Partner Institutions

Indiana University Libraries
University of Minnesota Libraries
University of Wisconsin-Madison Libraries

Project Overview

The goal of the Copyright Review Management System (CRMS) project was to build a system that would enable us and our partner institutions to investigate the copyright status of works published in the United States between 1923-1963, to further the national dialogue on making legal uses of our cultural heritage, and to increase access to public domain materials. We had originally hoped to identify and make available 60,000 public domain volumes over the course of the project. We are quite happy to report that we have greatly exceeded our original goal; as of November 1st, 2011, we have reviewed over 170,000 books, 87,000 of which have been determined to be in the public domain.

Project Activities

Project Activities (details included in previously submitted Interim Performance Reports):

- Define requirements, design and develop the CRMS – database, workflow management system, web interface; test iteratively and revise system as needed
- Release Version 1.0; 2.0, 3.0 in years 1, 2 and 3 of the project, release numerous updates to system between the major releases
- Perform copyright Review work using CRMS
- Develop methods for sharing findings on copyright status
- Continuously monitor progress and track areas where system falls short; work to resolve issues
- Monitor productivity and take steps to evaluate improvements in accuracy for copyright determinations – introduced Decision Tree to clarify precedence of various factors involved in making a reliable rights determination.
- Add functionality to system to accommodate training and integration of reviewers at outside institutions
- Develop training plan and materials; train reviewers at three outside institutions (Indiana University, University of Minnesota, and University of Wisconsin) to join in collaborative copyright review work
- Publicize efforts – website, regular email updates to stakeholder an interested parties, articles/interviews, presentations
- Contract with United States Copyright Office to evaluate CRMS determinations for three separate sets of volumes for which renewals were not found (one pre-CRMS; one after introduction of CRMS in phase II; one after reviewer at outside institutions joined the project in phase III); analyze results and produce report on findings

Project Audience
By identifying public domain works in our collections, and making those works freely available, the University of Michigan and our partners have vastly expanded the amount of public domain materials available to anyone who wishes to make use of them. As such, our target audience included scholars, students, educators, people whose family members wrote, now hard to access books, general browsers, and countless other categories of users.

In addition to these primary users, our project audience also includes any individuals or institutions that, in the future, wish to make use of our methods or process to undertake a rights determination of their own. As we move forward with the next phase of the work, we will be adding additional partners and training them in the CRMS determination process.

**Project Analysis**

**a. Targets**

Overall, the project has been a huge success. We exceeded our target of reviewed and opened books by over 30%, we created a reliable and efficient determination process, and we were able to recruit significant outside partners to work with us. While we did face challenges in the design and implementation of our project, such as difficulty in scaling training to accommodate the amount of interest in partner participation, we were able to successfully navigate those challenges over the course of the grant.

**i. Achievements**

At the outset of the grant period, we had hoped to be able to identify and make publicly available 60,000 public domain books over the course of the project. We have substantially surpassed this goal. As of November 1st, 2011, we have identified 86,755 public domain works by undertaking 170,174 reviews. In addition to the books we have been able to make available, we have also identified rights-holders and other renewal information for many of the other approximately 85,000 books. This information will allow us to ensure that all legal protections are provided to rights-holders, and that rights-holders who wish to make those works available in HathiTrust have a clear mechanism with which to do so.

We are also pleased by the extent to which we were able to successfully create an efficient system to facilitate making copyright status determinations. This process included developing systems to efficiently manage workflow and to prioritize productive reviews, incorporating a double review process to improve overall reliability, and refining our review process to continually improve consistency in our determinations work. We were particularly successful in this final element; two audits we undertook with the US Copyright Office’s own records showed marked improvements in the accuracy and reliability of our reviews.

Finally, we are pleased by the interest other institutions took in our project. Over the course of the grant, were able to successfully engage four outside institutions to contribute staff to the review work. It is of interest that we were able to scale our efforts upwards in a linear fashion as we added reviewers; that is, we found little, if any, losses to efficiency per worker as we added workers to the project. The project eventually became quite large: including staff at Michigan, there were approximately 20 reviewers across five institutions working in the CRMS to make copyright determinations. We feel that our successful collaborative work on this project demonstrated the viability of multi-institution collaborations to identify public domain materials, and provided the basis for the collaborative approach proposed in the new Copyright Review Management System-World grant.
Staff Turnover

