Overview

The University of Michigan Library has recently completed Phase I of the grant project, the goal of which was to design, build and release version 1.0 of the Copyright Review Management System (CRMS). We were scheduled to have moved into Phase II of the project by June 2009; however we faced some setbacks that delayed getting version 1.0 into production. Our primary setback occurred in mid-March, when the programmer originally assigned to the project resigned. An interim programmer was assigned to work on the project until we hired a replacement. Since that time we hired a new permanent programmer who started work on July 6, 2009. Despite these interruptions, we were able to complete testing of version 1.0 of the system during June and early July, and moved to full reliance on the CRMS for copyright review as of July 10, 2009. Therefore the project is about one month behind what was proposed in the Project Schedule of Completion as submitted with the grant proposal. We anticipate that with no additional major interruptions we will still be able to complete the project by the expiration date of the grant (November 30, 2011).

Summary of activities – December 2008-July 2009

Manual Review (scheduled December 2008 through May 2009) – Electronic Access (EA) Unit staff continued to perform copyright status determinations using the original “manual” methods (manual searching for works in HathiTrust and Stanford database, using spreadsheets to track their work) through June 30, 2009. Staff continued manual work for one month beyond what we had originally planned due to the delays in getting version 1.0 running in the production environment. Staff reviewed over 26,000 volumes over the seven-month period, almost 13,000 of which were determined to be in the public domain (see below for details).

<table>
<thead>
<tr>
<th>Month</th>
<th>Items examined</th>
<th>Items cleared</th>
<th>Time spent (min.)</th>
<th>Items Examined per Hour</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dec-08</td>
<td>2971</td>
<td>890</td>
<td>4405</td>
<td>40.47</td>
</tr>
<tr>
<td>Jan-09</td>
<td>2428</td>
<td>921</td>
<td>3281</td>
<td>44.40</td>
</tr>
<tr>
<td>Feb-09</td>
<td>2313</td>
<td>1,063</td>
<td>2933</td>
<td>47.32</td>
</tr>
<tr>
<td>Mar-09</td>
<td>3161</td>
<td>1,490</td>
<td>3433</td>
<td>55.25</td>
</tr>
<tr>
<td>Apr-09</td>
<td>5423</td>
<td>3,005</td>
<td>4703</td>
<td>69.19</td>
</tr>
<tr>
<td>May-09</td>
<td>4853</td>
<td>3,005</td>
<td>4850</td>
<td>60.04</td>
</tr>
<tr>
<td>Jun-09</td>
<td>5592</td>
<td>2,567</td>
<td>5630</td>
<td>59.60</td>
</tr>
<tr>
<td>Total</td>
<td>26741</td>
<td>12,941</td>
<td>29235</td>
<td>N/A</td>
</tr>
<tr>
<td>Avg./month</td>
<td>3820</td>
<td>1849</td>
<td>4176</td>
<td>53.75</td>
</tr>
</tbody>
</table>
Development of the CRMS Version 1.0

- Define Requirements (scheduled December 2008) – The Project Manager and Programmer scheduled hour-long observation sessions with each of the EA Unit staff performing manual review and related work. They also met with Library Core Services staff to discuss requirements for automating updates to the HathiTrust Rights Database. The Project Manager created a preliminary specifications document in early January.

- Develop Beta (scheduled January through February 2009) – The Programmer, with some oversight by the Project Manager, spent January and the first half of February building the Beta Version of the CRMS, which was completed ahead of schedule. Activities included:
  
  - Set up the tables for the CRMS Database
  - Built the CRMS Reviewer Web Application, including the review, expert review and preliminary administrative interfaces
  - Defined the selection criteria and process of selecting candidate volumes for review
  - Defined the criteria and ranking processes to control how items are loaded in the reviewer queue and expert reviewer queues
  - Defined the process for exporting determinations to the Rights Database

The EA Unit Staff and the Project Manager provided feedback to the Programmer on the various interfaces and basic functionality during the latter half of February.

