Digital Copyright and the Progress of Science

Jessica Litman

Keynote Address
Joint Conference on Digital Libraries 2002
Portland, Oregon, USA
July 14-18, 2002

Let me start with a truism: Networked digital technology has transformed information and the way we interact with it. Digital information is dynamic rather than fixed. What we think of as "documents" can change constantly. That's challenged our notions of what it means to archive material.

Digital information is ubiquitous rather than scarce. Digital information is extraordinarily accessible. If I have a question, I don't need to make up an answer that seems plausible or reason out what it's likely to be. I don't need to go to the library and ask the reference librarian if I can see the library's only copy of a reference book that ought to have the answer. I can just run upstairs to my computer and look it up on the web.

Digital information is *shared*. Ten years, even five years ago, it was conventional to talk about the Internet as a tool for disintermediation. Authors and musicians would be able to use digital networks for sending stuff directly to their readers and listeners. Remember Stephen King's *The Plant*?

There's some of that. People post content on their websites for the rest of the world to view. Academics exchange drafts of scholarly papers that way, and independent musicians and composers make recordings of their work available for sampling and download. The "blog" (or weblog) is an increasingly popular art form in which people post an ongoing, public diary of things they find interesting and want to share. But, while we've seen a small but appreciable amount of direct distribution, there's even more peer-to-peer distribution. Napster is a specific case of what's happening to information resources all over the Internet. Someone has the recording I want. If I can find her, she'll share it with me. I can copy it and pass it on. Someone knows the answer to my question. If I can find her, she'll tell me. I can learn it and pass it on. Someone has seen the source I want to consult. She can tell me where to find it.

Ten years ago, not only Washington but the entire journalism biz believed that the burgeoning digital network (which went by the name, back then, of the "National Information Infrastructure") would develop into a 500 channel interactive television system, with "interactive" meaning that it would incorporate a method for ordering and charging purchases and receiving targeted advertising. There are a number of businesses out there which are continuing to try to shove the Internet in that direction, but the fact that it isn't yet anything like a 500 channel TV is largely because of the way that lots and

lots of people have come to interact with information. We've discovered that if you make it quick and easy, people *love* to look stuff up. People want to know how old Steven Spielberg is. They want to know the history of early radio. They want to know what traveling musicians wore in 15th Century Europe. They want to know how to make Shaker Lemon Pie. They want to know what the Constitution actually says.

If it's quick and easy to do so, they'll look it up. The rate at which people have adopted the Internet as their research tool of choice is astonishing. People find the easy availability of all that information empowering. They really get in to discovering new stuff. And they *love* to share – they share information, sources, sites, newspaper articles, movies, music. The system has been evolving as we watch: all that interaction is leading to more information and better information and more accessible information, more complete and deeper archives, wider ranges of divergent sources.

There's a wonderful Internet archive named the "Wayback Machine" at www.archive.org that allows you to view old versions of documents on the World Wide Web. If you check out earlier incarnations of the pages you visit most, you can actually see some of the ways we're all getting better at this.

Most of us in this room would agree that this is the most important development in information space in our lifetimes. I suspect that one would need to go all the way back to the invention of the public library to find another development as significant.

Networked digital technology transforms information spaces. That poses particular challenges for information policy. Consider archiving. How do you archive dynamic information that is changing all the time? Do you save all of it? You can't. How do you pick? Think about access. Digital information exacerbates problems posed by the digital divide. It intensifies the assault on individual privacy and anonymity. The library tradition is deeply committed to giving access to anyone from off the street above a certain age and to refusing to keep track of what any individual reads. Is it appropriate to restrict access according to who is doing the reading? Is it appropriate to keep track?

And, of course, networked digital technology implicates copyright law, which is why I'm here.

