

Free Culture  
and  
the Digital Library  
Symposium Proceedings  
2005

Proceedings of a Symposium held on October 14, 2005 at  
Emory University, Atlanta, Georgia.

Martin Halbert (Editor)

MetaScholar Initiative  
Robert W. Woodruff Library  
Emory University  
Atlanta, Georgia

This work is published under the Creative Commons  
Attribution-NonCommercial-NoDerivs 2.5 License






## Copyright Information:

These proceedings are covered by the following Creative Commons License:

### Attribution-NonCommercial-NoDerivs 2.5

You are free to copy, distribute, and display this work under the following conditions:

	<b>Attribution.</b> You must attribute the work in the manner specified by the author or licensor. Specifically, you must state that the work was originally published in the <i>Free Culture and the Digital Library Symposium Proceedings 2005</i> , and you must attribute the author(s).
	<b>Noncommercial.</b> You may not use this work for commercial purposes.
	<b>No Derivative Works.</b> You may not alter, transform, or build upon this work. For any reuse or distribution, you must make clear to others the license terms of this work.

Any of these conditions can be waived if you get permission from the copyright holder. Your fair use and other rights are in no way affected by the above.

The above is a human-readable summary of the full license, which is available at the following URL:

<http://creativecommons.org/licenses/by-nc-nd/2.5/legalcode>

## Publication and Cataloging Information:

ISBN: 0-9772994-0-6

Editor: Martin Halbert  
Copy Editors: Carrie Finegan  
Katherine Skinner

Publisher: MetaScholar Initiative  
Robert W. Woodruff Library  
540 Asbury Circle  
Emory University  
Atlanta, GA 30322

LCSH Subjects: Freedom of information  
Digital libraries—Congresses  
Intellectual property—United States  
Copyright and electronic data processing  
Information society  
Cyberspace—Social aspects

## TABLE OF CONTENTS

Preserving the Information Commons of Society in the Library of the Future (Martin Halbert, Emory University).....	1
Between Pragmatism and Anarchism: The American Copyright Revolt since 1998 (Siva Vaidhyanathan, New York University) .....	13
It Is Easy for Universities to Support Free Culture with Digital Libraries: The NDLTD Example (Ed Fox, Virginia Tech) .....	43
The arXiv: Fourteen Years of Open Access Scientific Communication (Simeon Warner, Cornell University).....	56
The Librarian Revolutionary: Creating Cultural Exchange and Preservation on the Internet (Debora Halbert, Otterbein College).....	69
Catalysts for Change: Librarians and the Open Access Movement (Meghan Miller Brawley, University of Tennessee - Knoxville).....	82
Millennial Net Value(s): Disconnects Between Libraries and the Information Age Mindset (Charles F. Thomas and Robert H. McDonald, Florida State University) .....	93
Rights, Registries, and Remedies: An Analysis of Responses to the Copyright Office Notice of Inquiry Regarding Orphan Works (Denise Troll Covey, Carnegie Mellon University).....	106
Heritage Under Lock, but No Key: The Troubled Status of Unpublished Works in Digital Archives Projects (William J. Maher, University of Illinois).....	141
Government Information in the Digital Era: Free Culture or Controlled Substance? (Karrie Peterson, NCSU; and James A. Jacobs, UC San Diego).....	154
Will Fair Use Survive the Digital Age? (Marjorie Heins, New York University) ...	179
Greasing the Wheels of Regulation: the Google Print Library Project (Barrie Howard, Digital Library Federation).....	218
How Do We Sustain Digital Scholarship? (Bradley J. Daigle, University of Virginia).....	230
A Scholia-based Document Model for Commons-based Peer Production (Joseph Corneli; and Aaron Krowne, Emory University).....	241
Adapting CBPP Platforms for Instructional Use (Robert Milson, Dalhousie University; Aaron Krowne, Emory University).....	255
On Free Math and Copyright Bottlenecks (Raymond S. Puzio, University of Memphis).....	273
How Free Culture Will Save Digital Libraries (Aaron Krowne, Emory University) .....	300

## ACKNOWLEDGMENTS

I would like to express my enormous thanks to all the individuals and groups who helped to make this symposium a reality.

My thanks go to the institutions and departments that provided various kinds of support to enable this symposium, including the Robert W. Woodruff Library which provided the overall framework of support for the conference, the Emory Law School which provided the facilities for the event, and the divisions of Emory University which provided financial support, including the Emory Graduate Institute of the Liberal Arts, American Studies, and the Department of Mathematics and Computer Science.

I would like to thank all of the contributors to these proceedings for their input to the intellectual substance of this symposium. I would also like to thank all of them for their ongoing contributions to the causes of free culture and digital libraries. My thanks also go to the symposium presenters who were unable to contribute a paper, especially Dr. Lawrence Lessig; their contributions to the symposium were also key to making it a high quality event.

I would like to thank all of the scholars who participated (gratis) in the peer review of submissions to the symposium. The quality control and substantial feedback to contributors that these peer reviewers provided were critical to the success of the symposium.

I am grateful to Linda Matthews, Emory Vice Provost and Director of Libraries, who gave us permission to pursue this event and also provided so much guidance and financial support.

Finally, I would like to extend special thanks to the particular staff members of the MetaScholar Initiative who focused so much of their energy for so many months to make this symposium happen. I am privileged to lead this premier group of young Turks, who are each brilliant in their respective areas of technology and cultural studies. MetaScholar is Emory's ongoing initiative for digital library and e-scholarship projects. Key to making this symposium happen were Katherine Skinner, Sarah Toton, Paul O'Grady, and Liz Milewicz, who all worked in many capacities. My greatest thanks go to our copy editor and general symposium heroine Carrie Finegan, who carried all the details of this symposium forward for a full year, and who was the lynchpin in the whole endeavor.

Martin Halbert  
Atlanta, Georgia  
September 2005

Introduction to the Symposium

## Preserving the Information Commons of Society in the Library of the Future

Martin Halbert (Emory University)

**Abstract:** Outlines the themes and contributions of the Free Culture and the Digital Library Symposium. The article provides a summary of the conflict of interests between those who seek to preserve a shared commons of information for society and those who seek to commodify information. I introduce a theoretical framework called *Transmediation* to help explain the changes in media that society is currently experiencing.


### INTRODUCTION

It is my great pleasure to provide this introduction to the proceedings of the Free Culture and the Digital Library Symposium. You will find that these proceedings are an unusual blend of contributions inspired by a group of shared beliefs about freedom of information and digital libraries. The core convictions that inspired this symposium are 1) that the public has a right to freely access, preserve, and use shared cultural information; 2) that digital libraries (broadly construed) are key to providing for this right in the modern world; 3) that there are unfortunate trends afoot to constrain and deny this right; and finally 3) that we *must* mobilize efforts *now* to resist these trends and preserve this essential public right.

