2013

Developing Issues in Licensing: Text Mining, MOOCs, and More

Andrée J. Rathemacher
University of Rhode Island, andree@uri.edu

Citation/Publisher Attribution

This is a pre-publication author manuscript. The final published version is available at: Rathemacher, Andrée J. "Developing Issues in Licensing: Text Mining, MOOCs, and More." Serials Review 39, no. 3 (September 2013): 205-210. Available: http://dx.doi.org/10.1016/j.serrev.2013.07.016.

Developing Issues in Licensing: Text Mining, MOOCs, and More

Andrée J. Rathemacher

This report covers a program co-sponsored by the Collection Development and Electronic Resources Management Interest Groups of the Association of College and Research Libraries New England Chapter (ACRL/NEC), an independent chapter of ACRL. The workshop, titled "Developing Issues in Licensing: Text Mining, MOOCs, and More," took place on April 25, 2013 at Northeastern University in Boston, Massachusetts. Approximately 40 people attended.

Open Education on the Move: New Pathways for Learning

The morning began with M.S. Vijay Kumar (senior strategic advisor, Digital Learning and director, Office of Educational Innovation and Technology, Massachusetts Institution of Technology), who provided an overview of developments in online education. Kumar stated that his job is to help the community at MIT respond to the significance of all things open and all things digital and their impact on the university. The shift to "open" and "digital" has profound pedagogical implications, and what is exciting is that we do not yet know what these are. We need to figure out how to collect data so that we can ask the relevant questions and create preferred futures for education, futures defined by dramatic improvements in access and new ways of defining and realizing quality.

Kumar explained that on the demand side, there is a growing need for access to higher education, both in countries like Brazil and India that are trying to develop in a hurry and do not have an adequate educational infrastructure and in the United States, where cost has become a barrier to educational access. On the supply side, we have powerful new technologies, including tools like mobile computing, cloud computing, data visualization and analytics, augmented reality, and game-based learning. These technologies are contributing to the production and distribution of education through massive open online courses (MOOCs). Another significant factor in the transformation of education is openness, as expressed through open access scholarship, open educational resources, open source software, and open licenses like those from Creative Commons.

Open education has been expressed in different ways over the years through many educational initiatives, Kumar explained. Just over ten years ago, open courseware was launched when MIT made the content of all its courses available online for free for educational purposes. With this initiative, the MIT community began having discussions about MIT’s unique value proposition. Faculty articulated that what defined the value of an MIT education was intensity: a high level of interaction between high quality students and high quality faculty as evidenced in part through project-based learning and hands-on experiences. The question became how to maintain and extend this value proposition when offering online distance education to a broader set of learners who are more diverse in their levels of preparation.

While there are all sorts of open education initiatives at all levels, for example online courses and online tutoring, the biggest event so far in open education has been the launch of MOOCs. Kumar believes that one of the characteristics of MOOCs that makes them significantly different is that the community of self-learners enrolled in the MOOC plays a major role. MOOCs do not offer a one-to-many relationship so much as a peer-to-peer learning opportunity. For example, MIT has found that students in MOOCs have created tools and software to help each other learn; there is a whole ecosystem of production that is going on. Another characteristic of MOOCs that distinguish them from other online courses is that much of the process is automated, for example assessment.

Kumar mentioned that there are exciting new developments in online learning that are beginning to show up in MOOCs and have the potential to be dramatically transformative. These include tools that bring the practice of research to the process of learning, for example protein visualization, materials modeling, hydrology visualizations, and parallel programming opportunities. Through MOOCs students can be exposed to the discovery aspect of research and to the processes of doing research using interactive technology. The point is that MOOCs are not just about access to content like video clips and assignments posted online. MOOCs can enable end-to-end educational experiences, including hand-on experiences, at scale.

