

D-Lib Magazine
October 2001

Volume 7 Number 10

ISSN 1082-9873

Public Access to Digital Material

[Brewster Kahle](#), Internet Archive

[Rick Prelinger](#), Prelinger Archives

[Mary E. Jackson](#), Association of Research Libraries

brewster@archive.org

footage@panix.com

mary@arl.org

Introduction

The goal of universal access to our cultural heritage is within our grasp. With current digital technology we can build comprehensive collections, and with digital networks we can make these available to students and scholars all over the world. The current challenge is establishing the roles, rights, and responsibilities of our libraries and archives in providing public access to this information. With these roles defined, our institutions will help fulfill this epic opportunity of our digital age.

For the second time in history, people are laying plans to collect all information -- the first time involved the Greeks and culminated in the Library of Alexandria, which is attributed with being the centerpiece of the world's first university, illuminating human nature in a way never possible before. Now reports on how much total information there is [[Lesk](#) and [Lyman](#)] are leading many once again to take steps in building libraries that hold complete collections. The move to digital technologies for storage, access, and coordination of distributed efforts makes this possible.

While digital libraries are springing up in every corner, a recent experience may serve as an example of both advances and hurdles involved in building complete collections. Over the last five years, the Internet Archive has built some of the largest text and moving images collections in existence by developing cost-effective mechanisms of collection, cataloging, and preservation. By recording approximately 40 terabytes of text, 1000 movies, and television output, much of the digital collecting technology has been successfully tested.

Some of the challenges encountered involved issues of scale, funding, law, and access. The approach to the scale of the task of building the Internet Archive has been to enlist partners and the newest computer technology available -- and this technology gets better over time. Currently, the technology has reached the point where scanning all books,

digitizing all audio recordings, downloading all websites, and recording the output of all TV and radio stations is not only feasible but less costly than buying and storing the physical versions. As in traditional institutions, the funding for the Internet Archive has come through charitable giving, government funding, and in-kind donations. These funding sources form the support structure for most existing libraries, so these challenges are understandable. Legal issues, however, remain the main area where more work is needed.

Some believe that digital material is fundamentally different from print material and have erected legal barriers to using the new digital material in familiar ways. An article quoted then-President of the American Library Association Nancy Kranich as saying that library-goers should be able to duplicate limited amounts of information for educational purposes. "Suppose you want to copy a journal article, quote a section of a book or use a line from a poem," she says. "That is all permitted under the fair-use provision of the copyright law. In the digital arena, fair use has been narrowed to the point of disappearing... The publishing community does not believe that the public should have the same rights in the electronic world" [[Kranich](#)].

To try to accommodate those concerns, three proposals for a legal structure for public access to digital material have emerged. The first proposal is the donation of intellectual property rights to preserves or conservancies in exchange for tax credits. The second proposal is to use interlibrary loan as a legal structure for the exchange of digital material between libraries. The third proposal embraces the lending library concept in which the public can get direct access to digital material from their homes for a limited period.

The donation of 1,301 archival films by the Prelinger Archives to the Internet Archive provides a functional example of how a conservancy might provide widespread access to formerly unavailable cultural documents. While the donation approach is on fairly clear legal ground, it does not necessarily ensure comprehensive access to our cultural heritage because it relies on the good will of donors of digital content.

The existing interlibrary loan (ILL) system can offer a legal and organizational structure for rapidly expanding access to born-digital as well as digitized information. Interlibrary loan is an established mechanism to better serve the public through the pooling of collections to approximate a more complete library. While each ILL exchange in the physical world costs over \$25 and often takes two weeks or more, the digital ILL equivalent does not need to be as expensive and slow.

The Copyright Act of 1976 specifically outlined the last approach, the loaning of digital material, as a way to grant public access to television news recordings. A system to loan material over the Internet was demonstrated at the Internet Archive and Association of Research Libraries conference on Born Digital Material in March 2001.