It is not an overstatement to say that we are currently in the midst of a war over the future of social access to shared cultural information. This war is being waged at virtually all levels of our society; in courtrooms, in congress, in universities, in our workplaces, and in virtually all of our homes. We have in this country a tradition of a shared commons of cultural information that was historically predominant until the end of the Twentieth Century, when it was lost to legal maneuverings of commodifying

---

M. Halbert (Ed.): *Free Culture and the Digital Library Symposium Proceedings*. Atlanta, Georgia: MetaScholar Initiative at Emory University, 2005. pp. 1-12.

 This work is covered by the Creative Commons Attribution-NonCommercial-NoDerivs 2.5 License.

media interests. This loss has been marked by librarians, technologists, legal scholars, and social theorists such as the individuals represented in these proceedings. Simultaneously, the opportunities presented by networked digital technologies for enormously improved means of communication, expression of viewpoints, freedom of information, and the dissemination and creation of knowledge have been noted. We stand at a true social crossroads, looking ahead to two very different futures. On one path, a panoply of networked technologies that I broadly refer to as the digital library (although this is essentially a metaphor, as they include all manner of mechanisms that are technically distinct from digital libraries proper) offer the promise of a new era of common access to information far more revolutionary in scope than the previous historical advent of the printing press and all the practices and institutions that arose from it. Predicting this sounds like wild hyperbole, but I believe that we all know in our hearts that it could well be an understatement of what is happening in front of our eyes every day.

The other path that lies ahead of us is a pinched and crabbed future in which every drink from or offering to any well of knowledge, no matter how meager, is rationed and controlled by commercial overlords with utterly crass objectives. This path abandons our long-held traditions of socially shared libraries of cultural information. This is the path that our society is now set upon, and there is relatively little time to steer back in the other direction, or at least towards something approaching a middle course. This can still be done, but we have to collectively open our eyes to what is happening, and we must take action if we are to preserve the information commons of society in future digital libraries.

If you are mystified by my sense of urgency, I invite you to read the papers in this volume and come to your own conclusions. The contributors to these proceedings have provided a well-rounded examination of these issues.

### **Highlights of the Symposium Papers**

In these proceedings you will find papers representing many outlooks, including those of legal scholars, social theorists, librarians, and technological innovators. These perspectives often overlap in papers submitted by our multi-disciplinary contributors, who often approach these topics from several directions at once.

The noted social theorist **Siva Vaidyanathan** has here provided an insightful analysis of the battle for public culture in which we

find ourselves, and a pragmatic path for resisting attacks on free culture. Dr. Vaidhyanathan eloquently explores the connections between the core themes of this symposium in his paper, critiquing the return to robber baron ethics that the Copyright Term Extension Act (CTEA) has brought about, and the role of peer-to-peer (P2P) networks as a backlash against absurd copyright control excesses. I very much agree with him that P2P networks are emerging in the digital realm as libraries that preserve access to contested information.

The organizers of this symposium worked very hard to bring together both theorists and practitioners, including representatives of digital libraries that embody free culture principles. The arXiv has spearheaded open access to scientific information throughout the last 14 years, thereby inventing and championing the concept of the electronic pre-print (or “e-prints”) service. **Simeon Warner’s** paper provides a historical retrospective that analyzes key issues in the development of this key digital library service, which has literally revolutionized the way that scientists share information. **Ed Fox** describes the Networked Digital Library of Theses and Dissertations (NDLTD), a service which has fundamentally changed the way that academic institutions share the work of new scholars. **Daniel Mayer**, chief financial officer of Wikipedia will be speaking at the symposium, but was unable to contribute a paper. The tremendous impact of Wikipedia as a digital library of freely available content is emblematic of the intersection of free culture and digital library endeavors.

A trio of papers explores the role of librarians as change agents. **Debora Halbert** explores the librarian as an activist for change, building on their traditional roles and responsibilities in previous periods. **Meghan Miller Brawley** examines the ways that librarians can work to foster adoption of the Open Access model of publication, and the reasons for doing so. **Chuck Thomas** and **Robert McDonald** describe the issues librarians face as they provide for the information needs of the latest generation of students, whose information seeking behaviors and expectations are significantly different from previous generations.

Several of our contributors have written about legal issues, especially the standing of orphan works (items in copyright limbo). **Denise Troll Covey** has provided the first analysis of the hundreds of responses that the U.S. Copyright Office received when it posted a Notice of Inquiry regarding orphan works. This analysis yields a range of potential responses that, if implemented, would

greatly increase access to a vast body of information that will otherwise be walled off from access to posterity. **Bill Maher** elaborates on this issue from the practical perspective of archivists trying to provide access to important historical materials in the orphan works category. **Karrie Peterson** and **James Jacobs** take up the similarly disturbing challenges of public information provided by the government to the citizenry, a foundation of our concept of democracy which is rapidly eroding before our eyes. Finally, **Marjorie Heins** describes findings of the Free Expression Policy Project at the Brennan Center for Justice on the ways that fair use as a commonly understood concept and value in our society is under siege.

Some of our contributors examine the future of libraries as institutions. **Barrie Howard** has delved into the ramifications for libraries of the Google mass-digitization project, and the responses called for by this project. **Bradley Daigle** proposes new roles for research libraries in collaboratively fostering digital scholarship with faculty. The remainder of my own paper is a brief examination of libraries from a very broad historical perspective.

Finally, we have a cluster of papers that investigate Commons-based Peer Production (CBPP), a new model for collaborative authorship that has rapidly been spreading throughout society in the form of Wikis, Blogs, and other mutually authored systems. **Joseph Corneli** analyzes CBPP with reference to the historical practices of scholia (annotation) writing. **Robert Milson** reports on the pedagogical value of CBPP systems. **Raymond Puzio** considers the particularly vexing issues of the contemporary “fog of copyright” and CBPP created mathematical writings. **Aaron Krowne** (who has a hand in most of this cluster of papers) provides a final summation piece laying out the ways that he sees free culture intersecting with digital libraries in the future.

**Lawrence Lessig**, our keynote speaker, was unable to prepare a paper for these proceedings; however, his seminal book *Free Culture* was the single greatest inspiration for this event. Dr. Lessig’s work is cited by virtually all of our contributors, and we are deeply honored by his participation.