Kumar presented some of what MIT has learned through its experience offering MOOCs. One thing is the value of real-time feedback and correction that is the result of students in the MOOC helping one another. Along these lines, new tools and technologies are being developed that allow, for example, programs created for an assignment in a computer science course to be chopped into chunks and sent to many reviewers for grading, allowing for faster feedback. Another lesson learned from MOOCs is the value of online learning in enhancing face-to-face education on campus. Shifting information transfer (lectures) and assessments (tests) online allows for the “flipped classroom” in which scheduled class time is used for field experiences, labs, and other interactive experiences.

Kumar ended with a discussion of concept-based learning and modularity. Concept-based learning has as its goal to present students with a coherent sense of how the content and skills they are learning relate to specific concepts. These concepts can be linked to educational assets, for example labs and lectures, so that students seeking mastery of a given concept can chart their own path through the material required to master that concept. This in turn enables modularity, the ability to experience education in smaller chunks, which can make it easier to create opportunities for students, like internships or study abroad experiences, that do not interrupt the flow of education. Rethinking the entire curriculum based on concepts could play a large role in changing the ecology and economics of education. MOOCs and related technologies can offer an abundance of courses, content, and interaction opportunities. Access to courses can be blended with hands-on vocational opportunities that allow for a more customized and accessible education. The challenge will be in determining in this new environment what to discard and what to keep.

Why Humanists Need Data: New Uses for Electronic Archives

Speaking next was Ryan Cordell, assistant professor of English at Northeastern University and a core faculty member at the NULab for Texts, Maps, and Networks, Northeastern University's new center for Digital Humanities and Computational Social Science. He presented his research on nineteenth century U.S. newspapers. Cordell explained that he is interested in historical newspapers because he is interested in viral media. In the nineteenth century, before modern copyright had taken shape, newspapers in the United States were similar to today's blogs or aggregators. Newspaper editors combed through other newspapers to find material their readers might like and published it, sometimes with attribution, sometimes without. What Cordell is studying is how these shared texts moved around the country, changing as they did, and how they informed society at the time.

Nineteenth century newspapers included a wide variety of content, such as poems, short stories, and travel accounts. For example, one poem, "The Inquiry," was reprinted in newspaper after newspaper throughout the country, changing over time. The version that became the most popular, that "went viral," was one of the edited versions, not the original. The poem ultimately became so popular that it was parodied. Because nearly everyone in the country was experiencing texts such as this one, they can tell us a lot about the period.

Cordell's primary source for his research is the Library of Congress's site Chronicling America: Historic American Newspapers (http://chroniclingamerica.loc.gov/). The site contains the full text and page images of many American newspapers published between 1836 and 1922. Cordell explained that if he had been conducting his research two decades ago, he would have had to painstakingly read every newspaper he could, hoping to randomly encounter shared texts. Even conducting this research with simple full-text searching capabilities would be difficult, since searching relies on inputting known text. Cordell explained that what he needs for his research is the data itself, that is, the full text of the digitized newspapers generated using optical character recognition (OCR).

Cordell and his colleague David Smith of Northeastern University's College of Computer and Information Science have "scraped" the full text of all newspapers in Chronicling America published before 1860 in order to analyze it for matching passages. Smith's area of research is duplicate detection. They have created an algorithm that breaks the unstructured text into strings of five words (or *n*-grams) and then searches through the entire body of text for matching *n*-grams. If enough five-word sequences match between two or more pages, the algorithm identifies a possible matched text. Because the program is only looking for five-word matches, it is not affected by the frequently poor quality of the OCR. Each potential matched text is assigned an identification number and delivered to Cordell in a spreadsheet. So far, Cordell and Smith's research has identified 50,000 viral texts, though Cordell has focused thus far only on the top 5,000. The vast majority of these texts are items that literary scholars have never written about. Most are by anonymous authors or are minor pieces by major authors that we now realize were more influential than previously recognized.

