Libraries and Digital Property Rights

Mark Stefik and Giuliana Lavendel
Xerox Palo Alto Research Center
3333 Coyote Hill Road
Palo Alto, California 94304

Abstract. The realization of the digital library — a computer system to enable anyone with a workstation to have access to any of the published works of mankind — has stayed out of reach because of a presumed technical problem. Once a written work is digitized, it becomes so easy to make and distribute copyright infringing copies that publishers would go out of business. A technical solution to this problem based on trusted systems and digital property rights is now becoming available. The big issues for libraries — social and institutional policy challenges — are still ahead.

At least since Vannevar Bush's proposal for a Memex system in 1948 or J.C.R. Licklider's Library of the Future written in 1965, people have dreamed of augmenting libraries as we know them with computers, to create digital libraries. In a digital library, all the published works of mankind would be available on-line, accessible any time day or night to anyone with a computer workstation— as in Ted Nelson's Xanadu dream [Nelson 1988].

Despite many advances in computer technology over the last two decades, obstacles to creating comprehensive digital libraries have kept such developments well beyond reach. Several waves of technology — including videotex, CD-ROMs and now the Internet and World Wide Web have come along with high expectations. Each technology promised to transform the delivery of information and thus to change how libraries work for the better. However, even in those situations where computers and automation have proven useful, they have not provided the basis for operational digital libraries. There has been the ongoing challenge, for librarians mindful of time and budget constraints, to introduce new and imperfect technology into "legacy systems" — including aging computers, old media and formats, and established work practices.

The main issue standing in the way of digital libraries becoming a reality is the copyright problem. According to conventional wisdom, once a work has been put into a computer, it is very easy to make copies of it and to put those copies on diskettes or to transmit those copies anywhere on computer networks [Barlow, 1994]. The copies can be copied again so that, in principle, almost no one need buy a copy of a digital work again. You could get a copy of an interesting work from a friend or check it out from a library at no cost, at least in the United States. The potential for such copying and distribution to infringe copyright and eliminate income for publishers has been a substantial barrier to digital publishing, limiting it to special niche areas.

A technical solution to the copying problem, however, is now at hand and is becoming available from several vendors including Folio, IBM, Intertrust, NetRights, Xerox, and Wave Systems [Kahin and Arms, 1996]. This paper describes the technical solution briefly, and also prospects some of the coming policy and institutional challenges to libraries. These challenges potentially affect not only the operation of

lending libraries, but also the roles of libraries and the relationships with publishers, distributors, and the public in general.

1 Trusted Systems and Digital Rights

There are two main concepts behind the emerging approach to commercial digital publishing: trusted systems and digital property rights.

The term "trusted system" was originally used to describe computer systems for military and national security applications that could securely hold classified and secret information [Department of Defense, 1985]. More recently the term has been generalized to describe systems for electronic commerce and digital publishing applications [Stefik, 1997]. A trusted system can securely hold a digital work and can be trusted to enforce the rules governing its use.

How can a trusted system know what the rules are? Before publishers make a digital work available, they describe terms and conditions governing its use in a digital record or "contract" that accompanies the work itself. A digital contract specifies rules governing displaying or rendering of the work, copying of the work, and using portions of the same in further derivative works. It also specifies the time period over which the contract is offered, any restrictions on qualifications for using the work, and any fees associated with its use. The terms and conditions are expressed in a formal rights management language like the newly released Xerox's DPRL or digital property rights language [Stefik, 1997a]. Contracts written in such languages are intended to be clear, well-defined, and understandable to all parties; they allow signaling of agreement and acceptance by a digital signing process, and are potentially enforceable by trusted systems.

1.1 Levels of Trust

Much of the technical work surrounding trusted systems and digital publishing revolves around making computer systems secure enough to be trusted by the various stakeholders, including users of the systems, publishers of digital works, and financial clearing houses which handle the billing for digital commerce.

Trust depends on integrity. The integrity of a computer system rests on three technical foundations: physical integrity, behavioral integrity, and communications integrity.

Physical integrity refers to the capability of a trusted system to resist invasive physical attack, such as attempting to remove disks or to get information out of computer memories without authorization. Behavioral integrity refers to the correctness of the computer programs, which are expected to comply with established standards in enforcing the terms and conditions of digital contracts. Communications integrity refers to the capability of the trusted system to send and receive communications involving digital works, without being manipulated or compromised by other (possibly untrusted) computer systems.

