lowest proportion of permissions granted (13 percent of 158 requests). This group also represented the majority of copyright holders in the sample. These results are not surprising. Their need to profit from their materials is no doubt the main influence on this low rate of positive responses. Other influences include the difficulty in finding the copyright holder of the materials. Ownership had returned to the author in many cases and contact information was hard to find.

**Non-Respondents**

Analysis of the copyright holders who failed to respond revealed that Associations publishers, who returned the highest proportion of Yes responses, also had the highest proportion of non-respondents (30 percent) along with commercial publishers (30 percent). Though 75 percent of commercial publishers replied to our request, many of the replies indicated that they no longer held the copyright. As with the associations, the lack of response was due primarily to the inability to locate the correct contact information. In contrast, only 6% of the University Press category did not respond—that is, ninety-four percent of University Press publishers answered our inquiries.

**Response Time**

Response time was tracked after the contact information was found and the request letter was sent. On average, it took about three months after the initial request letter was sent to receive a yes or no reply. The time difference between receiving yes and no responses was noteworthy (yes averaged 101 days; no averaged 124 days). Several factors contributed to the length of the response time:

- Nearly 60 percent of the copyright holders who were contacted initially failed to respond and required a second or even third (16 percent) follow-up letter to meet the parameters of the study.
- Because no regular staff member was assigned to the study full time, and those who assisted had a limited amount of time to devote to the project, the length of time between the initial requests and follow-up letters was longer than originally anticipated. It was often eight to ten weeks or more between letters, rather than the four-six weeks that had been planned.
- Often (23 percent of the cases) the copyright holder that we contacted initially no longer controlled the rights and responded with a referral, in which case the request process was repeated.
- Because of changes in ownership of the copyright, some (23 percent) requests were sent to more than one and as many as three different addresses before a response was received or the request process was terminated according to the parameters of the study.
Discussion

Current project

The results of this project have produced some surprising, informative, and at times disappointing, results. Consider, however, that the motivation for this study was to learn more about issues related to the process and the complexity of the negotiations. By examining the results as they relate to the process of the copyright permission approval, modifications have been made to improve outcomes, and efforts have been focused in the area of best returns. The following are key findings.

Major conclusion

The most striking conclusion gained from this study is that without any additional efforts, such as personal contacts or focus of efforts on special collections and publisher information, activities such as are described in this study will result in response rates that remain low where there is no commitment from the copyright holders. Any library with this kind of plan should consider assigning some “purchasing agents” to carry out the complex procurement.

Copyright approval outcomes

The proportion of holders who provided permission to digitize their materials was low. About one-fourth of the total requests resulted in permission to digitize. Many of those who granted permission included restrictions such as limitations to access, length of time to for use of the material, and fees for use. However, most restrictions were reasonable; only a few were too great to make use of the permission. Consider that this study is based on a random sample of titles, not special collections, and that the procedures used the most basic means of contacting owner, that is, mailings and no personal contact. By focusing efforts and revising procedures, it seems reasonable to assume that the outcomes might be improved.

Responses by copyright holders

The rates varied among copyright holders. The most favorable rate of positive responses was from Associations, though the category, University Presses; and Museums, Galleries, etc, responded positively in nearly the same proportion. The Commercial publishers were the most reluctant to share their rights to the material and also represented the greatest proportion of copyright holders. The most probable reason for this reluctance is their concern that providing the material free-to-read on the Internet would reduce their own sales and profits. It seems reasonable to assume that results would be greatly improved by focusing efforts on non-commercial publishers.

Responses and response time

The process of gaining approval to use copyrighted materials proved to be time-consuming and difficult. Just receiving a yes or no response took considerable effort and time. On average it took approximately three months after sending the initial request letter before receiving a yes or no answer. Consider that nearly 60 percent required follow-up letters, and often the initial letter was returned due to an incorrect address or with a referral. In addition, only 52 percent, just over half, of initial requests resulted in a yes or no response. Over one-fourth failed to respond at all and some requests ended in a dead-end. In this study all contacts were made by mail, email or fax.

C. A. George, Testing the barriers…, New Library World
Preparing the requests

Obtaining permission to digitize copyrighted material is neither a quick nor easy procedure. Searching for contact information of holders, examining each record to determine which was reasonable to pursue, and creating the initial database that included relevant descriptive data required a considerable amount of time. Sending the request letters and follow-up letters, checking on responses and referrals, and record keeping added to the time for the entire process. To continue this process within a reasonable amount of time would require staff to devote more than fragments of time to the project. To improve the process it would be helpful to maintain a database of publishers’ contact information that is easily accessed. For future projects, monitoring the amount of time for each step in the process would be helpful information to support decisions related to the number of staff or percentage of staff time to allot for the process.