Cordell displayed a number of visualizations of his data, some involving mash ups with open data from other sources such as Railroads and the Making of Modern America from the

University of Nebraska, Lincoln, the David Rumsey Map Collection, and the Atlas of Historical County Boundaries from the Newberry Library. By taking a historical map, overlaying data about the railroad network at that time, and then adding data on reprinted newspaper texts, Cordell illustrated that the reprinting of newspaper content lined up neatly with the railroad network. Cordell also combined data on historical county boundaries with data about the founding of newspapers to show that, as the population expanded westward, newspapers appeared first, and then shortly after the establishment of a newspaper, political boundaries were established. He has also mashed up his data with historical census data to map the characteristics of populations near where certain types of viral stories appeared, for example what the population looked like in places where religious stories were reprinted.

Cordell also used network analysis to create a diagram on which circles represented individual nineteenth century newspapers and lines between the circles represented shared text. This type of data visualization shows which newspapers were the most influential, printing items that other newspapers chose to reprint. It also reveals which newspapers regularly shared stories, which was often the result of a shared religious or political affiliation or, in one case, a family relationship between the editors. Cordell pointed out that the diagram illustrates the prominence of some newspapers that we might not have suspected. For example the most prominent newspaper during the time period studied was the *Nashville Union and American* because at that time Nashville was the geographic center of the country.

Cordell startled the audience by revealing that his study includes no historical newspapers from Massachusetts. The Library of Congress Chronicle of America includes no Massachusetts newspapers because at this point in time only a commercial vendor, Readex, has digitized them, and their data is not available for text mining. Cordell stated that if he had access to Readex's America's Historical Newspapers and ProQuest's American Periodicals Series Online, he would have much more data to analyze. He and Smith have begun conversations with Readex and ProQuest about using their data, but the process has not been easy. They are trying to convince these vendors that if they were to allow the use of their text, Cordell and Smith could help them in return by providing corrections to the OCR identified through their research as well as increasing the visibility of these databases. Researchers who do text mining need the help of librarians to include data mining rights in license agreements.

Tipping Point: MOOCs, Copyright and the Challenges and Rewards of Wesleyan's Coursera Partnership

Jolee West, director of academic computing and digital library projects at Wesleyan University, spoke next about her experiences doing copyright research for Wesleyan's MOOCs. West explained that she is neither a lawyer nor a librarian; she has a PhD in anthropology and works as a technologist.

In the fall of 2012, Wesleyan partnered with Coursera, a for-profit educational technology company that works with universities to host MOOCs. Wesleyan's first MOOCs were offered through Coursera in February 2013. West described MOOCs as distance education combined with crowd-sourced learning. MOOCs have very large enrollments of students from around the world, many of whom are not native speakers of English. (At Wesleyan, the average Coursera enrollment is 30,000.) MOOCs are known for very high attrition rates (about 90%), but, as West explained, the students who remain are highly engaged. Students enrolled in MOOCs self-organize to help each other. Within hours of a MOOC opening, enrolled students form local study groups. Discussions do not need to be seeded.

West noted that faculty in the face-to-face classroom use a wide variety of copyrighted works, but this practice does not translate to MOOCs, which take place online and are open to anybody. Because the application of the distance education safe harbor in copyright law is questionable in the case of MOOCs, Wesleyan relies on fair use when including third-party copyrighted material, and a great deal of discussion and debate takes place around every copyrighted item used.

One issue that arose at Wesleyan around relying on fair use for MOOC content was that Coursera is a for-profit company, and fair use law favors "nonprofit educational purposes." For this reason, some other schools also offering MOOCs through Coursera will not depend on fair use when incorporating third-party copyrighted content. However, West mentioned, there are a number of legal cases in which fair use has been upheld for commercial entities, and she has spent a good deal of time scouring the Web for information on these cases “to get her head in the right place.” In then end, West believes that Wesleyan’s fair use with regard to Coursera MOOCs is not that different from fair use claims made by institutions using the not-for-profit EdX system.

