

The Digital Dilemma

Randall Davis

INTRODUCTION

The combination of technologies that make up the information infrastructure—information in digital form, computer networks, and the World Wide Web—has arrived accompanied by contradictory powers and promises. For intellectual property (IP) in particular it promises more—more quantity, quality, access, and markets—while simultaneously imperiling the rewards of those who create and publish. It is at once a remarkably powerful medium for publishing and distributing information, and the world's largest reproduction facility, running unchecked in practice if not in statute. It is a technology that can enormously improve access to information, yet can inhibit access in ways that were never before practical.

That technology has arrived in a world where our existing laws, policies, and practices around IP depend on a number of subtle, surprisingly complex, and at times conflicting elements of law, public policy, economics, and technology. Those elements are in relative balance today, but may well be thrown out of balance by the transformations resulting from the information infrastructure.

One core of the problem is illustrated simply enough. A printed book clearly can be read by one or perhaps two people at once, people who must, of course, be in the same place as the book. But make that same text available in electronic form, and there is almost no technological limit to the number of people who can access it simultaneously, from almost anywhere on the planet. At first glance, this is wonderful news for the information consumer and for society—the electronic holdings of libraries (and friends) around the world can become available 24 hours a day, year round and would never be checked out. These same advances in technology create new opportunities and new markets for publishers.

But there is also a more troublesome side. For publishers and authors, the question is, How many copies of the work will be sold (or licensed) if networks make possible planet-wide access? Their nightmare is that the number is *one*. How many books (or movies, photographs, or musical pieces) will be created and published online if the entire market can be extinguished by the sale of the first electronic copy?

The nightmare of consumers, on the other hand, is that the attempt to preserve the marketplace leads to technical and legal protections that sharply reduce access to society's intellectual and cultural heritage—resources that have long been seen as crucial to democracy and to science.

The information infrastructure thus has the potential to demolish the careful balancing of public good and private interest that has emerged from the evolution of intellectual property law. In the US, the public good is phrased as the betterment of society that results from the constitutional mandate to promote the “progress of science and the useful arts”; the private interest is served by the time-limited monopoly (a copyright or patent) given to one who has made a contribution to that progress. The challenge is in striking and maintaining the balance, offering enough control to motivate authors, inventors, and publishers, but not so much control as to threaten important public policy goals, such as preserving the nation's cultural heritage, providing broad access to information, or promoting education, science, and scholarship. As

usual, the devil is in the details, and by and large the past 200 years of intellectual property history in the U.S. have seen a successful, albeit evolving, balancing of those details. But the evolving information infrastructure presents a leap in technology that upsets the current balance, forcing a rethinking of many of our fundamental premises and practices.

The stakes involved in all this are high, both economically and in social terms. Decisions we make now will determine who will benefit from the technology and who will have access to what information on what terms—foundational elements of our future society.

It was in this context that the National Science Foundation, based on a recommendation by the Federal Networking Council Advisory Committee, commissioned a study of the issues by the Computer Science and Telecommunications Board of the National Academies. The report, *The Digital Dilemma: Intellectual Property in the Information Age* was published by the National Academies Press in February, 2000, and is the focus of this article.

ORIGINS OF THE PROBLEM

The difficulties we face today arise primarily from two sources. One source is a trio of technological advances that helped produce the infrastructure: the increasing use of digital information, the widespread reach of computer networks, and the creation of the World Wide Web. The second source is the emergence of computers as a routine part of everyday life. These two developments fundamentally altered the landscape; their consequences present a significant array of challenges.

Information In Digital Form

Information in digital form is orders of magnitude easier, faster, and cheaper to reproduce than is information in analog form (e.g., hard copy). Digital copies are also perfect, so each copy can in turn become the seed for additional perfect copies, quite unlike the situation with traditional media such as photocopies.

These properties of digital information are for the most part widely acknowledged. Somewhat less well appreciated is the fact that accessing digital information inevitably means making a copy, even if only an ephemeral copy. This copying action is deeply rooted in the way computers work. For example, when you view a page from the World Wide Web, several copies are made automatically—one so the document can be sent from the remote computer to your computer, a second when the document is loaded into memory, and yet another when it is displayed on the screen.

Such copying occurs with all digital information. Use a computer to read a book, look at a picture, watch a movie, or listen to a song, and you inevitably make one or more copies. Contrast this with the use of traditional media: reading a book does not involve making a copy of it, nor does watching a movie or listening to a song.