Did I hear a groan? My guess is that most of you are pretty frustrated with copyright. Copyright is frustrating for everyone who knows even a little bit about it, but I expect that the folks in this audience find it more frustrating than most. Some of you come to the digital library community as librarians and information specialists. Librarians have been fighting on the front lines of the copyright wars for more than 10 years. They "got it" before anyone else – including many copyright lawyers. Librarians have watched copyright lose its moorings and evolve from a legal system designed to encourage the broadest possible distribution and use of works of authorship to a system designed to restrict distribution and use in the name of the copyright owners' supposed right to "control" their works. Every time you turn around, the copyright law has changed again, for the worse. Some of you come to this community as software designers. For many of you, the most frustrating thing about copyright law is that nobody will give you a straight answer about what the law *says*. The policy debate is the

policy debate, but whatever the politicians do, you can design a system that mirrors the legal system of rights. If you could just get someone to give you a coherent account of what the rules are, you could just code them. Fair use, for example. If that means, say, "it's okay to make a copy for yourself, but you can't copy the copy, or pass it around" for example, you could code that. But, every time you ask, some lawyer says something like: "It's just not like that. It's deeply contextual. There are rules, of course, but I can't explain them in a way you'd understand."

Well, actually it depends on the lawyer. If you ask a lawyer who works for the recording industry or the film industry, she might tell you that the law says that copyright owners have the right to control every single copy of their works, period, and anything else is piracy. That, at least is easy to code. But it isn't the law. Copyright owners have *never* been entitled to control every copy of their works. That would defeat the whole purpose. The idea of copyright is to give authors an incentive to create new works and distribute them to the public, so that everybody else can use them, learn from them, build on them, and thus, as the Constitution puts it, "promote the Progress of Science and useful Arts." By "Science," the founders meant knowledge. Copyright is supposed to be designed to advance and spread knowledge. It incorporates a balance between incentives for authors and limitations on the control they might be tempted to exert. We can have lending libraries, archives, museums and concordances precisely because the copyright law puts significant limits on copyright owners' control of their works. The copyright system is intended to encourage the production and wide dissemination of knowledge.

Since copyright is intended to promote the progress of science-as-in-knowledge, then it would be reasonable to expect the copyright law to encourage the sort of sharing that's transforming our digital information space; to encourage it actively. If you read any recent news stories about copyright, though, what you read is that so and so has used copyright law to stop this, to prevent that, to turn this thing off and shut this other thing down and to prevent something else entirely from even getting off the ground.

By crying "Piracy!" at every opportunity, copyright lobbyists have done a pretty good job of selling much of the nation on the idea that either the current law entitles them to control all copying, distribution, performance or display of their works, or that it needs to be changed –right now-- so that it does enable them to exercise that control. Further, they are making the absolutely outrageous demand that systems need to be designed in a way that makes that control easy to exercise and as difficult as possible to defeat. By "systems", I'm not talking merely about proprietary file formats and digital watermarks. I'm talking instead about designing personal computers, hard drives, televisions, radios, cell phones, routers, IP protocols and the Internet itself to enable copyright owners to prevent any uses they don't want to permit. Which, as I said, is a right they don't happen to have under even the most expansive construction of current law.

In the past, whenever a new information or entertainment technology has appeared, it has developed along whatever lines made the most sense for it first, and then we've figured out a way to pay copyright owners some money to compensate them for use of copyrighted works. Early movies didn't get permission before writing screenplays based on stories; early radio didn't pay for the music it played. Early cable TV didn't pay anybody for anything. Neither did the manufacturers or operators of jukeboxes,

photocopiers, or videocassette recorders. Eventually, the new technologies matured, and figured out ways to pay for copyrighted content.

This time, copyright lobbyists insist, is different because digital is different. Networked digital technology enables anyone to make a million perfect copies and send them off to a million of one's closest friends. Networked digital technology is different because it has the potential to cause unprecedented harm. So digital systems need to have copyright enforcement built in – they must be designed so that people can't use them to infringe copyright owners' rights. If we permit digital systems to be designed in order to maximize their usefulness, their stability, their accessibility or their flexibility, we run a risk that those systems may end up being useful, stable, accessible or flexible, but may facilitate widespread digital copyright infringement. That would be an intolerable result.

Those arguments have proved to be surprisingly persuasive. How many of you are designing your digital library systems not by asking what design is going to be the most elegant, the most useful or what design will enable the most flexibility or interactivity, but by asking instead how to design it to accommodate the demands of proprietors that you control access to, copying of, and use of content they claim to own? It isn't that everybody agrees that copyright owners do have or should have that control: it's that it is becoming increasingly clear that they won't go away until they are satisfied that they have that control. Thus, the frustration: you're willing to code it, if someone would only tell you what the rules are.