One of the challenges we faced over the course of the grant was the loss of important grant-funded staff. The programmer/developer originally assigned to the project left his position at the Library in March 2009, just as we were in the midst of developing the first version of the system. An interim programmer was assigned to work with us on developing the first version of the CRMS (in addition to his regular responsibilities running Deep Blue, our institutional repository). In spite of these developments, we were still able to build the Beta version of the System, test, iterate, and release CRMS 1.0. During that time we also had Michigan reviewers using the system. Despite this complication, we were only about a month behind our initially scheduled date. After this initial release we were able to hire a new programmer, who worked with the interim programmer to take over responsibility for the system shortly thereafter. This staff turnover posed challenges on the development side of the grant; naturally, some programming decisions made by the initial programmer were different than ones that would have been made by the ultimate programmer; as such, some of the CRMS system had to be built in ways that were less than optimal. However, over the course of the grant we were able to correct many of these issues, and we are now quite pleased with the system as it stands.

Changes in outcome for sharing determinations

We had originally planned to share the CRMS copyright determinations with OCLC Worldcat and the OCLC Copyright Evidence Registry, as well as with Bowker/Global Books in Print, however circumstances arose that prevented us from doing so, and we had to come up with another strategy. From the Interim Performance Report Narrative report from 12/31/2009:

"Determinations will be shared with OCLC as part of the OCLC e-content synchronization project, which is delayed by OCLC at this time for reasons unrelated to this project. Sharing determinations with Bowker is pending the outcome of contract negotiations. OCLC has postponed further development and implementation of their OCLC Copyright Evidence Registry."

Unfortunately the issues with OCLC and Bowker were never resolved, so we developed alternate methods for making rights determinations available, as further described in the Interim Performance Report Narrative report from 12/31/2009:

"After deliberation, stakeholders determined that the best course of action for sharing rights determinations at this juncture was to add the information to the tab-delimited HathiTrust metadata files that are provided at http://www.hathitrust.org/hathfiles, beginning with the full metadata file produced on December 1, 2009. Fields added include the copyright determination reason code and the date the rights database entry was last updated. With this data included, the tab-delimited files will become an ongoing accessible source for information on how and when rights determinations are made. In addition, rights status determinations for volumes in HathiTrust are also currently being shared through OAIster, which was recently transferred to OCLC (as they pick up our records via OAI feed)."

To summarize, while we originally intended to share determination results with OCLC and

1 More information on how we currently share determinations can be found here: http://www.lib.umn.edu/sharing-rights-determinations.
Bowker Books in Print, we were unable to do so due to issues beyond our control. However, in order to ensure that the work we undertook over the grant period is preserved and made useful, we have enhanced our own records to better record this information and have made these records available to interested parties.

**Efficiently allocating resources and designing process to address conflicting reviews**

In creating the project plan and allocating staff resources, we underestimated the amount of time and effort needed to fulfill the role of tie-breaker/expert arbitrator when reviewers do not agree in their determinations.

From the time we released the first version of the CRMS and we instituted the double review process, we found that while review productivity rates among individual reviewers increased, the percentage of conflicting reviews also grew and was much larger than we anticipated based on the comparison testing we did during the pre-proposal proof-of-concept stage. It was originally anticipated that the project manager would be able to serve as the sole “expert” arbitrating reviewer for this project. With the introduction of the CRMS, we recognized that it was not possible for the central coordinator to serve as the sole person responsible for resolving the conflicting reviews. Conflicts made up more than 20% of determinations (and at times the percentage was as high as 35%). Under our process, those determinations could not become final determinations without expert resolution.

The details of this issue and our ongoing work to address it are included in the Interim Performance reports but, in summary, in order to manage staff resources/expert workload and be most efficient in our work, we reallocated staff to serve as experts to resolve the conflicting reviews and we made it a priority to reduce conflicts and ensure consistency so that we could produce more matching reviews. We were able to accomplish this by analyzing how conflicts tended to be resolved; our data indicated that a record that produced a conflicting review was overwhelmingly likely to not be in the public domain. By starting with this assumption, expert reviewers are able to rapidly resolve conflicts and we have been able to surpass our original productivity estimates.