All of these approaches can be used in combination to great effect. Additionally, while at the same time maintaining control over the number of copies made, the cost of

digitization and storage of our "analog" past can be shared among many libraries because of the linking of the computers via networks. An Internet-based library system containing the majority of our audio, visual, print, and digital cultural achievements would therefore cost less than the building of a new major, print-based library.

These approaches to building a library system of digital works can achieve the far-reaching goal of universal access to our cultural heritage. This article suggests how we can build this future.

Vision for Public Access to Digital Material

Imagine a high school student in Singapore writing a report on the life of Madame Curie based on documentary film footage and original photographs from the Curie family archives.

Imagine researching one's family roots in Europe by browsing the original birth and marriage records from the "old country" from a PC in the library.

Imagine being diagnosed with cancer and having the world's best medical research library and librarians less than one mile from your house.

Imagine a college student's documentary film about her grandfather's World War II battalion. This file could contain original military footage, current footage from towns involved in the battles, and interviews with surviving soldiers from the same battalion.

Anyone with access to a library can and should have access to this breadth of information. The digital divide between the information-age "haves" and "have-nots" can be overcome by making the majority of our culture's information available to everyone. We can help overcome this digital divide by upgrading our libraries to include a greater percentage of both nationally and internationally published information. Our libraries already offer free public access to a portion of our cultural heritage, but any particular library possesses only a tiny fraction of our history, which often excludes most audio visual and digital material. Worse yet, if libraries focus mostly on print material housed in physical libraries, then the library's essential role in society could diminish as society becomes more digital.

The current library and archive system has served the public well in the age of print media. We now need to extend these models, where possible, and invent new models where needed to offer access to our digital cultural artifacts. The following sections explain three methods for building public access.

Intellectual Property Preserves and Conservancies

To think about encouraging and sustaining public access to cultural resources is to consider basic questions of property and its privatization. Land use history offers precedents and a possible solution. In the late 19th and early 20th centuries, private

corporations exerted unprecedented pressures on the "public domain" -- American land and natural resources. The aggressive pursuit of extractive interests such as mining, logging, and agriculture threatened to exhaust public lands and encroach upon naturally or culturally significant sites. In response to this threat, the conservationist movement lobbied to organize a system of national forests, parks, and monuments. By preserving a limited public sphere not subject to the exercise of private property rights, the benefits of some wilderness and cultural sites were preserved for all.

In a similar way, preserves and conservancies attempt to create and nurture public space within the complicated and intangible territory of intellectual property. Presently, several people are proposing similar ideas. Among them are Eric Eldred, who is calling for a "copyright conservancy," [\[Eldred\]](#) and David Bearman, who has recently written a thoughtful paper describing models for "intellectual property conservanc[ies]" [\[Bearman\]](#).

How might something like "national parks for intellectual property" work? First, preserves would be repositories for intellectual property rights that had been donated by rights holders. These rights could include copyrights or, in the case of public domain material, the right to reproduce and disseminate them. The activities of the preserve would be closely coordinated with existing archives and libraries, which will likely still hold physical materials.

Preserves could hold published and unpublished, analog and digital material of all kinds. Assets could be acquired by purchase or donation. Why would copyright owners (or owners of public domain material) ever cede their properties to the preserve? First, and perhaps most important, are tax incentives. These would require amendments to the tax code to allow substantial deductions or tax credits for donating valuable copyrights or material. Second, following the precedent of public land acquisitions, key donors might be compensated with private funding. Third, it is recognized that such donations are prestigious deeds benefiting the national cultural heritage.

There is nothing particularly radical about the concept and practice of a preserve. It's an attempt to work within the system, a gentle expropriation, and a creation of incentives for property holders to do the right thing. Ultimately, though, its goals are to rebalance private vs. common property for mass benefit. The preserve aims to make a significant portion of our intellectual and cultural property available to one and all -- both individuals and corporations -- for nothing more than the physical costs of duplication and transmission, and in some cases, the local library might absorb those costs. Its concept supports freedom of inquiry and freedom of expression by implementing access rights in the broadest sense -- to quote, to duplicate, and to appropriate preexisting material.