Relying on fair use is often necessary because Wesleyan’s subscriptions to licensed electronic resources do not cover external students enrolled in MOOCs, and licensing rights for an additional 80,000-100,000 students would not be feasible. When Wesleyan first contacted the Copyright Clearance Center (CCC) about licensing an article for a MOOC, the CCC had no idea what a MOOC was and stated that the licensing fee for the article would be $3.00 per student, the same as for on-campus students. As a result, readings from journal articles and other copyrighted sources are off-limits for Wesleyan MOOCs unless they are available open access or unless the instructor wants to leave obtaining access up to the students themselves. When she is undecided about whether using a particular item would qualify for fair use, West confers with the university’s counsel. The fact that students need to register for MOOCs — that the courses are not totally open — mitigates some of the risk involved in invoking fair use, despite the fact that anyone can register for a MOOC and that participants have access to content for as long as the instructor leaves the course open.

West provided a number of examples of decisions about using third-party copyrighted content in MOOCs at Wesleyan. A professor for a Wesleyan Coursera MOOC on “The Language of Hollywood” addressed copyright concerns by only using materials openly available on the Web, such as movie posters, still images, and publicity shots. Nothing was taken from print publications, and no movie clips were shown. Instead, the instructor posted a list of movies and suggested that students enrolled in the MOOC obtain the films from their library, Netflix, or a video store. He linked to the IMDB.com page for each movie to reduce the amount of searching required of students in the course. For another class, West found that even clips from silent movies directed by Buster Keaton posed a problem, because although the motion picture content is in the public domain, modern releases of these early silent films have used musical soundtracks that might be protected by copyright.

In another example of working with instructors regarding MOOC content, West received a request to use an excerpt from a recording of a speech by Martin Luther King, Jr. Because West was aware that the family foundation that owns the rights to this content is very aggressive about copyright, the faculty member was not permitted to use the excerpt. This example illustrates the principle of avoiding the use of famous or aggressively monitored content in MOOCs if possible, a strategy that was echoed later by Kyle Courtney of Harvard.

Many of the decisions West helps faculty members make about including content in MOOCs concern images. At Wesleyan they had long discussions about whether images of book covers would be allowed and decided to use them on the basis that the advertising benefit to the rights holder by including the image in a MOOC would outweigh any possible market harm. In one case, a faculty member wanted to use an image of the label on a vinyl LP record. Because they were not comfortable with a fair use rationale for using the label, they did not use it. For an

Associated Press image used in a MOOC, Wesleyan decided not to rely on fair use but to license the right to use the image for five years at the cost of $150. When using fine art images, Wesleyan has found the terms of the Metropolitan Museum of Art to be fairly generous. The Met allows for the non-commercial and educational use of images as long as the images belong to the museum and are not subject to additional copyrights. The terms of use of images from the Museum of Modern Art, on the other hand, are more restrictive so West avoids using them. Wesleyan also makes a good deal of use of Creative Commons licensed images from Wikimedia Commons; students do the work of obtaining the images and recording the required attributions.

West provided an example of how one Wesleyan professor obtained content for his MOOC without having to conduct a fair use analysis, rely on open content, or pay a licensing fee. The professor of the "Social Psychology" MOOC, in which over 90,000 students had enrolled, wanted to use the same textbook that he uses for his on-campus version of the course (a book that was dedicated to him). He approached McGraw-Hill, the publisher, which offered to make a cheaper version of the book available for $100. The professor believed this was still too expensive for many students, so he convinced the publisher to allow him to use only three chapters of the text at no cost. He also wanted to use a photo of the Blue Man Group and to turn the photo green for a demonstration, and so he approached the organization and received permission. In the end, he acquired materials from many rights holders by asking them directly. In return, the main page for his MOOC (https://www.coursera.org/course/socialpsychology) thanks them and displays their corporate logos.

West concluded by stating that she believes we have reached a tipping point with regard to MOOCs, markets, and open access. She believes that MOOCs represent a significant new customer base for content providers and that in response they will adapt to serving this new market. Adaptation strategies might include new licensing schemes with a lower-tier price of entry, unbundled textbooks sold chapter-by-chapter, limited free access to content as in JSTOR's Register and Read program, and an increasing awareness among faculty of open access alternatives.

