TECHNICAL PERSPECTIVE

SHIFTING THE POSSIBLE: HOW TRUSTED SYSTEMS AND DIGITAL PROPERTY RIGHTS CHALLENGE US TO RETHINK DIGITAL PUBLISHING

MARK STEFIK[†]

TABLE OF CONTENTS

I. INTRODUCTION

In a widely circulated article about rethinking copyrights,[1] John Perry Barlow said that "everything you know about intellectual property is wrong." Ideas want to be free. Once works are digital, they will become free because anything put in a digital bottle will necessarily leak out. According to Barlow, in the digital medium commercial publishing as we know it is impossible.

Barlow was correct in thinking that we are on our way to a new economy of ideas. He was wrong, however, in thinking that copyright and other forms of author and publisher control over works in digital form are outmoded. Although he recognized some potential for technological protection of works, he greatly underestimated how great the potential was. With the development of trusted system technology and usage rights languages with which to encode the rights associated with copyrighted material, authors and publishers can have more, not less, control over their work. Barlow was in good company among those who did not understand the ideas, doubted them, or did not like them. In 1994 the ideas of digital rights and trusted systems were jolting and unexpected.

The situation today is much changed. First, there has been a technological shift that changes what is possible in property rights and commerce for digital works. Second, our society includes many stakeholders in digital property including authors, publishers, distributors, consumers, librarians, lawmakers, copyright lawyers and standards bodies in different countries. Collectively, society does not understand the shift yet. The ideas challenge common sense about computers and information. They need to marinate. This article aims to demystify the technological shift and to show how it enables new ways of thinking about computers, information, and digital publishing.

This is an interesting time to rethink digital publishing. The technology companies are, for the most part, still pre-competitive. The active digital publishers are early adopters, to follow Geoffrey Moore's term from *Crossing the Chasm*[2] and *Inside the Tornado*.[3]
Other people articulating positions include librarians and others interested in issues of copyright and fair use. People are trying to sort out rights and commerce associated with digital works and the picture is getting clearer.

II. BASIC TECHNOLOGY REQUIRED FOR DIGITAL PUBLISHING

A. Trusted Systems

A trusted system is a system that can be relied on to follow certain rules. In the context of digital works, a trusted system follows rules governing the terms, conditions and fees for using digital works. Suppose that you have a digital work stored on a trusted system, and you do not have a right to copy the work. Then if you ask the trusted system to make a copy, it simply will not do it. Instead, it will give you an error message. If you do have a right to copy and, for example, exercising the right requires paying a fee and certification that you are over 18 years old, then the trusted system would first make sure that the conditions are satisfied. Only then would it make the copy.

Suppose that a customer then wants to buy the digital work. In a typical case, he would use a network browser to select the digital work from an on-line distributor. At this point, the two systems-the consumer's system and the distributor's system-need to establish that they are both trusted systems and to determine their security levels and billing methods. One way to do this is with a challenge-response protocol. This protocol is similar to what you might imagine in a "spy versus spy" scenario when two secret agents who are strangers to one another first meet.

The security of the communications between the computers relies on the use of public key cryptography.[4] In public key systems, there are two keys used by a system for encryption: a public key and a private key. Each computer keeps its private key secret and its public key known. The keys are inverses. Anything encrypted in the public key can be decrypted by the private key. Anything encrypted in the private key can be decrypted by the public key. Assuming that the keys are long enough, decoding a message without having the proper key is very difficult, and it is difficult to derive one key from the other.

The consumer system begins by saying the digital equivalent of "I'm a trusted system and here is my certificate." The distributor system looks at the certificate. The certificate itself is encrypted in the private key

of a well known digital registry, such as a trusted system of the American Association of Publishers (AAP) or another appropriate body. The distributor system decrypts the certificate and obtains the public key of the consumer's system. Following the "spy versus spy" analogy, the distributor's system has now determined that there is a valid certificate, that it corresponds to a particular consumer system, and that the consumer system has a particular public key.

But how does the distributor system know that the system offering the certificate is bona fide? After all, the system on the other end of the communication channel may merely be masquerading as the consumer's genuine trusted system. The digital certificate might be copied, obtained perhaps through wiretapping or packet snooping. To remove these uncertainties, the distributor's system creates a "nonce", a one-time message made up for the occasion using bits from a suitable random source. It encrypts the nonce in the public key from the consumer system and sends it. Roughly, it says, "if you are really who you claim you are, then you can decrypt this message because only you will have the secret private key." The consumer system then decrypts the nonce and sends it back in the clear. There are several more exchanges in the protocol. These exchanges establish one-time session keys that the two trusted systems can use for efficient and secure communication on a possibly insecure channel.

There are variations on this approach. One way of organizing commerce in digital works is in terms of "digital envelopes" where the digital work is encrypted together with rights and sometimes a program ("applet") for interpreting the rights. This variation shrinks the boundaries of trust from enclosing the consumer's entire computer system to just enclosing particular software that accesses information inside the digital envelope. Other variations rely on a combination of hardware and software on the consumer's system. Different approaches have different costs, vulnerabilities and degrees of practicality.