- Test and Refine Beta Version (scheduled March – April 2009) – In late February the original Programmer assigned to the grant notified us that he was resigning from his position at the Library to run his family's lumber business. The interim Programmer was assigned and started working on the project in early March, so he was able to spend some time with the original Programmer before he left. We continued testing the Beta Version and working to further develop the system from March through early June. Activities during this period included:
  
  - Conducted User Testing - In March, the Library Usability Group (UG) evaluated the Review interface of the Beta Version by conducting focus groups with EA Unit staff. They then provided recommendations for changes to the interface
  - Enhanced and refined the reviewer interface based on UG feedback and additional input and ongoing review by the Project Manager and EA Unit Administrator (with periodic solicited feedback from EA Unit Staff)
  - Developed the process for selecting candidate volumes for review
  - Developed the process for updating the reviewer queue with new items
  - Developed the process to clear the queue of items that have been reviewed
- Developed the process to populate the Rights Database with final determinations
- Developed scripts to load rights determinations from pre-CRMS Excel spreadsheets

- Release Version 1.0 (scheduled May 2009) – By early June we released CRMS version 1.0 for testing in the development environment. The Programmer moved the system to Production in late June and the Project Manager and EA Unit staff tested for about two more weeks to ensure the system was stable and functioning properly. EA Unit staff ceased manual copyright status reviewing as of June 30th, and version 1.0 of the CRMS went live on July 10, 2009. From that point on, all rights determinations work has been done using the CRMS. Activities during this last period included:

- Project Manager and EA Unit Administrator developed testing plan and scripts. EA Unit staff and Project Manager tested the system in the development environment. We provided continuous feedback to the Programmer to address issues and bugs.
- EA unit staff used the system for copyright reviews for a specified period of time each day, with the Project Manager resolving all conflicts, during which we identified additional issues with the CRMS on the development server.
- Programmer completed preparations for version 1.0 release (fix bugs, package code, move to production servers)
- Project Manager repeated controlled user testing on the Production Server. EA Unit staff then used the system for copyright reviews for a specified amount of time each day. Staff reported any remaining bugs and issues and the Programmer resolved them until everyone felt confident that the system was stable and ready to go live.

System Details

The CRMS includes a database as well as a workflow management system that prioritizes and queues the volumes and allows them to be worked on by multiple reviewers. Reviewers are authorized to view the in-copyright page images in order to make rights determinations. Currently, only reviewers at the University of Michigan are using the system. As scheduled, during Phase III of the project, reviewers at other institutions will be able to access the web interface and work in it remotely via authentication.
Database

The CRMS includes a database that stores basic bibliographic information for all qualifying volumes (volumes published in the U.S. from 1923-1963 that are not U.S. federal government documents), plus a digital object identifier that links to the page images for each volume.\(^1\) In addition, it stores all determinations associated with a given volume, including detailed review information such as who performed the review, the date reviewed, details of the determination, current status of the review, and whether the determination was done pre-CRMS.\(^2\)

Web Interface\(^3\)

The CRMS review interface is made up of a vertical panel that presents reviewers with the digital object identifier and relevant descriptive metadata for a volume. Below is a section with buttons that control what appears in the main view of the review interface. Reviewers can automatically perform a search for a volume in the HathiTrust pageturner (default view), and in Mirlyn (the University of Michigan Library OPAC). There are also options (author, title, advanced) for searching the Stanford Renewal Database. The review panel includes a set of display options so reviewers can toggle views, go to a split view, or move the review panel to the opposite side of the screen. Reviewers enter information in the lower half of the panel, where they can indicate the their determination of the copyright status of the volume, and include any notes about the determination. The review panel includes a link to our “Use Cases” document, so reviewers can easily access that document when they have questions about guidelines for making a determination.\(^4\)

The CRMS home page includes separate links so that reviewers can go back and edit their reviews, view all the historical reviews, and view their reviewing statistics. There is a separate administrative interface that includes links to the expert review queues, a view of all active reviews, a summary of reviewer statistics, a page with detailed information about the status of volumes in the queue, and other functions for administering user accounts. It also includes a tool to add a volume to the queue “on demand” so it can then immediately be reviewed by an expert reviewer. All reviews (in each of the interfaces) are searchable by various categories of information (e.g., identifier, date, user, title, status, rights attribute, etc.), and search results can be ordered based on various criteria as well.

---

\(^1\) At this point we are only reviewing qualifying volumes that have been ingested into the HathiTrust repository, so authorized reviewers can use information on the title page, verso and other front matter to make a determination. In the proposal we also discussed reviewing works that were not stored in our repository (in such cases a determination would be based solely on review of information in the bibliographic record); however at this point we are focusing our effort on works in the repository.