**Ensuring accuracy and efficiency**

In addition to the above challenges, we worked to ensure that our review process was both efficient and accurate. For example, if a book incorporated a substantial amount of third party copyrighted materials (“inserts”) such as lengthy introductions, significant amounts of photographs, or chapters or other written materials obtained with the original author’s permission, we did not mark our copy of that work as in the public domain, even if the underlying work was public domain itself; the rights associated with those inserts required us to keep the work closed. However, as a legal and practical matter, it did not make sense to have our reviewers read every page of a book to ensure that there were not significant inserts, nor did a more time consuming review result in better accuracy. Instead, we identified areas where inserts would be mentioned and identifiable, such as the title page, the verso, and the table of contents. By focusing our inquiry into these areas we were able to ensure reliable and efficient reviews.

**Difficulty of making a copyright determination, difficulty checking with the US Copyright Office**

When we decided to work on making determinations for US 1923-1963 works we believed that the determination was going to be simpler than our experience now indicates. Our original process involved checking to see if the work complied with copyright formalities simply by
looking for a copyright notice on the piece and checking the Stanford Copyright Renewal Database renewals; however, we encountered works that were not in the Stanford Database, such as periodicals, foreign works, non class A works. As we became more sophisticated in identifying these works, we saw a marked increase in “undetermined/needs further investigation” reviews. For example, when we started doing copyright determination work in 2007 during the pilot phase of the project, we were unable to make a determination for only 7-8% of volumes. With the release of the CRMS, the occurrence of “undetermined” determinations immediately more than doubled to 17% (by the end of 2009), and over time it has continued to increase. As of November 2011, the overall “undetermined” rate is 28.5%. While this is a marked increase, and speaks to the difficulty of the task at hand, we plan to further investigate this class of works in the future.²

Our attempts to dig deeper into the undetermined class and to better analyze our determination process were made extremely difficult due to the absence of any authoritative and comprehensive source of information about renewal records. We chose to use the Stanford Renewal Database because it is reliable (in Stanford’s own reliability testing of the system they found less than 1% error rate), but it is certainly not comprehensive. And while the Stanford Renewal Database has proven immensely helpful, the lack of an authoritative, comprehensive resource for information about copyright renewals has been a hindrance to this work.

Originally, we had hoped to make use of the Library of Congress copyright renewal database, which we presumed to be the most authoritative resource. However, digital access to renewal records is not complete (electronic access is only provided to records from 1978 or later) and we found searching the database to be difficult. Accessing the paper records at the Copyright Office requires going to the Copyright Office and manually searching the records. The Copyright Office does have a service where one pays them to perform the manual searches; however, as we found with our efforts at evaluation, hiring the Copyright Office to check our accuracy was time consuming (in each case, the Copyright Office took nearly six months to return records to us), expensive (a search of approximately one hundred records costs five thousand dollars), and of limited utility (the results were difficult to interpret, were only provided in hard copy, and the Copyright Office refused to warrant the accuracy of their search). A usable and comprehensive database of copyright renewals would have been immensely helpful. As it was, we received only limited utility from hiring the Copyright Office to perform searches for us.

iii. Additional information on project impact

As detailed in our grant proposal, we sought to develop a process by which multiple institutions could reliably review the copyright status of works held in the HathiTrust. And in service of that goal, we reviewed over 170,000 works, made over 80,000 public domain determinations with our partner institutions. These books are now accessible for scholars, students, academics, and members of the general public. In addition, the successes over the first grant have advanced the national discussion on how best to make use of the vast cultural assets contained in digital collections around the nation. One example of this is the massive surge of interest we saw when we put a call for partners on our CRMS-World project. Whereas we had three partner institutions at the outset of this grant, we now have nearly twenty-three partner institutions. We attribute this in no small part to the progress we were able to make over the course of this grant.

² The “undetermined” class of works poses a series of interesting challenges that we are excited to tackle. For example, we suspect that many undetermined works were not renewed; developing another process to analyze these works is something we wish to undertake in the future.
What's next?

Copyright determination work will continue under a newly awarded grant from IMLS to (1) expand on the original grant by developing a process to make determinations regarding the copyright status of books published outside the US and (2) continue the work on the remaining population of works published in the US from 1923-1963. Under the new grant (“CRMS-World”), the University of Michigan Library will expand its current approach to copyright determination to international works, beginning with English-language books published in the United Kingdom, Canada, and Australia. With an Advisory Working Group of copyright experts, the Library will prepare responsible legal analyses and decision-making processes. The Library will also prepare a web-based training process and interface making possible a modified crowd-sourcing approach to gathering the information needed for copyright determinations and will prepare a web-based interface for making determinations. The process will clarify copyright status and identify books in the public domain that will be made fully available online through HathiTrust, a multi-institution digital repository tasked with preserving and making available the library collections of participating members.