This intimate connection between access and copying has considerable significance for intellectual property protection. One of the essential elements of copyright—the right to control reproduction—works as expected in the world of traditional media, where there is an obvious distinction between *access* and *reproduction*, and where the copyright owner's control of reproduction provides just that. But in the digital world, where no access is possible except by reproduction, complete control of reproduction would mean control of access as well, a consequence of considerable importance to all stakeholders.

Networks and the Web

Computer networks have radically changed the economics and logistics of information distribution, enabling information to be sent worldwide, almost for free, and (for items of reasonable size) almost instantaneously. The Web in turn has radically altered the economics and logistics of publication, allowing everyone to be a publisher with worldwide reach. Where reproduction and distribution put information in the hands of those who know they want it, publication makes people aware of information that is available, a function the Web performs well. The astonishing variety of documents, opinions, articles, and works of all sorts on the Web demonstrate that millions of people worldwide are making use of that capability.

This trio of technological developments—digital information, computer networks, and the Web—have together been the source of profound changes. Digital information radically changes the economics and character of reproduction, computer networks radically change the economics and character of distribution, and the Web radically changes the economics and character of publication.

For publishers, these technological developments have opened up new markets and new products, such as online music and books. But the same developments offer advantages to individuals or pirates making and distributing unauthorized copies: the process is orders of magnitude faster, easier, and less expensive. The important result is that the natural barriers to infringement have eroded significantly. Where unauthorized reproduction and distribution of hardcopy works is limited in part by the difficulty, expense, and loss in quality inherent in the process, with digital information, perfect copies can be made and distributed almost for free. Stakeholders on all sides of the issue wonder whether something can be put in place to restore the balance of forces, or whether the world has changed fundamentally and permanently.

Commonplace Technology

The second major source of difficulties in the digital dilemma is the routine presence of computers and the Web in work settings, and increasingly in households. What was not long ago found only in research laboratories is now a broadly available consumer product. One consequence is that individuals routinely have the means and opportunity to access and copy vast amounts of digital information, including software, text, and audio and video material, but have no clear picture of what is legal or acceptable. They are as a result unprepared to deal with the IP issues the technology brings. Corporations dealing with these issues turn to their legal staffs, but individuals are bewildered, if indeed they are aware of the law at all.

A second consequence of the diffusion of the technology into everyday life is that intellectual property law and its enforcement are becoming increasingly concerned with private behavior. Copyright has traditionally been concerned with public actions that have public consequences, such as public performance, public display, or the dissemination of copies. It has also focused on behaviors of organizations or individuals whose actions have large-scale public consequences. But with computer now commonplace in the home, individuals can do in private what once would have required substantial investment and perhaps criminal intent. As the potential impact of private behavior has grown, so correspondingly has interest in regulating that behavior. This shift in the focus of IP law represents an important consequence of information technology's emergence into everyday life and presents another social and policy challenge in managing the IP balance.

CONSEQUENCES OF LICENSING

One of the many problematic issues raised by the digital dilemma is the increasing use of licensing rather than sale as the primary mechanism for distributing information. Licensing is of course in some respects quite familiar: Mass market software has long been distributed via a (shrink-wrap) license that gives the purchaser a right to use the software, but provides no ownership in it. More recently, a wide variety of digital information is being licensed, including documents, databases, and images.

The difference between selling a work and licensing it is significant. The sale of a physical copy of a work has been the dominant model for transferring IP to the consumer for more than 200 years in the U.S., where it involves the complete transfer of ownership rights in a particular copy of the work. Copyright law explicitly anticipates the sale of intellectual property products and, by its “first-sale rule,” gives the purchaser a significant body of rights in the purchased copy.¹ The purchaser is, for example, free to lend, rent, or resell the purchased copy. In that sense, copyright law follows IP products into the marketplace and promotes the continued dissemination of information.

Licenses, by contrast, are contracts, private agreements that provide for limited transfer of rights to use an item. They can involve a wide range of terms and conditions, but, unlike copyright law, need not incorporate any public policy considerations beyond some basic limits on what constitutes an enforceable contract. To the extent that digital information is distributed by license, then, there is no statute, history, or tradition of incorporating public policy considerations such as fair use.

Licensing can have advantages—it may provide clarity on terms and conditions of access, and it may provide for increased rights for the institution that go beyond those provided under copyright (e.g., the ability to make unlimited copies for local use). Licensing may also increase the options for making information available. For example, a license may grant time-limited access to some part of a book or report, perhaps for far less than the purchase price for the entire work.

