Chapter 5

Digital Information and Copyright

Contents

Boxes

chapter 5

Digital Information and Copyright

Introduction

On an optical disc there might be recorded the sounds of a hit song, the pictures at an art exhibition, or the complete contents of an encyclopedia. In any case the disc looks the same--all that is visible to the naked eye is a silver disc with rainbow highlights. Information is encoded on the disc in digital form. Digital information is weightless, compact, and essentially invisible. To make it visible, people need tools—not just hardware, like a computer terminal or a disc reader, but also software tools. Some of the software runs the hardware to make the information visible or audible, but equally important is the software that sorts and navigates so that a person can find the particular fact that is wanted. A single disc might hold the equivalent of 275,000 pages of text.

People are becoming increasingly familiar with digital information and the software to manage it. Word-processing, spreadsheet, database, and graphics programs are among the most popular types of software used on personal computers; many people use them at home as well as on the job. Word-processing software helps a person create and manipulate text information, store it for future retrieval, or print it out on paper. Spreadsheets help people to manipulate numbers and perform calculations. Database management programs are used to organize collections of data—from something as small as a personal address file to something more complex like the billing records of a major corporation or the medical records of the patients in a hospital. Graphics software can be used to make free-hand drawings or graphs and charts illustrating mathematical data. Different kinds of digital information need to be stored and used in different ways, so many more sophisticated software tools are being developed to capture information, organize it, store it, and then find it, filter out irrelevancies, and put it in useful form. (See box 5-A on storing data and box 5-B on storing text and images.)

The line between "software" (or computer program) and the "data" to be manipulated by the software is sometimes hard to discern. For example, "knowledge bases," the databases used in artificial intelligence programs, are not mere lists or records of facts. The knowledge base itself includes structured rules and relationships needed for making decisions. New programming techniques, like object-oriented programming, further blur the distinction between program and data by putting more of the intelligence of a system into the organization of the data and by making the data "active." In object-oriented programming, an abstract data type, called an object, contains both data and methods (a method describes the algorithm the object uses to solve a problem).[1] Thus object-oriented databases can represent a blurring of the program and the data it manipulates (or better stated, object-oriented databases illustrate the essential equivalence of program and data).

Though the line between data and software is blurring, it is still possible to distinguish data or digital information and the concerns that have been raised with respect to intellectual property. Digital information is not just numbers or words, it can be text, motion pictures, or music. Some of the concerns closely parallel those raised for software in other chapters. Others are unique. To understand the issues, it is necessary to quickly review the growing field of digital information or electronic publishing to see the range of technologies and products that are becoming available.

Electronic Publishing

There is a growing commercial markct for digital information. The term "electronic publishing covers a wide range of processes, products, and services ranging from traditional books and printed materials to works that are available only in electronic form. Digital information and computer technology is revolutionizing the publishing industry. In addition to commercial producers, a growing number of businesses and government agencies are creating, storing, and using documents in digital form.

[1] Phil de Lisle, "Putting OOP to Work for You," *Data Based Adviser*, vol. 9, No. 3, March 1991, p. 87.

Box 5-A—Storing and Retrieving Data

A database management system is a collection of files and a set of computer programs that allow users to efficiently modify and retrieve the files. Data may be organized in a number of different ways, but the important point about a database management system is that it allows users an abstract view of the data; that is, most users need not understand the details of where and how each piece of data is stored and maintained. The software is designed to offer different users a view of the data that is appropriate to the task that user is performing. Many programming languages for database management systems use the notion of a ***record as*** **a** basic unit of organization. In a bank's database, for example, each customer would be represented by a record, and within each record would be several fields--name, account number, balance, address, zip code, etc. Even though the information is only stored once, different employees can retrieve and use it in many different ways; to notify all customers in a certain neighborhood of the opening of a new branch, to send overdraft notices to appropriate accounts, or to make an address change to a specific account.

There are a number of approaches to structuring database systems. The *relational model* represents data and relationships among data by a collection of tables, each with several columns with unique names such as name, account number, balance, etc. A number of separate computer programs are associated with the database to allow users to perform transactions involving the data, e.g., paying interest or deducting withdrawals.

An *object-oriented* database is based on a collection of *objects in* which are stored *instance variables* and *methods. In* a bank's database, an account object would contain instance variables for name, account number, balance, etc. An important difference between object-oriented and relational databases is that in object-oriented databases, the software instructions for making use of an object are contained in the object itself. These instructions are called ***methods.*** Methods are bodies of computer code that can act on the object or cause the object to behave in certain ways under appropriate circumstances. For example, an account object may contain within it a method ***called pay-interest*** which adds interest to the balance. If the interest rate changes, or the bank introduces a new policy on paying interest, the method called pay-interest can be modified without affecting other parts of the object.

SOURCE: Adapted from Henry F. Korth and Abraham Silberschatz, *Database System Concepts (New York, NY:* McGraw Hill, Inc., 1991), pp. 1-21.

Print-Based Electronic Publishing

Many books and periodicals are now written, edited, and typeset on computers; they are only committed to paper in time to be delivered to the end user. Computer-aided publishing systems offer many advantages to publishers: information only has to be typed once and then the captured keystrokes can be edited, corrected, rearranged, or updated with relative ease. Pages can be laid out and "pasted up" directly on the computer. Graphics can be inserted, either by using graphics software to draw them on the computer, or by using a scanner to make digital versions of printed photographs or drawings. Then all the text and graphic information can be converted to a form usable by computer-based typesetters and ***in*** some cases, computer-driven printing machinery.[3] Newspapers were at the forefront of implementing computer-aided publishing,[4] but the techniques have rapidly spread to magazine, journal, and book publishing as well. Fairly low-cost desktop publishing systems, using personal computers and laser printers, have even brought these abilities to small businesses, community groups, and schools.

The easier manipulation of digitized information means that many different products can be derived from basically the same information. For example, magazines and newspapers can more easily produce different regional editions of the same issue. While the bulk of the text may be the same, articles or advertisements of purely local interest can be

[2] Electronic publishing does not have a single definition. Some people insist that "print-based' electronic publishing does not exist, and that only processes that deliver information in electronic form directly to the end user should be called electronic publishing. More often, however, use of computer technology and digital information in early stages of creation of printed materials is considered to fall under the electronic publishing rubric. For a discussion, see Michael R. Gabriel, *A Guide to the Literature of Electronic Publishing* (Greenwich, CT: JAI Press, Inc., 1989), pp. 1-14, and Oldrich Standera, *The Electronic Era of Publishing (New* York, NY: Elsevier, 1987), pp. 6-10.

[3] Oldrich Standera, Th, *Electronic* Era *of Publishing: An Overview of Concepts, Technologies and Methods* (New *York, NY:* Elsevier Science Publishing, 1987), p. 157.

[4] Ibid., p. 110.

Box 5-B—Storing Text and Images

Text storage and retrieval: Text retrieval system, text base management system, or text data management system are terms for computer-based storage and retrieval systems that store documents in machine-readable, character-coded form so that they can be retrieved by a user or processed by a computer.[1] Text databases differ from other kinds of databases described in box 5-A mainly in that the fields are very large-often the size of a whole article or book. Text databases also differ from word processing systems, which also store text in character-coded form, in that text retrieval systems have much more powerful capabilities for creating indexes to significant words in the text and giving rapid access to text segments that contain the specified character strings. The software of text retrieval systems typically allows a user to search the entire text database (or text base) for all occurrences of a specific word or a phrase. More complex searches may be constructed using Boolean logic operators (e.g., using the terms AND, OR, and NOT to restrict or broaden a search), wild cards (e.g., searching for "creat**" to locate all instances of "creator", "creative", ''creativity", etc.), proximity matching (searching for ''nerve" only if it appears within 20 words of the word "brain' '), and other search tools.

Providers of on-line information developed proprietary text retrieval software for use on their own mainframe computers beginning in the mid-1960s and 1970s, with the earliest implementations limited to relatively brief documents like bibliographic citations and abstracts. Today text systems exist to handle documents of virtually any length, and software is commercially available to run on hardware of many kinds, including personal computers. Text retrieval systems are coming into use by corporations and government agencies to manage internal libraries of letters, reports, legal briefs, and other documents. Text retrieval software is also used in scholarly work to analyze electronic versions of books and other documents. Some systems allow users to create notes and annotations that can be electronically linked to specific areas of the text.