But we can't. Not "won't" but "can't".

We don't actually know what the rules are yet. It may be a number of years before we find out. Meanwhile, if some people in the copyright community can persuade everyone to design digital systems to allow publishers and record companies and movie studios to control how their works are used, it won't matter what the rules really are. So long as the systems are designed to allow copyright owners to enforce their preferences, the system in practice will function in practice as if the law said whatever they wished it to say.

It seems as if someone files a new digital copyright lawsuit every week. It seems as if Congress is debating a number of proposed laws that you hear described by different folks as likely to bring the apocalypse down on us all. It's very nasty out there. But for all that action in the courts and in Congress, there are not very many answers.

There are a few lessons we can draw from the copyright wars of the last ten years: First, we don't know who owns what legal rights in copyright works in connection with digital networks. Second, in the face of that uncertainty, copyright owners are insisting that the law gives them and should give them control over every appearance of a digital copy of any work anywhere. Third, it is as a practical matter nearly impossible to determine with any certainty who has the authority to grant the permissions you need in order to make a work available over a digital network. Finally, demands for excessive control over the exchange of information threaten to undermine essential values that have shaped our information policy for more then 200 years. Let me take those in turn:

<u>We don't know what the law is.</u>

Copyright lawyers may talk as if we're certain, but we don't know what the law is, only what we hope or fear that it will turn out to be. The current copyright law derives from a statute that's more than a quarter-century old. There are amendments here and there to deal with particular digital issues, but the general law setting out the scope of copyright owners' rights was written before anyone was thinking about networked digital technology. That statute, as many of you know, gives copyright owners the rights to make copies, to distribute copies, to make derivative works (or adaptations), and to perform or display the work to the public. It then subjects those basic rights to pages and pages of exceptions, like the first sale doctrine, and fair use, the provision allowing libraries to make interlibrary loan copies, and the provision allowing restaurants to play the radio for their customers.

How the language applies to digital networks is still unclear. Is storing a copy of a work in a digital archive making a copy? Is retrieving it making another copy? Is a new copy made every time someone views the file? Is allowing someone to view the file a distribution to the public? Is it a public display? What about if the patron gains access to the file over the Internet? Is that a copy? A distribution? A public display? What about the copies made in the computer servers between the patron and the digital archive? Are they copies too? Are they distributions? Are they public displays?

Those are good questions. We copyright lawyers don't know the answers yet. The statute is old; the Congress that passed it didn't have this stuff in mind. This problem is hardly unique to copyright law. The legal system's usual approach to interpreting an old statute, and we have many, is to litigate the uncertain questions so that our courts can tell us what the answers are. But that hasn't been happening here. Instead, what we've seen is an epidemic of copyright litigation that runs defendants out of business before the cases can get to trial. The lawsuit ends, but the legal questions remain unresolved.

Let's take the most famous copyright case of the 21st century, the *Napster* case. That case generated a couple of judicial opinions, but it never made it to trial. The Court of Appeals for the Ninth Circuit ruled in the *Napster* case that the 70 million or so individuals who were using Napster were breaking the law. That got a lot of press. What didn't get a lot of press was this: the court also ruled that Napster wasn't necessarily liable for what its users did. Napster had a good argument that even if Napster users were breaking the law, Napster itself was doing nothing illegal, because it fit squarely within a statutory exemption. The appeals court said "Napster may very well be right about that" and sent the case back to the trial court to decide on those defenses at trial. Napster also argued that the folks suing it didn't in fact own the digital rights to the music. The court was going to decide that at trial too. Napster ran out of money before the trial could even get started.

Napster, at least in the beginning, didn't want to pay anything. (Later it offered to pay handsomely, but the recording industry had already decided that it needed to bury Napster with a stake in its heart.) Several other online businesses tried to pay royalties at the outset, but found themselves on the wrong end of a lawsuit instead. I'll mention two of the most famous.