For the purposes of this discussion, I will lump all of these approaches to commerce in digital works together. Generally, the technical differences are much less important than the similarities. I will use the term "trusted system" generically since any system for commerce in digital works must have some part that is trusted.

B. Digital Rights Language

A trusted system is aware of the rights associated with a digital work because the rights come with the work. Every approach to digital property rights requires a means of expressing rights. They can be attached to the work itself or can be stored in a database. Digital rights fall into several natural categories. For example, transport rights include the rights to copy, transfer, or loan. Render rights include the rights to play and print. Derivative work rights include the rights to extract, embed, and edit.

Rights and conditions can be expressed in a formal language that can be precisely interpreted by trusted systems. Several companies, including Folio, International Business Machines (IBM), Intertrust (formerly Electronic Publishers Resources), NetRights, Xerox, and Wave Systems are developing digital rights languages.

To illustrate the use of digital rights languages, in a typical situation an author would create a digital work using any authoring tool of interest. Digital property rights are neutral to data format and interpretation; that is, they can potentially work with any digital representation of text, pictures, databases, music, or video. Once a work is created, a publisher could import it into a trusted system. He would decide the rights with which to associate the work, and encode them using the rights editor of a publishing program. He could then make the work available on a server for sale online.

What rights can be expressed in a digital property rights language? This is a key issue in understanding the technical possibilities and what could constitute fair use. In the Xerox digital rights language, for example, there are currently sixteen distinct kinds of rights. In the following section, I focus on a few rights that illustrate the possibilities created by the technological shift. I consider various transport rights, rendering rights, derivative work rights, and backup rights.

C. Digital Billing

Different approaches to billing have been proposed.[5] These fall into two broad categories: approaches for off-line use and approaches for on-line use. Off-line approaches are approaches that can support mobile and intermittently connected operation and billing. Off-line approaches typically involve secure storage and processing on the trusted system itself, using extra hardware such as SmartCards(tm) or PCMCIA(tm) cards. These cards can function as debit cards or credit cards and must communicate once per billing period with an on-line financial clearinghouse. An example of an on-line approach is NetBill, which coordinates purchases between a consumer system, a supplier, and a financial clearinghouse. Approaches like NetBill are intended to keep the overhead of transaction aggregation and billing quite low.

Digital licenses and digital tickets can also be used for billing. Digital licenses are digital certificates that indicate membership in a group, such as a book club, business organization, or university. Some transactions require presenting proof of membership before particular rights or discounts can be exercised on certain works. For example, one might need to have a subscription certificate before a certain work could be read. For licenses and tickets to work reliably, there must be some means to prevent counterfeiting and unauthorized use, such as secure storage, transport restrictions, and passwords.

Digital tickets are like the coupons found in a local paper that give discounts on the purchase of grocery products. Issued by a publisher, they correspond to prepayment or discounts for using works by the publisher. For example, a work might come with three tickets, each of which can be used for printing a hard copy of the work. Once the digital tickets have been "punched" by a digital ticket agent, they cannot be reused; a repository must obtain more tickets to exercise that right again.

A work can have different versions of the same kind of right, each with different fees and conditions. For example, a musical work could have a right to play it for a fee charged by the hour. Another right to play that piece may have a fixed fee for unlimited playing. Yet another right to play the piece may give special discounts to members of a music buying club. A publisher may give promotional tickets as part of an introductory offer. When a user elects to play the music, he exercises one of the rights matching his set of licenses and tickets, and his desires, against the various options offered by the publisher.

Some observers of digital publishing have predicted that prices and discounts for digital works may become as complex as the array of special prices and ticketing in competitive airline fares. They worry that this may lead to complex user interfaces on media systems. One way around this complexity is to include simple computational agents in trusted systems that simplify a user's choices. For example, a boom box interface may present nothing more than a play button, where the user has already established a policy of minimizing the cost of a play. A user may need only to choose between listening to the work once or buying the right to listen to it as often as desired.

D. Trusted Printers

Even trusted systems cannot prevent all copying. If you can print a digital page, you can photocopy it. If you can watch a digital movie, you can record it with a camcorder. If you can listen to a piece of digital music, you can record it on an audio cassette recorder. It seems that anything that can be rendered and experienced by the human senses can be recorded. But all of these copies are subject to noise and degradation in the process of rendering and re-recording. What trusted systems prevent is the wholesale copying and distribution of perfect digital originals. They can also embed hidden and visible "watermarks" in renderings that make it possible to trace unauthorized copies.

Most documents for sale over the Internet are short-under twenty pages. For some publishers, the risk of unauthorized print copies for lengthy documents is too high. They do not want to risk distributing valuable assets in digital form lest they lose control of them. If the situation could be shifted to reduce the risk of loss of control in printing, content providers would distribute many more documents digitally. This leads to a desire

for trusted printers.