## THE EVOLVING LIBRARY

In addition to highlighting the many contributors to these proceedings, I would like to provide a brief outline of my own observations and beliefs about the topics of this symposium. I am a librarian in a position of leadership, and I have spearheaded a



number of research projects assessing specific new innovations in digital libraries in recent years (Halbert 2003, 2004). These projects convinced me that there are indeed opportune new roles for research libraries to take up in the digital era, and that libraries need not feel marginalized or irrelevant to the information needs of contemporary society. Incorporating these opportunities into library operations will mean becoming *digital* libraries, because digital technology is what enables these new capabilities. Print materials will of course not disappear, but research libraries will increasingly focus efforts on implementing practices and workflows that center on computers and networks. The transformation of traditional libraries into digital libraries is not a simple transition, and will require the best efforts of the current generation of librarians. Research libraries around the country have been engaged in two decades of digital projects to become prepared for this transition. But reviewing this trend and undertaking such projects at Emory also made me aware of the enormous complexity of the surrounding social context within which this transformation of research libraries is taking place. I came to believe that this intricate context of previous social changes must be examined to understand the future of libraries.

In order to accomplish this, I undertook a doctoral investigation into how the recent changes in academic libraries fit into much larger patterns of societal change. The remainder of this paper will provide a sketch of some highlights of my conclusions.

### **TRANSMEDIATION STUDIES**

Libraries have to be understood as institutions in a great variety of contexts, but I think they especially must be seen in the historical context of institutional forms that arise based on particular media that enable social communication at any given point in history. I carefully examined the historical precedents for libraries, and was frankly surprised at what I found.

Not only have there have been myriad social structures in past eras that served the fundamental library functions of preservation and access to information, but these precursors to the library have died countless deaths, usually when the dominant means of communication in a society was challenged by a new form of communication. Most of these deaths were seen as apocalyptic at the time, a confrontation between one form of mediation and another. An often cited image in this regard is from Victor Hugo's famous book *Notre Dame de Paris* (translated popularly as *The*

*Hunchback of Notre Dame*), where a church official, gazing at both a printed book and the Notre-Dame cathedral, declares “This will kill that,” meaning that the printed book will overthrow the edifice of the church and its ways of thought. Or as Hugo says in chapter two of the fifth book,

It was the presentiment that in changing its form human thought was going to change its mode of expression, that the most important idea of each generation would no longer be written in the same material and in the same way, that the book of stone, so solid and so durable, would give way to the book of paper, even more solid and durable. (Hugo 1831; Krailsheimer 1993 translation)

Umberto Eco and a group of like-minded scholars have explored the parallels between this remark and the contemporary situation, in which digital technologies seem to be overthrowing the book and its associated institutions (Nunberg 1996). I begin with this quote in order to point out that a transition in the dominant medium of communication has inevitable impacts on the institutions which have arisen as supporting apparatus for the dethroned medium. These impacts are often seen as catastrophic to the mission of the relevant institutions at the time, as in the case of the Hugo quote, in which the church, an institution built up on the media of manuscripts and oral presentations, is seeing the sweeping advent of printed materials. While such media transitions are certainly disruptive of such institutions, they often simply transform the institutions into new configurations. Exactly how such transformations play out, and what set of values become institutionalized in the course of the transition, are a central issue for understanding how such social transformations occur.

Authors such as Eric Havelock (Havelock 1963), Elizabeth Eisenstein (Eisenstein 1979), and Walter Ong (Ong 1982) have analytically studied these transition periods when a new mode of social mediation overcame the previously dominant form. There are no commonly accepted terms for such periods or their study, so I have coined some vocabulary out of necessity. My term for the study of this historical phenomenon is “Transmediation Studies,” while I refer to individual transitions of this kind as “transmediational periods” or simply “transmediations.”

When historians look back on transmediations, these periods appear punctuated with critical moments with clear impacts and ramifications. Such critical moments are less clear to the people who live through them, because transmediations almost always encompass a great many confusing and disruptive events that are all intensely contested by different parties. Transmediations also occur at many scales, ranging from small and largely unnoticed to

gigantic and sweeping. Some of the big ones that the above authors studied were the transition to writing from the spoken word (Ong and Havelock), and the transition to printing from writing (Eisenstein). It is important to note that in any transmediation the previously dominant form of mediation does not disappear; the new system of communication simply overlays it. Yet, this has a very profound effect on how people think and express themselves. People think differently on paper than they do when speaking, for example.

A major observation made by these and other authors engaged in similar lines of inquiry is that virtually all transmediations enable both new forms of social freedom *and new opportunities for social control*. And the new modes of control seem to take hold more frequently than the freedoms. So what does all this have to do with my title?

What I mean by the “Information Commons of Society” is the broad category of social behaviors and expectations that nurtures freedom of communication by valuing and promoting free access to shared cultural information. The “Library of the Future” stands in for not only traditional libraries and new hybrid digital forms of the library, but also a range of particular institutions and practices in society that enable, protect, and foster the Information Commons, such as a free press, a free Internet, and the legal protections that support freedom of speech in many such venues. I realize that this tremendously loads these phrases, but I also think that these are fairly intuitive notions that are easy to understand.

I won’t elaborate what comprises the Library of the Future, because that is what this symposium is all about. Read the papers. Exploring this concept was the reason that we sought to hold this symposium.

To come back to the point about transmediations enabling new forms of social control, what I find so frightening about the transmediation we are currently experiencing is that the new digital media and the automata that make these media possible (computers and networks) provide radically more opportunities for control than any previous media transition. While new forms of control have always arisen in previous transmediations, it seems to me that digital technology provide a truly unprecedented range of options for social control, both governmental and commercial. The latest conceptions of control systems embody finely grained and sweeping frameworks of intellectual control euphemistically

termed *Digital Rights Management* (DRM). These systems are envisioned monitoring and mediating virtually every act of communication that an individual undertakes through digital media. These control systems are the direct descendants of the *Panopticon*, a prison for total surveillance and control dreamed up by Jeremy Bentham and famously analyzed as a concept by Michel Foucault (Foucault 1975). The Panopticon is an institution of constraint and punishment of human inquiry, the antithesis of the library.

There are innumerable legislations and control systems that I might list that are eroding social freedoms and leading us toward a modern day Panopticon, ranging from the USA Patriot Act to the Digital Millennium Copyright Act (DMCA). Any of these might or might not be the deathblow for the Library of the Future. As I've said, no one can say with certitude during the transmediational period whether such a moment has already occurred, will occur in the future, or will ever occur. However, I will single out for comment one recent critical moment among many, the Copyright Term Extension Act (CTEA) that was upheld by the Supreme Court on January 15, 2003.

Without delving into the details of this decision, I will echo the claims of our keynote speaker, Dr. Lawrence Lessig, who made the case against the CTEA on behalf of a coalition of activists, scholars and other interested parties that the CTEA decision deals an insidious blow to freedom of access to information, and is emblematic of the commodifying trends of social control in media today.