Preserves -- A Case Study

At the Internet Archive, we are moving in the direction of building an intellectual property (IP) preserve by creating a new paradigm for access to archives, in this case

historical film from Prelinger Archives, a private collection of 45,000 ephemeral films physically located in New York and San Francisco. What we've done is to select a body of 1,301 key archival films, films we have found to be most in demand, plus unknown films that experience suggests people will want to see and work with. Next, in creating the Internet Moving Images Archive, we have undertaken the expensive process of transferring the films to videotape and then digitizing them so that they can be stored and served online [[Internet Moving Images Archive](#)].

We offer these digital video files in two resolutions: the fairly high-resolution MPEG-2, suitable for exhibition, rebroadcast, and incorporation into other productions; and the smaller, lower-resolution MPEG-4, suitable for research, reference, and more casual viewing. All files are downloadable only. We have ruled out offering streaming video for technical considerations as well as out of a belief that users should have the right to view material repeatedly and do with it what they please. Stills grabbed from the films are also available.

In the nine months that our server has been up, over 150,000 movie files have been downloaded. Though the sample is still too small to analyze definitively, the preponderance of downloads are from educational institutions and DSL/cable modem users, as might be expected given the large size of the movie files (approx. 250 MB/10 minutes). As more films are converted into MPEG-4 files, which are about 12% of the size of corresponding MPEG-2 files, we expect many more downloads from a more diverse group of users.

Since all of the material is either owned by Prelinger Archives or is in the public domain, user restrictions are minimal. First, users are asked not to resell the data files themselves, or use them to go into the stock footage licensing business. Second, they are encouraged to propagate the material, but in all cases we ask that money not change hands. Third, users are asked not to convert the films into closed-source formats.

One key issue -- the effect of this initiative on the donor archive -- remains open. Will the availability of 1,301 key films from Prelinger Archives cause significant losses in stock footage licensing revenue, or will the concomitant publicity put the archives on the map and increase business? Ubiquity of an image has never been a deterrent to stock photography sales, just as repeated airplay has never reduced the value of a music copyright. It seems likely that high-end users will demand the best quality imagery available rather than settling for compressed video, but this is not yet clear. What is clear, though, is that Prelinger Archives, via The Internet Moving Images Archive, now has a long-desired means of providing mass access to a subset of its holdings without having to subsidize the costs of such access. In effect, it trades the risk of losing a measure of control over key material for the benefit of being able to provide cost-free access. And there is no doubt that the number of "access events" will be extremely high.

Already several users have responded to our initiative by offering to donate film or video of their own to the online collection. We hope that institutions as well as individuals will join this effort by agreeing to relinquish total exclusivity to the content

they hold in exchange for the visibility and reach that open access can provide them.

By propagating a canonical set of historical films, we also hope that popular authorship will expand to accommodate time-based media. While many people write, make visual art, compose and play music, relatively few people make videos and films. Is the comparative lack of work in moving image media a result of lack of imagination, tool complexity, or want of opportunity? Until recently, authoring tools were rare and expensive, and at least some of the problem has resided in the lack of access to content. Very few people have ever had significant access to primary moving image historical material, so it has had little chance to seep back into, and influence, the culture. There is much talk of media literacy, but access to the very archival material that might make exercises in literacy dig deeper has always been difficult. Here is an initial step toward making it easier.

While it may require the support and expertise of elite circles to organize something like an intellectual property preserve, and though the activities and even the existence of such a preserve might be invisible to most people, a preserve could mount a fundamental challenge to our definitions of public and private property. In so doing, it would be a greater force for change than any possible reform of copyright law. The Internet Moving Images Archive is a small, but functional, step in this direction.

Interlibrary Loan as a Model for Digital Material

Interlibrary loan (ILL) is the process by which a library requests material from, or supplies material to, another library. ILL expands access to needed material by providing a system that permits libraries to share books, microfilm, and other returnables as well as supply copies of journal articles, book chapters, and other non-returnables. Although some librarians use the phrase "document delivery" to describe the requesting and supply of copies, the more encompassing term of ILL is used to include both returnables and non-returnables. The ILL process has traditionally been mediated. That is, the library, acting on behalf of its user, requests an item from another library, receives the item, loans the user the item to use, and then returns the item to the owning library.