Documents may be put into the text base through direct keyboard entry, by transferring text files from other computers and word processors, or by converting printed documents to digital form through optical character recognition (OCR). Of these three input methods, transferring files is the easiest—most text retrieval systems are designed to ' 'import' digital text from any source. For documents that exist only on paper, scanning with OCR can often turn out to be less expensive than keyboarding, but it is not yet a problem-free method of converting printed text to digital text. Changes in type style or blemishes on the paper can cause scanning mistakes that must be found and corrected by keyboard. Despite advances in sc arming technology in recent years, the conversion problem prevents many organizations from replacing paper archives with digital text bases.

Although online document storage saves space compared to storage of paper letters, legal documents, books, or reports, it still requires disc space. Data compression algorithms and more sophisticated indexing algorithms are aiding in reducing these storage requirements. Some compression techniques can reduce data storage space requirements by as much as 75 Percent.* Improved indexing algorithms allow both for faster searches and reduced storage needs for the index database (a complete index of all occurrences of all words in a document can be nearly as long as the document itself unless space-efficient methods are used).[3] Text bases can be stored on both magnetic and optical discs.

Image storage and retrieval: Image storage captures a document's appearance, rather than its content. An electronic document imaging system uses a scanner to convert documents to a form that can be stored digitally. The most widely used scanners divide the document into many tiny areas called pixels (picture elements), measure the light reflected from each pixel, and send a corresponding electrical signal to image processing circuitry which converts the signal to a stream of digital code.[4] The scanners are fairly reliable and simple to operate compared to OCR. This easier input task is an advantage, as is the ability to copy the exact appearance of a page of text, including any associated photos or graphics. A disadvantage is that text stored in image form usually can not be directly manipulated, analyzed, or searched by text retrieval software; each image must be properly categorized and linked with index terms when it is entered into the system. Images also require a great deal of disc storage space even after

1 William Saffady, *Text Storage and Retrieval Systems:* A *Technology Survey and Product* Directory (Westport, CT: Meckler Corp., 1989), p. 3.

2 Ibid., p. 25.

3 See, e.g., Dennis Allen, ''Text Retrieval Witb a Twist," *Byte,* July 1989, pp. 201-204.

4 For more information on image storage, see William Saffady, *Personal Computer Systems for Automated Document Storage and Retrieval* (Silver Spring, MD: Association for Information and Image Management, 1989), pp. 19-27.

Continued on next page

Box 5-B—Storing Text and Images-Continued

processing by compression algorithms. High-capacity optical discs are often the storage medium of choice for image databases. Not only pictures of documents but also drawings, photographs, and images produced by some medical diagnostic equipment can be stored in this way. Images are retrieved by searching the index database in many of the same ways mentioned above (words, phrases, Boolean search strategies, etc.). Retrieved images may be viewed on a high-resolution screen or printed out on a laser printer or other appropriate printer.

Storing and exchanging compound documents: There are many different types of documents and different formats for storing data. Documents produced by one word processing software package, for example, may not be readable by another because they use different conventions for indicating format changes (type styles, paragraph indents, boldface, etc.). Format becomes increasingly important for documents produced by sophisticated publishing software or by multimedia systems—these documents may include extensive format information about type fonts, photos, graphics and other non-text information. If a document is stripped of this format information and reduced to its simplest terms (e.g., when a text document is converted to a file of ASCII[5] characters) it may be read by a wide variety of software packages, but so much formatting information maybe lost that the document is useless for certain purposes. Even if it is still readable, a document so treated may no longer be revisable, that is, it cannot be edited or updated.

There are a growing number of situations in which fully formatted, revisable digital documents need to be exchanged by organizations that use different software. Much architectural and mechanical design work, for example, is done on computers. Exchanging and storing drawings, specifications, manuals, and other documents in electronic form rather than on paper can reduce storage and maintenance costs. (For example the B-1 Bomber reportedly has over 1 million pages of documentation, 200,000 pages of which must be updated yearly. The Department of Defense, through its Computer-Aided Acquisition and Logistics Support strategy—CALS--is in the process of requiring that all information submitted by contractors developing weapons systems be submitted electronically. [6])

Interchange of complex documents, independent of particular software or hardware environments requires another level of standardization-standard ways to describe how data was handled by the originating software so that the receiving software can handle it in an appropriate and compatible manner. An approach to this problem is use of a‘ ‘document description language’ or ‘metalanguage’; the ‘‘tags’ or ‘‘labels’ generated by the document description language are included in the document, A person reading the text mayor may not see these tags; their purpose is to describe the document to the receiving software. Several such languages have been developed and are competing in the marketplace; no clear standard has yet emerged. Use of document description tags within the document greatly increases the size of the document and thus the storage space requirement on disc or other storage medium,

5 ASCII stands for American Standard Code for Information Interchange. It is the standard 7-bit code for transferring information On local and long distance telecommunications lines.

6 Brooke Stoddard, ‘‘Shedding the Burden: The Federal Government Spent $130 Million on Electronic Image Processing Last Year, ” *Government Computer News,* Apr. 29, 1991, p. S4.

inserted in the different editions. The completed electronic text can be sent via telephone lines to several printing and distribution facilities throughout the country, thus reducing mailing costs and speeding delivery to subscribers.

In the book publishing arena, materials can be customized ‘ ‘on demand. For example, publishers can respond to requests from college professors to create textbooks that only include those chapters that are actually to be used in their courses. By selecting text chapters from an electronic database, publishers can make several different versions of a standard text for different universities. Authors and publishers also find it easier to keep books up to date or to add new chapters if the books are stored electronically until it is necessary to print them.

Once information has been placed in digital form for publishing, it is also possible to use it for other purposes. Some newspaper and magazine publishers also sell electronic versions of their publications in some of the forms discussed below.

Nonprint Electronic Publishing

The other type of electronic publishing, and the one that will get the most attention in this chapter, is the type in which information is delivered to the end user in electronic form.

On-line Information

One segment of this market is the on-line information retrieval service where end users retrieve information from databases housed in some distant computer via telephone lines or computer networks. In 1988 it is estimated that the four largest such services had a total of 1.3 million subscribers and annual sales of over $560 million.[5] Costs to subscribers are usually based on a subscription fee plus usage charges for actual use of particular databases, citation charges for printing or downloading some results, and telephone line charges.

The vast majority of subscribers to these commercial services are organizations—businesses, libraries, universities, etc.—whose employees need access to such services. Ten minutes of on-line searching might cost anywhere from $3.50 to $14.00[6] or more. Costs can mount up quickly if searches are complex, if expensive databases are used, or if users are not thoroughly trained. Other services aiming to give home users affordable on-line information services (as well as home shopping and entertainment) are currently a small part of the market in the United States. In Europe, Australia, and a few other countries, on-line information services intended for the home user are provided by the telephone company or a government agency and are usually called videotex.[7]

Most on-line information services offer their subscribers access to databases provided by a variety of publishers. The services negotiate contracts with the owners of the databases for the right to distribute them, and pay royalties based primarily on how much the database is used. Having access to many databases through one service is an advantage to users, allowing them to search more economically. For example, the largest service, DIALOG Information Retrieval Service, offers over 400 databases in four categories numeric data, directories bibl bibliographic records. and full-text records. Examples of numeric databases include, for example. stock and bond price quotations as well as many kinds of statistical and financial information, some based on government statistics. Directory databases include many standard reference works and handbooks, many of which are also published in hard copy. Bibliographic databases have citations to journal, magazine, or newspaper articles and sometimes also include abstracts of the articles cited. Full-text databases contain electronic versions of magazine, journal, and newsletter articles; they may be collections of articles from many sources, or actual electronic editions of full journall or magazine,

While many elcectrnically published journals are digital versions of the 'paper" journal, electronic publishing is gradually growing and changing the way informal ion is exchanged among research communities. Ten refereed electronic journals are now available on the Intcrnet,[8] a collection of research and information networks through which many universities. businesses. and government agencies share services 1ike electronic mail, file exchange, and access to distant computers,[9] Within the Internet community, and on other computer networks, them is a growing camp of researchers who view electronic publishing as an ongoing interactive process. Some experimental journals on the lnternet use publication of text as a tool to e1icit comment from other researchers.) One on-line project on the Genome Project at the Welch Library at Johns Hopkins University mounts text related to the Genome Project on an on-line database. Geneticicists students, and critics from around the country can

[5] Mary Lu Carnevale and Julie Amperano Lopez, "In Electronic Publishing, All Eyes Are Fixed on AT&T," *The Wall Street Journal*, Aug. 2, 1989. The four services are Lexis-Nexis-Medis, owned by Mead Corp.; Dialog Information Service, owned by Knight-Ridder; Compuserve Information Service, owned by H&R Block; and Dow Jones News/Retrieval Service, owned by Dow Jones & Co.

