Digital Rights Management and Digital Libaries – A User's Perspective

Christopher Ashley
Glenda Morgan

This paper addresses a significant aspect of digital rights management (DRM): the unique challenges of building a working DRM infrastructure for networked digital libraries or repositories.[1] We outline how, based on our experience with building, populating and operating a digital library, many of the existing problems for the current generation of DRM technology are magnified in a distributed, networked environment. In particular, the evolution of standards in the DRM industry, the questionable adaptability of DRM technology to consumer preferences, and daunting technical issues such as authentication and rights specification are all made more difficult when content is distributed through a network, especially when the content is broken down into smaller pieces to increase access and utility. Moreover, we believe that, to the extent one or more rights specification standards appear to be gaining a foothold in some market sectors, they may not adequately address the needs of the educational community that will be a significant user set for digital libraries.

We will use the Academic Advanced ADL Co-Lab, a joint project of the University of Wisconsin System and the Wisconsin Technical College System, for illustrative purposes throughout this article. One of the main projects of the Co-Lab is a prototype repository with features that make it a suitable case study for many of the issues discussed in this paper. In particular, the repository operated by the Co-Lab features:

- A wide array of content in various formats from a range of contributors including video assets provided by Wisconsin Public Television, publications provided by various commercial publishers, and instructional content provided by faculty and others;

- A state-of-the-art digital asset management[2] system provided by Artesia Technologies (www.artesia.com).

- A large but bounded set of users from several academic institutions using the content in diverse ways.

I. Background

[1] For ease of discussion we use the term "digital library" interchangeably with "content repository" although we recognize that, in some circles, the terms are not considered equivalent.

[2] Digital asset management (DAM) is distinct from digital rights management (DRM). Very broadly speaking, DAM refers to systems that allow digital indexing, storage and retrieval of media assets within a repository. DRM focuses on the distribution of these assets and enforcement of rights attached to them. Most DAM systems have DRM capability in that they allow users to associate rights information with content through metatags.

As in industry, libraries hold an ever-increasing portion of their assets in digital form. Whether it is new content such as on-line journals that are purchased that way, or analog content – videotapes, microfilm, etc. – that is converted to digital form, libraries are implementing digital asset management systems to sort, maintain and distribute their assets. In higher education, faculty and curriculum designers are increasingly drawing upon digital libraries, repositories or "referatories" to find content for course development for on-line delivery or for enhancement of face-to-face instruction. In practice, higher education has focused on three different levels of content: electronic textbooks, or textbook supplements; online courses or parts of courses; and small chunks of digital content known as learning objects or shareable content objects.

A. Electronic Books

The market for electronic books has been slow to develop but is starting to expand rapidly and diversify. Much of the early focus was on electronic material that came packaged with a conventional text in the form of a CD-ROM or a website accessed via a password purchased with the text. These sorts of electronic texts produced their own kinds of challenges, for, among others, libraries, used textbook dealers and textbook rental systems (both the CD-ROMs and the websites essentially ruled out resale and rental of textbooks, at least as far as the additional digital material was concerned, though departments at a number of UW institutions have had some success at negotiating around some restrictions). Slowly textbook publishers are starting to produce fully online texts accessible through readers such as Microsoft's and Adobe eBook reader. Here too there are real obstacles to rental and resale, but also questions about flexibility and usage. Many of the new electronic texts are designed to be used in conjunction with a course management system such as Blackboard and WebCT[3]. However, publishers such as Pearson have shown some sensitivity to the way that academics and students use texts, and allow cutting, pasting and the extraction of particular chapters. Nevertheless, in the context of digital libraries, electronic books that depend on a particular reader or viewer are problematic, especially if the library is designed to be used by several institutions which may have different course management platforms.

B. Modular On-Line Courses and Learning Objects

Increasingly, digital libraries are being used to store and index modular components of content for use by faculty and curriculum designers in creating on-line courses.[4] These modular components can be either "learning objects" that are discrete activity-based chunks (often thought of in terms of Lego™ blocks or atoms) linked to specific course objectives or separate content objects that are not linked to specific objectives. The theory is that using shareable learning and content objects would greatly increase the efficiency and cost-effectiveness of creating, maintaining and updating internet-based instruction. Rather than creating each new course from scratch, an

[3] For example http://www.webct.com/content.

