

Managing Access to Digital Information: Some Basic Terminology Issues

by Patrice A. Lyons

Patrice A. Lyons is an attorney with the Law Offices of Patrice Lyons, chartered, Washington, D.C. This paper was prepared for the UNESCO International Congress on Ethical, Legal and Societal Aspects of Digital Information, held in Monaco, March 10-12, 1997.

Often, a marked technical advance stimulates a period of intellectual progress. It is widely recognized that the printing press was such a development. Whereas, before this invention, only a few books were laboriously produced, and fewer still were available to the public, the printing press opened the doors for sharing information with a much larger audience. There is little doubt that this new procedure for communicating ideas had a major impact on civilization. Other data structures besides books, such as newspapers, monographs and journals, also emerged to take advantage of the capabilities of the printing press.

In this century, radio and television technology ushered in a yet more diversified medium of communication. In addition to expressing ideas with printed text and illustrations, information could be widely shared in a dynamic form consisting of a series of related sounds and images. While the data structure understood as "the book" played (and continues to play) a leading role in the print-on-paper world, a unifying structure, known as "the transmission program," facilitates the origination and transmission of information in the broadcast, cable and satellite communications industries. This unit for organizing and identifying information has generally been regulated under communications and trade laws, but it also has implications for the application of copyright law in a communications environment. For example, the North American Free Trade Agreement (NAFTA) makes provision for the protection of "encrypted program-carrying satellite signals."

Like books and transmission programs in the past, what logical entities are most appropriate to facilitate commerce in creative works in a digital environment?

Over the last decade, there has been substantial growth in the use of computer networking capa-

bilities for the creation and dissemination of copyright works. Of particular note is the emergence of the Internet. For definition of *Internet* see

http://www.fnc.gov/Internet_res.html

This phenomenon is not a unique situation in the history of intellectual progress. It has been a distinguishing feature of human potential to challenge existing assumptions, to reconceptualize given knowledge and to generate diverse informational materials and artifacts for entertainment, educational, industrial and other purposes. Technology has simply helped to accelerate the process.

The widespread availability of global information systems like the Internet carries with it the potential to generate and share information at a degree of complexity and pervasiveness that was unimaginable until recently. Already, information is being posted on the 'Net that would otherwise only be available to a restricted group, if anyone knew of its existence. Unlike transmission programs consisting of sounds or images that are produced solely for communication to the public in sequence and as a unit, digital information is inherently malleable. Information expressed as sequences of binary digits (or bits) may be accessed interactively, data streams from widely distributed sources may be intermingled and new works dynamically generated and processed.

There is a growing perception in the research community, and increasingly by leaders in copyright-dependent industries, that data structures are needed to enable the organization and identification of units of digital information for purposes of managing rights and interests in a network environment. Efforts in this direction are well underway. Of particular note is a framework under development that will enable copyright works and other information resources, once configured as "digital objects," to be reproduced,

stored, accessed and disseminated over computer networks in this new form of data structure. This architecture grew out of a program organized and led by the Corporation for National Research Initiatives (CNRI) under the sponsorship of the U.S. Defense Advanced Research Projects Agency and with the active participation of the U.S. Copyright Office of the Library of Congress. Fundamental aspects of this information infrastructure were described in a paper entitled "A Framework for Distributed Digital Object Services" by Robert Kahn and Robert Wilensky. It is available on the Internet at

<http://www.cnri.reston.va.us/home/cstr/arch/k-w.html>

Digital objects (sometimes referred to as packages, containers or, more generally, structured bit sequences) and their supporting technologies have emerged as a focus of experimentation. In this context, a digital object is understood as one or more sequences of bits or sets of such sequences that contain "typed data" (to allow the sequences to be interpreted), and include a unique, persistent identifier for the object known as a "handle" (or, in certain instances, a "DOI"). The digital object is intended to be a generic means of structuring information in the digital world. A digital object may incorporate information in which copyright, patent, trade secret or other rights or interests may be claimed, although this need not always be the case. Key infrastructure components of an open architecture that supports digital objects are discussed in a Cross-Industry Working Team (XIWT) white paper entitled "Managing Access to Digital Information: An Approach Based on Digital Objects and Stated Operations" that is available at

<http://www.xiwt.org>

Digital objects may be deposited and stored in a network-based computer system or "repository" for possible subsequent access. Repositories may be operated in a variety of ways, spanning the range from individual storage depots to bulletin boards to broadcast stations on the Internet. From a copyright perspective, it is important to stress that a "handle" identifies a particular logical entity, i.e., a data structure, in which a work or other information has been embodied, but not the underlying information itself.