[6] These examples include discounts for high volume use. From *The DIALOG Service*, user catalog, spring 1991 (New York, NY: Knight Ridder, Inc., spring 1991).

[7] Gabriel, op. cit., footnote 2, pp. 77-88.

[8] Association of Research Libraries, "Statement of the Association of Research Libraries to the Subcommittee on Science, Technology, and Space, Senate Committee on Commerce, Science and Transportation for the Hearing record of March 5, 1991 on S. 272—The High-Performance Computing Act of 1991," March 1991, Washington, DC. In "refereed" journals, submitted articles are reviewed by a committee of scholars before publication. Publication in refereed journals generally carries more prestige in scholarly communities than publication in other journals or magazines.

[9] For more information on the Internet and on the future evolution of a National Research and Education Network, see U.S. Congress, Office of Technology Assessment, *The National Research and Education Network* (Washington, DC: U.S. Government Printing Office, forthcoming).

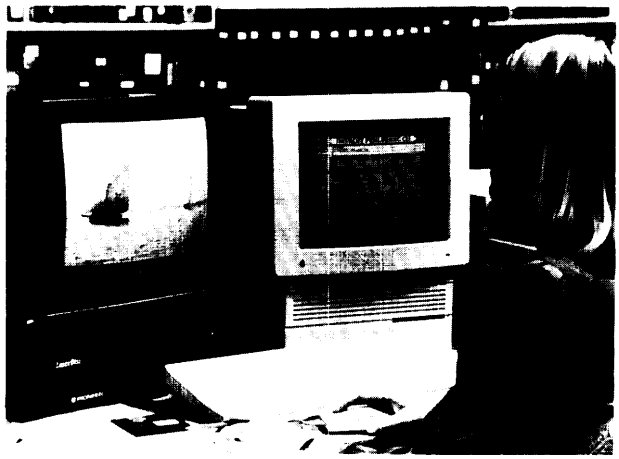

Photo credit: U.S. Library of Congress, American Memory Project

Multimedia databases can give users access to information in several forms-e. g., text, graphics, motion pictures, or sound recordings.

access the text and respond to the authors via electronic mail. The output in this case is an on-line textbook which changes constantly to reflect advances in the field.l" (See box 5-C for a discussion of how electronic publishing may change scholarly publishing.)

CD-ROM

Another increasingly popular way of marketing digital information is through publication of CD-ROMS. CD-ROM (compact disc, read-only memory) is an optical storage medium. CD-ROMs are essentially the same as the compact discs now used for musical recordings, although the data are stored in a different format and require a different player.[11]

Many of the databases available through on-line services are also available on CD-ROM. CD-ROMs are often made available to library patrons. Some of the databases on CD-ROM offer more ''hiendly" user interfaces than do on-line services and inexperienced or occasional users can search at their own pace without accruing huge bills for 'connect time' and telephone usage. However, the user might have to worry about holding up another user, since licensing requirements or equipment limitations often mean that only one person at a time can use a particular disc. Licenses are becoming available that allow multiple users to access CD-ROMs on local area networks; while a few vendors do not charge extra for network licenses, in most cases these are much more expensive than single user licenses.

CD-ROMs will become increasingly popular as more titles are offered and as the price of players and discs fall. Recent developments that simplify creation of CD masters (from which individual discs are stamped) will probably make this format more accessible to small publishers.[12]

Using Digital Information

Using digital information has both disadvantages and advantages compared to books or magazines. Information displayed on a computer screen is often not as comfortable or convenient to read as the printed counterpart. Certainly the computers most people use today cannot be conveniently used at a bus stop or on the beach, and research shows that people read 20 to 30 percent more slowly from screen than they do from a printed page.[13] Perhaps more important, the traditional search and retrieval aids we use with books (page numbers, indexes, tables of contents, visual memories of how something looked on the page) are not used in the same way with digital information.

But information in digital form has powerful advantages over printed documents. For example, retrieval software can search through and sort information to help a user find the specific information he/she wants, rather than reading through a whole book or using a (usually inadequate) printed table of contents. With a few keystrokes, a user can use an electronic index, receive a report on how many times a requested term appears in the text, and then actually look at each instance in context. This feature alone may not always be sufficiently helpful if it turns out that there are many instances (hundreds or even thousands) in which the desired term

[10] Ibid.

[11] For a discussion of different optical storage formats, see U.S. Congress, Office of Technology Assessment *Copyright and Home Copying: Technology Challenges the Law, OTA-CIT-422* (Washington, DC: U.S. Government Printing Office, October 1989), pp. 45-48.

[12] Tom McCusker, "CD-ROM Production Power!'' *Datamation*, vol. 37, No. 4, Feb. 15, 1991, pp. 26-29.

[13] For a discussion of the problems of typographic design for computer screens, see Richard Rubenstein, *Digital Typography: An Introduction to Type and Composition for Computer System Design (Reading, MA:* Addison-Wesley Publishing Co., 1988), pp. 189-193. For research on human factors of reading from screens, see also John D. Gould, ''Reading Is Slower From CRT Displays Than Paper: Attempts To Isolate a Single-Variable Explanation, ' *Human Factors*, vol. 29, No. 3, 1987, pp. 269-299.

Box 5-C—Digital Information and the Scholarly Publishing System

This view of the effects of electronic publishing on the current system of scholarly publication was taken from an OTA contractor report:

In the scholarly world today, the printed version of knowledge has the function of creating an archive of knowledge more than it serves the function of the exchange of knowledge. This is more true in scientific and technical fields that have access to networks and computers than to the humanities that do not. Scholarly exchange on the network occurs through ''affiity group computing, " such as the 2,000 Listserv protocols on Bitnet. This kind of exchange is a very good example of gift giving, since it creates a scholarly community (remembering that the function of a gift is to create a social bond, not a profit).

The exchange of knowledge as a gift exchange system among scholars creates "the invisible college" of researchers. The strength of this culture is that it is governed by a search for truth; the weakness is that access to it is restricted. The invisible college traditionally occurred through ''old boy networks' meeting in face-to-face interactions, such as conventions. But today, the use of digital networks has expanded it greatly.

(It must be stated that this "gift culture" is possible because the rewards of scholarly research are not given by a market, but nonetheless they exist in the economic rewards of promotion and tenure. However, this is deliberate. The system of scholarly communication was setup through a deliberate system of subsidies, such as the creation of university presses and the higher rates that libraries paid to subscribe to journals. This system of scholarly communication has been destroyed as book and journal publishing moved out of universities and became profitmaking enterprises in the marketplace. Today no research library can afford to pay for the full range of scholarly journals; the price of scholarly journals is rising twice as fast as any other research cost. The destruction of the print-based system of scholarly communication is an excellent case study of what happens when the fine balance between a gift exchange system and a commercial market exchange system is destroyed. However, this very destruction may be driving the development of digital-based scholarly communication. See below.)

The invisible college of research activities today exists on the network. Since nearly every scientific and technical field is growing and changing much faster than the print publication process can reflect, the real exchange of knowledge occurs long before the publication process. Most scientists must actively seek "preprints" in order to find out the current state of research in their field; the actual publication in printed form only validates the contribution for historical reasons and creates an archive. The most interesting experiments in digital publication reflect this; for example, the Online Mendelian Inheritance in Man (OMIM) and Genome Data Bank (GDB) projects at Johns Hopkins University, in which current research findings are peer-reviewed online in databases, are available electronically throughout the world. In essence these projects have done away with the print publication process altogether. Similarly, there are now about a dozen online peer-reviewed scholarly journals which essentially do away with the print publication process.

This evolution has several causes. One is the rapid increase in the cost of scholarly print publications, remembering that scholarly communication was conceived of as a gift exchange process that has been distorted by the marketplace. But another is that digital media have some of the qualities of an oral culture, and oral cultures have traditionally been more effective in providing natural homes for gift activities.