American libraries have shared material with each other since 1876 when the director of the Worcester, Massachusetts Free Public Library suggested that libraries lend books to each other for short periods of time. The requesting portion of the ILL process was automated in the late 1970s with the introduction of the OCLC and RLIN ILL messaging systems, which now use the Internet rather than proprietary networks to transmit ILL requests. Libraries began using fax machines in the mid 1980s to expedite the shipment of photocopies. In the early 1990s, the Research Libraries Group (RLG) introduced the Ariel[®] software to permit libraries to scan copies of journal articles or book chapters and send those scanned images via the Internet to the requesting library. The 2001 revision of Interlibrary Loan Code for the United States no longer limits what can be requested on ILL to scholarly or research material, as was the case with earlier

codes.

In today's environment, the majority of the ILL process is still mediated by library staff. Users submit requests to their ILL department, and ILL staff confirms the accuracy of the information, selects one or more potential lenders, and orders material on behalf of the users. The library owning the item sends it to the requesting library via the U.S. Postal Service, a commercial carrier such as UPS, or a state or regional courier. For photocopies, libraries send the copy by mail, fax, or Ariel. Within the past several years, requesting libraries have begun using Prospero, free software that converts Ariel documents into PDF files and then mounts them on a secure web server. Prospero emails a notice to users informing them of the URL where their documents are temporarily stored.

The legal basis of interlibrary loan is grounded in the U.S. copyright law. The borrowing portion of the ILL process is supported by Section 108(g)(2). This section permits libraries to request copies of articles for their users as long as they do not use ILL to avoid purchasing the book or journal. Section 108(g)(2) is further supported by the CONTU Guidelines describing the phrase, "in such aggregate quantities" used in that section. The CONTU Guidelines permit a borrowing library to receive five articles from the most recent five years of a journal title. The CONTU Guidelines do not have the force of law, but are widely followed by libraries.

The lending portion of the ILL process is supported by the first sale doctrine (Section 109) that permits libraries to purchase books and other material and lend them to their own users or to other libraries. Lending is also supported by Section 108, which permits libraries to make copies under certain circumstances such as preservation, replacement, or ILL. Libraries may make copies for another library as long as the copy becomes the property of the user and the library filling the ILL request has no indication that the copy would be used for any purpose other than private study, scholarship, or research.

As lenders, many libraries (research, college, medical, and special) have imposed charges to lend a book or supply a photocopy; lending fees charged by public libraries are less common. Most ILL fees are transaction-based. One interesting variation to transaction-based charging is the Center for Research Libraries (CRL). CRL was formed by a group of Midwestern universities to provide remote storage for lesser-used material and has evolved into a collection of specialized material selected by members. Libraries join CRL and commit to help build and maintain CRL's collections. One of the benefits of CRL membership is the ability to request any item held by CRL via ILL. The variety and depth of CRL collections is an incentive to join, but even the Center does not house all material needed by its members.

The interlibrary loan model offers an attractive model for providing access to digital material. For the past decade, libraries have used digital technologies to share print material. Libraries have used and continue to use digital fax machines or Ariel workstations to scan and send articles and other excerpts from print-based material. Because contract law, not copyright law, governs electronic (born-digital) resources

being increasingly acquired by libraries, libraries are not able to rely on Section 108(g)(2) to make and send copies to fill ILL requests. The license must specifically permit libraries to use the licensed digital resource to fill ILL requests, but not all licenses provide the same level of use as currently supported by the U.S. copyright law and CONTU Guidelines.