A unique and important attribute of a digital object embodying a copyright work is the capability of the object to incorporate data about itself. This information or metadata may include conditions for accessing the digital object and/or its underlying content, or an indicator to where such information may be available. The digital object may also enable a negotiation to take place where a user wishes to go beyond any conditions previously set forth in its metadata. This capability is an essential ingredient to enable and encourage the growth of commerce in copyright works in a digital environment.

Several organizations are now building testbeds to implement the digital object framework. These include two at the U.S. Library of Congress and another in the publishing community sponsored by the Association of American Publishers. Information on the publishers' initiative is available at

<http://www.doi.org>

A key goal in these efforts is to provide an open architecture that allows the identification and management of access to digital information. They seek to make both proprietary and non-proprietary information available in a structured and well-known way with open interfaces, protocols and object structures.

A digital object as a structured package of encrypted information may also facilitate the development of flexible and efficient mechanisms for managing rights or interests in a computer network environment. In this context, the keys can be managed and distributed independently from the digital object itself. This capability for managing rights or interests also applies where intelligent agents, structured as digital objects, act on behalf of rightsholders in a network environment to protect works embodied in such objects.

What is the copyright status of original works of authorship structured as "digital objects"?

When Congress revised the United States federal copyright statute about 20 years ago, it restated the two fundamental criteria of copyright protection: *originality* and *fixation in tangible form*. From the first U.S. copyright statute, which designated only "maps, charts and books," the copyright law has grown to include new forms of expression as creative and worthy of protection. The wording of the definition of *fixation*, however, limits this expansive intent. It specifically provides that a work is "fixed" in a tangible medium of expression when it is embodied in an authorized "copy" or "phonorecord." Generally, a *copy* for these purposes is a material object (other than a phonorecord). This limitation is not just a matter of passing interest in the context of U.S. law. The concept of fixation is important, since it represents the dividing line between the application of the federal copyright statute and any protection that may be available under State common law or statute.

What it means to be a *copy* also came up at the Diplomatic Conference on Certain Copyright and Neighboring Rights Questions convened by WIPO and held in December 1996. Specifically, the following text appears under the Agreed Statements concerning Articles 6 and 7 of the WIPO Copyright Treaty adopted by the Diplomatic Conference: "...the expressions 'copies' and 'original and copies' being subject to the right of distribution and the right of rental under the said Articles, refer exclusively to fixed copies that can be put into circulation as tangible objects." While the Conference thus clarified the intended meaning of *copies*, the meaning of *original* may require further analysis. In the United States, an original may be deemed to apply to the first fixation of a work in a tangible form; however, many countries extend copyright protection to what are sometimes termed *original* works without a fixation requirement.

This topic is particularly interesting to consider where "original works of authorship" for purposes of U.S. law (or what are sometimes termed "original works of the mind" under other bodies of law) are created wholly within a global information system like the Internet, and where, in this environment, there may be no material fixation (or copy)

generated, much less distributed. A novel interpretation of *materially fixed* might include a capability that supports "fixation on demand"; however, there would still be some inherent ambiguity about the status of such works prior to their fixation.

The development of a digital object infrastructure may enable the expansion of copyright protection to accommodate works that are not first fixed in a tangible medium of expression, or, in the case of material such as live broadcasts, that are not recorded simultaneously with their transmission. Introducing the notion of a structured, logical unit, i.e., a "digital object," may better accommodate the emerging capabilities of digital technology. These include, in particular, the deployment of such dynamic resources as intelligent agents. It may also avoid the use of ambiguous and oxymoronic terms such as *intangible copies*.

