Overview

The Copyright Review Management System (CRMS) project continues to progress in accordance with the Schedule of Completion submitted with the grant proposal. We are currently halfway through Phase Three of the project, which is scheduled to run from June 2010 through May 2011. As of December 15, 2010, staff from the participating institutions have reviewed 45,816 volumes in the HathiTrust Digital Library using the CRMS. Of those volumes, over 24,636\(^1\) (approximately 54%) have been determined to be in the public domain. Those volumes can now be discovered publicly via Mirlyn (Michigan’s OPAC - http://mirlyn.lib.umich.edu) as well as in the HathiTrust Digital Library (http://catalog.hathitrust.org/).

Over the past six months project staff at the University of Michigan have focused the majority of our efforts on integrating new reviewers from the three CIC (Committee on Institutional Cooperation) Libraries that committed staff as part of the CRMS grant proposal. We made some processing adjustments to the CRMS to significantly reduce the workload of “expert” reviewers at the University of Michigan. In addition we received and analyzed results from the U.S. Copyright Office searches that were included as part of the Evaluation portion of the grant proposal.

Goals and Accomplishments to Date for the Reporting Period – July through December 2010

1. Staff use CRMS to perform copyright status review work; Project Manager works with Technical Services Staff to monitor issues

   Accomplishments: Staff at the University of Michigan (and CIC Institutions) continued to rely solely on the CRMS to perform and manage copyright review work. As of the end of Phase I we have implemented the “double-blind” review process, store all reviews in the CRMS database, and export rights determinations automatically to the HathiTrust rights database on a nightly basis.

---

\(^1\) This number does not include the volumes reviewed during the grant period prior to the release of the CRMS Version 1.0. From December 2008 through June 2009, staff “manually” reviewed 26,741 volumes and 12,703 volumes were determined to be in the public domain. Therefore, project totals to date are 72,557 volumes reviewed and 37,339 now available as full text.
2. Project Manager works with Programmer to resolve system-related issues

Accomplishments: The system has remained relatively stable with few significant updates over the past six months. Aside from minor bug fixes and backend server issues that required attention, our main programming efforts focused on implementing automated processing of certain combinations of reviews in order to reduce expert workload (see Section #4 below). We also added new reporting mechanisms so that Institutional Administrators at the outside institutions can view statistics for their own institution's reviewers. We are also working on implementing rights inheritance for volumes that are duplicate copies of volumes that have undergone copyright review. The initial round of rights inheritance for non-legacy CRMS determinations was implemented in November 2010, which resulted in the full text availability of an additional 3,751 volumes.

3. Train CIC Institutions in order to engage collaboratively/CIC uses CRMS

Accomplishments: Over the summer months, staff from the University of Michigan trained reviewers at the Indiana University, University of Wisconsin and University of Minnesota. There are now 18 people across 4 institutions contributing reviews to the project. The increase in staff has resulted in a larger number of public domain volumes being identified and opened up in HathiTrust on a monthly basis, from 470 public domain volumes in June to over 2800 volumes in October. Total determinations rose from approximately 1100 in June to over 5800 in October.2 There has been some turnover in reviewing personnel at the outside institutions but we have been able to accommodate adding new people by setting them up to work self-sufficiently in the Training CRMS and under the supervision of their institutional administrators. This has allowed us to add new people without adding significantly to the workload of the experts at Michigan.

4. Central Coordinator will monitor productivity and take steps to evaluate improvements in accuracy as described in the evaluation plan and refine processes as needed.

Accomplishments:

- Changes in review guidelines - We made several minor adjustments to the Decision Tree and Decision Tree Guidelines (copies attached) to clarify the determination process based on internal and external reviewer questions and

2 Numbers were slightly lower in November due to ongoing back-end technical issues affecting but unrelated to the CRMS, as well as the Thanksgiving holiday. We expect December numbers will be similar, but are likely to be back up in January.
feedback.  