The example of scholarly journals shows that if the marketplace is allowed to define knowledge solely as a commodity, the system of dissemination of ideas and subsequent intellectual innovation can break down. This is happening now, as libraries cancel journal subscriptions, and are not allowed to share subscriptions because of the limitations of print copyright. Exchange systems governing intellectual products on the network must be devised that encourage use by "the invisible college. "

SOURCE: Steven W. Gilbert and Frank W. Connolly, ''A Wealth of Notions: Regaining Balance as New Information Technologies Collide With Traditional Controls and Incentives for Intellectual Work" contractor report prepared for Office of Technology Assessment, July 31, 1991.

appears. Users of databases usually find that a major problem with digital retrieval is 'getting too much information. Most on-line database retrieval systems allow the use of several strategies to narrow the search to retrieve a reasonable amount of information. (See box 5-B.)

Retrieval and text analysis software are becoming increasingly sophisticated. Still, getting the precise information one wants without reading through lots of irrelevancies, or worse yet, missing something important, can sometimes be a daunting challenge; the challenge can be even greater if one doesn't

know which of several databases offered by different publishers has the information. In today's market, different publishers have different user interfaces and search protocols. Learning to use them all can be time consuming and expensive. A class of software tools called ' 'agents' or '' filters are coming on the market. These can collect information from multiple sources, including electronic mail, on-line news services, and internal corporate databases, and sort it according to the user's priorities and interests.[14] Some database providers are offering more sophisticatcd software tools that w will allow users specify a search and then have it automatically performed on a number of different databases offered by that provider. In some cases these new tools controvcrsial because~ owners of the databases disagree about how royalties should be calculated when such software is used. [15] As use of digital information grows, people are going to need even more sophisticatcd search tools. One group envisions ''knowbot programs' that will act as personal librarians in the future. These artificial intelligence tools would accept the users requests for information, search many different sources, and then ret urn the results in a form (hat would be most useful to the reader. [16]

Hypertext

One great advantage of digital information is that it does not have to appear to the user in the same order in which it was written by the author, nor indeed does it need to appear to different users in the same way if they have different needs. ' [7] The ability of computer software to link different pieces of information also allows information to be presented in innovative formats, ' 'Hypertext and "hypermedia' are generic terms for systems that link related pieces of information for presentation in a nonsequential manner. Hyperlinks give the reader the power of ''subjective linearity."[18] When the material is read, the reader chooses the particular items to be presented, and the order of presentation, depending on his or her needs or level of interest.

Dictionaries and encyclopedias, with their many discrete entries and heavy dependence cm crossreferencing are obvious candidates for hypertext, and some have been published in this format. With a hypertext encyclopedia, the user may begin reading an entry on ''elephants and upon seeing the mention of ''ivory may ' 'click on' (select) that topic. The hypertext software will then usually open a window with a brief discussion of the term; by clicking again, the user may then choose to get more details on the new topic or go back to the original topic. In this way, the user can search through many related topics without toting half a dozen volumes from shelf to table, and without trying to mark a place in several books at once. Electronic bookmarks and other aids help the reader navigate through the information.

Hypertext is useful for on-line help systems for computer users. Hypertext provides the ability for a user to go directly from an error message to the relevant section of the on-line user's manual or even to a tutorial program giving computer-aided instruction on how to avoid the problem in the future. CASE (Computer-Aided Software Engineering) systems have been created in hypertext, allowing soft ware developers to link and navigate through the various versions of the reports, documents, and code objects developed during a major software development project.[19] In the commercial world, prototype hypertext systems have been developed for financial auditing, a field that is heavily dependent on crossindexing and relating information from different sources. Hypertext also lends itself to advertising, product catalogs, and tourist guides.

Hypertext is also used for educational and scholarly work. For example, Harvard's Perseus project brings together information from ancient Greek

---

14 David S. Marshak, ''Fi]tcrs: Separating the Wheat From the Chaff, ' *Patricia Seybold's Office Computing Report*, vol. 13, No. 11, November 1990, pp. 1-15.

15 For example, see Mick O'Leary, ''Dialog and the American Chemical Society Play a High Stakes Game, *Online*, January 1991, pp. 15-20.

16 Robert E, Kahn and Vinton G, Cerf, Th *Digital Library Project, Volume 1:* The *World of Knowbots* (Mc Lean, VA: Corporation for National Research Initiatives, 1988), p. 60. ''Knowbot" is a registered trademark of CNRI.

17 A few novelists, e.g., Juno Cortazar in *Hopscotch* (New York, NY: Macmillan, 1972), have experimented with giving readers alternative sequences in which to read chapters in printed books. Nonfiction writers often give prefatory *warnings,* such as those cautioning lay readers to review explanatory appendices before starting difficult text. In general, however, it is difficult to overcome the linear nature of printed information.

18 Rick Gessner, "Building a Hypertext System: Hypertext for Every Programmer's Toolbox, " *Dr Dobb's Journal, vol. 15, No. 6,* June 1990, p. 22.

19 Ibid., pp. 44-45.

history, archaeology, and literature, including the original Greek text, translations, dictionaries, and scholarly articles. Hypertext links allow students to open windows where the original Greek writing and its translation can be viewed simultaneously, or to check the validity of scholarly interpretation by going directly from the citation to the underlying text.[20]

Mixed-Media Databases

Nor is digital information limited to data or texts. Anything that can be seen or heard can be digitized, including voices, music, photographs, and moving pictures. Databases that combine items from several such sources are called mixed-media or multimedia databases. In many cases, multimedia packages are complex affairs that may link together computers, compact disc readers, and several different display screens. Mixed-media databases are still in their infancy, but the software to make them work and the display technology to make them appealing and easy to use are becoming available.

One example of a mixed-media database is the *American Memory Project*, a prototype digital library developed by the Library of Congress. *American Memory* is a collection of photographs, pamphlets, films, and sound recordings dating from the late 1800s and early 1900s. The information is stored on a set of video discs (with photos and films in analog form) and CD-ROMs (with all other material in digital form) and packaged together with hypertext search software.[21]

When mixed-media databases are combined with hypertext techniques the result is the ability to combine and link many different types of information on the same topic. The Perseus system mentioned above, for example, contains many photographs of ancient Greek art and architecture, in addition to text. In some cases, students can see more detail in the screen image of the photos than they would be able to if they had access to the original items. Further, the hypertext links allow related works to be compared. A hypertext treatment of Beethoven's Ninth Symphony includes a CD recording of the Vienna Philharmonic performing the symphony along with different versions of the score, discussions of the music and of Beethoven's life, examples of his writings, translations of the choral movement, and a quiz in the form of a video game. Students can listen to the whole symphony, study one part in depth, or listen note by note as they move the cursor through the score.[22]

Digitized video and audio require a lot of storage spacc compared with text. Optical media such as CD-ROM are good storage media for such types of information, but even here there are limits. The capacity of one CD-ROM easily accommodates a 26-volume encyclopedia and leaves room for the text retrieval software, 15,000 illustrations. 45 animation sequences, and one hour of audio.[23] A C can normally hold about 540 megabytes or 275,000 pages of text,[24] but only about 74 minutes of high-fidelity digital audio and far less of full-motion digitail video .[25] Several digital compression methods can be used to reduce the space required for digital audio or video by 2 to 10 times, though most of these methods are ''lossy,'' that is, the playback version is not of the same quality as the original because some information is lost.[26]

Digital Libraries

If digital technology is changing publishing it is also changing libraries. Libraries have been experimenting with and investing in computer technology for 25 years. Patrons at many libraries across the country have been either thrilled or dismayed to find that computer terminals have replaced the card catalog. High school students no longer thumb through musty index volumes at the public library in

[20] Ibid, p 68.

[21] GaryH. Anthes, "Llt)riU} Releases Data via Laser Disc, ' *Computerworld*, Sept. 10, 1990, p. 53. Also Carl Fleishaur, Director, American Memory Project, [~. S Library of Congress, personal communication, January 1991.

[22] Robert Haawind, "Hypertext" The Smart Tool for Information Overload,'' *Technology Review*, November-Dcccmber 1990, p. 47.

