

Copyright Investigation Summary Report

Participating RLG Program partner institutions and staff:

Cornell University
Peter Hirtle

Stanford University
Mimi Calter

Emory University
Lisa Macklin

University of California, Los Angeles
Angela Riggio

New York Public Library
Tom Lisanti

University of Michigan
Judy Aronheim

Rutgers University
Grace Agnew

University of Texas at Austin
Dennis Dillon

Report prepared by:

Merrilee Proffitt, Arnold Arcolio, and Constance Malpas, RLG Programs

Background:

To support the RLG Programs work agenda, RLG Partners collaborate to identify new service areas and structures. This collaboration helps libraries, archives, and museums achieve a common understanding of the processes for which they should be responsible.

In this project, staff from eight partner institutions participated in copyright investigation interviews between August and September 2007 to share the ways in which their institutions currently obtain copyright permission to provide users with access to high-risk or special collection materials.

A few things to note about the participating partner institutions include:

- Some are involved in mass digitization projects, while others are not.
- Some are concerned with determining copyright status in order to provide access to digitized materials.
- All of these institutions are concerned with identifying rights-holders and securing specific permissions when scholars or businesses wish to use materials, often special collections materials, particularly photographs.
- In some cases, the same staff are responsible for both mass digitization and activities around archival or special collections materials.

It's also important to note that staff who participated from almost every institution expressed a sense of "just getting started" or "realigning efforts to be more consistent across campus and across library units." Almost all of the staff interviewed were in newly created positions; several noted that conducting copyright investigations in a centralized fashion was a new area of focus for their institutions.

This report summarizes the results of these partner interviews. RLG Programs is grateful to the partner institutions and staff who participated in this project to share their knowledge and perspectives.

What work is being done?

- When institutions provide access to published materials, the task is to identify high-risk materials and eliminate them from the pool of items to which access is being provided. The goal is to reduce risk, rather than to achieve certainty—determining items or classes to be risky will end the investigation of those items. These determinations ought to be true for all uses and users of an item.
- When institutions gain permission for the use of special collections materials, the task is to determine who holds the rights, and obtain permission for from the rights holder or likely rights holder. The rights secured in this process are likely to be limited to specific user and use of the item. The burden of this work is generally understood to fall on the patron, but staff we spoke with at several institutions do substantial work to support patrons in this endeavor.

What types of materials?

For mass digitization:

- Books, mostly published in the U.S., although some institutions are experimenting with published outside the United States.

For unpublished or special collections materials:

- Many institutions pursue obtaining permission for these materials. However, it is important to note that rights pertain to a particular use by a particular individual and do not necessarily scale.

How much effort?

- Except for one institution (University of Michigan), copyright investigations are pursued by a single person who devotes, at most, a half-time effort to this work. One or two other people may also assist in the effort, on a part-time basis. Sometimes campus counsel may provide assistance as well, but this represents a small amount of time on an occasional basis.

How much is being produced?

For mass digitization:

- Numbers ranged from around from a goal of 100 books a month to 1,000 books a month. Many institutions are still building capacity, so there is not a lot of data and practice to draw from. Looking at institutions that are already at a high level of productivity, it's clear that doing risk-based analysis is a considerable investment of time and energy.

For unpublished or special collections materials:

- Obtaining permission for the use of this work is very time consuming and the numbers are very low—in the hundreds of items per year, at best.

The University of Michigan shared its information on throughput. From May through November 2007, University of Michigan Library staff reviewed 16,896 post-1922-1963 U.S. imprints for copyright status. They found:

- 9,719 to be in the public domain,
- 4987 to be in copyright, and
- 988 to be U.S. government documents.

The remaining items were more problematic and required further investigation.

What triggers action?

For mass digitization:

- The decision to digitize items, or to give access to a swath of already-digitized materials.

For gaining permission for the use of unpublished or special collections materials:

- A request by a patron.

What data is in hand?

For mass digitization:

- Institutions are working with MARC data from their local catalog. This information allows them to rule out high risk materials, and work with a narrower pool of items. Beyond this, institutions need to examine items for information not contained in MARC record.

For gaining permission for the use of unpublished or special collections materials:

- Institutions use a broad range of paper files, curatorial input, and often need to do research. This information not likely to be in an electronic form.

What problems come up?

- It is impossible to know from the work (or be certain from the local catalog) if there was no immediately previous foreign edition of a work published in the United States. Information missing from item in hand (both in terms of a defective item, where another copy could provide information, and also in terms of contextual information from title pages, such as whether or not a copyright notice appears, which is simply lacking). Information is not present in MARC records both because local MARC record sometimes doesn't contain information it should, and also because cataloging rules don't call for including some information that is necessary for this purpose.
- Completeness or certainty may not exist. Interviewees recognized that completeness of information or total certainty about the status of an item or collection may not be achievable. Although this was frustrating, being certain was not an ambition expressed by interviewees.