Lessig and other legal scholars have extensively described the ramifications of the CTEA. My conclusion after reading the work of these authors is that the decision basically upholds the right to indefinitely extend copyright. Think about that. The essence of copyright is the ability to regulate who has the right to communicate (by making printed copies or transmitting digital copies) any particular expression of ideas. By extending this ability indefinitely, the potential exists to make every attempt to freely reproduce, transmit, or access previously published content a crime.

If it remains unchallenged, this act may be a mortal blow to the Information Commons of Society as I define it. My conception of the Commons is a broad public space of information that is held "in common" for the benefit of and unencumbered use by our

entire society. If no information ever reaches this space, then all information becomes locked up in the Panopticon. As Siva Vaidhyanathan and Marjorie Heins argue in this symposium, our society has long held that under certain conditions and after a certain reasonable period of time, published content should pass into a shared cultural space. Obviously, content producers should have the right to recoup investments; this is the entire point of the justifiable traditional concept of copyright. But the point of this control over content is ultimately only for the good of society as a whole, not particular investors. Permanent control of information dissemination serves no one's best interests except those of commercial media conglomerates.

Criminalizing the long-term free exchange of content completely polarizes the discussion of reasonable recouping of investment and fair financial support mechanisms. Further, it drives underground innovative forms of what I consider the essence of the library (open information exchanges) into unscrupulous uses of file sharing technology. Worse, it obscures our socially shared conception of what constitutes fair use of information by declaring almost any form of free information exchange a criminal act.

It also destabilizes the free exchange information. When guerilla usage of P2P technologies becomes the only way to preserve access to information, we lose the structure and validating practices of libraries. While a forum like KaZaa is certainly a dynamic bazaar of communication, I also think it lacks much of the stability, organization, and authority that the library provides that is useful in fostering long term serious inquiry among communities of thinkers.

There are a great many ways that the cultural identity of scholars will change in connection with the decline of both the traditional library and my extended concept of the library as a public information space. I do not have time here to explore all the ramifications that I believe the current transmediation holds for academia, but I generally think that many of these changes in the near term will unfortunately serve to further limit, constrain, and diminish open scholarship. Some examples:

- If online mechanisms for enforcing strong copyright restrictions are successful, commercial publishing conglomerates like Elsevier will be better able to tightly commodify and commercialize scholarly communication in the online world as they have in print media (this is the

ruinous pattern of charging universities for their own intellectual output that has led to the serials crisis that has basically crashed the whole approach of libraries to scholarly journals, at least scientific ones).

- If copyright bludgeons such as the CTEA and DMCA intimidate and constrain scholars' explorations of the beneficial possibilities of disseminating their content in digital media, then academia may never have the opportunity to even grasp the online freedoms that are being contested, and such freedoms of expression will be stillborn. This will disenfranchise scholars further, leaving only commercial channels of expression for scholarly communication through media and technology conglomerates (think AOL and Microsoft). This would leave academia triply at the mercy of such conglomerates, paying for the salaries of scholars, paying for library acquisitions of the content scholars produce back from the conglomerates, and paying to express oneself in the first place. If you are skeptical, note that this phenomena is already commonly happening in the sciences, where scholars and their institutions pay so-called "page charges," basically a per-page fee to be published in a scholarly journal (forget about them getting any payments or royalties from the publishers).
- The scale at which media conglomerates act dwarfs academia. A simple example: All U.S. libraries together spent roughly \$6 billion dollars in 2002. AOL alone lost \$99 billion dollars in the same year! The relative power of the library and the conglomerate may already be so skewed that the library may die simply by disappearing from our cultural experience in comparison with the overwhelming output of commercial media channels. The library has conceptually been the level playing field in which free scholarly discussion takes place, at least ostensibly free of bias. Commercial channels are anything but free of bias.

### **The Library of the Future**

While I think things will get worse before they get better, I do believe that they will eventually get better. Transmediations come and go, and forms of control give way to freedoms in cycles. We are currently seeing a great erosion of the Information Commons of Society that may take some amount of time to reverse. This erosion of freedom places a great responsibility on all of us to act

in support of a future that embodies the values of free culture, and to work to mobilize efforts in support of such a future.

I do not doubt that some new form of institutions comprising the Library of the Future (as I've defined it) is going to arise. We do not know exactly what set of organizations, technologies, or practices will eventually come to symbolize the trusted, open access to information that the concept of the library today embodies in the public perception. We may or may not call it the "digital library" and, for that matter, it may not have the word "library" anywhere in it. It may solidify tomorrow, or it may take a century to take form. As a librarian, I very much hope that the values and skills of contemporary libraries inform the Library of the Future, but I am realistic enough to acknowledge that it might not happen. If we librarians do not act promptly, we may very well be totally marginalized. The Library of the Future will come about, but somebody else will build it.

What disturbs me the most is the uncertainty that we inevitably face looking forward, wondering whether this is going to be a particularly long and severe period of increasingly repressive controls, or if there may actually be a great deal of gathering strength in the trends that foster freedom of communication and access to information. If the traditional library has declined, we can hope that it will be transformed and reborn quickly this time around. Whichever possibility comes to pass, this is truly a critical moment for the Information Commons of Society.

## REFERENCES CITED

- Eisenstein, Elizabeth. *The Printing Press as an Agent of Change: Communications and Cultural Transformations in Early-Modern Europe, Volumes I & II*. London: Cambridge University Press, 1979.
- Foucault, Michel. *Discipline and Punish: the Birth of the Prison*. New York: Random House, 1975.
- Halbert, Martin, with Joanne Kaczmarek and Kat Hagedorn. "Findings from the Mellon Metadata Harvesting Initiative," in *Research and Advanced Technology for Digital Libraries, 7th European Conference, ECDL 2003*, Trondheim, Norway, August 17-22, 2003, Traugott Koch, Ingeborg Sølvsberg (Eds.), Proceedings. Lecture Notes in Computer Science 2769 Springer 2003, ISBN 3-540-40726-X. pp. 58-69.
- Halbert, Martin, with Aaron Krowne. "Combined Searching of Web and OAI Digital Library Resources," in *Joint Conference on Digital*

*Libraries 2004 Proceedings*. New York: Association for Computing Machinery, 2004. pp. 343-344.

Havelock, Eric. *Preface to Plato*. Cambridge: Harvard University Press, 1963.

Hugo, Victor. *Notre-Dame de Paris*. 1831. Translated and with an introduction and notes by Alban Krailsheimer. Oxford University Press, 1993.

Nunberg, Geoffrey. (ed.) with an afterword by Umberto Eco. *The Future of the Book*. Berkeley, California: University of California Press, 1996.

Ong, Walter J. *Orality and Literacy : the Technologizing of the Word*. London. New York: Methuen, 1982.