[4] For example, Academic Systems (http://www.academic.com/) and McGraw-Hill's Aleks (http://www.mhhe.com/math/devmath/aleks/) are two products that provide online course materials or modules in remedial mathematics.

instructor can incorporate learning and content objects from a range of different sources. Although there may be situations in which content is inextricably bound to its context and therefore cannot be made into shareable learning or content objects, the early returns suggest quite strongly that in many areas they will be a valuable tool for faculty and curriculum designers.

A number of projects have been developed to provide access to learning objects in higher education. Some of these are collaborative and more open-source in nature,[5] while others are for-profit[6] or are closed. The open and collaborative projects generally speaking do not address copyright issues in any detail. The materials in the repositories or referatories are contributed by the faculty that created them, or by others. Where copyright is discussed at all, it usually consists of one of the metadata fields and specifies either that the material is covered by copyright or not. Occasionally there is more detail, for example a description of who the copyright holder and under what kinds of conditions the material may be used, or a hyperlink to a site that describes the terms of use. This allows great flexibility in the use of the learning objects. Faculty use these materials in their teaching, in classrooms, on web sites and by either incorporating the objects into course management systems (such as WebCT and Blackboard) or by linking out to the objects.

However, significant challenges face for-profit repositories or repositories that try o track (and control) the copyrights in learning objects. First, there is an assumption that needs to be made that the material placed in the repository does not contain any copyrighted aspects or that those have been cleared. Second, the system used needs to ensure maximum flexibility. The goal in the use of learning or content objects is to provide content at as a low a level of granularity as possible. The hope is that faculty will use large numbers of different learning objects in their teaching and that content can be reused by different faculty and in different contexts. Thus it is likely that a repository will contain learning objects with a wide range of copyright restrictions and permissions. The repository needs to be structured to allow these different kinds of uses in a relatively simple way. If there are charges attached to objects for usage, these charges need to be low and there needs to be some effective way of collecting them. Many models currently being designed rely on student payment of these fees, but this does not seem viable as the prices drop and as the number of objects used increases. The structure of the repository also needs to allow for browsing (at least on the part of the faculty member) as well as fair use.

[5] For example MERLOT (http://www.merlot.org/), SMETE.org (http://www.smete.org/), Digital Library for Earth Systems Education (http://www.dlese.org), iLumina (www.iLumina-dlib.org).
[6] For example, Lydia http://www.lydialearn.com. Interestingly, recent attempts to reach the Lydia project site on the web have been unsuccessful.

II. Hurdles to Application of DRM Technology to Digital Libraries[7]

The challenge of balancing security and business needs of content providers with consumer needs is especially acute in the context of digital libraries for which researchers and educators are the primary users. While DRM technology may someday be part of the solution, to a large extent, DRM technology is not yet ready for "prime-time" deployment in the digital library context. There are significant technical hurdles that the current generation of DRM needs to overcome to gain traction in any market. This would include development of user interfaces that are sufficiently adaptable to consumer preferences to survive in the market. In this regard, the DRM industry is well-advised to consider the experience of the software industry in marketing copy-protected software.

DRM also needs to evolve to a point where it can ensure content providers that content distributed through digital libraries is sufficiently secure, and it must do so in a way that avoids reliance on end-to-end use of a single platform. Where digital libraries are shared among numerous institutions, this emphasizes the importance of robust authentication/authorization structures that work across institutions and for diverse categories of users. The task is further complicated by the necessity of accommodating the research and educational needs of the user community, in particular fair use rights, and the corresponding need to protect the users' privacy. Although there has been considerable progress in these areas, a robust system has yet to be implemented.

Finally, one must expect that the use and transformation of content distributed through digital libraries will be richer and the corresponding rights issues more complex compared to usage scenarios that contemplate only reading, downloading and copying digital content. While metadata tagging and rights specification languages will continue to adapt to different uses of content and while XrML, ODRL and other XML-based rights expression languages may someday meet these challenges, using digital content in creating and delivering on-line instruction raises the bar significantly in terms of how usage scenarios will need to be modeled and rights expressed.

A. Adapting to Consumer Preferences

New efforts at digital rights management come against a context where efforts at copy protection or access control have been tried with varying degrees of success.