Making the Legal Case for Transformative Fair Use: MOOCs and Text Mining

Up next was Kyle Courtney, a librarian and attorney who works at Harvard Law School as the manager of Faculty Research and Scholarship. Courtney began his talk with text mining and the fact that Judge Baer's decision in the recent HathiTrust case (The Authors Guild, Inc. et al v. Hathitrust et al) confirmed that text mining is a transformative fair use.

The HathiTrust Digital Library is a collection of digital texts from over 70 research libraries around the world. It contains over 10 million volumes, including books digitized through the GoogleBooks project as well as other digitized materials. In 2011, the Authors Guild sued HathiTrust for copyright infringement. At stake were three issues: storing scanned materials for preservation purposes, enabling text mining of full text files, and making the full text of books available for print-disabled users. In December 2012, Judge Baer ruled in favor of HathiTrust, stating in his opinion, "I cannot imagine a definition of fair use that would not encompass the transformative uses made by Defendants… and would require that I terminate this invaluable contribution to the progress of science and cultivation of the arts that at the same time effectuates the ideals espoused by the [Americans with Disabilities Act]." (http://www.scribd.com/doc/109647288/Ag-v-Ht-Opinion-Order).

"How did we get to this amazing sentence?" asked Courtney. He noted that the emphasis on transformative use in fair use cases is relatively new. Transformative use is the use of copyrighted material in a manner, or for a purpose, that differs from the original use. The use develops in such a way that the expression or meaning is essentially new. Courtney traced the history of this notion of transformative fair use through a close examination of two court cases. In the case Campbell v. Acuff-Rose Music, decided in 1994, the Supreme Court ruled that the rap group 2 Live Crew's parody of the Roy Orbison song "Oh Pretty Woman" was a fair use. The Court's decision stated that "the goal of copyright, to promote science and the arts, is generally furthered by the creation of transformative works." (https://supreme.justia.com/cases/federal/us/510/569/case.html). This case was the first mention of the concept of transformative fair use. In another landmark case, Bill Graham Archives v. Dorling Kindersley, the U.S. Court of Appeals ruled in 2006 that publisher Doring Kindersley's reproduction of concert posters was a transformative fair use because thumbnail images of the posters were used to illustrate a timeline of events related to the Grateful Dead, a use which was different from the posters' original aesthetic and promotional purposes. (http://law.justia.com/cases/federal/appellate-courts/F3/448/605/637042/).

Courtney connected this important notion of transformativeness to the text mining of copyrighted works by examining how Judge Baer applied each of the four fair use factors in the HathiTrust case. Courtney explained that the first factor, the purpose and character of the use, favors non-profit educational uses. In addition, the judge found HathiTrust's uses to be transformative because the copies of the books were used for an entirely different purpose (superior search capabilities) than the original works (actual access to the copyrighted material). Factor two, the nature of the work, was not significant because the use was transformative. The judge also found factor three, the amount used, in HathiTrust's favor, because searching and access for the print-disabled require copying the entire work – copying the exact amount to "serve the purpose." Finally, factor four, the effect on the market, was also in HathiTrust's favor, because a use that "falls within a transformative market" does not cause a copyright holder to "suffer market harm due to the loss of license fees." Courtney pointed out that following their victory in court, HathiTrust is moving forward with providing users access to their content for transformative uses. On April 22, 2013 they announced the availability of data mining and analytics tools for large-scale analysis of HathiTrust's contents (http://ovpitnews.iu.edu/news/page/normal/24146.html).