In addition to the existing requirement under U.S. law that an original work of authorship be "fixed in a tangible medium of expression" for federal copyright protection to attach, an alternative criteria may prove very useful in a network environment:

an original work of authorship structured in a persistent, uniquely identifiable medium of expression from which it may be reproduced, perceived, performed or accessed by any device or process for a period of more than transitory duration.

For purposes of this proposed new provision, *structured* may be defined to include digital objects and other equivalent data structures.

A digital object with its unique persistent identifier thus serves much the same purpose as a material fixation under U.S. law. Moreover, this concept may also prove of assistance in countries that extend protection without the need for a fixation. A capability of persistently and uniquely identifying a data structure in which copyright works, or performances of works, are embodied may encourage the development of a new marketplace for copyright works in a digital environment. Of course, where an original work of authorship structured as a digital object is actually fixed in a tangible medium of expression, copyright protection would subsist in accordance with current U.S. copyright law. My proposal would simply offer an alternative basis for protection to attach.

Should the processing and communication of bits be viewed as a distribution and/or a performance?

Questions have been raised about the classification of new creative works like MIDI sequences for purposes of copyright. Are they literary works? Musical works? Computer programs? Sound recordings? Further, what happens when users access a network-based repository of such works on an interactive basis, and the results of such access are disseminated over the Internet? Depending on the nature of the access request, the dissemination may not represent any particular sequence of bits that previously existed in that, or indeed, any repository. This situation is also likely to become increasingly prevalent where complex works, such as

knowledge-based systems, are made commercially available over the Internet to provide advice and guidance on a wide variety of topics.

Many information resources (configured as digital objects or not) that are now accessible to the public over the Internet may look and sound like conventional copyright works. Often, the term *multimedia* is applied to these capabilities, as if these resources were simple compilations of several traditional works, such as music, photographs, films or text, to be treated as what might generally be called *data*. It may be appropriate to regard these works as a whole as either computer programs or computer databases, or some combination thereof. However, a more accurate, comprehensive and flexible terminology to describe this emerging area is needed that reflects the realities of the underlying technology.

Information in digital form (whether of a purely symbolic or numeric character) is a purely conceptual entity; however, it may be represented as a real entity in the form of symbols or numbers fixed in a material object, where it is usually considered a "literary work" for copyright purposes. In light of the developing capabilities of digital technology, Committee No. 702 of the American Bar Association explored whether it might be helpful to establish a subcategory of literary works capable of behavior, to be called "digital works." In its 1996 report, the Committee proposed the following definition for discussion purposes: "Digital works" are literary works consisting of an ordered set of symbols from a discrete alphabet, such as computer programs or knowledge structures, that are capable of behavior when processed."

Such a provision is particularly important where a patented process may be involved in the performance of a digital work subject to copyright or where there may be patents involved in the methods used for structuring data.

If a consensus can be reached on what it means to be a "digital work," it may lead to a better understanding of what occurs from a copyright, patent and communications law perspective where information represented in some digital format is mapped into a waveform. Terms such as *digital communication* or *digital transmission* may not be adequate to describe the situation fully. It was the Committee's understanding that, strictly speaking, there are only continuous waveforms (or analog signals) in the real world. A "signal" is meant to be "digital" only in the conceptual sense that it is understood to contain a sequence of discrete symbols or bits.

Any sequence of discrete symbols that corresponds to the expression of certain information may be mapped into one or more continuous waveforms. For purposes of copyright, where this ordered set of symbols is viewed as a "digital work," the mapping of the information into a waveform by any device or process may be viewed as a performance of the work. There may be other performances of works that take place, not just at the source, but at the point of reception and within the network itself, where intelligent agents may be tasked with performing various operations. Certain of these performances may be deemed exempt from copyright liability.