A music site named MP3.com negotiated and paid royalties for a license to stream recorded music to consumers who demonstrated that they owned the CDs the recorded

music was from. MP3.com rolled out the service. Music publishers and record companies sued MP3.com, arguing successfully that it broke the law when it copied CDs it had purchased into file formats required for streaming so that it could perform the music under the license it had paid for. The court ruled that MP3.com couldn't make those copies simply because it had paid for the right to perform the music; rather, it needed separate permissions from different copyright owners to copy the music in order to perform it.

A Canadian site named "ICraveTV" applied for a statutory license under Canadian law to stream local Toronto broadcast TV signals over the Internet to viewers in Toronto. The site was shut down in Pittsburgh by U.S. movie producers who argued that streaming might be legal under the law of Canada, but it was illegal under the law of the U.S. Unless ICraveTV could guarantee that nobody in the United States could use a ruse to gain access to its webcasts, it had no business taking advantage of its rights under Canadian law. The court in Pittsburgh agreed.

I could go on telling stories about a very long list of cases. They reveal a pattern: Some Internet upstart comes up with a compelling idea and a threatening business model. This upstart is sufficiently well-capitalized to be worth slapping down, but not enough to withstand scorched-earth litigation. Copyright owners file a lawsuit. The upstart raises a couple of compelling legal defenses. Then it runs out of money and goes out of business. Case closed.

Here's the bottom line: we don't know the answers to the legal questions, and we're unlikely to resolve them soon.

Every copy needs a license

In the face if that uncertainty, the position I've heard from publishers, movie studios, record companies is, in essence, that digital is different. A digital copy on a network is a copy *and*, they say, it's also a distribution *and*, it's also a public performance or display. So, for every copy of every work (including those computer copies that no human ever sees) you need a license to make the copy *and* a license to distribute copies *and* a license for the public display. Moreover, digital copies have the potential to be shared with a million close friends. Digital copies are as good as the original. All of the exceptions and limitations on a copyright owners' rights, like the first sale doctrine or fair use, apply to analog, not digital. Digital is different.

Take fair use. Most people in the United States have a pretty firm seat-of-the-pants sense of what fair use is: Noncommercial use for criticism, analysis, research, education, parody, private individual copying are all fair uses. Ask a lawyer for the film industry or the recording industry or the music publishing industry or the book publishing industry what fair use is, though, and you'd get a different story: fair use is not a right at all, but a limited narrow privilege to use a little bit of a work in circumstances in which your use will cause no harm whatsoever to any of the copyright owner's interests, and, moreover, there's no practical way to ask permission, either because your infringement is de minimis or because it's critical. The idea that there's some right to make private copies, they continue, is a myth.

What about the *Sony Betamax* case? Didn't it hold that it's fair use when consumers tape television shows? Well, okay, they say, there was the *Sony* case, but, the *Sony* case was wrong, and anyway, the only thing the Court actually held was that consumers could make temporary analog copies of free broadcast TV in order to "time-shift" them and then erase them. It would be a mistake, they insist, to try to expand the privilege to encompass any other kind of work (pay TV, books, records), or any other kind of use. There is emphatically no general right or even privilege to make a personal copy. But in any event, they say, there is no fair use privilege to make a digital copy of anything, because a digital copy could potentially generate millions of perfect digital copies.

Thus, the motion picture industry is suing Sonic Blue for marketing Replay TV – a digital video recorder. We're heard some particularly intemperate language about skipping commercials being the same as theft.

Where are they getting this law? They're making it up, or, rather, they're describing what the current statute might say if every single one of its ambiguities were resolved in their favor. A law requiring everyone to get permission for every digital copy, even the transient ones in random access memory, and to get permission for every transmission from computer to computer, and to get permission for each person who views, reads, hears or gains access to every copy each time, would be seriously unworkable. I frankly doubt the courts would go along. But since these cases never make it to trial, the courts aren't being given much of a chance.

The costs of designing systems to maximize access control

So, maybe I've persuaded you that we copyright lawyers have a screwy way of doing business. I probably haven't convinced you yet that your best course is *not* to try to stay out of trouble by writing code that will enforce any limitation the owner wants to impose, whether the law gives it that right or not. There are at least two reasons why that's a bad idea. One is philosophical and the other is practical.