- Decrease Rate of Conflicting Reviews - The CRMS process requires that two reviewers contribute matching determinations of copyright status in order for that determination to become final. If the two reviewers disagree in their assessment, the volume goes to a separate queue of “conflicts,” which must be resolved by an expert reviewer. We are happy to report that even with the influx of new reviewers working in the CRMS, our overall rate of matching reviews has been on the rise over the last few months, and the percentage of conflicts has continued to decline. As of the end of 2010, the rate of conflicts was down to a project low of 10% from a high in May 2010 of almost 35% (when we first introduced the new determination guidelines to veteran reviewers at Michigan, and as reported in the previous Interim Narrative Report). We attribute the continual decrease in conflict rates to the consistency in how we are applying the guidelines and responding to reviewer inquiries, along with the ongoing increase in cumulative experience among the reviewers across all four institutions.

- Efforts to reduce Expert Workload - Although the conflict rate in the CRMS is currently at a historical low, the increase in the number of reviews means that there is still a lot of work being generated for experts to resolve the conflicting reviews. The conflicting determinations almost always involve some level of ambiguity and therefore require additional time and effort to resolve. As the volume of reviews has increased substantially, even with the decreasing conflict rates, it was a challenge for the four expert reviewers at the University of Michigan to keep up. As of late September 2010, we implemented automatic processing of what we would formerly consider “conflicts” in order to reduce the strain on expert resources. We analyzed sets of conflicting reviews for the likelihood that they would be finalized as public domain, in copyright or undetermined. For those sets of conflicts that had a high likelihood of always getting the same final determination, we put in place automatic processing so that an expert no longer has to spend time on conflicts that are not likely to benefit from the extra effort.

---

3 We carefully deliberated before making any changes in part to avoid confusing new reviewers still adjusting to the process, but also as we have found that with the larger number of reviewers primarily at remote locations the ability to assimilate and adjust to any change is variable across reviewers. Therefore any change has resulted in a temporary increase the rate of conflicting reviews, and experts have had to spend additional time providing feedback to reinforce the latest rule changes.

4 For example, if two reviewers agree that a book is in the public domain but for different reasons, (e.g., one indicates it is due to non-renewal, the other notices the work was actually published prior to 1923), the work is no longer reviewed by an expert but instead will get a third automatically-generated review with rights determination “public domain” with a reason code that reflects that reviewers agreed on the rights status for different reasons.
This automatic processing has reduced expert workload dramatically (see table below - total CRMS determinations increased slightly from September to October, however the number of determinations requiring expert time and the amount of time spent decreased by almost 40%).

<table>
<thead>
<tr>
<th>Total Determinations Exported</th>
<th>September 2010</th>
<th>October 2010</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Determinations Exported</td>
<td>5449</td>
<td>5853</td>
</tr>
<tr>
<td>Determinations requiring Expert Review</td>
<td>1888</td>
<td>1187</td>
</tr>
<tr>
<td>Expert Time spent Reviewing (minutes)</td>
<td>1565</td>
<td>970</td>
</tr>
</tbody>
</table>

5. Publicity

Accomplishments:

- Interview with the Principle Investigator was featured on the Stanford Copyright & Fair Use website ([http://fairuse.stanford.edu/blog/2010/09/rising-into-the-public-domain.html](http://fairuse.stanford.edu/blog/2010/09/rising-into-the-public-domain.html))

- Project Manager was interviewed for piece that appeared on Library Journal’s website ([http://www.libraryjournal.com/lj/home/887388264/hathitrusts_copyright_detectives.html.csp](http://www.libraryjournal.com/lj/home/887388264/hathitrusts_copyright_detectives.html.csp))

- Project Manager presented on the project at Coalition for Networked Information Fall 2010 Membership Meeting ([http://www.cni.org/tfms/2010b.fall/Abstracts/PB-copyright-karle.html](http://www.cni.org/tfms/2010b.fall/Abstracts/PB-copyright-karle.html))

- Information about the project was posted to various listservs (i.e., DIGLIB, open-bibliography)

- Continue to maintain project website that includes Interim Performance reports, project updates, presentations ([http://www.lib.umich.edu/copyright-review-management-system](http://www.lib.umich.edu/copyright-review-management-system))

- Continue to send out performance reports and project updates to relevant UM Library staff, CIC partners and colleagues at other institutions who have expressed interest in the project