# Between Pragmatism and Anarchism: The American Copyright Revolt since 1998

Siva Vaidhyanathan (New York University)


**Abstract:** Since 1998, questions about whether the United States has constructed an equitable or effective copyright system frequently appear on the pages of daily newspapers. Calls both for stronger and looser copyright systems have grown in volume and furor. Such debate echoes around several important court cases. For example, the U.S. Supreme Court ruled in early 2003 that the foundations of American copyright, as expressed in the Constitution, are barely relevant in an age in which both media companies and clever consumers enjoy unprecedented power over the use of works. Such tensions and conflicts have been narrated through the frameworks of binaries such as “protection vs. piracy” and “property vs. commons.” Some accounts of recent copyright battles have emphasized the excessive, often absurd, level of protection and vigilance by copyright holders. Others have explored the influence of scholar-activists in the rise of a reform movement. In contrast, this paper argues that the best way to explain copyright trends and battles in recent years is to examine the struggles that individuals and groups have mounted on behalf of their rights and abilities to control their cultural and information ecosystems. This is a pragmatic analysis—focused on what people can and may do with their culture and the information available to them. The struggle has been about local and private autonomy over what gets sung, played, and made—about who gets to generate the soundtracks of American life in the 21st century.

## FIRST PRINCIPLES

About a year before the U.S. presidential election of 2004, a group of students at Swarthmore College in Pennsylvania posted on a university-provided website some information that they considered essential to public debate in a democratic republic. They offered a collection of 15,000 e-mail messages and memos generated by Diebold Election Systems, one of the leading manufacturers of controversial electronic voting machines. The collection of

---

M. Halbert (Ed.): *Free Culture and the Digital Library Symposium Proceedings*. Atlanta, Georgia: MetaScholar Initiative at Emory University, 2005. pp. 13-42.

 This work is covered by the Creative Commons Attribution-NonCommercial-NoDerivs 2.5 License.

documents from Diebold revealed that in 2002 elections the company's proprietary software had suffered from many alarming problems, ranging from security weaknesses to miscounting to plain failure. Upon learning that Swarthmore students had posted these internal documents on college servers, Diebold sent a "cease-and-desist" letter to Swarthmore, demanding that the college immediately remove the copyrighted material under the "notice and takedown" provisions of the 1998 Digital Millennium Copyright Act (DMCA).<sup>1</sup>

Swarthmore officials complied with Diebold's requests, thus protecting the college from the potential civil judgment that its students would face instead. Diebold underestimated the will and means of this group of students, however. The student group, then called Swarthmore Coalition for the Digital Commons, was well acquainted with the DMCA and copyright law in general. The students were aware of many other stories of corporate copyright intimidation intended to limit criticism. They later changed their group's name by adopting the title of Lawrence Lessig's influential book of 2004, *Free Culture*. The book has since lent its title to the entire global movement of copyright critics.<sup>2</sup> So, under the direction of their leader Nelson Pavlovsky, the students sought the aid of the Electronic Frontier Foundation (EFF), which agreed to represent the students in their effort to defend themselves against Diebold. Meanwhile, word quickly spread around Internet sites and communities devoted to fighting copyright expansion—and the collection of memos did, too. Within days one could acquire copies of the memos from peer-to-peer interfaces such as Kazaa, Gnutella, and Freenet. Many other websites openly posted the memos and challenged Diebold to try to stamp out every source.<sup>3</sup>

After much adverse publicity, Diebold backed down on its threats to the Swarthmore students. But the students were not done with Diebold. With the help of the EFF, they sued Diebold in federal court, issuing a rather untested claim, "copyright misuse." Less than a year after Diebold shut down the site, Judge Jeremy Fogel wrote in his decision in favor of the Swarthmore students, "no reasonable copyright holder could have believed that portions of the e-mail archive discussing possible technical problems with Diebold's voting machines were protected by copyright." In addition, Fogel ruled that Diebold had "knowingly materially misrepresented" its copyright claims and had misused the DMCA "as a sword to suppress publication of embarrassing content rather than as a shield to protect its intellectual property."<sup>4</sup>

By fighting back rather than backing down, the students at Swarthmore did more than foster a better climate for debate and criticism of voting methods in the United States. The very distribution of the Diebold memos accomplished that. By using methods both within the law (via the federal courts) and beyond the law (by facilitating anarchistic distribution of the memos regardless of threats from Diebold and cowardice by the Swarthmore administration), they set an example for activists and citizens' groups in many areas of life to follow. After decades of being shut out of information policy decisions in Washington, D.C., many Americans have banded together as both formal organizations and informal information networks to push back for greater democratic control of culture and information.

The Diebold case exemplifies what is at stake. Will only powerful institutions, those with adequate legal representation and capital at their disposal, be able to enter debates about important public issues? Or will legal belligerence chill critics who lack resources to defend themselves? Will copyright law act as it was intended—to foster a richer public sphere—or will it work against its expressed aims by retarding speech and criticism? What may citizens do with these powerful new media systems at their disposal? Must citizens be receptors of content or may they speak back in texts rich with reference yet devoid of reverence?

At the very moment when inexpensive communicative technologies, widespread literacy, universal public education, and civil rights allow us to live in the sort of democratic culture that John Dewey could only dream of, expansive copyright laws and an ideology of vigilance and surveillance undermine efforts to foster the richest possible environment for democracy to flourish in America.<sup>5</sup> In his seminal debate with Walter Lippmann concerning the potential for democratic governance in a modern, technocratic nation, Dewey focused on capabilities—on what was possible or reasonable to expect from citizens. In the 1920s, it was hard to imagine that the United States might some day develop its information infrastructure and cultural habits in such a way as to foster real democracy. Dewey had faith and hope. If he had had e-mail, he might have seen the route to that realization.<sup>6</sup>

Copyright is supposed to perform a pragmatic role in the cultural economy of the United States. It was designed to generate confidence among creators, distributors, and investors so that they believe they might reap returns on certain cultural and intellectual endeavors.<sup>7</sup> To foster such confidence, the United States

Constitution instructs Congress to create a system of laws that would create limited monopolies. Wary of the censorious and corrupting potential of monopolies (especially state-granted monopolies), the founders explicitly limited the scope and duration of copyright.<sup>8</sup> For most of the ensuing 200 years, American courts and Congress maintained that healthy respect for the negative externalities of powerful copyright protection, and thus designed and redesigned the system to work while respecting the rights of citizens to use and build upon works already in circulation.<sup>9</sup>