Copyright issues aside, one of the remaining challenges of using the ILL model as the basis of access to digital resources is determining where the needed item is located. Staff in ILL departments currently spend a significant amount of time searching OCLC and/or RLIN union catalogs to find libraries that own the item. ILL staff also search a variety of other sources, including Web-based library catalogs, an effort that is not sustainable as more material is found only on the Web. Having a universal, comprehensive, and single point of access to *all* digital resources is an essential component for this model to succeed.

In summary, the ILL model for non-returnables for providing universal access to cultural heritage would encompass the following characteristics:

- The material requested is not owned locally;
- The process of requesting and receipt would be mediated by library staff;
- The requesting library would have limits on the quantity of non-returnables it can request;
- The digital copies would become the property of the user;
- The process would not serve as a substitute for an item if the level of local use was sufficiently high; and
- The range of material would be greatly enhanced to include print and digital resources.

Lending Digital Material to the Public

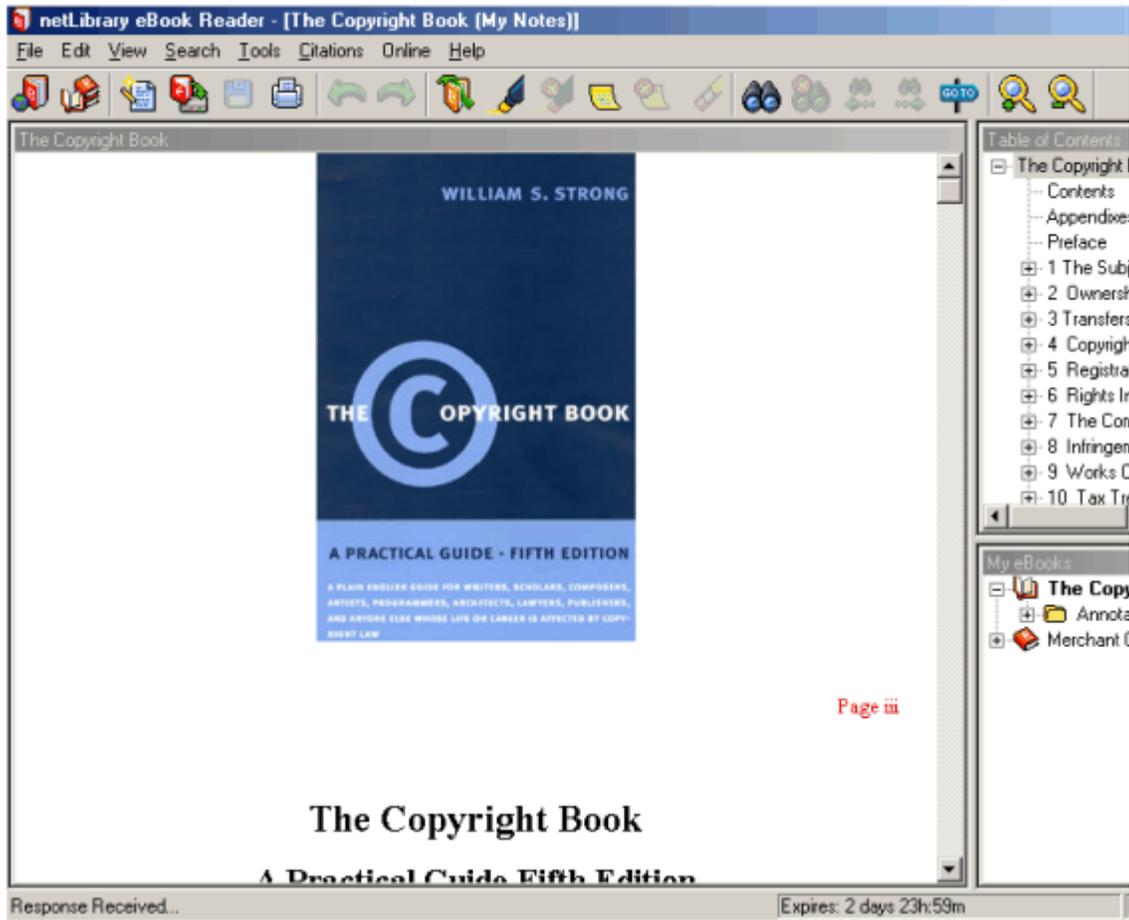
Public libraries acquire large collections of books and other works generally at retail price and then lend them out to their local citizens. Ben Franklin and Andrew Carnegie put this in action in the United States to promote democracy, education, and the advancement of the underprivileged. As a society, we have supported these values in our laws and our tax dollars. Six billion dollars was spent in 1997 on public libraries, of which 15% was spent on acquiring material. Where printed objects must be lent out physically, the digitized and digital library material can be lent out digitally. Using this model, libraries can continue to serve the public good in the future without a major change to the institutional structure as we digitize the collections.