Modifications to Subsequent and Future projects

The problems encountered in this study and overall low positive response from copyright holders, primarily publishers, might lead some to believe that the process of building digital libraries will be very slow indeed. Little historical information is available to provide an argument to the contrary. Digital libraries are still in the early phase of development by providers and use by researchers, students, and citizens. However, feedback from students indicates overwhelming preference for digital materials. Rather than discouraging the digital collections’ efforts, this study has provided information that can be useful in the development digital collections and improve the process of seeking copyright approval.

As a result of this study the establishment of best practices in the pursuit of copyright permission has been explored, a database has been established with variables that are needed to track the process, and methods have been defined to scan the material and provide access to the digitized works. In addition, this study has led to a change in procedures that have been used in subsequent efforts involving copyright permissions. The following illustrate what we have learned

Revised focus

In succeeding projects, the Libraries have focused efforts on the copyright holders who are the best candidates for request approval. Based on the previous study and current literature best candidates include holders of out-of-print materials, earlier publication date and non-commercial publishers, reasoning publishers of these materials will be more agreeable to providing copyright approval to their materials. In addition, the Libraries have had considerable success with permission requests for special projects, i.e. fine and rare books.

In order to address the time-consuming and tedious process of seeking permission for each individual publication, the Libraries, have focused efforts on all titles owned by each publisher asking publishers for consent for all out-of-print materials. This approach has led to considerable decrease of time spent on each publication request and a significant increase in number of consents.

Tracking database

A considerable amount of time was spent previously on searching for copyright holders. To address this problem, a publisher database has been designed for use with subsequent projects. While the previous database record focused on each individual publication, this publisher database record focuses on the publisher. Records include all the necessary
information needed to track the process and record the outcome of requests for all titles owned by each publisher. Not only does this database provide easy access to publisher information for future projects, it also enables analyzing the outcome data in order to evaluate the copyright permissions effort and make further improvements to the system.

Designated staff and personal contacts
In addition, this study demonstrated the considerable amount of time necessary to attend to the entire process of acquiring permission and the need to follow-up on initial contacts in a timely manner. It also demonstrated the difficulty of securing permission by using only a written request. For this reason on subsequent projects, a regular staff member has been designated to work specifically with the requests. Follow-up to non-responses included personal contact by phone or email and occurred within a few weeks of the initial contact. Both measures significantly increased the rate of positive responses.

Outcomes of current efforts
The number and variety of ongoing digital projects speak for themselves. There is growing demand and a world-wide need for digital information. A great deal of energy is being directed at solving problems and making a critical mass of high quality academic material available online. After a critical mass is achieved and online materials become the norm, proponents argue, current barriers to creating digital libraries—including the difficulties of copyright that this study explored—will fade. The changing environment in the realm of copyright permissions is evidenced in the changes in the subsequent digital projects at Carnegie Mellon Libraries and the outcomes of current efforts. The following are a few examples of the continuing work at Carnegie Mellon Libraries.

- **One Million Digital Pages**—The Carnegie Mellon Libraries became the first Library to digitize one million pages of scholarly material and provide web-based access.

- **Posner Memorial Collection**—this collection, on loan to the Libraries from the Posner family and currently housed at the new Posner Center on campus, is now available electronically via the Internet as a searchable database. The collection of 622 early and rare literature, decorative arts, and history of science materials had 46 percent of the titles protected by copyright. To date 71 percent of copyrighted titles have been approved for digital access.

- **Million Books Collection**—initially funded by the National Science Foundation, this project includes Carnegie Mellon University, other academic libraries; India and China as collaborators in developing this collection. The goal is to provide free electronic access of a one million titles of diverse materials via the Internet. This collection is in the early stages of development, however early results indicate that 54 percent of requests have resulted copyright permission approval.

- **DIVA**— Digital Image Versatile Archive, developed at the Libraries, enables users to search, browse, view, and print collections of archival materials. This project has focused on archival collections donated to the Libraries. Current collections include the H. John Heinz III Archives, Allen Newell Collection, Herbert Simon Collection, and coming soon the papers of the Nobel Laureate Clifford Glenwood Shull.
Summary
Carnegie Mellon University Libraries conducted a two year study to explore the issues related to acquiring permission to digitize copyrighted material and provide digital access via the Internet. The goal of the study was to determine a realistic estimation of time, complexity, and issues related to this process. Based on a random sample drawn from library shelves, 273 titles were the focus of efforts to acquire copyright permission. The study provided insight into the process, problems, and obstacles confronting libraries seeking to develop their digital collections. However the study yielded some discouraging results: Less than one fourth of copyright holders granted permission to digitize their books. Nearly one-third of copyright holders did not respond to queries, even after diligent follow-up. Addresses could not be found for about eleven percent of the copyright holders. However, the study also yielded valuable strategies that have made subsequent copyright permission projects quantifiably more successful. In the long run, this and other projects will be rewarded by the development of robust digital libraries.

References

Acknowledgements
Many people contributed their time and energy to this project including Denise Troll Covey, Tracey DePellegrin Wagner, Valarie Wadyko, Leigh Caskey Schenk, Martha Knott, Gloria Henning, and Cindy Carroll.