But there are also concerns about the impact licensing may have on public access. Consider libraries as the archetypal example. In the print world, a library's failure to renew a subscription or buy an updated version of a book has no effect on the availability to patrons of earlier volumes or editions. In the world of licensed information, however, ending a subscription to an electronic publication may mean the end of access to earlier volumes or editions as well. While some libraries and publishers have worked to negotiate licenses that preserve fair use and other public access features, concerns remain about the use of a mechanism that lacks any of the built-in protections for public access embodied in copyright law.

Questions also arise about the interaction of licenses and copyright. For example, copyright law currently gives owners of copies of computer software the privilege to make backup copies. Can that privilege be taken away by a shrink-wrap license? Or can a license term prohibit disclosing flaws to other potential users? These and related questions are far from resolved.

If licensing becomes the dominant means of distributing information in mass markets, additional concerns arise for works that are considered part of our intellectual and social heritage.

¹ The first sale rule is contained in Section 109 of the copyright law, and says, in part: “... the owner of a particular copy [of a copyrighted work] ... is entitled, without the authority of the copyright owner, to sell or otherwise dispose of the possession of that copy...” It's called the first sale rule because the rights of the copyright owner over a particular copy of the work are extinguished by the *first sale* of that copy.

Imagine a world in which novels, poems, and paintings, for example, are available only (or mostly) by license rather than sale. The consequences for public access are far from clear.

Finally, the trend toward licensing means that increasing amounts of information are delivered as experiences, rather than as artifacts. This is not entirely unfamiliar—where books are information artifacts, first-run movies are information experiences. We have lived with both for some time, but the difference matters. Buy a book and you own it forever; pay for access to a digital book and when the period of service is over you often retain nothing.

RESPONSES TO THE DILEMMA

There are a variety of actions that can be taken in response to the challenges posed by the dilemma. We focus here on four: technological protection mechanisms, innovative business models, taking a broader perspective on the problem, and rethinking the nature of copyright.

Technological Solutions

The key technical problem in large-scale management of digital information is determining how to provide access without giving up all control. A variety of clever schemes have been proposed, many of which rely on a combination of encryption and rights management software. Encryption encodes information so that it can be accessed only with the appropriate key; rights management software enables fine-grained control of access, specifying such things as the number of accesses permitted, whether the material may be printed, etc.

A common scenario for using encryption involves making it both machine-specific and persistent. Encryption can be made machine-specific by incorporating into the encryption process some reliable property of the decryption/playback device. For a computer, for example, the serial number of the hard drive or CPU might be used. Then, even if both the file and encryption key are passed on, the information remains inaccessible because it cannot be decrypted on a machine with different identifiers.

But if the original purchaser decrypts the information what prevents that person from passing on the decrypted file? Persistent encryption tries to address this, narrowing as much as possible the window of opportunity during which the decrypted information is available. In this approach information is decrypted just in time, that is, just before it is used, making it available as briefly as possible, and then only in small chunks at once. It is never stored, even temporarily.

Rights management involves providing some indication of the rights the consumer has purchased; this information is either encrypted along with the content itself or possibly maintained on a separate “rights server” accessed over the Web. Software on the user’s machine then controls access to the information, consulting the rights listing to determine whether a requested action is permitted.

Technological solutions can be useful, but are limited in a number of important ways. One constraint arises from need for consumer devices to be simple and fast. In some cases this precludes the use of industrial-strength encryption, which, though nearly impossible to crack, may be too slow for some consumer uses. The less powerful encryption systems used in commercial products, however, have routinely been cracked (e.g., the content scrambling system used to encrypt DVDs).

Second, anchoring content to a specific machine raises an interesting problem for consumers: What happens when you upgrade your computer (or other playback device)? Must you repurchase everything you bought previously?

Finally, there is substantial difficulty in attempting to provide end-to-end protection within a general-purpose computer. PCs have been successful to a significant degree because they have open architectures; that is, components of the machine (e.g., the hard drive, the video card, etc.) are accessible to the consumer. As long as the machine is designed this way, decrypted information can be intercepted and captured as it passes from one place to another inside the machine. Hardware and software designers could make such steps progressively more difficult, but the effort they would have to expend, and the consequential costs for all involved, would be substantial.