The system maintained a healthy equilibrium until the mid-1970s, when two disruptive technologies—the photocopier and magnetic audio tape—threatened to lower the cost of copying and distribution to such a level as to allow real democratization of communication. In subsequent years the availability of inexpensive high-quality cassette audio tape, personal tape players such as the Sony Walkman, and home video cassette recorders (VCRs) amplified commercial anxieties and consumer opportunities. From about 1975 through 2005 copyright battles took on a new dimension. For the first time, citizens could control their personal media spaces. They could create and distribute their own or others' work over long distances to many people at low marginal cost. New consumer markets developed. But so did extra-market or non-market global discursive and creative communities such as punk rock, hip hop, ska, militant Islam, and computer hacking. While much recent commentary on the relationship among copyright, culture, and technology has focused on the Internet and such formats as MP3, the real global democratic technology involved the nexus of the cassette tape and the battery-powered tape recorder.<sup>10</sup>

In 1998, the U.S. Congress radically revised American copyright laws without much public scrutiny or protest. Copyright was too arcane, too technical, too boring, to break through the headlines about political sex scandals and celebrity murder trials. With the Sonny Bono Copyright Term Extension Act (SBCTEA) and the Digital Millennium Copyright Act (DMCA) the United States abandoned 200 years of moderate, successful copyright traditions. Copyright used to balance the public's interests and private needs. Now it only serves large, established copyright holders. Yet while Congress was considering these radical changes, newspapers and thus the public, scarcely paid attention to the changes.<sup>11</sup> Only in recent years, with the accumulation of horror stories about copyright abuses and bullying, have we seen sufficient attention

paid. As a result, we are finally seeing a critical mass of public interest activism. Between the spring of 2001 and the winter of 2003 the following events kept copyright in the news:

- Eric Eldred, a World Wide Web publisher, found that his practice of publishing public domain works on the Internet is thwarted by Congress' radical extension of the duration of protection for works created in the 1930s and after. After both a district court and an appeals court ruled that Eldred's claim that the extension was unconstitutional (in violation of the requirement that copyright last "for limited times"), the Supreme Court considered the merits of his case in October 2002. Then, in January 2003, the Supreme Court ruled 7-2 to uphold the lower court ruling allowing the copyright term extension.<sup>12</sup>
- The National Writers' Union, led by their president Jonathan Tasini, won a landmark case before the U.S. Supreme Court in 2001. The court ruled that freelance writers who had not explicitly assigned their rights to electronic versions of their work were due compensation from major newspapers and magazines that had sold these rights to electronic databases such as Lexis/Nexis and ProQuest. This case somewhat redressed the balance between creator and publisher in the copyright system, although in most media and in most fields the creator still operates from a very weak bargaining position.<sup>13</sup>
- In the summer of 2001, a federal court issued an injunction against the publication of a novel by Alice Randall called *The Wind Done Gone*. This new novel was a revision and retelling of Margaret Mitchell's *Gone With the Wind*, published originally in 1935. Despite the fact that the original novel should have entered the public domain some time in the 1980s, Congress kept its copyright alive through retroactive copyright extension—the very issue the Supreme Court considered in the Eldred case. Appealing the injunction against the publication of *The Wind Done Gone*, lawyers for the publisher, Houghton Mifflin, argued that the new novel was a parody of the original, and thus the use of similar characters and events constituted "fair use." The appeals court agreed with the parody argument and allowed the novel to be published.<sup>14</sup>
- Also in the summer of 2001, a federal court freed choreographer Martha Graham's legacy from the hands of her

friend, a pretend executor, Ronald Protas. The court ruled that Protas only controlled the rights to a single dance, "Seraphic Dialogue," after he claimed to control most of her oeuvre, and had exercised exclusive rights over many of her dances, thus preventing many companies from performing the works. In addition, a court ruled that Protas could not prevent the Martha Graham Dance Company from using its founder's name.<sup>15</sup>

- The Federal Bureau of Investigation handcuffed Russian computer programmer Dmitry Sklyarov in the Las Vegas airport after he had given a presentation on the security vulnerabilities in Adobe Corporation's E-book Reader software. The company Sklyarov worked for in Moscow, Elcomsoft, soon faced federal criminal charges in the United States (even though the DMCA is only a United States law) after Sklyarov agreed to testify in exchange for immunity. First Sklyarov, and then Elcomsoft, were accused of violating the Digital Millennium Copyright Act by distributing a program that willfully violated the act by allowing readers to make private copies of e-books. In December 2002, a San Francisco jury found the company not guilty.<sup>16</sup>
- The recording industry moved its attention from the distributors of peer-to-peer software to those actually offering copyrighted music files on those networks, filing civil suits against hundreds of individuals.<sup>17</sup>

Since 1998 questions about whether the United States has constructed an equitable or effective copyright system frequently appear on the pages of daily newspapers. Calls both for stronger and looser copyright systems have grown in volume and furor. Such debate echoes around several important court cases. For example, the U.S. Supreme Court ruled in early 2003 that the foundations of American copyright, as expressed in the Constitution, are barely relevant in an age in which both media companies and clever consumers enjoy unprecedented power over the use of works. Such tensions and conflicts have been narrated through the frameworks of binaries such as "protection vs. piracy" and "property vs. commons." Some accounts of recent copyright battles have emphasized the excessive, often absurd, level of protection and vigilance by copyright holders.<sup>18</sup> Others have explored the influence of scholar-activists in the rise of a reform movement.<sup>19</sup> In contrast, this paper argues that the best way to explain copyright trends and battles in recent years is to examine the struggles that individuals and groups have mounted on behalf

of their rights and abilities to control their cultural and information ecosystems. This is a pragmatic analysis—focused on what people can and may do with their culture and the information available to them. The struggle has been about local and private autonomy over what gets sung, played, and made—about who gets to generate the soundtracks of American life in the 21st century.

### **IRRATIONAL EXUBERANCE**

The copyright system used to be brilliant and effective. It attracted massive investment. It generated the rise of the motion picture, recording, and software industries during the 20th century. It filled libraries with books. But these metrics are not as revealing as the fact that until 1998 there did not exist in U.S. history a copyright rebellion—a widespread and somewhat organized set of practices aimed at resisting the power of copyright holders. There was always infringement. But the cost of infringement was built into the price of goods and it was considered either an acceptable leak or a positive social good (when employed for the purpose of education or criticism).<sup>20</sup>

Something changed fundamentally in the 1990s. Policy makers saw copyright as something very much beyond and unlike its traditional role and scope. They saw copyright and the industries that depend on copyright as the source of much future wealth for the United States. And they disregarded the democratic safeguards that had kept copyright users and future creators satisfied (or, at least unconcerned) with the system. So they advocated policy changes that were meant to maximize the potential revenue from copyrighted goods by minimizing the leaks in the copyright system—even if those very leaks were what made copyright work so well and so quietly.<sup>21</sup>

The radical changes of the late 1990s attracted the core of copyright activists who now make so much noise. Generally, these activists lament the erosion of the democratic safeguards that made American copyright such a brilliant and effective system and filled our libraries with books. Copyright can censor. It is a prohibition on what we may reproduce, quote, perform, and distribute. Through both statutes and the common law over the past 200 years, the copyright system developed four democratic safeguards that mitigated the potentially censorious power of its prohibitions:

- The principle of fair use—at its base a legal defense against an accusation of copyright infringement. If you are accused of infringing, you can make an argument that your use of the

protected works is “fair” because of some combination of the following four factors: the nature of the original work is important to public discussions or concerns; the nature of your use of it is important because of teaching, research, or commentary; you did not use very much of the original work; your use did not significantly affect the market for the original work. In the public discourse about fair use, it has served as a term representing a collection of uses that consumers could consider “fair,” such as recording television shows for later viewing, making cassette tape or MP3 mixes from compact discs, and limited copying for private, noncommercial sharing.