In general, there is a tradeoff between having high levels of trust and having wide availability. The most sophisticated trusted systems are instrumented to detect tampering; they have special hardware and software to provide the highest practical integrity, while the simplest trusted systems just add special software to existing personal computers and workstations. Systems at intermediate levels of trust augment software with special hardware for such functions as holding encryption keys and billing data. The expense of systems with the highest security levels precludes their wide distribution, thereby making them uninteresting as a marketing and publishing

platform for most publishers. At the other extreme, trusted systems at the lowest cost offer very little security for holding digital works, creating greater risks that the digital works they are supposed to protect will be copied without compensating the rights owners. Therefore the marketplace which is evolving for trusted systems is likely to be stratified, so that the most expensive digital works will require higher levels of security than the least expensive ones. In general, the price/security-performance characteristics of trusted systems are expected to improve over the next few years, making systems of higher trust levels more widely available.

1.2 Digital Rights

Digital rights are operations that can be carried out on trusted systems under specified terms and conditions. From a system design perspective, exercising a digital right corresponds to carrying out a transaction. The transaction involves checking various conditions such as whether the right has expired or whether the parties possess particular valid digital certificates; it entails encrypting content for secure communications and storage, and reconciling bills with a financial clearing house.

There are three main categories of rights governing the use of digital works: transport rights, rendering rights, and derivative works rights.

Transport Rights

Transport rights govern the digital communication of a digital work from one trusted system to another. Three main transport rights are distinguished: the right to copy, the right to transfer, and the right to loan. These rights correspond to familiar operations performed on physical works, such as buying a book, giving it away to a friend, or loaning it to a friend for a period of time.

The right to copy creates a new usable copy of a digital work [Stefik, 1997b]. This process is invoked when someone purchases a copy of a digital work. In a simple scenario, someone selects a digital work from an on-line distributor or bookstore, where certain conditions might need to be satisfied for some particular right -- for example, there may be a discount for members of a book club, or restrictions regarding the age of the purchaser. If the conditions are met, then an encrypted copy of the work can be sent to the purchaser's computer and an invoice for the purchase price can be directed to a financial clearing house and destined for the user's credit card.

There are several interesting variations on copy transactions. At one extreme the usage right conditions can specify that the seller of the digital work must be a certified distributor. At the other extreme, the right can specify that anyone with a copy of the work can make another copy. When consumers sell copies of works to others, the process is called consumer-based distribution or superdistribution.

For printed works such as books and magazines, it is much more common for a person to give a previously-purchased copy to a friend rather than to create a new copy. In U.S. copyright law the first sale doctrine grants this right to consumers. For digital works, the analogous operation is called a transfer right. A transfer transaction mimics the operation of handing someone a book; after the transfer, the recipient has the book and the original owner no longer does. From the perspective of computer operations, this is similar to transferring money from a savings account to a checking account at an automated teller machine. At the end of the transaction, the checking account has the money and the savings account no longer does.

The third category of transport rights is a loan right. This right may prove fundamental to digital libraries because it allows digital works to be checked out from a library for a period of time. This right permits a work to be copied to a second repository, together with information about the rights that can be exercised and the duration of the loan. During the time while the work is checked out, the work cannot be used at the library. The trusted systems at the library and client both have clocks.; at the end of the loan period, the digital work on the client's trusted system is deactivated and the digital work at the library is reactivated -- amounting to a sort of "auto-return." There are other interesting variations on this right that we will consider later.

Render Rights

Render rights govern the creation of digital work representations beyond the control of trusted systems. According to the dictionary, to render something is to represent it in a verbal form or in a drawing or a painting. In the context of trusted systems, we use the word render to refer to processes that transform a digital copy of a work into a form where it can be perceived and used. For example, a digital book or digital photograph may be presented on a computer display, a piece of music may be played through a loud speaker, a movie may be rendered in sound and video on a computer with speakers, or a page of a book may be printed on a printer.