[23] Jakob Nielson, *Hypertext and Hypermedia* {Boston, MA: Academic Press, Inc , 1990), p. 53

[24] David C. Miller, *Publishers, Libraries & CD-ROM Implications of Digital Optical Printing* (Portland, OR: Library and Information Resources for the Northwest, 19X7), p.7.

[25] For a discussion of compact disc, CD-ROM, and other optical formats, see (-1. S. Congress, Office of Technology Assessment, *Copyright and Home Copying*, op. cit., footnote 11, pp. 44-48 and Nielsen, op. cit., footnote 23, pp. 123-126.

[26] Edward A. Fox, ''ACM Press Database and Electronic Products--New Services for the Information Age,'' *Communications of the ACM*, vol. 31, N{). 8, p. 948.

search of magazine articles; they search CD-ROM databases.

For most libraries, the first step was to use computers to streamline acquisitions, circulation, and other administrative operations. An early development was the Library of Congress's MARC (Machine Readable Cataloging) system, begun in 1966. The MARC system provides standards for encoding bibliographic information in machine readable form; many libraries use it to create their own machine readable records as a basis for on-line catalogs. By the early 1980s, many research libraries had some on-line cataloging capability; in same cases it was available for patron use. Cooperative agreements among libraries have led to major bibliographic databases such as OCLC (On-line Computer Library Center) and RLIN (Research Libraries Information Network) which receive contributions from, and process inquiries for, thousands of libraries.

On-line catalogs are a great aid to library users, allowing them to search large collections, or multiple collections, for the resources they need. For example, a user can now search the catalog of the entire University of California system at any time, from home or office. Ten years ago, the task would have required visits to over 1100 card catalogs on nine different campuses across the State. In addition, the cross-referencing possible with an on-line catalog is far superior to what was possible with cards. For example, a user can easily identify all holdings of the library published in a given foreign language during a given time period—a search which would be impossible with a card catalog.[27] Some university and research libraries have also begun providing users with on-line access to journal literature, either citations or full-text form, by obtaining licenses to some of the commercial databases or on-line services discussed above. (See box 5-D.) The next step for some research libraries in the forefront of automation is the automated delivery of library materials that have been identified through a search of the on-line catalog; materials might be delivered either in either in hard copy or in electronic form. An experimental project at Carnegie Mellon University will provide links from catalog citations to either full-text records or digitized images of printed pages available on-line .28 The same project is also attempting to make extensive digital resources such as databases and electronically published journals available at any terminal on campus. Part of this project is the enhancement of the university's electronic catalogs to give more useful search information. Examples of enhancements include listing of chapter titles, separate listing of authors of stories, essays, or chapters in books, and abstracts of technical papers.

What will be the digital library of the future? Though many of the building blocks are in place, many say that its potential has not been realized, primarily because so much of the world's knowledge is still not in digital form and will not be for some time to come.

Copyright Issues for Digital Information

The previous sections pointed up some major differences between digital information and information in more traditional forms. These differences have been summarized as a list of six characteristics of digital information[29]:

1. Works are easily copied.
2. They can be easily transmitted to other users or be accessed by multiple users
3. They can be easily manipulated and modified
4. Works are essentially equivalent: text, video, or music are all reduced to a series of bits and stored in the same medium.
5. Works are inaccessible to the user without hardware and software tools for retrieval, decoding, and navigation.
6. Software also allows for new kinds of search and linking activities that can produce works that can be experienced in new ways.

These characteristics of works in digital form have implications for copyright because they change how authors create, the kinds of works they can

[27] Clifford A. Lynch, "Library Automation and the National Research Network, ' *EDUCOM Review*, fall 1989, pp. 21-27.

[28] Denise A. Troll, *Library Information System II: Progress Report and Technical Plan,* Mercury Technical Report Series, No. 3 (Pittsburgh, PA: Carnegie Mellon University, 1990), pp. 5-16.

[29] Adapted f'rem: Pamela Samuelson, ''Digital Media and the Changing Face of Intellectual Property Law, '' *Rutgers Computer & Technology Journal*, vol. 16, *No. 2, 1990*, pp. 323-340 and from discussion at the OTA workshop on ''Digital Libraries, Electronic Publishing, and Intellectual Property, " Feb. 11, 1991, Washington, DC. A similar set of issues was also developed in *Copyright and Home Copying*, op. cit., footnote 1 I, especially ch. 2.

Box 5-D-One User's View

The availability of digital information offers many new opportunities for people of all kinds, but it also gives rise to some uncertainties and confusions, as discussed in this first-hand account taken from an OTA contractor report:

When I was in high school and college, I took handwritten notes of the library material I would use for research papers. Often the teacher would put material on reserve and I would trudge over to the library and laboriously copy stuff out by hand. By the end of college in the late 1960s, copy machines were installed in college libraries, but the copy quality was not particularly good and the price was pretty high, at least compared to handwritten note taking. A few years later, in graduate school, copy machines were common in college libraries, copy quality was pretty good, and the price of a copy was coming down. Colleges had even put in copy service centers in the library so I didn't have to stand at the copy machine; I could take my stuff to a central location and someone else would copy it for me. Still, as in high school and undergraduate days, I did my research manually using print indexes. But with the cheaper price of making copies and the better quality, my professors began to compile packets of material which lessened the number of trips to the Reserve Room.

When I began my doctorate, on-line search services were available but expensive for a graduate student without a research grant, so again, most of my research was done manually, using print indexes. In contrast to my undergraduate days, I did little manual note taking, using the copy machine instead. Most of my doctoral research was done on ERIC (Educational Resources Information Center—a bibliographic database), which I poured through volume by volume, year by year, About a month before I finished my research, my library installed ERIC on CD-ROM—too late to be of much use to me. Now, however, I'm not even sure where the paper copy of ERIC is in my university library. Should I so choose, I can download from CD-ROM ERIC and use the information in my home or office computer. In addition, I can access the university/consortium catalog of holdings from my home or office computer. No longer am I bound to the physical location of the library as I was as an undergraduate. Nor am I bound to paper. Should I want to compile a bibliography from ERIC, I can simply reformat the information I've downloaded to a format acceptable for whatever purpose I'm using it for.

To me, at this point, that 'stuff' that I've gotten from ERIC is similar to music coming over the radio or a television program. Either the music or the television program I can tape-it's coming into my home; I have the technology to tape it and so I do. And that's legal (I think). But what if I use that tape in my classroom? Is that legal? What if I simply reformat the information from the ERIC CD-ROM I've downloaded, and distribute that information to my students? Is that legal? Do I even stop to wonder whether it's legal? Suppose I'm working on a video disc presentation using Hypercard. I have these television programs I've taped, radio music I've taped, the bibliography I've downloaded, and the technology to put it all together into a video disk that will be used only by my students to help them learn better. Do I stop to wonder whether it's legal? Probably not, I have the tools (technical) to do the job and the information and material to put into the tools, so I probably just go ahead and make the video disc without too much concern about legality. I want to use the most effective tools and resources available as quickly as I can to help my students learn—and I could be frustrated waiting for every legal clarification and permission-if I could even find the sources.

SOURCE: Essay by Judy Ann Pearce in Steven W. Gilbert and Frank W. Connolly, "A Wealth of Notions: Regaining Balance as New Information Technologies Collide With Traditional Controls and Incentives for Intellectual Work", contractor report prepared for Office of Technology Assessment, July 31, 1991.

create, and the ways that readers (or users) read (or use) the works.

What Is a "Work"?

As mentioned in chapter 2, copyright protection attaches to an ''original work of authorship' when it is ' 'fixed in any tangible medium of expression, Thus, when an author writes a novel on a computer or word processor, it is clear that a printout is fixed and tangible and protected by copyright. It is also fairly clear that the words on the cathode-ray tube are evanescent and therefore unprotectable.[30]

A new kind of work that is increasingly being produced today is the electronic mail message,

30 Stanley M. Besen and Leo J. Raskind, '' An Introduction to the Law and Economics of Intellectual Property, ' *The Journal of Economic Perspectives*, winter 1991, vol. 5, No. 1, Case law has held that the fixation requirement for computer programs is met when the source code is written on paper or when the object code or microcode is fixed in a computer chip.

which usually exists only in digital form (fixed in the magnetic disc of the computers where it is sent or received) unless it is printed out. Users of electronic mail on nationwide systems like Internet, Bitnet or CompuServe send millions of messages a year. In addition, many agencies, corporations, and universities have internal electronic mail systems. Some types of electronic mail communication are intended by their senders to be private, others are public. However, there are currently no well-established rules of etiquette for electronic mail nor is there a clear distinction between public and private communications. Most messages are of an ephemeral nature and their writers may or may not care whether their rights under copyright are protected. Other users of electronic mail use this medium to contact and collaborate with colleagues, to express ideas, to exchange drafts of work in progress. In these cases, people would be interested in retaining the rights to their writings.