Trusted printers combine four elements: print rights, encrypted on-line distribution, automatic billing for copies, and digital watermarks for marking copies that are printed. When assigning rights to a digital work, a publisher uses a digital property rights language to distinguish between viewing (or playing) rights and printing rights. Play rights are used to allow the making of ephemeral, temporary copies of a work such as an image of text on a display or the sound of music from a loudspeaker. Print rights allow the making of durable copies, such as pages from a laser printer. Furthermore, a publisher can assign particular fees and conditions to the rights. For example, printing might be allowed only by certified members of a subscription club or by employees of a particular institution.

To reduce the risk that a digital copy will be stolen by wiretapping or packet snooping, a trusted system encrypts the document when sending it to a trusted printer.

Whenever a document is printed, the trusted system automatically logs the billing transaction. This contrasts with the situation of a typical patron at a photocopy machine in a library. If a fee is required to make a copy of a work, the patron faces the inconvenience of trying to locate the rights holder, determining the fee, and making acceptable payment. This makes it difficult for honest people to act honestly. In contrast, trusted printing handles billing automatically, making it easy to log and pay for uses of a work.

Finally, a trusted printer can mark each copy with watermarks as it is printed.[6] Watermarks can either be highly visible or hidden. They can contain information identifying both the rights holder and also describing the printing event. Watermarks can give transaction data about the printing of the work and also indicate whether additional copying is permitted. Visible watermarks serve a social function, reminding people about copyright infringement and fair use. Glyph technology, an approach to marking paper with tiny symbols, can be used to carry hundreds of bits per square inch in various gray patterns on a page. With careful design, glyphs can be integrated as graphical elements in a page layout. To foil unscrupulous users who would cover up visible watermarks so as to prevent tracing a copy back to them, hidden watermarks can also be employed. There are many techniques for hiding information in the spacing of words and lines of text, or hiding it redundantly in the dot patterns of images. Information in both visible and hidden watermarks can be recovered by a scanning program. Furthermore, special patterns of material can be made by printers that are distorted in predictable ways by scanners and copiers, making it possible to tell whether a given sheet is a printed original or a copy. Thus, watermarks can be used as a social warning, to carry information, and to leave digital fingerprints for detecting and tracing unauthorized copying.

Trusted printers offer a new way to read large documents of commercial interest that currently do not appear on the net. They change the practical usefulness of networks for distributing large documents digitally. The risks of theft of digital works is reduced by encrypting all transmissions of content to the printer and by embedding watermarks in the printout. Since payment systems are automatic, a publisher can rely on a revenue stream that is more dependable than the stream from reprint requests today. The required level of traceability of copies can be obtained by specifying in the rights the kind of trusted printer which may be used and watermarking.

III. SOCIAL AND LEGAL IMPLICATIONS OF DIGITAL PUBLISHING

I have now reviewed the basic technology required for digital publishing: trusted systems, attached rights, and electronic billing. The promise of digital publishing has significant social and legal implications. Digital publishing not only raises new issues, but also can make moot old ones. I will address these issues in the following sections.

A. Legal Standard for a Copy

Much of the confusion about digital publication centers around the question of what it means to make a copy.

For works on paper, the steps in the operation of a photocopier offer an analogy. The content is expressed by marks on each page. A photocopier creates an image of the marks and puts corresponding marks on another piece of paper.

The idea that copying marks on a medium amounts to making a copy has led to confusions in electronic media. Although this analogy works pretty well for describing what it means to make a copy of a video cassette, it does not work so well for digital copies in computers. The extrapolation of the analogy says that even copying bits into computer memory amounts to "making a copy" and thereby infringes copyright. But modern architectures of computers do not have one simple, undifferentiated "memory." In a typical case, when a computer receives data from some input device, the bits first are loaded into an input buffer. They might then be copied to main memory, and then into a high speed cache that parallels main memory. They would then typically be copied into a display buffer prior to rendering on a screen. In an extreme interpretation, copyright would be infringed several times by any computer with a reasonably modern architecture before any person could even see the work.

Although the legal force of this extrapolation is open to question on the issue of whether marks in computer memory are "fixed,"[7]

what is really needed is a way of talking about copying that takes into account how the information is used.[8] To return to the example of paper, what makes a new copy on paper a potential infringement is that it can be used like the original. If I take a published work on paper and run it through a modern copier, I can create a physical artifact that is just as useful as the original. Unregulated use of a copier increases the number of usable copies in the world without compensating the publisher or creator. This copying potentially reduces sales and undermines the ability of the publisher or creator to make a living, thereby undermining their incentive to create new works. Therefore, the usability of a copy and its potential for destroying potential revenues of rights holders need to be taken into account when defining what it means to make a copy on a computer.