Business Models

In general, then, while technology can play a useful role as a deterrent to unauthorized copying, it is far from a panacea. Consider, then, a second possibility: the use of innovative business models. By selecting an appropriate business model, a rights holder can at times significantly influence the pressure for and degree of illegal commercial copying and unauthorized reproduction.

Several general principles offer insight into the sorts of business models that can help. One principle suggests making the product cheaper and easier to buy than to steal. The basic point is to reduce the motivation to deal with unauthorized sources from the outset. Music provides one example: single tracks can be bought online for 99 cents. But why pay for something you might be able to get for free? Because, even with peer-to-peer programs (like Napster and Gnutella), the music isn't really free; there's an investment of time and effort in finding and in downloading it. With the informal, all-volunteer effort that is at the heart of all the file-sharing schemes, the difficulty of finding what you want may be substantial² and the time required to download (from an overloaded personal computer sitting in a dorm room somewhere) may be significant. Would you pay a dollar to avoid the hassle? Quite likely. Hence even beyond the obvious issues of ethics and legality, the answer to the question "Why pay?" is, for service, reliability, and speed.

A second general principle suggests reconceptualizing the basic product. The digital music business may in fact be primarily a service business, rather than a product business. As the value in the digitized music itself becomes increasingly difficult to protect (because digital information is so difficult to share without losing control of it), the value may reside in the service of providing speed, reliability, and ease of access, rather than in artifacts like CDs or tapes.

A third principle is illustrated by writer Stephen King, with the online publication of *The Plant*. It appeared online in installments, and while downloading an installment was unrestricted, the author warned that if too few people paid for their downloads, the installments would stop coming. This is intellectual property approached not as law, but as economics (pay or it dies) and as sociology ("No stealing from the blind newsboy" is King's elegant description).³

These general principles suggest thinking about the digital information business in innovative ways and lead to a number of other nontraditional business models. Each of these models, importantly, creates an environment in which there is significantly less need for IP protection:

² Looking for a song by Neil Young and can't find it? Try looking up "Niel Young."

³ The experiment was suspended after a few months for a variety of interesting reasons; Mr. King offers his explanation of events at http://www.stephenking.com/sk_120400_2.html.

- Give away the product; make money from an auxiliary service. The Linux operating system, for example, is given away, yet a number of companies are in the business of providing service, consulting, customization, and extensions.
- Give away the product; sell upgrades. Many antivirus vendors make available free, fully functional versions of the programs on their websites. They give away the program in order to sell subscriptions to the regular updates they make to the virus database.
- Give away one piece that promotes another. Adobe gives away their Acrobat Reader software to popularize the Acrobat PDF format and to create a market for all their other programs that create PDF files.
- Offer extreme customization. Custom CDs with a particular customer's selection of audio tracks are likely to be a less appealing target for reproduction.
- Offer a mass market product at a low price and high volume, with frequent improvements. Many software products fit in this category. Keeping the price low reduces the pressure for piracy, while constant improvements mean that the damage from unauthorized reproduction is time-limited.

None of these solves the problem completely, but each one can sharply reduce the need for IP enforcement.

There is also a more general point here about the relative power of law and business models: although legal prohibitions against copying are useful against large-scale pirates, they are unlikely to be nearly as effective against individual infringers, where detection and enforcement are problematic. Where such private behavior is concerned, business models may offer a far more effective means of dealing with IP issues.

The Need for Multiple Perspectives

The problems that arise from the interaction of IP and the information infrastructure need to be considered in a context that encompasses not only law, technology, and markets, but economics more generally, as well as psychology and public policy.

This multiplicity of views is important in three ways. First, each of them brings a fundamentally different approach and mindset to the problem, phrasing the basic question differently. As all of them are relevant, there is power in considering each of these different conceptions of the problem. Second, some disagreements arise because the positions are grounded in different perspectives (e.g., law rather than economics), and are thus in effect asking (and answering) different questions. Finally, being aware of the multiplicity of perspectives may open up additional routes for dealing with issues; not every problem need be legislated into submission.

Is Copy Still the Right Concept?

One final response to the digital dilemma asks a fundamental question about intellectual property. All of the preceding discussion accepts a fundamental perspective that underlies copyright, namely, the concept of *copying* as a foundational legal and conceptual notion. As the very name of the law indicates, the right to control reproduction is central to copyright. Deciding whether a work has been copied has, as a result, been a fundamental question underlying much of copyright history and analysis.