This model is different from the interlibrary loan model, because it assumes that the

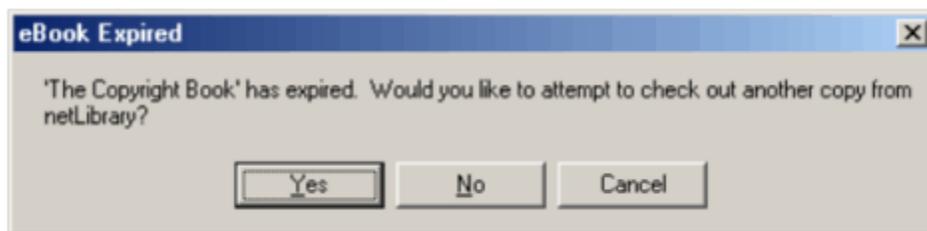
local library owns or licenses the needed material.

Several technologies are now available to allow lending. One is netLibrary, which is both a technology and a service. Another is eBook by Adobe Inc. (see appendix 1), which is just a technology. A third is a public domain technology called "Lending Library Format" that is starting to be used for a few collections. These copy protection technologies are never ironclad since if a file is viewable, it is recordable, possibly with some degradation in quality. On the other hand, library users can copy items owned by their libraries; they have been making handwritten notes and photocopying legally -- and perhaps illegally.

Using netLibrary.com, the San Francisco Public Library (SFPL) system has started lending out electronic books to anyone who holds a San Francisco library card. San Francisco Public determines how many copies of a particular title should be made available and the loan period for each. With a library card, a user can search and download books to his or her hard drive. If the library had selected only one copy, another SFPL user would be told that the title is in use. After three days, if that is the "due date" set by SFPL, the user is not able to open the file and is prompted to check the book out again.



After a netLibrary book has expired, the user who tries to launch the book gets this notice:



While there are not very many books in this format currently, most users will understand the procedure because it is closely modeled on checking a book out of a library. Rather than returning the physical book to the library, the electronic equivalent

is for netLibrary to disable the copy on the hard drive.

Other software publishers are making technologies that can be used in such systems.

Lending Library Format

A public domain system for lending digital files has been written called the Lending Library Format [[Lending Library Format](#)], which is being tested with limited collections. With this system, a library can lend a digital file in much the same way as netLibrary, but without having to have a central company prepare the material. With this capability, libraries could digitize rare and fragile holdings or record material only available in digital form and then lend them out. Because of the reduced cost of digitization or recording digital files, comprehensive collections could be made available in the near future.

Where lending physical copies guarantees that only one person has a book at a time, achieving the same restriction with digital files requires extra work. The 1976 US copyright law made provisions on how this can be accomplished for one electronic collection, television news programs, and put in provisions of what lending such material means (see sections 108(f)3 in Copyright Law of the United States of America [[Copyright Law](#)]).

Therefore, the analogue to lending physical material is to lend digital material. Libraries would then make their digitized and digital material available to their users. Audio, video, scanned books and periodicals, scanned photographs, and ephemera could be made available with limitations on who may have access to these collections and for how long. Even beyond what is possible with physical objects, different restrictions could be imposed such as username/password, library card number, IP address, etc.

Digitization prices have dropped, so making unique collections available from libraries is becoming more affordable. Scanning a page is now under \$0.10 [[1](#)]; even in fairly small quantities of material, scanning an hour of videotape is about \$10.00 [[2](#)]; and recording an hour of a radio broadcast costs about \$1.00.