B. Copy Versus Transfer

The Xerox digital property language distinguishes between two different rights, both of which transfer information between repositories. These rights are a right to copy, and a right to transfer. To copy a digital work is to make a new, usable digital copy on a repository without deleting the original. To transfer a digital work is to make a new usable copy on a repository and then delete the original. In copyright law, the term "distribute" is closely related to "transfer." The crucial difference between exercising a right to copy and a right to transfer is that the former increases the number of usable copies in the world and the latter does not. A transfer right mimics what happens when I give a friend a copy of a book purchased at a bookstore. Once I have given the copy away, I can no longer read it. Similarly, once a repository transfers a copy to another repository, the copy on the first repository can no longer be read. A transfer transaction does not increase the number of usable copies in the world.

A transfer transaction of a digital work is like a bank transaction with money. When you use an automatic teller machine and transfer funds between accounts, the money simultaneously disappears from one account and appears in the other. A transaction has the property of completing its operation entirely, or not performing the action at all. If a user interrupts a transfer process, he does not end up with two partial copies. Rather, both repositories go through a standard cleanup procedure. The sending repository restores the work to usable status and reports the transaction failure. The receiving repository deletes its partial copy and reports the transaction failure.

This distinction between transfer and copy rights gets to the heart of what matters about copying for copyright infringement. It's not so much the communication of bits that matters; it's the creation of useable copies.

For typical use of paper documents, the transfer right is governed by the "first sale doctrine."[9] In the absence of contracts saying otherwise, if you buy a book at the store, you can generally give it away. What you own is a

particular physical copy and you can dispose of it as you please. The first sale doctrine modifies the distribution right by dictating that after the first sale of a copy to the public, the copyright owner cannot control further distributions.

Not all printed documents are governed by the first sale doctrine. For example, the purchase of an expensive industry report from a consulting company can be conditioned on the proviso that it not be distributed to any organization beyond the one that purchased it.

With digital works, it need not be the case that transfer rights are free or universally granted. Publishers could charge a small fee whenever a work is transferred between repositories. The same mechanisms that prohibit copying without a right to copy could prohibit transferring without a right to transfer. The same billing mechanisms for a copyright work for a transfer right. The technology itself is neutral on this point. Trusted systems could enforce either policy.

C. Shades Of Gray In Fair Use

Issues of fair use[10]
are sometimes argued as being black and white. Digital property rights introduce shades of gray. Convenience of billing and a moderated scale can turn a fair use issue into a negotiation about price, rather than an all-or-nothing confrontation.[11]

In the past, when there has been a dispute about whether a use is fair, adherents of the two positions could line up their forces for a legal fight. Someone uses a work in a particular way, a publisher challenges the use as being an infringement, and the user defends the use as fair. If the court judges the use as "fair," then publishers will never realize any revenue from that use. If a use is judged not fair, then users are required to get permission and potentially pay for the use, no matter how incidental the use, or how inconvenient and expensive it may be to identify and locate the rights holders. This can present a substantial burden to potential users.

With low overhead, it is practical for publishers to establish very low fees for simple and even rare uses. With automated billing, they can make compliance relatively inexpensive and convenient. Since the fee to exercise a right can be large or small, the gap between fair use (free) and paying to exercise a right (possibly expensive) can be populated by many positions in between: nominal fees, low fees, medium fees, pretty high fees, and so on.

In a broad and healthy market, the availability and price of rights can become a point of commercial competition rather than a point of legal interpretation. Consider a hypothetical digital newspaper. One newspaper might disallow printing on the rationale that unregulated distribution of stories would undermine digital sales. Another newspaper might allow printing of any article more than one week old, on the rationale that printing old news amounts to advertising the newspaper. In this way, providing the most attractive suite of rights becomes an element of competition among publishers.

D. Digital Libraries

The right to loan illustrates another variation on how new digital rights influence how we think about fair use. It is common practice to loan books to friends, even if it is also a common experience that such books often are not returned. Lending is crucial in the operation of public lending libraries. In the digital realm, there has been much interest in understanding the role and operation of digital libraries.[12]

A digital property language can specify a loan right, which is illustrated by the following example. Suppose that Larry buys a digital book. After he has read the book, Larry is approached by his friend Bob, who wants to borrow the book. Assuming that the publisher has granted the loan right, Larry (the loaner) agrees to loan the

book to Bob (the borrower) for a week. Their two repositories communicate, and an encrypted copy is sent to Bob's machine, together with a specification of the digital property rights to be associated with the loaned copy. During the next week, both repositories have copies. However, Larry's copy cannot be read because it is marked as loaned out. Suppose that Bob goes on vacation and initiates no action to return the work. Both repositories have clocks. At the end of the week, Bob's repository deactivates his copy and Larry's repository reactivates his copy. No communication between the repositories is necessary for the work to be returned.

After the loan period is over, Bob may still want to read some unfinished portion of the book. Whether Bob's repository actually deletes the digital book is a policy question for how Bob's repository manages its storage. What matters from the perspective of preserving the number of usable copies is that the loaning and borrowing repositories not allow simultaneous use of the loaned work. However, it can be in the publisher's interest to allow Bob to purchase additional time to read the book after the loan period ends, or even to purchase a copy of the book-converting Bob the borrower to Bob the buyer. The same machinery that enforces fees, terms, and conditions could be called upon in this case to allow Bob to exercise a right to copy for a fee-essentially converting his expired loaned copy into a usable copy. In this scenario, any loaner or digital library is potentially a distributor.

