



Evidence Summary

Obtaining Copyright Permission to Digitize Published Works Remains a Significant Barrier

A review of:

George, Carole A. "Testing the Barriers to Digital Libraries: A Study Seeking Copyright Permission to Digitize Published Works." *New Library World* 106.1214/1215 (2005): 332-42.

Reviewed by:

Susan Haigh
Senior Policy Officer, Library and Archives Canada
Ottawa, Ontario, Canada
E-mail: susan.haigh@lac-bac.gc.ca

Received: 20 March 2006

Accepted: 3 May 2006

© 2006 Haigh. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Objective – To assess the effectiveness and efficiency of the copyright permission-seeking process and to suggest improvements in order to improve outcomes.

Design – Workflow study.

Setting – Carnegie Mellon University Libraries, Pittsburgh, PA, USA.

Sample – A random sample of titles published 1999-2001 was selected from the library's circulating collection. After eliminating duplicates, technical reports, theses, dissertations, and missing items, the sample comprised 337 titles. Of these titles, 70% were books, and 56% were from commercial publishers. From this a

working sample of 273 titles was derived, comprising those titles protected by copyright and with the rights owner clearly indicated. About 73% of this working sample appeared to be out-of-print; their median publication year was 1981.

Method – In this two year study (1999-2001), a random sample of books was selected, and pertinent bibliographic and copyright holder information researched and recorded. Permission letters were sent and, six weeks later, follow-up letters were sent to non-respondents. The letter allowed respondents four options:

1. Grant full permission to digitize the work and provide unrestricted Web access;

2. Grant permission to digitize the work and provide read-only Web access, limited to Carnegie Mellon University users;
3. Declare that they do not hold the rights, and hopefully provide information to identify and locate the actual rights holder;
4. Deny permission for digitization.

Results were then recorded and analyzed.

Main results – Of the 273 letters mailed, a clear ‘yes’ or ‘no’ reply was obtained for just over half (52%) of the documents. Sixteen percent of the rights holders could not be found (the letter was returned, or a referral proved impossible to locate and contact). Another 25% of the copyright holders simply did not reply, and 7% were otherwise problematic. Of the 143 ‘yes’ or ‘no’ responses, 54% denied permission, while 46% granted permission. (Note: these percentage figures appear to be erroneously reversed in Table 1 of George’s article.) Therefore, of the overall working sample of 273 titles, permission to digitize was obtained for only 24% of the titles.

A substantial portion of the permissions (41 of 66, or 62%) carried some restriction. This represents 15% of the total working sample. Only a few restriction requests were deemed too great to make use of the permission.

Commercial publishers who made up 58% of the working sample granted permission at the lowest rate (13%).

Response time averaged three months from the time the initial letter was sent until a ‘yes’ or ‘no’ response was received. Negative responses averaged a bit longer than positive responses (101 days to 124 days). However, some of this time was attributable to delays in issuing follow-up or

redirected request letters (a step required in 60% of cases), owing to the limited staff resources at Carnegie Mellon.

The copyright ownership had changed in 23% of the sample, requiring more than one and up to three different addresses to be contacted before a response was received or the effort was terminated.

Conclusions – The study concluded that the permission rate would remain low unless additional efforts were made in the permission-seeking process (e.g., personal contacts in addition to letters and emails), or unless more selective approaches were employed (e.g., targeting non-commercial publishers). It also concluded that the process to seek copyright permissions was neither quick nor easy, suggesting the need for dedicated staff time and a readily accessible database of publisher contact information. As a result, subsequent projects have improved their permission-seeking process, focusing on more non-commercial publishers or older publication dates, and asking publishers for blanket consent for all of their out-of-print titles.

Commentary

The useful data developed by this study is based on a broad working sample of published material and helps to identify the least amount of staff effort needed to obtain permissions. Some context as to U.S. copyright law might have been useful. There is an assumption that all readers are aware of U.S. copyright law, including length of copyright.

While the study methodology was appropriate, there were a couple of gaps in data, pertaining particularly to staff time. First, the time it took to examine the original random sample of 337 titles and research copyright holders (in order to conclude that some would not be able to be located) did

not seem to be factored into the overall consideration of staff time required. Then, once the working sample was determined, the first step in the process was to determine the identity of the copyright holder and obtain contact information. How much time did this take on average? It is described as 'considerable', but not quantified. The author does mention that it would be helpful in future projects to monitor the staff time expended on each step of the process. But without such data, the study's aim to determine efficiency cannot be fully explained.

The low rate of permission to digitize (24%), the relatively high rate of restrictions (62%), and the high rate of non-response (up to 30% for some publisher types) are all notably discouraging findings. As Google has discovered with its Google Print Library project and the U.S. Authors Guild lawsuit which it occasioned, copyright remains a major obstacle to building comprehensive retrospective digital collections of published material.

And yet, as the author notes, there is a growing demand for digital information. Carnegie Mellon has determined to continue

its efforts, but with a honed approach. Over time, this will build critical mass, but inevitably (and unfortunately) the resulting collections will fall short on more recent and more commercial titles. There is no discussion of whether the demands and priorities of students and faculty will be met by the revised scope of the library's digitization effort.

The author does not comment on the fact that many publishers denied permission, even, presumably, for long out of print titles, or that many publishers simply chose to ignore permission requests. The conclusion does mention that one of their subsequent projects, the Million Books Collection, which presumably employs their refined approaches and strategies, achieves a 54% permission rate—an improvement, but hardly ideal.

Those seeking to digitize copyright library material may want to think twice about doing so after reading this study. If they persist, its findings provide some guidance toward setting realistic expectations and timeframes, as well as pragmatic workflow processes, for optimal success.