Courtney presented some additional examples of text mining to show its value. A project at Harvard is text mining articles related to climate change from the "prestige press" in the U.S. and Great Britain to gauge the emotional tone of the articles in order to determine bias. This material is copyrighted. Another project of which he is aware analyzed the text of 23,000 articles in an attempt to identify proteins that might relieve a mouse model of multiple sclerosis so that potential new drug targets for the disease could be developed. On the other hand, publishers are pushing back against text mining, as described in a recent article in Nature, “Text-mining spat heats up.” (http://www.nature.com/news/text-mining-spat-heats-up-1.12636). This publisher resistance leads to delays in research. For example, it took Max Haeussler of University of California, Santa Cruz three years to get the rights to download 3 million articles from which he is extracting DNA data to annotate an online map of the human genome. Such a wait is ridiculous, Courtney believes. In this case, copyright is interfering with the progress of science and useful arts, not promoting it. We should be able to text-mine all licensed resources. Fortunately, there are signs of progress. Later this year, the United Kingdom will exempt text mining for non-commercial purposes from copyright, and JSTOR’s terms of use allow users to “perform research activities involving computational analysis.” (http://www.jstor.org/page/info/about/policies/terms.jsp).

Courtney then shifted gears to discuss MOOCs and copyright. He explained that one of his roles at Harvard is copyright advisor for HarvardX, which includes Harvard’s participation in the MOOC platform edX. Courtney acknowledged that MOOCs offered through HarvardX cannot use third-party copyrighted materials as might be done in a face-to-face classroom; with as many as 150,000 students registered in a MOOC, the damages for copyright infringement would be “unimaginable.” Courtney presents faculty with four options when they ask him about using third-party copyrighted material in a MOOC. First is simply the option to not use the material at all: Is it really necessary? Second, he asks faculty whether the material can be replaced with something else, for example an open access version of an article found in a repository instead of the publisher’s version, or a Creative Commons licensed image instead of one with all rights reserved. The third option is the heart of Harvard’s strategy with regard to MOOCs: Can the professor rely on transformative fair use? Finally, if the first three options are not available, the option remains to seek permission from the copyright holder. Harvard, however, has no budget for permissions. Courtney exclaimed, “We’re not paying for third-party materials!” He explained that Harvard’s focus on educational, transformative fair use and refusal to pay licensing fees improves the educational experience. Any third-party copyrighted material used under these guidelines is more likely to engage students as opposed to being used for aesthetics, or “window dressing,” with no educational purpose.

Courtney provided examples of the transformative fair use of material used in HarvardX courses. In the course HLS1x Copyright, the professor wanted to illustrate the provision of copyright law that pertains to compulsory cover licenses of music and to show that a cover version may differ noticeably from the original. The course made use of about 30 seconds of “Little Wing” by Jimi Hendrix and then about 15 seconds of the same song by Santana featuring Joe Cocker. Courtney pointed out that in both cases, transformative fair use applied, because the song was being used not for its original purpose to entertain, but to illustrate a point about copyright. In addition, the entire song was not used, but only the amount necessary for the transformative purpose. Finally, HarvardX professors are instructed to carefully place the copyrighted material in the context of the course. Each song clip was introduced by the professor, who explained the point that the song clip illustrated. After the clip played, the professor again commented on its purpose. These “bookends” on third-party copyrighted material help establish the transformative nature of the use.

In another example, a professor wanted to use an Associated Press image of smoke in Moscow during the 2010 Russian wildfires for the HarvardX course PH278X Human Health and Global Environmental Change. Courtney determined that the use of this image would not be a transformative fair use because the purpose of its use in the class was identical to the original purpose: to show the impact of wildfires on air quality in Moscow. As an alternative, the professor was able to substitute a similar Creative Commons licensed photo from Wikimedia Commons. Courtney noted, though, that using the Associated Press photograph would have been transformative fair use if it had been used in, say, a photography course to illustrate depth-of field, focus, image composition, or a similar topic not related to its original purpose. In conclusion, Courtney pointed out that the four-factor test for fair use really boils down to two factors: 1) Is the use transformative, that is, does it add value to or repurpose preexisting material for a new audience?; and 2) Is the amount of material taken appropriate to the re-use? These are the two factors they are relying on for MOOCs at Harvard.