The loan right gets to the heart of some of the issues that have been raised about the operation of digital libraries. One of the concerns about digital libraries is that in the absence of copy controls, digital libraries would amount to free distribution centers. The trusted system approach addresses that issue head-on. If some publishers do not desire works to be loaned out, they can simply not grant loan rights. Alternatively, suppose that they want to allow loaning but do not want copies to be made of any works that are loaned, based on the rationale that this competes unfairly with their other sales channels. In that case, they would arrange for works on loan to not have rights to copy. Suppose that publishers want to allow library patrons to copy the works, but for a fee and subject to certain conditions. The same trusted system elements that control fees, terms, and conditions for other copy transactions work the same way for loaned library copies. Thus, library patrons may be able to make copies providing that they can pay the fees and meet any other digital licensing requirements.

Libraries are consistently under pressure to reduce their expenses. Digital distribution offers some ways both to reduce costs and to provide new sources of income for libraries. Cost reduction comes from the automation of information services. New income comes from the ability of libraries to charge fees for on-line services and for distributing works.

Increasingly, libraries face issues of deciding which services to freely offer to the public and for which services to charge. The decision of how many copies of a digital work to buy for loaning illustrates one way to balance these values. Suppose that a library buys ten digital copies of a best seller to distribute for free to its patrons and ten more copies to distribute for a fee. Patrons that are more economically-minded can wait for free copies to become available. Patrons in a hurry and willing to spend more money can access the for-fee copies in lieu of waiting. Revenues from the for-fee copies can subsidize more for-free copies.

E. The Backup Issue

When I first presented the ideas of digital rights on trusted systems to colleagues at PARC, one of the criticisms was that the approach provided no means of making backup copies. Like most veteran computer users, my critics viewed the making of backup copies of software as an essential defense against media failure.

At first, I was perplexed by the backup issue. I felt that making backup copies was a legitimate user interest.[13] Backups can protect access to expensive information. Because the quality and reliability of the storage system is in the control of the user rather than the publisher, this should be a user responsibility.

The problem was that backups seemed to undermine the ability of repositories to maintain trust. A person could make a backup copy, sell the original, restore the backup copy, and then do it again. Uncontrolled backup copies could become an inexhaustible supply of unauthorized free originals.

The resolution of this apparent dilemma came when we realized that we could treat the making and restoring of backup copies in terms of rights, with the full range of fees and conditions. Making a backup copy is defined as making an encrypted copy that can be stored off-line. When a backup copy is read back into a repository, the only right on it is a restore right. The backup copy is not useful for anything else.

A restore right can have fees and conditions. A restore right can require contacting a publisher for permission. A publisher might give each user three "restore" tickets. A user who exhausts these tickets may still exercise a different restore right that requires the user's repository to communicate with the publisher's permissions server to obtain permission. When a permissions server sees too many requests for restoring the same copy, it can refuse to grant permission.

The backup and restore rights provide another example of how digital property rights convert polarized black-and-white arguments about fair use into shades of gray. By adjusting the fees and conditions, the interests of publishers and consumers can be brought into a point of balance that depends on the situation. This is another place where market competition may lead to a de facto right to make and restore backup copies.

F. Digital Reuse

Works are often made up of other works. A simple example of this is a set of readings for a college course, such as a collection of case studies used in a business school. Each case is written by a different person. Another example is a book of poems by different contributors. In music, there are collections of performances by different artists. A newspaper gathers elements from different sources-stories and photographs from the wire services, from archival services, and from local reporters.

Digital works can be composite, that is, made up of other works. For example, a digital newspaper might have stories from a wire service combined with photographs from various sources, local stories, and advertising. Different parts can have different rights. This means that the attachment of rights relates to components of a work. When a work is used, the repository must check the rights on each of the incorporated parts.

When a book editor wants to assemble a collection of articles, the process of obtaining permissions can be onerous. The rights holders must be located to obtain the necessary permissions and potentially to negotiate a fee. Often the permissions process operates out of a back room at the publisher. In many cases, common practice includes getting permission from both the publisher and the author. The general attitude is that reuse does not generate much revenue for the publisher so it gets little priority. The process of obtaining permission to use video or photographs can be much more complicated, potentially requiring not only permission of the photographer and publisher, but also permission from people who appear in the picture.[14]

In digital multimedia, photographs, sound recordings, video and text are all available in digital formats. In this regime, reuse of elements is potentially easy ("cut and paste") and potentially desirable. Imagine a group of artists, or for that matter, advertisers, creating a multimedia work: "Give me some ocean sounds here." "Do you have a picture of the Eiffel Tower?" "I need a video clip of a biplane flying over Manhattan." "I need a photograph of an idealized American housewife cleaning house in the 1950's." For such cases, the cost of developing the element from scratch would be high, creating a market opportunity for vendors of photograph and audio collections.