When a digital work is rendered, information moves outside of the control of a trusted system, which is designed to protect against the uncontrolled making of perfect digital copies. Presentation of a work as image or sound makes it possible for someone to see or hear the work. Of necessity it also makes the work accessible for analog copying; you can record anything you can hear, as you can photograph anything you can see. Many techniques for rendering from digital originals allow for the embedding of tracing information, making it possible later to identify the work from a recording, and also the person who rendered it. This information may be hidden and distributed in the rendering using spread-spectrum techniques. The marking is called "watermarking," in an analogy with the translucent marks put on paper to certify its source. For many publishers, the goal of security and watermarking is to help honest people stay honest. By combining automatic billing with potential traceability of infringers, trusted systems raise the bar high enough to reduce the expected risk substantially for many publishers.

There are three kinds of render rights: play rights, print rights, and export rights. Play rights refer to ways for making a transient or ephemeral copy of a work available for use, such as displaying a page of a book on a computer screen. Print rights refer to making more permanent rendered copies of a work, such as printing hardcopies on a laser printer. An export right makes a digital source copy of a digital work outside of trusted system control, such as making a clear (unencrypted) copy of an older work on a diskette after it has passed out of copyright.

Derivative Work Rights

Derivative work rights govern the reuse of a digital work, in whole or part, to create a new composite work. There are three kinds of derivative work rights: extract rights, edit rights, and embed rights.

The extract right allows removing a portion of a digital work to create a new work. The edit right grants a limited power to modify a work. The embed right allows to include a work as part of a larger composite piece. These rights form the axes of control for repackaging digital works. For example, a course instructor in a business school could collect together electronic business case studies into a digital reader or

coursepack. Whenever a student purchases or prints the coursepack, all of the contributors would be compensated with fees as assigned.

Other rights are also possible. For example, Xerox's DPRL language includes several additional rights, such as rights for controlled making and restoration of encrypted backup copies and file management rights.

2 Library Loans of Digital Works

The relationship between libraries and publishers is symbiotic. From a publisher's perspective libraries increase publishers' income by buying books and decrease it by loaning books to the public. Loaning books arguably undermines publisher profits when people, who would otherwise buy a book, get it from a library instead. However, the unsharable nature of the book as a physical object limits the size of the loss, since no more than one person can use a book at a time.

The loan right for digital works mimics the serial reusability of physical documents. Implementing the loan right on trusted systems opens the door to public access in libraries for many digital works. Modeled on physical loans, the digital loan right can affect the operation of a library, without changing the relationship between libraries and publishers. We also consider digital variations on loans that alter these relationships.

2.1 Loans with Physical Affordances

In most public libraries in the United States, no fee is imposed for borrowing a book. From a library policy perspective, maintaining this approach for digital works is the most straightforward but not the only possible course. Just as it is possible for the fees and conditions on a digital work to assign a fee to copy a digital work, it is possible to assign a fee to loan a work. The revenues for a loan transaction can be assigned in various ways. The revenues could go to the library, to the publisher, or in principle, to third parties.

From a publisher's perspective, assignment of revenues to the publisher could be used to compensate for losses due to leakage from imperfect trusted systems, or losses from a model of lost sales, if the usage and buying patterns for some classes of digital works are different from that of (e.g.) books, CDs, videogames, or videotapes that can currently be borrowed from many American libraries.

From a library's perspective, loan fees could be a revenue source for a library. For example, there could be two queues for borrowing works. Patrons in a hurry could check works out of a shorter for-fee queue, while economy-minded patrons could wait in a longer for-free line. Revenues from for-fee loaning could be used not only to buy more copies for the for-fee queue, but could also subsidize the for-free line. In this way, the revenues could benefit all.

2.2 Loans with Digital Affordances

Another perspective is that the digital loan right modeled on physical loaning is unnecessarily restrictive. Why carry the limitations of the physical world over into the digital world? Are there ways to serve the public better without undermining the interests of publishers?

For printed works. the purchase of copies has implications on library resources and patron usage that go beyond the affordance of serial reusability. Physical books must be manually inventoried, checked for damage, and shelved. Each copy of a physical

book requires additional shelf space. In contrast, for digital works, a thousand copies requires little more space than a single copy. Copies can be automatically inventoried and copies can be made essentially invulnerable to physical wear and tear.