- The principle that after the “first sale” of a copyrighted item, the buyer could do whatever she wants with the item—including lending, reselling, or burning—save publicly performing the work or distributing unauthorized copies of it for sale. The first sale doctrine is what makes the lending library possible.
- The concept that copyright protected specific expression of ideas, but not the ideas themselves. This is the least understood but perhaps most important tenet of copyright. You can’t copyright a fact or an idea. Because you can’t, anyone may repeat your idea to criticize it or build on it. Journalism, along with many other forms of common expression, depends on this principle.
- The promise that copyright would only last—as the Constitution demands—“for limited times,” thus constantly replenishing the public domain. The public domain allows for low-cost scholarship, research, and revision of formerly copyrighted works. The reason that bookstores are filled with high-quality yet affordable scholarly editions of Mark Twain’s *Adventures of Huckleberry Finn* and John Stuart Mill’s *On Liberty* is that they are in the public domain. The reason there is no annotated scholarly edition of Ralph Ellison’s *Invisible Man* is that it is not.

Copyright, when well balanced, encourages the production and distribution of the raw material of democracy. It is supposed to be an economic incentive for the next producer, not a guarantee for the established. But the new rhetoric of copyright, in the wake of the infamous Department of Commerce “White Paper” of 1995 no longer reflects that subtle dynamic. The general message delivered by copyright maximalists is something close to “if some copyright is good, more is better.”<sup>22</sup>



So in general critics of the corrupted copyright system share an agenda that would restore those democratic safeguards. They have been fighting for a hacker magazine's (and thus everyone's) First Amendment right to describe certain illegal algorithms and create hyperlinks to other pages that describe and offer these algorithms. And they have been playing defense in the halls of the Capitol against legislation that would create a new and dangerous property rights in facts and data and other, more odious legislation that would require all producers of electronic hardware and software to include anti-copying devices in their products. In addition, they have engaged in disobedience, both civil and uncivil. Pranks, hacktivism, satire, and unmasked outrage have all played a part in the copyright rebellion.

### **PARACOPYRIGHT**

Inspired by the vision outlined in the "White Paper" and influenced by the motion picture and recording industries, Congress in 1998 added a power to copyright holders that went far beyond the general right to exclude others from making copies of works without payment and authorization. Over and above real copyright, these new rights restrict a variety of uses and regulate access to the work itself. This new mode of regulation, what Peter Jaszi has called "paracopyright," has been the source of much anger and consternation, and has served as the locus for much of the copyright rebellion.

Consider the action that the Church of Scientology took in the summer of 2002 against the search engine Google.com. The Church of Scientology used a "notice and takedown" letter (authorized under the DMCA) to persuade Google.com to block links to a Norwegian site that includes some criticism of the wealthy cult. Back in the 20th century, if someone accused you of copyright infringement, you enjoyed that quaint and seemingly archaic notion of due process. You would be warned and perhaps sued. And if you wanted to defend yourself in court, you could appear at a hearing, present evidence and arguments, and have a judge render a ruling based on statute and precedent. We no longer have those rights in the digital world. The DMCA puts the burden of proof against an accusation of copyright infringement on the accused. And it makes the owner of every Internet service provider, search engine, and content host an untrained copyright cop. The default action is censorship.<sup>23</sup>

The conflict between Scientology and Google is just one in a string of DMCA-supported copyright abuse. The collection of evidence of the “chilling effect” on website authors has grown into a major concern among activists for Free Culture.<sup>24</sup> Besides limiting due process and empowering private parties to enforce censorship through the “notice and takedown” provision, the Digital Millennium Copyright Act has another major provision that upends more than 200 years of democratic copyright law. It forbids the circumvention of electronic access controls that protect works—even those portions of works that might be in the public domain or subject to fair use. It puts the absolute power to regulate access to information in the hands of the companies that distribute the material.

More to the point of intellectual freedom, in the spring of 2001 the music industry prevented a computer scientist from presenting a scholarly paper at a conference because the paper dealt with encryption algorithms that the recording industry hoped to use to protect its digital content. The Recording Industry Association of America sent a “cease and desist” letter to Princeton professor Edward Felten, accusing him of violating the provision of the DMCA that makes it illegal to “make available” any technology that might be used to circumvent access controls to digital material. The Felten case is merely the best known of several efforts the content industries have made to prevent researchers from discussing certain technologies and algorithms.<sup>25</sup>

The Digital Millennium Copyright Act is the legal backbone behind the move to install “digital rights management” (DRM) technologies to digital content. As a result of putting the power of the United States government behind encryption and other similar technologies that govern users’ ability to use the work, the DMCA grants an alarming amount of power to allow or deny access to a work with the producer or publisher of that work. The producer may prohibit access for those users who might have hostile intentions toward the work. This power could exclude critics and scholars. Most likely it could exclude parodists and satirists as well. The anti-circumvention provision shifts the site of setting terms of use from the user (and the courts in the case of likely infringement) to the producer. The producer has no incentive to grant access to any user who might exploit the work for fair use—including scholarship, teaching, commentary, or parody. Under this regime, a user must agree to terms of a contract with a

monopolistic provider before gaining access. One must apply to read, listen, or watch.