The digital property rights approach allows publishers to assign rights to works in their collections. These rights set forth various fees and conditions bearing on the use of the work. Specifically, the rights set forth the fees for making a copy or printing the work, as well as conditions on extracting from context, changing it, or embedding it in a larger work. An author or editor interested in reusing a work need only locate the work and use a trusted editing system to embed it in a larger composite work. Automatically, the rights travel with the work. When a consumer later buys a copy of the composite work, the rights holders of all the included works get paid automatically.

This example diverges from current practice in many ways. The most profound change is that it enables publishers to automate and streamline the permissions process. In current practice, when an editor wants to incorporate a previously published article in a book, the contract is typically a fixed fee in which the rights holder tries to guess how many copies of the article will be printed and to charge accordingly. The rights holder argues that the number of copies will be high, whereas the conservative editor argues that the number of copies will be low. In the digital rights approach, it is possible to eliminate the guessing by assigning a flat fee per copy.

Finally, it is worth mentioning another issue that often comes up in discussion of reuse. What is the smallest unit of a work?[15]
Is it the entire work? A page? A paragraph? What are the units for a picture? Can a piece of a photograph be used? How small a piece? What about sampling in music? Copyright law refers to fair use of short sections of works for review purposes.[16]
Across industries, various practices have evolved to set limits on the amount of material that can be used for quoting. Trusted systems can contribute to the balance of control by providing concrete means for extracting and editing portions of works and embedding them in compound documents. If the dimensions and parameters of such use can be worked out by the various stakeholders, rights languages can express the terms and conditions and trusted systems can enforce the rules.

In summary, trusted systems enable publishers to pre-specify conditions and fees for derivative works and enables even small fees to be automatically aggregated and collected. Although automatic permissions in support of reuse are not appropriate for all situations, digital rights for reuse could unleash some niche markets and greatly increase the quantity of works available.

G. Copyright Expiration

There is some debate about how trusted systems should behave when the copyright in a digital work expires. The law addressing copyright term has evolved over time. At the time of this writing, copyrights for written works created after January 1, 1978 extend for a period of fifty years after the author's death.[17] There are special cases in which the term of copyright varies from this standard, for example, when there are multiple authors,[18] works published under a pseudonym,[19] and works for hire.[20] But the significant issues are about the technology rather than the special cases of the law. What should happen at the end of life for a copyright for works distributed on trusted systems? Could trusted systems effectively nullify the automatic passage of a work into the public domain after the period of time proscribed by law?

This concern frames the issue in terms of the behavior of trusted systems and the good will of publishers. In this framework, one answer is that free access to the work should become available on every trusted system, so that the work could then be used without charge or restriction. But this is not the only possibility. Another possibility is that different unencumbered copies of the work could become available through public institutions such as the Library of Congress, while fee-based copies of the work are distributed commercially as before. One argument for the second possibility is that convenience of access is worth something. People can choose to locate a free version or they can use a for-fee version promoted by a publisher. Furthermore, even leaving aside the issue of how a trusted system determines when an author died, programming copyright law into trusted systems raises its own issues.

Suppose that vendors program trusted systems to model copyright law and to release digital works after a particular interval. Who is liable for the behavior of the systems if the copyright laws change during the lifetime of a work? Who is liable if a user of a trusted system liberates an unencumbered version of a digital work in one country when it goes out of copyright under local laws, and then distributes it in another country where the period of copyright is longer? These examples remind us of potential issues that can arise because of the dynamic nature of law, and the absence of national boundaries on the Internet.

H. Putting Boundaries into Cyberspace

The net makes the world seem smaller. People can interact with each other and exchange messages and works, whether they are in the same building or separated by great distances. In this way, the net overlays the physical and political world, crossing over national, city, school district, corporate, sales district, and other boundaries. Because the network crosses so many boundaries, it brings people from diverse cultures together who have different interests and values.

The crossing of boundaries leads to issues about restricting the copying of certain digital works beyond the considerations of copyright. To allow anyone to make a copy of any digital work, even for a fee, would violate many expectations of business practice as well as various laws and community standards. For example, sale or import of some works may be banned in certain countries.[21] Some works contain sensitive data, such as company trade secrets. Some countries have restrictions on the percentage of local content on certain kinds of works.[22] Some communities may choose to restrict sales of certain categories of works to children.[23]

As described already, the terms and conditions that govern the exercise of rights are expressed in a digital property rights language. These terms and conditions can restrict who can exercise a right on a particular document. They can also express requirements on a trusted system itself, such as its security level or its ownership. Since questions of identity must often be resolved over a communication channel, signed digital certificates are typically used as part of a secure process for establishing identity. In this way, licenses as digital certificates can put boundaries into the network, acting as identifiers for groups of people and gating the flow of information.