The cost of serving digital material is also dropping if libraries are careful in the hardware they purchase. A terabyte server that cost under \$7000.00 will store the text of one million books or about one thousand hours of video [[3](#)]. Therefore, libraries can bring parts of their own collections online cost effectively to serve their patrons.

By borrowing a work in Lending Library Format [[Lending Library Format](#)], library users would utilize a system similar to netLibrary, but instead of using a custom reader, it would use the normal reader for that file format. The reason this would work is that the Lending Library Format is an "envelope" around a file that indicates the start and stop times during which it is legitimate to view the file. If it is within that period, then a simple client program decodes the file and passes it to the appropriate viewer; if it is not

within that period, it prompts the user to check the work out from the library again.

Conclusion: Roles Rights and Responsibilities

Since the roles and responsibilities of libraries for providing public access have not changed simply because the format of material has moved from analog to digital, it is reasonable to assume that the rights to perform these societal tasks have not significantly changed either. Leveraging the existing system of library and archives, with its existing approaches to serving users, can be easily extended to include providing public access to digital material. A combination of the three structures discussed here -- donations, interlibrary loan, and lending libraries -- have served well in the past and can continue to serve in the future. Required are leaders to build and test examples of the technologies and models. These leaders in libraries, archives, and public policy groups can have a far-reaching impact on the future of education and scholarship by offering universal access to our digital cultural heritage.

This material is based upon work supported by the National Science Foundation under Grant No. 0109524. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Related Work and Related Organizations

Appendix: An Example of E-Book Restrictions

Notes

[1] See, for example, <http://www.archive.org/arpanet> and <http://www.roottech.com/>.

[2] See, for example, <http://ftp.archive.org/html/index.html>.

[3] See, for example, <http://www.alex.com> and <http://www.aslab.com>.

References

[Bearman] Bearman, David. "Intellectual Property Conservancies," *D-Lib Magazine* Volume 6, No. 12, December 2000. Available at <http://www.dlib.org/dlib/december00/bearman/12bearman.html>

[Copyright Law] Copyright Law of the United States of America and Related Laws Contained in Title 17 of the United States Code, Chapter 1, Section 108. Available at <http://www.loc.gov/copyright/title17/92chap1.html#108>.

[Eldred] Fonda, Daren. "Copyright crusader," *The Boston Globe Magazine*, August 29,

1999. Available at <http://www.boston.com/globe/magazine/8-29/featurestory1.shtml>

[Kranich] Weeks, Linton. "Pat Schroeder's New Chapter," *The Washington Post*, February 7, 2001; page C01. Available at <http://washingtonpost.com/wp-dyn/articles/A36584-2001Feb7.html>

[Internet Moving Images Archive] The Internet Archive. "Internet Moving Images Archive: Movie Collection", March 2001. Home page at <http://www.moviearchive.org>

[Lending Library Format] Kahle, Brewster and Rod Hewitt. "Lending Library Format (.llf)," Version 1.0.5, February 22, 2001. Available at <http://www.archive.org/llf.html>

[Lesk] Lesk, Michael. "How Much Information Is There In the World?" Available at <http://www.lesk.com/mlesk/ksg97/ksg.html>.

[Lyman] Lyman, Peter and Hal R. Varian, "How Much Information", 2000. Available at <http://www.sims.berkeley.edu/how-much-info/>.

(Corrected coding on link to Note 3 8/31/05.)

Copyright 2001 Brewster Kahle, Rick Prelinger, and Mary E. Jackson

[Top](#) | [Contents](#)
[Search](#) | [Author Index](#) | [Title Index](#) | [Back Issues](#)
[Editorial](#) | [Next Article](#)
[Home](#) | [E-mail the Editor](#)

[D-Lib Magazine Access Terms and Conditions](#)

DOI: [10.1045/october2001-kahle](https://doi.org/10.1045/october2001-kahle)