Technology allows a person to forward an electronic message received from someone else very easily to any number of other people, Is this kind of distribution the same as ''publishing' (a right which the copyright law gives exclusively to the author)? A message can also be modified before forwarding; does this create a derivative work (for which permission from the author should be gained)? Most people would probably agree that mail messages belong to the writer and that publishing them without attribution or modifying them without permission is a breach of manners, at best. However, whether it is an infringement of copyright has not yet been tested.

A further complication in the definition of a work arises because computers make collaboration and multiple authorship so easy. Many electronic mail messages are generated as a part of computer ''conferences. Conferencing is a method whereby people can communicate about topics of mutual interest, even though they are geographically separated. Conferencing software on the host computer records and organizes incoming messages so that each participant can read what has been written by others and then add his or her own responses. Conferences can be of short duration (a day or two) or they can go on for years; they can be limited to a few authorized members or open to anyone with access to the host computer.

Are the "proceedings" of a computer conference one joint or collective work, or many separate works? If it is a collective work with may contributors, the individual contributors can claim authorship in their respective contributions, but who can claim authorship in the collection as a whole? If it is not a joint work, does each individual message constitute a separate work, or do all the contributions of one author constitute a work? The question of what constitutes the work, and the identity of the author or authors, will determine the right of various contributors. For example, if the conference is a joint work, each contributor would have the right to publish the whole conference (subject to accounting to the other joint authors for their pro-rata shares of any royalty). Each joint author would have the right to sue for infringement of any portion of the conference. On the other hand, if the conference is composed of separate works of authorship, each individual author could exercise exclusive rights only over his or her own portion.[31]

In addition, the question of the size of a work might be important in determining if infringement has taken place and if a "fair use" defense against infringement is appropriate. Fair use is determined by four criteria (discussed in chapter 2), one of which is the amount and substantiality of material used with respect to the whole. If a computer conference is one work, then using a single message *in toto* is a small part of the whole; if each message is a work in itself, then copying a single message would be appropriation of the entire work and the fair use defense would be on shakier ground.

Mixed-Media Works

The fact that digital storage makes all works essentially equivalent complicates the definition of a digital work. Current copyright law treats works according to the category to which the work belongs. Categories defined by the law include: literary works; dramatic works; pictorial, graphic, and sculptural works; audiovisual works; motion pictures; musical compositions; computer programs; and sound recordings. These different categories sometimes have different implications for protection of the work. There is no category for a mixed-media work that combines examples from each of these categories.

[31] Morton David Goldberg, partner, Schwab Goldberg Price & Dannay, personal communication, Oct. 3, 1991.

One school of thought holds that a mixed-media work should be considered to be a series of different works, with each type of work being treated according to its class. Another approach would be to consider the whole package as if all the parts were of the same category.[32]

Converting Works to Digital Form

Developers of mixed-media products encounter copyright questions not only in protecting their works, but in trying to create them. Getting permission to put copyrighted works into digital databases has sometimes been so difficult as to prevent projects from getting underway. Because the medium is new, most people have never dealt with it before and the channels for copyright clearance and agreed upon rates for royalties have not yet been worked out. And because many mixed-media projects are large collective works, many different rights owners often need to be satisfied.

In the field of music, songwriters and music publishers collect royalties through ASCAP (American Society of Composers, Authors and Publishers) and BMI (Broadcast Music International), who grant licenses for public performances of music, and through the Harry Fox Agency, which grants mechanical licenses for incorporating music into recordings or movies .[33] The fee structures of these organizations are geared to traditional uses of music, and permission to use music typically costs a few percent of the expected sale price of a published disc. The price structure is reasonable for traditional recordings that use a small number of complete songs on each disc, but prohibitive, for example, for a multimedia library intending to use small parts of hundreds of songs. Until new structures are developed, mixed-media producers are generally limited to works in the public domain, or works composed especially for use in multimedia presentations. Several companies are developing libraries of such works.[34]

Another layer of complication arises if multiple licenses are needed for each work, for example, if multimedia presentations are deemed to be public performances of the copyrighted works they include. It is fairly common in the music industry for performance rights to reside with one entity, while reproduction and derivative use rights rest with another. If one user sits before a computer terminal and hears part of a copyrighted song in the presentation, has a public performance occurred? One royalties-collection agency has taken the position that each use of computer-based presentation is a public performance.[35] The definition of public performance is not clear in the case of computer-based works.

Similar, or worse, difficulties exist in other areas, particularly in areas like images and writings where there are no collective organizations like ASCAP, and negotiations must be made with many individuals. A project to copy baseball cards on CD-ROM was scrapped when it was realized that the publishing company did not have, and could not grant to a licensee, the rights to make digital versions of its own printed cards. The rights would have to be obtained through separate negotiations with more than 500 individual players or their lawyers, but the royalty that could reasonably be expected from sales of the CD-ROM product would be far too small to justify such an undertaking.[36]

Getting permission to convert whole works, such as books, into digital form is generally easier. A number of books have been converted to digital form (some packaged with text analysis software to facilitate scholarly research). Many of these are in the public domain, but for those that are copyrighted, a typical contract follows the model of a contract granting translation rights.[37]

Originality and Authorship

Copyright attaches to ‘ ‘original works of authorship. . . ."38 Original, in this case, means that the work was independently created by the author and

[32] American Association of Law Libraries, ‘ ‘Copyright Considerations for the Use of Mixed Media in Libraries' discussion draft, appeared as an appendix to *A-V & Micrographics SIS Newsletter*, vol. 10, No. 2 , May 1990, and *Automation,* vol. 9, No. 2, winter 1990, pp. 12-23.

[33] For more discussion of royalty structures for music, see *Copyright and Home Copying*, op. cit., footnote 11, especially ch. 5.

[34] Jack Shandle, ‘ ‘Multimedia Computing Hits a Sour Note, ’ *Electronics,* June 1991, pp. 48-53.

[35] Ibid.

[36] Ibid., p. 50.

[37] Michael Newman, Director, Center for Text and Technology, Georgetown University, personal communication, June 21, 1991.

[38] 17 U.S.C.102(a).

not copied from another work. Original does not mean novel—two writers could conceivably create identical works, but as long as neither copied from the other, the works would be original. In earlier cases, the U.S. Supreme Court has also held that some degree of creativity must be involved, and that protectable writings are the "fruits of intellectual labor.' '[39] The U.S. Supreme Court has also defined ''author' as ''he to whom anything owes its origin; originator; maker. ' '[40]

A lot of digital information is in the form of compilations of facts. Facts themselves are not copyrightable. However, an author's selection, arrangement, and organization of facts may be sufficiently original to make the compilation copyrightable. Many publishers, for example, compile and resell information available from the Federal Government (e.g., court decisions, laws and regulations, economic and financial statistics). The database publishers add value to the government material by organizing it, adding indexes, packaging it with search and retrieval software, etc. Government information is in the public domain, not covered by copyright. Yet the publisher's selection and arrangement of the public domain information can be copyrighted .41

How much of the publisher's contribution should be protected is sometimes subject to controversy. In a lawsuit involving two major legal publishers, West Publishing Co., claimed that a competitor, Mead Data, infringed by offering Mead's subscribers an electronic information service with citations including the page numbers on which legal opinions appear in West's publications. The district and circuit courts found that the organization of the information, including page and section numbers, was copyrightable expression and that a competitor's unauthorized use of them was an infringement.[42] The case ultimatey ended in settlement and the decisions remain controversial.[43]

A recent case decided by the U.S. Supreme Court, *Feist Publications v. Rural Telephone Service Co.*, found that telephone White Pages are not copyrightable, and that copying them into another compilation (a regional telephone book) was not an infringement. Lower courts had ruled, in accordance with earlier ''sweat of the brow' or ''industrious collection' tests, that Rural Telephone was entitled to copyright because of the effort it expended to created the White Pages directory. However, the U.S. Supreme Court held that the proper test for copyrightability of a compilation is originality-in this case, the intellectual work of selection and arrangement of facts. Rural Telephone did not select the facts (it was required to list all subscribers with published numbers) and the arrangement was the same as is traditionally used in White Pages telephone directories. Thus, this compilation did not have the minimal spark of creativity to warrant being called an original, copyrightable work.[44]

Database publishers also consider their user interfaces and search and retrieval aids to be copyrightable expression, but it is not always clear how much of this is actually protectable.[45] The same controversies apply here as with the discussions over protectability of user interfaces as discussed in chapter 5.