By certifying that someone is an employee of a particular corporation or a member of a work group, digital certificates can control the transporting and rendering of documents. Only someone who has the right kind of certificate can read documents with the restrictions. Similarly, certificates can be used to indicate membership in a book club, such as an on-line book club with purchase discounts for its members. Certificates can be used to certify someone as a member of an organization having a site license. They can be used to identify people with up-to-date subscriptions to a journal. They can be used to indicate professional status, such as a researcher at a library. They can be used to give cheap or free access to members of the public that would otherwise not be able to afford access to information.

In some commercial publishing settings, it is appropriate to control who can distribute works. For example, some distributors can be licensed to sell all copies of a work in a geographical region. For paper-based books, the definition of regions often follows national and geographic boundaries, respecting expenses of either language translation or government regulations. One rationale for granting a distributor a monopoly for a region is that it compensates for the costs of local advertising. With regard to computer software, some publishers demand that distributors be qualified to provide training and support.

The relationship of national boundaries to particular computers presents some interesting issues. Suppose that a piece of software is supposed to be restricted in some way when it moves between country A and country B. For example, an import tax might be required. If a person from country B travels to country A and loads the software onto his laptop computer, has he imported the software? Or is the software imported when he carries his laptop across the border? Does importing take place when a document is electronically transmitted across the border? What if a person from country A is visiting country B with his laptop and electronically loads software over the network onto his laptop, and then carries the laptop back to country A without ever giving a copy to anyone in country B? Has he exported the software twice, or not at all? All of these examples raise questions about enforceability of import and export restrictions on computers.

One approach to the import/export issue, not yet sanctioned by law or practice, would be to register computers in the same way that we register ships. Thus, a laptop registered in country A would be considered to be "country A registered" no matter where it is physically located at a particular moment. This is a bit like a

"Panamanian registered ship" and a bit like a "Panamanian embassy." Importing between country A and B then occurs whenever a digital work is moved from a trusted system registered in country A to a trusted system registered in country B, independent of the physical locations of the two systems. The registration of a trusted system would be certified by a non-transferable and not-easily counterfeited license that it carries.

Recapitulating, many issues about transport and use of digital works lie outside the scope of copyright law. Digital rights rely on a combination of laws and contracts. Digital certificates can be used in a practical way to introduce representations of boundaries and identity into cyberspace, making control practically enforceable. There are several technical means to make digital certificates non-transferable, dated, and hard to counterfeit.

IV. COMPETITION AND COOPERATION

A. Achieving Balance

Trusted systems shift the balance and put more power in the hands of publishers. Publishers decide what rights to assign to a work and what fees to charge. If someone buys a work and wants to put it to a use not available as part of the rights attached thereto, they are blocked. It is not likely that designers of these trusted systems will put a button in the user interface which says "Just trust me. I intend to use this work in some different but fair way. Just give me a copy in the clear." It is significant that the user is denied a fair use defense because he cannot get a copy of the work. Reminiscent of the "Laws of Robotics" from Isaac Asimov's robot novels, we can ask what are the rules of good behavior for trusted systems that take into account the public good? And who determines the rules?

Stepping back from the argument about a shift in power to publishers, however, it is interesting to see how publishers view digital publishing in the absence of trusted systems. From a publishing perspective, uncontrolled digital technology itself shifts the balance in the social contract between those who create and distribute works and those who use them. For many kinds of digital works, it has become very easy to use and duplicate a work without having authorization or providing compensation. Untamed, the digital frontier is so wild that publishers cannot imagine how to make a living. The fundamental challenge is to provide appropriate checks and balances for the interests of the various stakeholders.[24] The public is now getting its first experience with trusted systems. Trusted systems do not exist in a vacuum. They exist in a social framework. The search for balance involves the design of appropriate social institutions.

Many stakeholders are unwilling to assume that publishers will always act in the public interest. After all, copyright law itself has evolved over time to reset the balance of power between the public, authors, and publishers. In some historical cases, the law was changed because the public believed that publishers were charging too much. In a healthy market, competition may guide publishers toward policies that are in the public interest. In other markets, cartel behavior remains a possibility.

I believe that dedicated social institutions will be needed together with trusted systems in order to best serve the market and the public good. In discussion with various stakeholders, I have used the name "Digital Property Trust" (DPT) to refer to such an institution. The DPT would have as its objective the promotion of commerce in digital works. The DPT would augment the social framework so that stakeholders would have representation in a system with a high degree of automation. Given the greater efficiency that automation and trusted systems create, it is reasonable to handle the rare cases with human labor and human judgment. In my mind, the DPT itself would be governed by representatives of the various stakeholders-including publishers, trusted system vendors, financial institutions, lawmakers, librarians, and consumers-and would interact in an appropriate and organized way with governing bodies and law enforcement agencies in different countries. It would be funded by contributions from the stakeholders and by a small tithe on trusted system transactions.