Consider the situation where a new work becomes available and there is a great demand for it at the library. For physical works, the library must determine how many copies to purchase, balancing the expense of the purchase and the ongoing cost of shelf space even for books where the public interest may either be sustained for years or may die out quickly in a few weeks. The straightforward approach to loaning digital works forces librarians to make the same inflexible commitments that they would make for physical works.

There are other possibilities. Consider, for example, the approach of "metered use" or rental on digital copies. The library could act as an unlimited source of digital copies for its patrons, with the proviso that it pay the publisher a fee dependent on the number of copies in use. Consider the case where 10 copies are in use the first week, 7 copies the second week, 5 copies the third week, and 3 copies the fourth week. One way to assess the usage is that there were 25 usage-weeks in that month. To compensate the publisher, the library could pay an amount for the digital work that depends on this usage. To keep the library rental budget predictable, the library could arrange matters so that it buys (say) 3 copies for free public use and charges a fee for usage beyond that amount. The accounting for this approach would be carried out automatically by the library checkout system.

Another variation of the loan transaction has the library acting (possibly unknowingly) as a distribution channel for the publisher. During the time when a work is checked out, copies of the work exist on the trusted system of both the library and the borrower. When the checkout time ends, the borrower's copy must be deactivated but it need not be deleted. If willing, the borrower could pay a fee to own a copy of the work. This is similar to consumer-based distribution in that it provides a new sales channel for the publisher. In this example, the "bookstore" is nowhere in sight, and the library effectively acts as a distributor. Since the loan transaction need not involve the library's trusted system, the library need not even know that a copy of the work was ultimately purchased. Alternatively, the library could get a percentage of the sale price of the work.

3 Policies and Institutional Challenges

With advances in technology, the balance of power between authors, publishers and consumers has shifted back and forth. When computers first became available and were considered as a platform for delivery and distribution of works, publishers saw them as shifting the balance too far away for comfort. Compared with the effort of copying a book, even with a modern photocopier, copying a digital work is trivial. With a single keystroke someone can copy a digital work that represents a lifetime's work for an author, and that work can be sent to thousands or millions of consumers easily, without compensation. This situation has been far too dangerous for publishers to risk their intellectual property assets.

With trusted systems, publishers can name the usage terms and conditions for the works they publish. They can potentially monitor the use of the works to a fine grain of detail. Since the use of the work is governed by terms and conditions, publishers can base the rights on contract law rather than just copyright law, potentially at the expense of the public interest. Although this oversimplifies the situation, the argument is that trusted systems shift too much power back to authors and publishers.

It is the nature of libraries that they are positioned in the middle, between those who create works and those who use them. To use a metaphor currently popular among librarians, they are flying the airplane at the same time they are building it [Imperato, 1994]. Being in the middle, libraries are expected to understand and advise on the policies and social contracts for those on both sides. Since the technology of trusted systems and digital rights is just coming into use, the policy issues they raise are just now coming into focus

3.1 Fair Use

Fair use is a concept in US copyright law that governs the use of published works. Technically, fair use is a kind of legal defense, stating that the material has been used for personal, research, or non-commercial aims. A publisher may challenge the use as being an infringement; and the user can defend the use as fair. When there is a dispute on whether the use of a work has been fair, adherents of the two positions line up their forces for a legal resolution. A court then decides.

An argument against the trusted system approach is that it precludes exercising a fair use defense. A trusted system just follows the rules in enforcing a digital contract. If you wish to extract a portion of a work for a derived work and claim that this use is "fair," you are blocked from doing this if the rules associated with the work do not allow it. In other words, a person is unable to use a work in a potentially unfair way and to then defend that use because the trusted system prevents the use in the first place, even if it could ultimately be judged as fair.