But is the notion of copy still an appropriate conceptual framework in the age of digital information? Two reasons suggest it might not be. One reason, discussed earlier, is that legitimate use of a digital work *requires* making a copy. Hence, noting that a copy has been

made tells far less about the legitimacy of the behavior than it does in the hardcopy world, where there are few legitimate rationales for copying an entire work.

A second reason is that, because copying is so bound up with the way computers work, controlling the act of copying, in the view of some legal scholars, provides unexpectedly broad powers, considerably beyond those intended by the copyright law. In the world of physical works, once a work has been published, the rights holder cannot in any pragmatic sense control access to the copies distributed. Social institutions (such as bookstores and libraries) and individuals with copies enable any motivated reader to gain access to the information in the work.

But when access requires reproduction, the right to control reproduction is the right to control access, even to an individual copy already distributed. Authors would not, of course, routinely deny access to their published digital works. But because access requires reproduction, control of reproduction provides control of access to individual published copies, a right not conceived of as part of copyright, and hence not to be embraced lightly, whether or not routinely exercised.

Considering that control of reproduction is a means, not the goal, can we find some other mechanism that is more tightly connected to the goal, whether in the digital or analog world? This will not easily be done, but one suggestion may help promote serious consideration of the issue.

It may be useful to start not by asking whether a copy has been made, but by considering what the law is attempting to achieve—ensuring progress in the sciences and arts—and ask instead whether a use being made of a work is substantially destructive of a common means of achieving that goal, namely, providing incentive to authors. This approach is similar in overall spirit to the concept of fair use, which requires consideration of the impact on the market for the work or on the value of the work. But it is somewhat broader in scope, as incentive arises for authors in more than the marketplace alone, coming as well, for example, from the ability to control the time, place, and manner of publication.

This view would not conflict with all of the other traditional exclusive rights in copyright. Creation of derivative works, distribution, public performance, and display of the work can all be conceived of and protected on grounds independent of whether a copy has been made. They also have impact on incentive, whether via economic effects in the marketplace or other factors, and hence would be consistent with an incentive-based analysis.

Any such substantial change would of course also bring problems. There would have to be a substantial period of familiarization, and a means of dealing with the tension between trying to make such a law easier to follow (by drawing a sharp line defining what constitutes incentive-destroying use) and keeping the criteria more general, as is the case for fair use, so the law can address unanticipated situations in the future. Nevertheless, the discussion and the examination of what constitutes a well-grounded model of IP protection in the digital world may well be worth the effort.

LOOKING AHEAD

The development and deployment of the information infrastructure presents a variety of challenges, ranging from the pragmatics of enforcing laws that can be casually broken by individuals (in private, inexpensively, and almost undetectably), to the possibility and perhaps the need to rethink some of the foundational concepts underlying those laws. Yet, from its origins in patent laws of Venice in the 1400s through to modern times, intellectual property law

has evolved and changed in response to such challenges. What is likely to happen this time? Several things.

To some degree, society will simply adjust to the new reality and carry on in familiar ways. Recall how software vendors gave up on the awkward technical mechanisms used in the 1980s to defeat piracy (e.g., distributing on disks that could not be copied in the ordinary way), and continued to do business and prosper in a world where there is nontrivial piracy. So it is likely to be with the distribution of digital content. Customers will grow used to subscriptions to online information, and authors and publishers will continue to do their work in the presence of some unauthorized reproduction.

But as the turmoil in the music, publishing, and movie industries suggests, the upheaval created by new technology may mean that accommodation is not enough. New business models will need to be explored and tested. New approaches to IP issues will need to be founded on more than law and technology, embracing as well an understanding of the economics of information, sociology, and psychology. And IP itself may need to be conceptualized to some degree, in recognition of the changes we now face.

The issues are difficult, but inescapable. We are all unavoidably engaged in an experiment that tests our ingenuity and resourcefulness in finding ways to accomplish the apparently paradoxical goals of motivating individual creation while still reserving the ultimate benefits of that creation for the good of society. And we need to do this in a world where replication, distribution, and publication are astonishingly easier and less expensive than they have ever been. The issues are as important as they are difficult: The decisions we make will shape foundational elements of our future society.

Randall Davis is a professor in the computer science department at MIT. He chaired the National Research Council committee that produced the report, *The Digital Dilemma: Intellectual Property in the Information Age* (National Academy Press, 1999), upon which this article is based. This article appeared in *Communications of the ACM*, Vol. 44, No. 2 (February 2001), Pages 77-83.