\(^2\) Reviews performed under the previous manual routines from January 2008 through June 2009 will be exported from spreadsheets and stored in the CRMS database, but will be marked as “Legacy” reviews so we can distinguish them from those that have gone through the double-review process.

\(^3\) Note: various screenshots of the interface are included on the CD submitted along with this report.

\(^4\) In the proposal we mentioned including a link to Live Help in the review interface, so reviewers could instant message or chat with the Project Manager/expert reviewer. At this point, everyone working on the project is in close proximity, so we have not felt a strong need for this tool in version 1.0.
Workflow/Process

The CRMS uses “rules” to ensure that each volume in the queue is reviewed by at least two different “regular” (as opposed to “expert”) reviewers. Volumes in the queue are “inactive” until a reviewer makes an initial determination (the system includes record “locking” to ensure that only one person at a time can work on an item in the queue). The determinations are then stored in the (active) reviews table. The system prioritizes the queuing, so if one reviewer has reviewed the volume, it gets put at the top of another reviewer’s queue to obtain a second opinion. If the determinations of the two reviewers match, the determination becomes final. If they do not match, or if the conclusions of the two reviews both are “undetermined,” the volume is queued up for an expert to review. The expert’s determinations are always final.

The system assigns a status to each volume in the queue as follows:
- the volume has not been reviewed or has been reviewed by one reviewer;
- the volume has been reviewed by two reviewers but their determinations conflict;
- the volume has been reviewed by two reviewers but both could not make a determination;
- the volume has been reviewed by two reviewers that agree; or
- the volume has been reviewed by an expert.

As review work takes place, the system updates the status for each volume, and the reviews for a volume are moved to the various queues. Ultimately, when a final determination is made, the system moves the volume out of the queue.

The CRMS runs several processes on a nightly basis, during which the statuses of the volumes are updated, and the final determinations are exported to a file that is used to update the HathiTrust Rights Database. Finalized reviews are then moved from the (active) reviews table to a “historical reviews” table. The system updates the table of “candidate” volumes with any new volumes that have come into the HathiTrust repository that qualify for copyright determination work. The system then repopulates the queue with new volumes from the candidates table for each year from 1923-1963, repeating the process as necessary, to bring the queue back up to 800 volumes.5

Using the CRMS - Next Steps

During the final testing phase, and since staff started using the CRMS officially for copyright determination work, it was clear that the reviewers were able to more efficiently make determinations using the system than they were under the manual process. Although their productivity rates have been high, we have seen a larger percentage of conflicting reviews than we expected based on past comparison tests. Thus far, keeping up with reviewing the

---

5 We originally proposed prioritizing review work based on reverse chronological date of publication, on the assumption that earlier works were less likely to be in copyright. We subsequently decided that it would be preferable to review works from all years so that we are not simply opening up the earliest works.
conflicts has been a challenge. At times we have had to limit reviewers to a certain number of volumes per day so we can analyze the nature of the conflicts and determine when reviewers are consistently making mistakes and why. For instance, is the reason for the error that they are not following the guidelines set forth in our Use Cases document (in which case retraining is needed), or are they missing things because they are going too fast? We anticipate some general retraining of the group will be required, combined with work with the individual reviewers to point out where they are consistently making mistakes. Determining retraining needs and clarifying any ambiguity in our use cases is a priority at this point so we can proceed to take full advantage of the system.

As we move into Phase II of the project, we also plan to continue to monitor the system in general, to resolve any further technical issues, and define priorities for features and enhancements that we did not work on before releasing version 1.0. We will begin to focus efforts on developing methods for making our copyright status determinations available for sharing and distribution. We will continue to look for ways to improve the accuracy of our determinations, and to clearly document our guidelines, in preparation for training reviewers at other CIC institutions next year.
Addendum to CRMS Interim Performance Report Narrative – 7/31/09
Summary of Personnel Changes 12/1/08 through 7/31/09

Joshua Santelli, Application Programmer/Analyst Intermediate, resigned as of March 17, 2009. Brian Hall, Application Programmer/Analyst Intermediate, commenced employment as of July 6, 2009. During Year 1, the Programmers devote 92% of their time to the project.

Senovia Guevara, Information Resources Assistant Intermediate, has replaced Erik Saran, Information Resources Cataloger Specialist, as of April 1, 2009. Both Saran and his replacement Guevara devote 40% of their time to the project.

Resumes for Hall and Guevara are attached.