5. Contract with United States Copyright Office (USCO) to evaluate CRMS determinations

Accomplishments: As part of our efforts to evaluate our work, the IMLS grant stipulates that we will engage the US Copyright Office to undertake comparison searches of works we have determined to be in the public domain in order to evaluate the accuracy and effectiveness of the CRMS process. We sent the first
set of volumes for evaluation in early 2010, and received results from the Copyright Office in September. Based on the budget allotted, the Copyright Office was able to evaluate 96 volumes for renewal status. Only four of our determinations were determined to be incorrect (i.e., a work we determined was not still under copyright was actually renewed). Further analysis revealed only one of these volumes to be a true mistake on the part of CRMS reviewers. The others were the result of ambiguous or missing information in the Stanford Renewal Database.

While pleased with the results of this evaluation, we realize that a comparison of results of pre-CRMS determinations would provide a more meaningful measure of our progress in increasing the accuracy and effectiveness of our process. Therefore, while not officially part of the project plan or budget, we submitted a new request for the Copyright Office to research the same number of volumes that were determined to be in the public domain prior to the release of the CRMS. We hope to have results on those searches back in early 2011. (A second round of CRMS evaluations is scheduled/budgeted for next year, in order to evaluate any change in accuracy now that work is spread across several institutions.)

6. Ingest metadata records from CIC institutions for qualifying (in-copyright) works in their collections for collaborative checking of copyright status

Explanation: This was included in the grant proposal and scheduled to take place during Phase III, however we did not implement for several reasons. As the project and system have evolved we have realized how important it is to have the digitized volume available in order to make an accurate determination of copyright status under the CRMS process. Relying solely on bibliographic information to make copyright determinations would involve a high level of risk that we would miss important information included in the front matter of the book and likely result in significantly less reliable determinations. In addition, possible arrangements for the CIC libraries to obtain copies of the digitized in-copyright works for purposes of copyright review work have been on hold pending a final decision regarding the Google Books Settlement. Nevertheless, we were able to secure a significant number of Indiana University volumes deemed to be in-copyright and we reviewed these in the CRMS, finding 40% to be in the public domain and opening access to these books.

7. Sharing Rights Determinations

Accomplishments: For the past year we have been making rights determinations for all works in HathiTrust publicly available via tab-delimited files. Among other data about the works in HathiTrust (identifiers, source of original volume, access information), these files include rights information for each volume, including the rights attribute (e.g., public domain, in-copyright, etc.) as well as the rights determination reason code (e.g., copyright not renewed, no copyright
notice on the piece, etc.). More information about HathiTrust Data Distribution & APIs and the HathiTrust Metadata is available here:

http://www.hathitrust.org/data
http://www.hathitrust.org/hathifiles_metadata

Evaluation of Issues/Efforts for Current Reporting Period– July through December 2010

Training of reviewers at outside institutions has been generally deemed a success and was completed ahead of schedule. All new reviewers completed training and were working in “Production” CRMS as of September 1, 2010, a full two months ahead of what we had projected in the Project Schedule of Completion.

October was a record month with over 5,800 rights determinations exported. While the numbers were a bit lower for November and will likely be again for December due to the holidays, at this point we estimate that we will be able to maintain a level of on average 5000 determinations per month in 2011.

As predicted in the previous Interim Narrative Report, we have seen that consistency in approach and application of our review guidelines has resulted in relatively quick integration of new reviewers into the process, while veteran reviewers have also now had time to adjust to the Decision Tree model and the updated guidelines.

We are extremely pleased with the changes to CRMS processes that have helped us to address the issue of limited expert resources and still keep up with much higher volume of reviews.

At this point it is highly likely that we will be able to achieve our goal of opening access to 60,000 public domain volumes by the end of the grant period.

Next Steps

Our main priority for the rest of Phase III is to maintain the high number of determinations and low conflict rates. Starting in January we will begin to define requirements for CRMS version 3.0. We will also continue to work to implement rights inheritance for duplicate volumes into the CRMS determination and export processes. We will send the second round of CRMS determinations to the Copyright Office for comparison searches and will analyze the results of the supplementary search of pre-CRMS determinations to further ascertain the effectiveness of the project. Finally, we are in talks with the University of California/California Digital Library about training staff from one of their institutions to participate in the CRMS project starting in 2011.