---

[7] In discussing the uses and limitations of DRM, it is important to remember that the "digital management of rights" aspect of DRM – mechanisms for securing content and enforcing the usage rules applied to it – is not equivalent to DRM's "management of digital rights" aspect – i.e., identifying and expressing the rights attached to the content [E.g., *Digital Rights*, Commission of the European Communities Staff Working Paper (February 14, 2002) available at www.europa.eu.int/information_society/newsroom/documents/drm_workingdoc.pdf]. In some respects, the "management of digital rights" conception of DRM is closer to the related concept of digital asset management (DAM), tools for indexing and storing media assets and describing and associating usage rights with them. [E.g., *What is Digital Asset Management*, Artesia White Paper, available at www.artesia.com/what_dam.html (last visited 4/8/02)]. This aspect of DRM is pure contract management – a mechanism for authenticated users to discover and review the rights attached to content and procure permission based on the underlying license.

Among the most well known efforts at restricting unauthorized use or copying was that associated with the Lotus 1-2-3 spreadsheet program. Originally products used key discs – discs with errors purposefully written into them and the application software used the error codes as a signal to start the application. Things got more sophisticated with high-end applications like drafting programs from Autocad and administrative systems from SAP, which started to rely increasingly on dongles. These are devices that attach to one of the computer ports, usually a parallel port and the application communicates with it.

These forms of protection were very unpopular with consumers. Key discs were lost and dongle solutions were notoriously unreliable. The claim is frequently made that copy protection schemes make life very difficult for the average, legitimate user and play little role in deterring piracy (for example Schneier 2000). That may be so, but with these early efforts at copy or use-protecting mainstream applications, methods of evading the protection became widely available and appear to have been extensively used including by those with legal copies of the programs who merely wanted backups or a more reliable system. A small cottage industry grew up providing work-arounds for these copy protection methods. For example, in the days of key disc, Central Point Software came out with a utility called Copy2PC that allowed users to make backups of their software. As software distributors became more sophisticated, so did the solutions. The makers of Copy2PC came out with the Copy2PC ISA board which was installed inside the cpu in such a way that all the data traveling between floppy drive and floppy connector went through it, making copies along the way (Sullivan 2002). Alternate dongles became available for products like AutoCAD. Customer dissatisfaction ultimately caused most software manufacturers to stop using copy protection on all but the most high-end applications. By 1987 Lotus had abandoned their key diskettes.[8]

It remains to be seen whether consumers will tolerate mechanisms that regulate their access and use of digital content. What is clear now is that DRM measures have yet to achieve the transparency and efficiency that consumers will likely demand in the marketplace.

B. Securing Content in a Distributed Environment

A further challenge to incorporating DRM solutions into digital libraries is that they may rely on a single platform end-to-end and thus do not work well in a distributed

[8] One notable exception is the computer gaming world where copy protection was never really abandoned and remains strong. One of the most common forms of protection used is simply requiring the game CD to be in the CD-Rom drive in order for the game to be playable. SafeDisc from Macrovision manufactures a popular form of this. Similarly Sony playstation copy protects its discs mostly using invalid checksums (Wallach 2001). These have caused complaints on the part of consumers, not only because of the inconvenience but also because it means that CDs frequently don't work on certain CD drives (see Kay 2000). However, despite quite considerable customer dissatisfaction, the game industry shows no signs of retreating from its use of copy protection schemes.

environment.[9] The major purpose for creating repositories is to share content. In order for this to occur, inter-operability needs to be maximized. The higher education and training communities have begun vigorously to support technical standards (such as the IMS, IEEE and SCORM) that will ultimately serve to maximize this inter-operability and the repositories and referatories described above are being built around these. The end-to-end structure of some current DRM systems, while possibly secure, would conflict with the objectives of any digital library seeking to provide content to users on a platform-neutral basis, such as the Academic ADL Co-Lab.

To get around this problem, the next generation of DRM solutions will have to look at other ways of securing and providing access to digital content. Considerable work is being done in establishing protocols for authenticating users across institutions, such as is necessary with projects like the Academic ADL Co-Lab. There are several ongoing projects that may produce useful results in this area. One is the National Science Foundation – supported National Middleware Initiative (NMI) which is focusing on methods for inter-institutional sharing of web resources subject to access controls. The related Shibboleth project also involves methods for authenticating users and authorizing use across several institutions that may be part of a digital library project (middleware.internet2.edu/shibboleth/shibboleth-project.html). What is particularly encouraging about these projects is their emphasis on building user privacy protection into secure authentication systems. As Burk and others have noted, a user's ability to access content without compromising his or her anonymity is a necessary element of any fair use infrastructure for digital content. [Burk & Cohen 2001; Feigenbaum, Freedman et al 2001).