Here's the philosophical one: Coding a system to maximize copyright-owner control undermines one of the greatest strengths of our culture, and backs away from what libraries and librarians have stood for for many years. Uncontrolled, indeed, wild information does more for the "Progress of Science" than almost anything else I can think of. When you visit a bricks-and-mortar library today and look at a book on a shelf, nobody checks you ID. When you go to the reserve desk and ask to see a reserve book, the librarian doesn't phone the publisher to say, "*I've got a class-three middle-aged female here in blue jeans who wants to consult your book. May she see it, and, if so, should I give her the upscale edition or the version sponsored by Kmart?"* Even when you check out a book, the library records your name and address in case you fail to return it, but it doesn't keep the date in a long-term file, and it certainly doesn't pass it on to the book publishers or sell it to marketers. You don't receive letters that say "*Dear Ms Jones, On the basis of your library records, we're confident that you'll be interested in this special offer…*"

Here's the practical reason. Let's imagine that you've decided that you don't want trouble – you'll just take the copyright owners' word for what its rights are and code that into the system. Even that won't work.

You can't actually find out who owns the rights you need. Why not? Well, for starters, there isn't anywhere to look it up. For the first 185 years in which the US had a copyright system, the overwhelming majority of potentially copyrightable material was in the public domain. As recently as 25 years ago, when you saw something you wanted to use or share, the default rule was that you were entitled to do so. Unless the object had a copyright notice -- was marked "do not copy" -- you were, with some modest exceptions, entitled to assume it was in the public domain. Why? Because the absence of a copyright notice ensured that it was in the public domain (even if it hadn't been before). Not only that, but the notice had to be accurate, had to tell you when the copyright was scheduled to expire, and had to tell you to whom you needed to address any request for permission.

The overwhelming majority of potentially copyrightable works didn't have this notice and entered the public domain the minute copies were publicly distributed. Of the ones that bore the prescribed copyright notice, only a fraction were registered, and of the fraction that were registered, only 15% were renewed, so for most of the copyright-protected works that had the requisite notice, copyright protection lasted only 28 years.

Formalities like notice and renewal get a bad rap these days. We got rid of them about the time we realized we had become a copyright exporting nation rather than a copyright importing one. We no longer condition copyright protection on a bunch of technical requirements. What we miss when we dismiss the formalities as characteristic of a provincial and outmoded attitude is that the formalities were the principal method embodied in US copyright law for preserving the public domain and encouraging the public to use, reuse and share potentially copyrightable material. Having abandoned them, we have created a system in which copyright rights may be owned by a variety of different people and there is simply no reliable way to find out who they are.

There's no place to look it up. More fundamentally, we don't actually know who owns the digital rights in particular works --- publishers and record companies have been brazenly assuming that they do, but the couple of cases that have reached the courts have concluded that they are often mistaken. University Microfilms and Lexis/Nexis both purchased licenses from the New York Times to produce CD Rom and online versions of the New York Times. Both were held liable for copyright infringement in a suit bought by free-lance authors, because the Times had licensed rights it didn't have. While there are legal systems that would take well to coding, this isn't one.

So, what is it you *can* do? You can do what innovators before you have done. Figure out how to design your systems so that they meet the information goals you have. You're the experts, but I'd think that things like accessibility, reliability, interoperability and comprehensiveness would figure in there. I would not be in a rush to help anyone turn your digital libraries into digital copyright kiosks. I'd ask you instead to focus on designing information spaces that are optimized for the storage and use of information. Copyright owners need to be paid, and they will be, you may be sure. There are lots of ways to accomplish that. We don't need to build a world in which information is tightly controlled and access to it is closely monitored. Some copyright owners – the big ones who litigate – have focused tightly on control, sometimes to the exclusion of

compensation, because networked digital technology frightens them. They can't be expected to focus on advancing the "Progress of Science."

Making as much information as possible as widely available as feasible is what libraries are *for*. So do that. The copyright uncertainties are going to take a while to shake out, and it's a mistake to assume that any of us can predict what they will look like when the dust settles. By the time they begin to get that settled, you'll have a better idea of the sort of exemptions and limitations you need and the sort you can live with. You'll be able to design systems that look and work like digital libraries rather than digital shopping and credit bureaus.