In the question-and-answer period, Courtney was asked about using copyrighted reading materials in a MOOC. Courtney responded that Harvard attempts to find an open access version of the material. In some cases they write for permission, but they will not pay. This has resulted in not being able to use material from the *Boston Globe* and Oxford University Press, for example. Most publishers do not yet understand the exposure that MOOCs can bring to their content. In one case, for a computer science HarvardX MOOC, Elsevier agreed to donate jpeg page images of a complete textbook written by the course instructor so that any students enrolled in the course who could not afford to purchase the book could access it. The MOOC also included a link to Amazon for purchasing the book. During the time that the MOOC ran, sales of the book increased by 2000%, and every copy in every warehouse in the world sold out. Partly as a result of this experience, Courtney has decided that in the future he will consider approaching publishers' marketing departments before their licensing departments.

The Consortial Arena: The Challenges of Negotiating Cutting Edge Licensing Provisions

The program's final speaker was Celeste Feather (senior licensing program account manager, LYRASIS). In her role at LYRASIS as the lead negotiator for the Association of Research Libraries (ARL) Licensing Initiative, she is engaged in conversations with librarians and publishers about developing new license terms and how they need to be implemented. She discussed several current "hot topics" in licensing.

The first is the need for license provisions that allow for text mining of the licensed material. Libraries need to be specific about what is required in order for researchers to mine content easily, and these terms should be included in the license. Often a simple clause permitting text mining is not enough, because there are technological barriers to mining the content that were not addressed in the license. We need to educate publishers about text mining; many are not familiar with the concept and would not know how to respond if they received a request from a researcher. For example, when LYRASIS negotiated with university presses about the ability to text mine their e-book collections, the publishers pushed back because they had never been asked about this before. What will be more challenging is negotiating text mining rights in the STEM and business areas, because vendors of this type of content are developing separate products to allow researchers to mine their data. Feather recommends that libraries work together to develop model license language for text mining.

Another hot topic in licensing is the issue of interlibrary loan (ILL) versus short term loans. Feather noted that this is particularly an issue with e-books, for which ILL tends to be allowed at the chapter level only, not for the entire book. Sometimes chapter-level loans work well, but for the average humanities book, which tends to be read as a unified whole, interlibrary loan by chapter does not meet the borrower's needs. Vendors have introduced the concept of a short term loan (or lease), in which, for a fee paid by the library, a borrower can access an entire e-book for a short period of time. Feather believes that short term loans, if the price is appropriate, can be faster and cheaper than traditional ILL when staff costs are considered. There is the potential for both the library and the publisher, which receives an additional revenue stream, to benefit. Feather has found, however, that ARL libraries are adamant about not giving up their ILL rights in favor of short term loans. As a negotiator, she faces the conundrum of requiring ILL rights in the license even when the publishers offer what might be a more cost-effective alternative for libraries. Feather mentioned a new service by the Greater Western Library Alliance (GWLA) as a possible alternative to both chapter-by-chapter interlibrary loans and short term loans. The GWLA has developed a product called Occam's Reader, an ILLiad add-on that combines chapter PDFs into a single "dumb" image file of an entire e-book for easy transmission to a borrowing library in a single transaction. The file cannot be printed or downloaded. Occam's Reader is currently in beta testing. Publishers have told Feather that they do not see the need to invest in the technology to support full e-book files on their sites. Feather wondered what their response will be when they learn that libraries might have solved the technical problem through Occam's

Reader. It also remains to be seen whether borrowers will be satisfied with file that they cannot manipulate.