C. Rights Expression Languages

Finally, the task of describing rights and modeling usage for content distributed through digital libraries presents is more complex that in the kind of one-on-one transaction that currently predominates in the market. This is especially true of usage scenarios that contemplate content passing through a workflow, combining and being incorporated into other content objects that could be either derivative works of the original or new works. While rights expression languages have gained considerably in their ability to model and describe increasingly complex transactions, it remains to be seen whether the current XML-based languages are sufficiently flexible and robust to work in a distributed environment.

XrML or Extensible Rights markup Language is emerging as the dominant form of digital rights language (although there are doubts about the extent to which it can claim to be a standards based solution, see McAllister 2002). Originally developed at PARC conceptually as Mark Stefik's "trusted system," later as DPRL, and finally as XrML in 1999. The most recent version, Specification 2 was released toward the end of 2001. In developing XrML the goal is develop a language that is at the same time

[9] For example, Microsoft's Windows Media Rights Manager™ v. 7.1 appears to require Windows™ end-to-end on both the provider and customer sides. [Neil McAllister, *Freedom of Expression: Emerging Standards in Digital Rights Management*, New Architect (March 2002) (www.newarchitectmag.com)].

comprehensive (ie able to describe both simple and complex multi-layered ownership rights) and precise (able to describe exactly the ownership structures and rights to all the participants in the System). XrML claims to be business model neutral. It is intended to be structured in such a way that, as a language, it can describe any combination of rights and usages for any kind of digital content in any kind of medium. There is certainly flexibility within XrML in that there are multiple levels to which rights can be assigned for the same piece of digital content depending on whether one is, for example an author, publisher, distributor or end consumer. Other features such as pattern matching allow rights to be assigned to classes of users (or even devices), rather than specific individuals

XrML works on the basis of rights that are assigned to chunks of digital content. The rights cover actions such as copying, printing reading, embedding and making backups.[10] XrML also specifies the conditions under which the specified rights can be exercised, for example a particular period of time. The core of XrML consists of a license which grants a principle the rights to a particular piece of content.

Despite the apparent flexibility of XrML, there remains some concern whether it will sufficiently accommodate the needs of the research and education communities. Although XrML and another commercially-developed rights language XMCL have apparently been submitted to an international standards body, they were developed in a commercial context by a single vendor with its primary focus on deployment in a commercial context [See Neil McAllister, supra]. Further, there are questions about the ability of the current generation of rights language to model more complex usage of digital content. For example XrML's RIGHTSLIST element includes many different forms of the "rendering" of content objects. However, it is not clear how the RIGHTSLIST would model re-uses of content objects such as the creation of transformative uses and aggregate and derivative works. There are certainly also many questions to be asked about the extent to which rights languages like XrML can model academic uses and needs such as fair use. For example, one of the primary goals of XrML is to be as precise as possible and it is hard to see how this drive to precision can accommodate inherently fuzzy, contextual imprecise concepts such as fair use.

D. Licensing and Business Rules

While waiting for standards and consensus to emerge in the DRM field, digital libraries continue to actively license content for research and educational use. The landscape is beginning to change in this arena because of the emergence of DRM technology. Libraries often are able to negotiate broad access to digital content that protects uses of the content for educational, non-commercial uses, consistent with fair use principles. Rights management in this context is primarily an issue of capturing and expressing the rights associated with a particular piece of content and authenticating users

[10] The full list of rights includes accessing the folder information; backup; copy; delete; edit; embed; execute; export; extract; install; loan; manage the folder; play; print; read; restore; transfer; uninstall; verify; write; issue; obtain; posses property and revoke.

who are confirmed to have permission granted under the license. The description and association of rights with content in a repository is a necessary function to be served by DRM technology, leaving aside, as described above, any possible inherent biases or values in the rights description language employed.