Networks and network servers can generally be either

Digital Object Identifiers

active or passive entities in any communications system. As passive entities, they typically serve to communicate bits without essential change from a source to one or more destinations. As active entities, they have the ability to process the information in arbitrary ways. When the information is encrypted at its source, the processing options along the communications pathways are inherently more limited, but it is still possible to perform a limited set of functions within the network, such as aggregation, selective filtering and disaggregation. Thus, the extent of copyright liability for any given situation should be based upon the nature of the service being provided. There may be classes of operations performed on digital objects that have only a minimal, if any, impact on any underlying copyright works. While strictly speaking performances, such operations might be deemed to encompass the "distribution" of digital objects embodying copyright works. Complex operations would most likely bring into play the copyright right of public performance.

There may be rules and procedures developed for access to digital objects, or repositories of digital objects, that may overlap and impact in practice any copyright and other rights or interests that apply to the underlying information content. In the context of a digital object infrastructure, there has been some discussion of the notion of "access to per-

form stated operations on a sequence of bits." Whether, and under what circumstances, such operations should be accommodated under communications laws, and how protection at the digital object level will interact with any copyright, patent, banking, privacy, trade secret and other rights or interests in an object's contents, is an important area for continued discussion and experimentation. Where a copyright work is configured as an encrypted digital object, a new set of capabilities is introduced having great potential for the management of rights or interests in a network environment or even for indicating that there are no restrictions placed on access to digital information.

In summary, this paper has introduced the digital object as a logical structure for organizing information expressed as sequences of bits (like the book or the transmission program in other media). It compares the characteristics of digital objects, i.e., unique persistent identifiers, network accessibility and typed data, to the attributes of fixation in a material object and shows them to be generally equivalent. In addition, it introduces a notion of a digital work as a literary work that is capable of behavior and discusses some of the attributes of encrypted digital objects that may bring into play the copyright rights of distribution, as well as public performance, in a network environment.

The MIT Press

AFFECTIVE COMPUTING

Rosalind W. Picard

"Picard makes a compelling case for building computers that will recognize your emotions and use a shared emotional vocabulary for more natural, entertaining and effective human-machine interactions. She has judiciously sifted through the rich and varied research on emotion, and has combined that with a review of current technology to present persuasive evidence that affective computers will be a reality in the foreseeable future. You and your computer will one day have a different kind of relationship because of the ideas put forth in this groundbreaking book."

— Peter E. Hart, Chairman and President, Ricoh Silicon Valley, Inc.

"Do you have more physical contact with your computer than with other people? From digital voices to 'smart' clothing, computers that recognize, express, and 'have' emotions can 'get to know' you intimately, adapting to your needs instead of you to theirs. Picard explores boldly how emotion-capable machines are efficient and flexible, can prioritize, avert information overload and, through sensor feedback, hold up a mirror to teach you things about yourself that you otherwise might never learn. Affect in the machine may be a long-sought key to synthetic intelligence."

— Richard E. Cytoiwic, author of *The Neurological Side of Neuropsychology*

275 pp., 30 illus. \$27.50

TECHNOLOGY AND PRIVACY

The New Landscape

edited by Philip E. Agre and Marc Rotenberg

"Commerce, bandwidth, equity of access, copyright—these are some of the battles that are taking place in the Internet environment, but the one that will affect all of us is the battle over privacy. *Technology and Privacy* is an excellent survey of the different issues on the different fronts. This is important reading for journalists, legislators, activists, and citizens."

— Steve Cisler, Advanced Technology Group, Apple Computer

"[These essays are] brilliant and clearly written. They offer the most innovative thinking about this issue in a decade and a half. Agre truly moves the privacy debate ahead...a significant new contribution to the privacy debate, even among professionals who deal with these issues regularly."

— Robert Ellis Smith, Publisher, *Privacy Journal*

"This collection represents work of many of the most interesting voices in the area of privacy and information technology, skillfully woven together. For anyone interested in privacy, this is essential reading."

— Helen Nissenbaum, Princeton University

280 pp., 13 illus. \$25

<http://mitpress.mit.edu>

To order call 800-356-0343 (US & Canada) or (617) 625-8569. Prices higher outside the U.S. and subject to change without notice.