Libraries find themselves in the middle of discussions on fair use, since they offer photocopiers in the library space and make digital works from the Net accessible from library workstations. [Bennett, 1994], [Stanford Fair Use Page]. In an attempt to balance the interests of publishers and consumers, an appropriate body has been proposed for issuing fair use licenses to people who can demonstrate an understanding of the concept [Stefik, 1997b]. Publishers can then include, among the rights concerning digital works, additional rights that can be exercised by anyone with a fair use license. This proposal gives people with licenses potentially broad rights for using a work, such as extracting sections and embedding them in other works, trusted or otherwise. The risk to publishers, given the escalation of the Net, is that a licensed user will distribute a digital work beyond what would be considered fair use, potentially undermining sales or value of the work beyond the user's ability to compensate the publisher. How could this risk be lessened? What should be done if the damage caused by fair-use licensees exceeds their ability to pay for it? One proposal is that the licenses come with insurance. In releasing a work, a publisher declares an insurance value for it. The insurance is paid by the publisher and by a tithe on transactions involving the work. In the event that the legal system assesses culpability and damage, insurance will compensates the rights owners as needed.

The institutional arrangements for fair use licensing and insurance are projected in the future. Librarians will be among the experts who are consulted on such issues, as trusted systems become more established in our society.

3.2 Privacy and Archiving

The privacy and archiving issues are too large to be addressed in depth in this paper. However, there are connections with trusted systems that are worth noting here.

Trusted systems can potentially keep billing logs for information use at a fine grain of analysis, by means of billing techniques called "microtransactions," which keep track

of instances that cost pennies or even fractions of pennies. Logging usage information at this grain presents an opportunity to collect very detailed data, potentially violating privacy [Cohen, 1996]; at issue is how usage information is collected and aggregated. Trusted systems, as defined above, are largely designed to protect the interests of owners of digital works. It would be possible to have "trusted billing systems" which are expected to protect the interests of consumers with regard to privacy and fair billing issues. Standards and principles for such systems, however, are yet to be developed.

One of the key roles of libraries is the archiving of information for posterity. A review of recent thinking on digital archiving is available [Research Libraries Group, 1996]. From a librarian's perspective, the making of backup copies is a legitimate operation intended to guard against loss due to media failure. From a publisher's perspective, uncontrolled production of backup copies opens trusted systems to a potential scam. A consumer buys a digital work and makes a backup copy; then uses the trusted system to give away the digital work — claiming that media failure caused loss of the work — and restores the work from a backup file. By repeating this operation, the consumer can have an unlimited source of free copies of the work to sell. This liability can be avoided by making sure that backup and restore are both controlled rights, that is, that the backup operation always makes an encrypted copy., while the restore operation can have conditions, such as requiring the use of a digital ticket, or a payment, or authorization from the publisher. A publisher who sees a pattern of repeated requests to restore a work could become suspicious and start charging a fee.

3.3 New Forms of Documents

Digital documents on the Net differ from printed documents in many ways. As Negroponte of MIT's Media Lab and others have pointed out, the publishing, entertainment, computing, and communications industries are converging. From a document-centric point of view, this leads to documents that are interactive and multi-media; it leads potentially to documents that grow as people add to them and documents that link to other documents on the Net. Today's literacy is not just "reading," it is also viewing, listening and interacting.

When Vannevar Bush described his Memex for the library of the future, he imagined a device that gracefully mixes personal and public information. A Memex user could create "trails" in the library, annotating various documents and linking portions of documents together. Although annotations in books are generally discouraged for printed documents, the creation of linked works envisioned by Vannevar Bush has been a staple of hypertext systems for many years, and has been in wide public use on the Net since the advent of browsers.

The current generations of browsers on the Net are not trusted systems and do not provide rights-enabled access to documents, but it is reasonable to expect that in the next few years, some browsers will be trusted and rights enabled. In the current stage, the Net is a place where a lot of unrefereed documents circulate, accessible mainly for free; there are few controls or organizational principles governing "publication," and the quality is so low that libraries are concerned, but have not tackled archiving issues with their customary thoroughness.

In a second, future stage, we imagine works on trusted systems published on the Net; they are available for a fee and often are superior in quality to free uncontrolled works. They are treated much the same as their printed counterparts , and for the most part priced and packaged separately.

In a third stage, people may want to share private annotations on published works, while digital rights on the works themselves govern how they can be combined in derivative works. This third stage finally goes beyond Vannevar Bush's vision, since he did not explore copyright, or envision trusted systems for rights enforcement.