It is when DRM measures associated with content go beyond describing and associating rights with content to actual enforcement of usage rules through technical means that raises more significant barriers for digital libraries and other consumers. However, the current uncertainty in the DRM market and the resulting hesitancy of content providers to adopt particular DRM solutions for rights enforcement and transactions in digital content has, to our knowledge, insulated digital libraries from application of DRM enforcement mechanisms in negotiating acquisition of content from publishers and other providers. At this time, it may be possible for digital libraries with sufficient market influence to limit the kind of DRM enforcement mechanisms that publishers might seek to apply to content acquired by the library. They can do this because content for which libraries represent a significant part of the market allows the library room for meaningful negotiation of license terms with the content providers (Lynch 2001). This is largely what has happened during the pilot phase of the Academic ADL Co-Lab project. Publishers in the pilot project have granted broad access and use rights to the content contributed including the rights to use, copy, host, distribute, transmit, sublicense, perform and to make derivative works from the materials. The provider agreement also includes a representation and warranty by the contributor that the content does not include digital rights management mechanisms or devices that could disable or restrict access to the content in a manner inconsistent with the agreement.

Of course, publishers and others who provide content to a digital library may not always have full control of DRM measures applied to the content, especially if the content is multimedia or other kind of composite work. Although we are not aware of situations in which this has happened yet, a provider using a particular asset management system that incorporates a particular DRM solution, could be subject to restrictions on how the content is packaged or used by downstream users. [See Bechtold nd). Content providers might not even know if DRM “hooks” have been applied to content they provide to a repository since such hooks could be very hard to detect. Digital libraries might consider covering this situation by obtaining a separate warranty from the contract provider to take responsibility for fixing any content that is later found to have a DRM mechanism that interferes with appropriate use under the license.

In addition to adapting licensing practices to the current state of DRM technology, we have found that content providers, in determining how to market and distribute content through digital libraries, will likely have to revisit traditional business models, especially to the extent the content is meant to be combined into composite or aggregated works by faculty or curriculum designers. For example, the common practice of providing a free sample or preview of a larger work that the user must pay for is not compatible with an economy of content objects distributed through digital libraries. A course designer may not want the complete work, only the preview, so that it can be combined with other objects to make a course. If a content provider seeks to derive

revenue from digital objects, it must determine the price rules that apply to the objects themselves at a sufficient level of granularity.

Content owners may need to consider relinquishing control of usage rights of some content once it is aggregated into higher-level digital objects. For example, an individual publisher who provides a single book chapter or a single media clip to a digital library cannot control the user's ability to combine it with content from other publishers. Once such an object is incorporated into aggregation, the provider of the individual content objects would no longer control the use of the object as part of the aggregate, although the provider might receive compensation when the aggregate is used or sold. The aggregation itself could be identified and registered in the digital library as a separate content object. Digital libraries will need to develop systems for tracking and storing objects in this way, and an e-commerce infrastructure for transactions in these content objects.

References

Bechtold, Stefan *From Copyright to Information Law – Implications of Digital Rights Management* (www.star-lab.com/sander/spdrm/papers/bechtold.pdf) (last visited 4/4/02)].

Burk, Dan & Julie Cohen, 2001 *Fair Use Infrastructure for Rights Management Systems*, 15 Harvard Journal of Law and Technology 41 (Fall 2001).

Feigenbaum, Joan, Freedman, Michael J., Sander, Tomas & Shostack, Adam, 2001 *Privacy Engineering for DRM Systems*. 2001 ACM Workshop on Security and Privacy in Digital Rights Management.

Kay, Russell 2000 *Copy Protection: Just Say No*. Computerworld September 04, 2000. http://www.computerworld.com/cwi/story/0,1199,NAV47-68-84-91_STO49358,00.html (accessed April 8, 2002).

Lynch, Clifford 2001 *The Battle to Define the Future of the Book in the Digital World*, 6 First Monday No. 6 (June 2001) (http://firstmonday.org/issues/issue6_6/index.html).

McAllister, Neil 2002 *Freedom of Expression: Emerging Standards in Digital Rights Management*, New Architect (March 2002) (www.newarchitectmag.com).

Schneier, Bruce, 2000 *The Fallacy of Trusted Client Software Controlling*. Cryptorhythms Column, August 2000. http://www.infosecuritymag.com/articles/august00/columns2_cryptorhythms.shtml (accessed 8 April 2002).

Sullivan, Paul, 2002 *Have Consumers Gone Spineless? Part 2*. Firing Squad: Hardware and Games News and Reviews, January 25, 2002. http://firingsquad.gamers.com/features/spineless2/ (accessed April 10, 2002).

Wallach, Dan S. 2001 *Copy protection technology is doomed*. IEEE Computer, 34, (10), October pp. 48-49. http://www.cs.rice.edu/~dwallach/pub/copy-protection.pdf