Referring back to the fair use defense, the DPT could offer an arbitration and licensing service. Members of the public with special needs-such as librarians, researchers, and teachers-could have special licenses. To get such a license, a person might need only to demonstrate an understanding of the principles of fair use and copyright

to a suitable authority. These licenses could have associated special discounts or free use for certain kinds of works and perhaps fewer limitations on rights. To balance the risk of substantial financial losses due to unauthorized digital publishing by licensed individuals, there could also be insurance associated with licenses that grant less-controlled use. Works released in this way might have special watermarks embedded in the digital data, used to track any source of unauthorized publishing or alteration. The licenses, insurance, and watermarks are mechanisms that can be used by the DPT to balance the competing interests and risks of the stakeholders.

B. Achieving Interoperability

An important concern being raised by publishers is whether trusted systems by different vendors will be compatible. Can digital works with rights assigned using one approach be used by a consumer who mainly uses another vendor's equipment? Publishers are aware of market confusion over competing standards, such as the Sony/Betamax battle to establish the dominant standard in the video cassette recorder market. Publishers are worried that incompatible standards will increase expenses and slow the growth of the digital publishing market.

At present, the trusted systems and rights encoding systems being developed by vendors are proprietary and incompatible. Providing for interoperability requires substantial cooperation among the vendors. One way that interoperability can be achieved is through the use of an established standard language for terms and conditions that works across platforms and systems. However, the real issues of interoperability go much deeper than the use of common formats.

Consider the following example. Someone buys a publisher's work using a trusted system supplied by vendor A. Later he copies it properly for a friend who uses a system supplied by vendor B. There is a chain of such transfers and loans. At some point, the work is copied without proper payment between a system supplied by vendor F to one supplied by vendor G, and furthermore, the consumer of G now begins to distribute the work for free without authorization.

This example raises several issues. What are the losses or potential losses for the publisher? To what level should the publisher be compensated for loss of revenues? Which consumers are liable? What vendors are liable? To what level of risk and liability is each vendor in the chain exposed?

There are many possible answers to these questions. The answers in force as the market develops will affect its growth and development. I think one approach that would be conducive for a lively market would involve an important role for the DPT, and I will now outline that approach.

The participation of a trusted system in commerce in digital works implies a fiduciary relationship between the vendor and several parties, including the rights holders, the consumers, and the financial clearinghouses. Offering a computer system to participate in this commerce involves more than just building systems that can interpret certain data formats. Trusted systems need to be certified. They cannot be certified just by their vendors, and market competition gets in the way of certification by other vendors. The certifying organization needs to have certain skills and standards in order to verify that a system is compliant and can be trusted to follow instructions in the digital property language on all digital works. As a certifying organization, the DPT needs to be technically competent and independent.

The established way to manage and amortize risk is insurance. Each step in the chain of transactions in our example adds a separate risk and should have a separate insurance component. A practical way to manage this process is to have a small premium associated with each transaction that pays for the insurance. The premium would vary with the level of risk and with the amount of insurance coverage required by the rights holder. The DPT could act as a central clearinghouse for the insurance, as well as an issuer of the digital certificates that mark each repository as "trusted." The DPT would also be involved in maintaining and distributing hotlists of rogue repositories and would carry out other measures to ensure the smooth operation of commerce. This

approach bounds risks faced by both publishers and vendors and provides a scaleable income stream to pay for the insurance. Restated, the DPT backs trust with insurance.

Over time, the standards for defining rights and conditions in the digital property language would change. An important role of the DPT would be to manage both the evolution of the standards, and also the expansion of the DPT into new areas. The governance of the DPT should reflect the strong interests of the different kinds of stakeholders: publishers of books, movies, newspapers, and music; vendors of computers, trusted players and trusted printers; financial clearing houses; governments; and consumers.

At present, no organization is empowered to act as the DPT. Indeed, individual vendor companies have been building patent portfolios in support of their own technological approaches. The path ahead could involve market confusion and patent fights or it could involve enlightened self-interest in establishing a healthy market in digital property. A well-organized movement to create the DPT, with appropriate representation and investment by stakeholders, would be the single most effective step in creating a lively market in digital publishing.

V. CONCLUSION

I am sometimes asked by publishers when I expect digital publishing to take off. At a recent publishing industry roundtable, representatives from several technology companies were asked this question and the answers ranged from late 1996 to 2001. I have been looking for a definitive moment to mark the beginning of digital publishing. When I look closely, what I find is a fractal. Each period of time has many small beginnings.

At the time of this writing, several companies have trusted systems and software in alpha-test stages. Some early publishers are building their first applications. Some publishers are holding back, predicting that other models of revenue generation will dominate, such as advertising-based distribution. There are many conferences and discussion groups in which the issues of fair use, billing models and platform availability are being discussed. Momentum is building. More people are sensing the possibilities and are moving up the learning curve. More bits of the infrastructure are falling into place. The DPT does not yet exist physically, but increasingly, publishers and vendors are beginning to understand the issues that the DPT would need to address. There is little doubt that the market will develop. How quickly the market will grow and how well it will thrive depends on whether competitors in the relevant industries merely compete in offering their products and services, or whether they collaborate on actually creating the market for trusted systems, promoting standards and the use of the developing technology discussed in this article.