''Authoring' is a technical term used in the process of constructing works in hypertext. In this case the ''author' is the one who turns a lot of different objects (words, paragraphs, pictures, sounds) into hypertext by establishing the links among them. This ''author" may or may not be the

39 *The Trade-Mark Cases,* 100 *U. S., at* 94.

40 *Burrow-Giles,* 111 U. S., at 58.

41 For a thorough discussion of electronic publishing and government information see U.S. Congress, Office of Technology Assessment *Informing the Nation: Federal Information Dissemination in an Electronic Age,* OTA-CIT-396 (Washington, DC: U.S. Government Printing Office, October 1988).

42 *West Publishing Co.* v. *Mead Data Central, Inc.*, 616 F. Supp. 1571 (D. Mire. 1985) (grant of preliminary injunction "copyright "sSue")* *aff d,* 799F.2d 1219 (8th Cir. 1986), *cert. denied 479* U.S. 1070 (1987). Trial was held on April 5-15, 1988 in the U.S. District Court for the District of Minnesota. Prior to a decision on the merits, parties resolved their dispute and entered into a confidential settlement with approval of the District Court. Order No, 4-85-931, (D.Minn. July 21, 1988).

43 For example, see L. Ray Patterson and Craig Joyce, ''Monopolizing the Law: The Scope of Copyright Protection for Law Reports and Statutory Compilations, " *UCLA Law Review*, vol 36. April 1989, pp. 719-814. The authors note that the defendant did not copy the numbering system, but merely cited it; in addition they believe the courts gave too much weight to economic effects on the plaintiff rather than to the purposes of copyright.

44 *Feist Publications, Inc. v. Rural Telephone Service Company, Inc ,* *No.* 89-1909,59 U. S.L.W. 3243 (U.S. Oct. 1, 1990).

45 Pamela Samuelson, "Some New Kinds of Authorship Made Possible by Computers and Some Intellectual property Questions They Raise, " presented at the Intellectual Property and Authorship Conference, Case Western Reserve University, Cleveland, OH, Apr. 19-21, 1991

same as the one who actually wrote the words that appear on the screen. Yet the establishment of the hyperlinks can be a significant intellectual effort; one that greatly contributes to the usability of the final product.

What are the implications of "authoring" if one person establishes new hypertext linkages within a system to which copyright is already held by someone else? If a scholar, working with ordinary print materials, were to make a new discovery, he or she would report it by writing an article; the article would undoubtedly be a copyrightable work of authorship, even though it contained many quotations from other works. An annotated bibliography, in which a scholar cites many references and adds his or her own comments, is also considered a copyrightable work.

However, an electronic library offers scholars new ways of publishing articles and guiding readers through relevant literature. A scholar could develop a set of hyperlinks that directly leads readers through the referenced materials in just the order the author wishes to make a point or demonstrate a discovery. Such a work might represent considerable intellectual effort, and might be considered a work of scholarship (the scholarly communities will have to work out their own standards about publishing in an electronic environment), But is it a writing? Could a set of hyperlinks be considered a copyrightable work? Or is it an idea or discovery, and therefore unprotectable? Could an electronic article consisting of a set of hyperlinks be considered a "derivative work' based on the underlying works, in which case permission should be obtained before it is created?[46]

Use of Digital Information

Book authors ultimately seek to collect financial rewards for their work by selling copies of their work to readers (often through publishers). A reader who has purchased a copy of a book is free to do whatever he or she wants with it—read it aloud to a child, make notes on it, give it to a friend, or return it undamaged to the store for a refund. The book is property that the reader owns, and under the "frost sale doctrine, ' the owner is free to sell it to someone else.

Electronic publishing is also about delivering works to readers and returning royalties to copyright holders. However, several characteristics of digital information make the delivery system different and also lead copyright owners and their publishers to want more control over the readers' uses of the information.

Photo credit: Mark G. Young

A CD-ROM database containing images of magazine and journal articles. Users can read from the screen or printout an authorized copy.

When Is Information Used?

In using an on-line information service, a reader does not purchase any piece of property; rather he or she buys *access* to the electronic information. Once that access is permitted, the information is out of the control of the copyright owner and the publisher. The user might decide the information is useless and do nothing further with it; on the other hand, he or she may download it (store it in the user's own computer) for future use. For the most part, publishers have no way of knowing the final disposition of the material. For this reason, publishers consider information as "used' as soon as it reaches the reader. They wish to be paid in advance. In the case of on-line vendors today, most fees from users are paid as periodic subscription fees plus use charges related to the amount of time spent searching each database, and sometimes charges for specific documents retrieved. The various schemes for digital libraries usually postulate charging for use of

46 *Ibid*

documents based on how much information a user has retrieved.

From the user's point of view this means that some amount of useless material is paid for. A partial remedy for this is to improve search and retrieval software and to offer means to browse through information (with tables of contents, abstracts, free or low-cost views of a portion of the document, etc.) before a reader coremits to requesting a whole document.

To get access, users generally have to agree to certain limitations on their use of the information. People sometimes purchase a copy of a work on CD-ROM or floppy disc;, but in many instances, close reading of the wrapper may show that it is leased or licensed, not purchased. In these cases, the first sale doctrine does not apply; the use of the material is subject to the terms of the license agreement. The license may state, for example, that users may not resell the disc, or alter it, or place it on network where more than one person can use it. Users may have to return old discs when new ones are supplied, or when the subscription period ends.

Digital information often comes to the end user through a long chain of intermediaries-the publisher, the database service, a library. Contracts govern the rights and responsibilities at each link of this distribution chain.

If there is a long chain of suppliers, there can additionally be many "layers" of users. It is sometimes hard to actually identify the end user of information in a real sense. A student researcher downloads an article from a CD-ROM database to a floppy disc and gives the disc to a teacher who posts it to an electronic bulletin board. Someone sees it on the bulletin board, makes a printout, and faxes a copy to a colleague who hangs it on a physical bulletin board. Each layer of use here has adapted the article to a new medium and involved a new end user. Users may or may not be aware of how the article got to them or what happened to it after it passed from their hands. Issues like copyright infringement and breach of contract may be involved, but who is at fault—the maker of the first copy or of the second or the person who received the last one? Often infringement is so widespread and diffuse that it is difficult to determine damage.[47]

Traditionally, copyright law does not give copyright owners rights to control the access that readers have to information.[48] Copyright owners in the electronic world use contracts to impose restrictions to make sure that they are paid for every instance of access or use. Still, as a practical matter, these restrictions do not prevent unauthorized copying. Once a user has paid for one legitimate copy of something, there is often not much except moral suasion to prevent his or her making other copies. Digital information is easily copied and easily transmitted to many locations. These characteristics make electronic distribution an attractive publishing medium; but they have a flip side; almost any reader is a potential ''publisher' of unauthorized copies.

Unauthorized Copying

Unauthorized copying and distribution is not a problem unique to digital information. Over the past 20 years, the photocopy machine has made copying of books, articles, and other printed works very easy. The introduction of the fax machine has even made it easier to deliver photocopies over long distances. Still, there are limitations to the distribution of unauthorized copies on paper: copy quality degrades with each generation; fax machines, at least at the present time, take some effort to program for large distribution lists; a copied document is still in the same format as the original and can be easily identified as a copyrighted work; photocopying large amounts of material can be inconvenient and time-consming.

Digital copies, on the other hand, do not degrade; each copy is of the same quality as the original. Distribution is easy; the copy could be posted on a computer bulletin board, or distributed to a list of users on a computer network. If one wants to disguise the origins or authorship of the document, format changes can be made with a few keystrokes, Scanning technology now allows one to turn information on paper into digital information so that it can be changed or manipulated.