The possibility of public annotated works raises interesting issues for libraries. As long as the Net contains in the main low quality documents, and documents on the Net are mostly isolated and unconnected, archiving policies can follow familiar criteria, selecting material to be archived among works obtained from traditional sources, and ignoring random and ephemeral information from the Net. However, as documents on the Net become interlinked and annotated, authorship and boundaries of documents from an archiving perspective become more challenging. Should we archive the classic book by Professor XYZ by itself, or should we also archive the influential on-line debate by members of the relevant scientific society? Should we also archive the annotated version by Professor ABC, written three years after Professor XYZ's death, to address some new discoveries? This is one of the many dilemmas which will face the library of the future, viewed both in its collection and its social obligations. There is an interesting parallel with the "annotated" editions of classics the early Renaissance humanists left behind, and are now in some fortunate libraries — several of them in Italy — where posterity treasures them.

4 In Closing

Trusted systems and digital rights will help bring works of value on-line; they make it possible to mimic and extend the physical affordances of printed works, allowing controlled purchases, transfers, loans, rendering, and other operations. Properly designed, these systems may serve to balance the interests of both publishers and consumers [Stefik and Silverman, 1997]. Now becoming available, these technologies will come into wide use over the next few years.

The arrival of works of value in digital form means that libraries will find themselves in an advisory role, as institutional and social challenges around fair use, privacy, and archiving policies arise. Furthermore, as works move on-line, libraries will evolve in ways that reflect digital affordances, rather than the physical limitations of documents that have guided library policies so far. Taken together, these changes will challenge established library policies, leading us to rethink the relationships among libraries, publishers, and the public.

References

Barlow, John Perry. The Economy Of Ideas. *Wired*, March 1994, page 85.

Bennett, Scott. The Copyright Challenge: Strengthening The Public Interest In The Digital Age. *Library Journal* , Volume 119, NO. 19, pages 34-37, Nov 15, 1994.

Bush, Vannevar. As We May Think. *The Atlantic Monthly*, July 1945.

Cohen, Julie E. A Right To Read Anonymously: A Closer Look At Copyright Management In Cyberspace, *Connecticut Law Review*, Volume 28, Number 4, Summer 1996, pp 981-1039.

Department of Defense, Department of Defense Standard: Trusted Computer System Evaluation Criteria. DoD 5200.28-STD (Orange Book), December 26, 1985.

Imperato, Nicholas, and Harari, Owen. *Jumping The Curve; Innovation And Strategic Choice In An Age Of Transition*. Jossey-Bass, 1994.

Kahin, Brian and Arms, Kate (Ed), Forum On Technology-Based Intellectual Property Management: Electronic Commerce for Content., Volume 2, *A Journal of the Interactive Multimedia Association*, August 1996.

Licklider, J.C.R. *Libraries Of The Future*. Cambridge, Mass. MIT Press, 1965.

Nelson, Theodor H. Managing Immense Storage. *Byte*, Volume 13, No.1 , January 1988, pages 225-238.

Waters, Donald, and Garrett, John. Preserving Digital Information: *Report of the Task Force on Archiving Digital Information*, Washington, DC: The Commission on Preservation and Access, 1400 16th Street, NW, Suite 740, May 1996.

Samuelson, Pamela. Regulation Of Technologies To Protect Copyrighted Works. *Communications of the ACM*, Vol. 39, No. 7, pages 17-22; July 1996.

Stanford University Libraries. Copyright And Fair Use Page. <http://fairuse.stanford.edu/>

Stefik, Mark. Letting Loose The Light: Igniting Commerce In Electronic Publication. In Stefik, Mark (Ed.), *Internet Dreams: Archetypes, Myths, and Metaphors*. Cambridge, Mass. The MIT Press, 1996.

Stefik, Mark. Trusted Systems. *Scientific American*. March, 1997a.

Stefik, Mark. Shifting The Possible: How Trusted Systems and Digital Property Rights Challenge Us to Rethink Digital Publishing. *Berkeley Technology Law Journal* 12:1, pages 137-159, 1997b.

Stefik, Mark, and Silverman, Alex. The Bit and the Pendulum: Balancing the Interests of Stakeholders in Digital Publishing. *American Programmer*, (in press), 1997.

Carol Peters   Costantino Thanos (Eds.)

Research and Advanced Technology for Digital Libraries

First European Conference, ECDL'97
Pisa, Italy, September 1-3, 1997
Proceedings

Springer