The next topic Feather addressed was Google Scholar's Library Links program. Library links are "article-level links to subscription full text for patrons affiliated with a library" from within Google Scholar search results. (http://www.google.com/intl/en/scholar/libraries.html) In order to set up this service for libraries with OpenURL-compliant link resolvers, Google needs for the library to provide them with their IP ranges and institutional holdings or to ask their OpenURL vendor to do so. Though Google has offered this service for a number of years, only 150 libraries are participating, which Google considers to be a dismal failure. As a result, Google wants to go directly to vendors like EBSCO, ProQuest, and Gale for library subscription information and IP addresses. Some vendors have been compliant, while others have refused or said that they will only do so upon request by a library. Feather agrees that the easiest way for Google to get this information is from publishers and aggregators, and that Google Scholar does provide a unique service. Should libraries be requiring vendors in licenses to provide this information to Google? If consortia were to include license language to this effect in their licenses, Google would quickly be able to obtain this information for most vendors and libraries. Does cooperating with Google in this way undermine libraries' missions, or could this be a help for underfunded libraries that can not afford discovery systems?

Feather next moved on to the issue of accessibility. Institutions and libraries have been sued because their content is non-accessible to the print-disabled, which comprise approximately 20%

of the population. Claims of ADA compliance do not necessarily mean that products are usable by those who are visually impaired or who cannot process what is on a screen because of learning disabilities. ARL is taking the lead on this issue and has issued a number recommendations (http://www.arl.org/storage/documents/publications/print-disabilities-tfreport02nov12.pdf), one of which states, "Licensing must be done deliberately to protect the values and meet the legal requirements of accessibility, particularly in light of libraries' increasing reliance on licensed content in the digital environment. Research libraries should negotiate for more favorable terms in order to permit broader latitude to adapt content to meet the needs of patrons. With copyrighted works, research libraries should aggressively assert fair use in support of accessible services for the print disabled." Feather believes that consortial licensing agreements are the best tools with which to push for these goals.

Another licensing hot topic Feather mentioned was archiving. ARL licensing specifications require that vendors allow archiving by third parties. The problem is that many publishers are only offering third-party archiving through one service, typically LOCKKS or Portico. Libraries are divided about how they want to archive their licensed materials: LOCKKS, Portico, or self-archiving. Publishers, particularly smaller ones, are challenged to maintain their primary hosting platform and mirror sites, provide local loading options for libraries, and participate in multiple third-party archiving services. Feather hopes that a more focused consensus among libraries will develop.

The final hot licensing topic Feather addressed was MOOCs. Feather believes that it might be possible for libraries and publishers to develop licensing language related to MOOCs that works for both parties. First, libraries will need to figure out what it is that they want: Reduced permissions fees for using content in MOOCs? Limits to the number of articles or chapters that can be placed in a MOOC during a given time period? Feather stated that a library-by-library approach to licensing content for MOOCs will lead to inconsistency, confusion, and inefficiency. Feather believes that in this case, too, consortia are best placed to raise these issues, as they have the ear of the sales people who can put pressure on legal staff. If nothing else, raising the issue of MOOCs in license negotiations serves as an awareness tool to educate vendors.

In connection to MOOCs, Feather mentioned the Stanford Intellectual Property Exchange (SIPX) (http://www.sipx.com). This is automated system to which libraries contribute their holdings data and licensing terms and copyright owners register their content and pricing. The system, which can be embedded into learning management systems and MOOCs, also includes royalty-free and public domain content. Faculty members can search within SIPX for content and embed links to the content on a syllabus. Students who click on the links will access the content at no charge if the item is a library-licensed resource. If not, the student can pay a required fee to access the content. SIPX offers a single, seamless user experience and might be able to help faculty to identify no-cost or low-cost options for MOOC content.

Feather concluded her talk by reiterating that consortial licensing remains the most effective way to press licensors to consider new issues and ideas, and that libraries should continue to work together to create model licensing language.

Contact info:

Andrée J. Rathemacher

Professor
Head, Acquisitions
University Libraries, University of Rhode Island
15 Lippitt Road
Kingston, RI 02881-2011
Phone: (401) 874-5096
Fax: (401) 874-4588
E-mail: andree@uri.edu
http://www.uri.edu/library/