What data is produced at the end of the process?

- Machine readable local copy of information: some in spreadsheets, some in MARC records. Permissions information stored in electronic and paper forms.

Batch or one-at-a-time?

For digitization/access:

- Batches of items are pulled for consideration, but determination is made on an item-by-item basis. Essentially, items are assessed in a bulk fashion for de-selection, and then further assessed for digitization or display.

For existing documentation

- Staff at almost all institutions had recently developed documentation or were in the process of developing documentation. Everyone who had documentation was willing to share, even in draft form.

What constitutes success?

- Interviewees had a hard time answering this question. Success is judged by adding to what's available online; and by not getting sued or getting into trouble. In either case, success is not quantifiable.

What is unsatisfactory about the current process?

- Results still leave staff with ambiguity. Interviewees also felt ambiguity in terms of open questions about due diligence and fair use.

What is the most time-consuming part of the process?

For mass digitization:

- Checking items for information not in MARC record is done on an item-by-item basis and is very time-consuming.

For gaining permission for the use of unpublished or special collections materials:

- This work is also very time-consuming—it requires a significant amount of time of consulting with sources, communicating with rights holders, etc.

What's missing that would improve the process?

- We need to find foreign editions of English-language materials. In order to do this, we need to find the nationality of authors (as an indicator of both foreign editions and foreign born authors publishing in the U.S.). We have the sense that without a different framework for orphaned works, people would not be motivated to invest more effort than they are now.

What external tools are used?

A number of specific tools were mentioned by interviewees, including:

- Stanford University's Copyright Renewal Database
<http://collections.stanford.edu/copyrightrenewals/>
- Cornell University's Copyright Term and the Public Domain in the United States Chart
http://www.copyright.cornell.edu/public_domain/
- The University of Texas at Austin's Harry Ransom Center FOB (Firms Out of Business) Database
<http://tyler.hrc.utexas.edu/fob.cfm>
- The University of Texas at Austin's Harry Ransom Center WATCH (Writers, Artists, and their Copyright Holders) Database
<http://tyler.hrc.utexas.edu/>
- The Visual Arts and Galleries Association (VAGA)
<http://www.vaga.co.uk/>
- The Society of Authors
<http://www.societyofauthors.org/>

In addition, institutions had licensed, or were considering licensing, data from the Library of Congress Copyright Office.

Is there currently collaboration between institutions?

- Very little, probably because there is not a large community of practice. There was also the sense that, because there is likely to be little overlap in collections digitized (especially between institutions doing focused digitization), collaboration isn't as fruitful.

Were external conclusions made about copyright status?

- Staff interviewed said they would only trust conclusions that they felt resulted from processes in line with what they understood to be “diligent.” However, their institutions are responsible for these conclusions and staff have to defend their processes. The general consensus is that the legal standard for due diligence is unclear. Institutions are more comfortable carrying out this work on their own.

What areas should OCLC Programs and Research focus on that would be of most value to the copyright investigation process?

1. Disclosing earlier foreign editions of U.S. works (“views” suggests that a person will look at something to make this determination, and it could be a batch process, a table or something else).
2. Providing information about the nationality of authors.
3. Gaining consensus about an adequate definition of due diligence. Helping to define what is “reasonably diligent” will help institutions be more confident about the steps they should take to show they’ve taken appropriate action. However, this may not help institutions share conclusions without going through the steps on their own.

What is the value of this work? What’s missing?

- In digitization projects, interviewees talked about the satisfaction of “bringing more materials into the light” and “making available online what Google will not.” Over the course of these interviews, we began to wonder about the anticipated aggregate value and nature of what is being exposed (material published in the U.S. pre-1923, and a partial band of materials published in the 1923-1963). Although these materials could form a useful resource for some disciplines, investigation would need to be done to identify which ones. Almost certainly, these materials do not compose a firm platform for basic or comprehensive research and scholarship or teaching.
- What will be important for research and teaching is how the works that will be made available (or determined to be available) through these present efforts will be combined with works not in the public domain, and by whom.

Next steps

The work undertaken by RLG Partner institutions for this project was an early step in contributing to the development of the Registry of Copyright Evidence. This registry will ultimately lead to the creation of a consistent, accepted process for libraries, archives and museums to gain permission to provide access to high-risk or special collection materials.

RLG Programs is working with the Society of American Archivists Intellectual Property Working Group to further explore community practice and issues around unpublished materials held in special collections and archives.