Some proposals have been put forward to use technology to control unauthorized copying in the

47 Thanks to Rosemary Talab for articulating the concept of layers of use. Rosemary Talab, Kansas State University, personal communication, Nov. 28, 1991

48 Samuelson, "Digital Media and the Changing Face of Intellectual Property Law,' op. cit., footnote 29, pp. 323-340.

context of a digital library. One option is to assign intelligent software agents, for example the knowbot programs mentioned above, to the job of representing the copyright holder's interests. A special type of knowbot program is called a courier. A courier is assigned to a specific item of information (a database, a document, or a paragraph). Depending on the wishes of the owner, the courier can record all uses of the information so that charges can be applied or it can immediately request permission before releasing the information to the user (and deny access if permission is not granted.) The system could also allow for users to make derivative works or to extract parts from a protected work while still giving full credit (and paying royalties) to the original owner. When a user includes a piece of protected information in another document, the courier will create another version of itself to accompany the extract and to represent the owner's potential interest in the new work.[49] Some proposed systems hope to encourage users to do all their reading, writing, and adapting electronically and to discourage unauthorized copying with a pricing structure that makes working within the system and using authorized copies less costly than making unauthorized copies.[50]

In any case, technological proposals for limiting unauthorized copying generally seem to work only within a closed system. Once a user moves an authorized copy out of the system there seems to be no way to prevent further copying. Some writers suggest that there is no solution to the problem of unauthorized copying and that the problem is sufficiently grave that electronic publishing will never thrive as an industry because authors and publishers will not release works in digital form.[51] However it is possible that, as in the case of the photocopying of books or home taping of musical recordings, a viable market will persist despite the presence of unauthorized copies.[52]

Special Concerns of Libraries

Libraries, as mentioned earlier, have been actively making use of computers and digital information for two decades. Digital information allows libraries new ways to offer services, and completely new services to offer, but some uncertainties still need to be worked out.

Many of the rules under the copyright law regarding lending and sharing library materials or making preservation copies or replacement copies of damaged works were developed with printed books and journals in mind. For example, for purposes of preservation or security (or to deposit with another Library) a library has the right to make a "copy in facsimile form" of an unpublished work.[53] Neither the law nor the legislative history define "facsimile, ' but the dictionary definition is ' 'an exact copy, ' ' which may indicate that conversion of a printed work to machine readable digital text is not permitted.

Some provisions in the copyright law also deal with copying and other use of ''computer programs," but do not specifically extend to digital information. For example, the copyright law gives the owner of a computer program the right to make an archival (backup) copy under certain conditions. There are two points here. In the first place, the library may not be the owner of the computer program. Vendors often say that programs are licensed, not sold. The library, as a licensee rather than an owner, does not have the rights described in the copyright law; these are abrogated by the terms of the license. There is considerable controversy over the enforceability of many of these contracts where the vendor has enough barg aining power to force terms on the user.[54] At present, there is a wide variety in the terms and conditions of software and database licenses. An institutional user like a library or university computer center often uses hundreds of different program and data packages, and to ensure

49 Cerf and Kahn, op. cit., footnote 16.

50 Daniel Gross, Magnetic Press, Inc., discussing the pricing structure of the proposed Xanadu system under development by Autodesk and its partners at the OTA workshop on "Digital Libraries, Electronic Publishing, and Intellectual Property, " Feb. 11, 1991.

51 See, e.g., Robert Weber, "The Clouded Future of Electronic Publishing, " *Publishers Weekly*, vol. *237, No. 26,* June 29, 1990, pp. 76-80.

52 *Copyright and Home Copying,* op. cit., footnote 11, especially ch. 7.

53 See 17 U.S.C. 108(b).

54 American Association of Law Libraries, op. cit., footnote 32.

compliance with all of their different requirements is difficult.[55] (For more on licenses, see chapter 2.)

The second point is that the copyright law currently refers only to computer programs and not to data or digital information. Computer data is stored in the same medium as computer programs, and it would seem logical to treat them in the same way, but the argument remains that digital data does not fit the definitions currently set out in section 101 of the Copyright Act so owners have no right to make archival copies.[56] The two points raised here become even more complicated for libraries in the case of mixed-media works where printed material, digital data, computer programs, microfiche, and other forms might be packaged and used together.

Libraries have a long tradition of resource sharing. Several libraries may cooperatively purchase material, and some libraries may refrain from making certain purchases in the knowledge that the material can be obtained through interlibrary loan when needed, Resource sharing practices have long been viewed as prudent use of both funds and storage space, especially for items for which demand is low. Lending of materials among libraries is institutionalized both by tradition and under the provisions of the Copyright Act (section 108), and interlibrary loan usage has increased dramatically in recent years. However, resource sharing practices have recently come under fire from some publishers who see them as depriving information providers of sales.[57] Publishers strengthen their position by leasing, rather than selling materials, thus denying libraries the rights which ownership permits under the ''fust sale doctrine.' Contracts with electronic information providers sometimes limit or forbid sharing or lending of materials. Libraries, particularly public ones, have an obligation to balance the interests of users and producers—a balance which the Copyright Act is intended to maintain.

It has been suggested that the growing use of electronic information, and the tendency of information providers to control the uses of this material through contracts, will lead to greater distinctions between for-profit and not-for-profit libraries, in terms of their library operations, cost differentials, and access .58 Not-for-profit libraries may find themselves placing heavier reliance on free or lower-cost databases, and there may be less ability to share materials between libraries. Profit-based libraries will have access to more expensive information resources, but will also have great controls on their abilities to share resources or to network with other libraries.

Many libraries are examining their own role in offering digital information services to patrons. The shift to digital information introduces new kinds of costs. Public libraries are struggling to determine fair allocation of resources between digital information and printed library materials; in addition there is the question of whether, or how much, to charge patrons using some expensive data services. Public libraries have traditionally been free (supported by taxes) so taxpaying users have already paid for services. In addition, public libraries have an obligation to provide information services to those who cannot get them otherwise. Some libraries are developing usage charges for access to some databases, or are trying to allocate use of scarce resources among users by imposing time limits on the use of workstations with access to certain databases. Over the years the balance in cross subsidy between traditional and electronic services may change several times.

Another question is remote access to library services. The technology exists to allow users at home, office, school to use essentially any computer-based service they could use within the library walls. That many libraries are not now offering such services reflects both the costs of starting up such a service as well as questions that must be resolved concerning license policies. One observer points out that remote access could be a ''boon to the user and a bane to the supplier. ''[59] In many cases, libraries would be passing on to users access to data that has been obtained through a contract between the library and an information vendor. Many contracts now forbid remote access or make it prohibitively expen-

55 Comments at OTA Advisory Panel meeting, Sept. 11, 1991.

56 American Association of Law Libraries, op. cit., footnote *32.*

57 Rosemary Talab, "General Trends in New Technology Usage: Stages of Copyright Development on a National Level, " in *Advances in Library Resource Sharing,* vol. 3 (Greenwich CT: Meckler, in press), pp. 82-83.

58 Ibid., p. 78.

59 Ibid., p. 79.

sive. Libraries and vendors will have to work out the pricing of such services.

Also to be worked out are policies about the use of material obtained by patrons. Some libraries already offer on-line information, and other services such as access to electronic bulletin boards,[60] to their patrons; they therefore become an additional link in a complex of transactions. To what extent are libraries responsible if users make unauthorized copies, post copyrighted material on electronic bulletin boards, send obscene messages, or otherwise infringe copyrights, violate contracts, or break laws? These problems are not new. The advent of the photocopier eventually caused libraries to evolve a policy of providing copiers, posting a notice about relevant aspects of the copyright law, and then leaving users unsupervised to follow their own consciences. Policies regarding digital information—what can be downloaded, number of printouts allowed, etc.—will also need to be worked out, but the policy setting process may be more complex since contracts with information vendors will also be involved.

Another area of uncertainty is in the creation of information based on library holdings. On-line catalogs can be made more useful by adding more information about the works being cataloged—tables of contents, lists of illustrations, etc., but there may come a point where enhancements to the catalog infringe the copyright on the underlying works. As libraries increasingly work to create new information these questions may arise. Some envision libraries, especially research libraries of major universities, as eventually becoming electronic publishing centers for scholarly work.[61] If this is to happen some of these questions will have to be settled.

[60] See Nancy Cline, ''Information Resources," *EDUCOM Review*, summer 1990, pp. 30-34.

[61] Ibid.