The absurdity of digital rights management, and thus the DMCA, has never been clearer.<sup>26</sup> As technology advocate and novelist Cory Doctorow has argued, DRM does not work, it is bad for society, bad for business, and bad for individual artists. Rights management systems involve futile commitments of resources to the installation and re-installation of DRM systems, largely because they break so easily. As Doctorow writes, “DRM systems are usually broken in minutes, sometimes days. Rarely, months. It's not because the people who think them up are stupid. It's not because the people who break them are smart. It's not because there's a flaw in the algorithms. At the end of the day, all DRM systems share a common vulnerability: they provide their attackers with ciphertext, the cipher and the key. At this point, the secret isn't a secret anymore.”<sup>27</sup>

Doctorow's assessment matches the record. Every effective digital rights management scheme released commercially to the public since 1998 has been cracked or easily evaded. So they have had no positive effects, in the sense that they have spectacularly failed to limit the unauthorized distribution of digital materials, what the copyright industries call “piracy.” But the very presence of such “electronic fences” in the digital environment has had two negative effects.<sup>28</sup>

First, and perhaps most significantly for the long-term prospects of the copyright system, DRM schemes have frustrated consumers and put them in an oppositional and rebellious position in relation to the firms that distribute protected products such as copy-protected compact discs and electronic books. Copyright users have few qualms about cracking and evading limits on products that they have purchased or materials that they consider to be parts of their culture. Digital rights management and the frustration it has generated have undermined the social norms that a healthy copyright system needs to function. As a result, the copyright industries that fought for the DMCA in 1998 have done more harm to the principles of copyright than their opponents have. In fact, critics of excessive copyright such as Lawrence Lessig have been clear about their belief in real copyright as an engine of free expression. Alas, the motion picture industry, the recording industry, and Congress gave up on real copyright in favor of paracopyright in 1998.<sup>29</sup>

The second effect has been censorious and monopolistic. The DMCA has prevented non-infringing, socially beneficial uses of copyright material. It has restricted researchers like Ed Felten from doing their jobs without fear and mobilized many academics to join the “Free Culture Movement” and engage with its intellectual branch, Critical Information Studies.<sup>30</sup> It has driven librarians to call for exceptions and exemptions from its restrictions so they could deal with technological failure that might prevent access to their collections.<sup>31</sup>

In the private sector, the DMCA has enabled the rise of “technology cartels” among firms that sign on to licensing terms that tether digital files to particular digital rights management schemes—thus cutting competing technology firms (and non-firms, such as Open Source software projects) out of certain markets. And it prevents people who use Open Source software from legally using digital content they have lawfully acquired.<sup>32</sup> The DMCA even encouraged (ultimately futile) attempts to limit competition in after-market goods such as printer cartridges and garage door openers.<sup>33</sup>

The DMCA is the fulfillment of a Robber-Baron-era (or, at least pre-New Deal) proprietarian ideology, one that fundamentally relies on private ownership and strict privacy rights in spite of the negative public externalities they create. In a political and legal environment in which the DMCA represents the clearest statement of such values, it has been difficult to assert a different vision of a good information ecosystem, of information justice.<sup>34</sup>

## **THE PUBLIC’S DOMAIN**

The case of *Eldred vs. Ashcroft*, inspired by the efforts of independent publisher Eric Eldred, who wanted to use the powerful distribution technology of the World Wide Web to offer public domain works in usable form, ended in disappointment for the Free Culture movement. Even though the ruling in *Eldred vs. Ashcroft* was a blow to efforts to immediately open up more democratic breathing space in copyright, the decision itself offers seeds that might grow into something good. Justice Stephen Breyer wrote in his dissent: “It is easy to understand how the statute might benefit the private financial interests of corporations or heirs who won existing copyrights. But I cannot find any constitutionally legitimate, copyright-related way in which the statute will benefit the public.” This is the key to any public interest movement: show

that narrow special interests are profiting and the public interest is suffering.

In her majority opinion, Justice Ruth Bader Ginsburg herself aided the public's rhetorical cause even while ruling against its interests. While dismissing the notion that excessive copyright expansion has severe First Amendment implications, she invoked two of the classic democratic safeguards of American copyright: the idea/expression dichotomy and fair use. Because of these two concepts, Ginsburg concluded, the court need not take the censorious power of copyright seriously. Ginsburg's expression of faith in the power of the idea/expression dichotomy and fair use did not recognize that both these rights are under attack in Congress and lower courts. The motion picture, music, publishing, and software industries are trying to expand their control over the home use of machines in order to limit the uses people might make of material they have lawfully purchased. Ginsburg made one more statement that public interest advocates can take to heart and use for their purposes. While dismissing the petitioners' First Amendment concerns, she wrote, "But when, as in this case, Congress has not altered the traditional contours of copyright protection, further First Amendment scrutiny is unnecessary."<sup>35</sup>

As a matter of fact, the 1998 Digital Millennium Copyright Act did just that. By outlawing technologies that could break through access controls around digital materials, Congress created a whole new technological regime and a new set of powers for copyright holders to use against scholars, librarians, students, and artists. This shift in the locus of enforcement from human relations to hard technology has certainly "altered the traditional contours of copyright protection." As Yale Law professor Jack Balkin has argued, these words could be used to render the most pernicious parts of the Digital Millennium Copyright Act unconstitutional. In the wake of this decision, if Congress and later courts are going to take Ginsburg's words seriously, they must take fair use and the idea/expression dichotomy seriously. They cannot take them for granted, as so many have in recent years.

The Eldred decision, in the words of University of Buffalo law professor Shubha Gosh, "deconstitutionalizes" copyright, pushing it further into the realm of policy and power battles and away from the principles that have anchored the system for two centuries. That means public interest advocates and activists must take their battles to the public sphere and the halls of Congress. They can't appeal to the founders' wishes or republican ideals. They will

have to make pragmatic arguments in clear language about the effects of excessive copyright on research, teaching, art, and journalism. Because of both the publicity and the result of *Eldred vs. Ashcroft*, the Free Culture movement grew in volume and determination.

### **THE PARODY PARASOL**

The one users' right that has grown stronger in recent years involves the use of copyrighted material for parody. Building on the 1994 Supreme Court case of *Campbell vs. Acuff Rose*, in which the court ruled in favor of hip hop group 2 Live Crew after it issued a parody of Roy Orbison's classic "Oh, Pretty Woman," and the 2002 case involving *The Wind Done Gone*, a revision of the *Gone with the Wind* story, many copyright rebels have sought refuge under the parody parasol. Simply stated, parody is fair use.<sup>36</sup>

Yet when two Web cartoonists calling themselves Jib Jab released a satirical version of Woody Guthrie's song "This Land is Your Land," Guthrie's musical executors considered the work to be beyond the rather narrow legal definition of parody. Parody is supposed to target the original work, not some third party or society in general. This is the legal definition between parody and satire. In this case, Jib Jab had used the song to make fun of the two men running for president in 2004. They had rewritten the words to the song and placed alternating stanzas in the mouths of caricatures of George W. Bush and John Kerry. Emboldened by their misunderstanding of parody, yet willing to fight instead of relent to the pressure of a cease-and-desist letter, Jib Jab asked the EFF to take its case. After some quick exploration about whether this case could serve to expand the definition of legally protected parody, EFF lawyers instead pursued the idea that "This Land is Your Land" is in the public domain. Research on both its origin (based on an older Carter Family song in the public domain) and the fact that Guthrie never renewed his copyright on the song proved that this song is our song.<sup>37</sup>

### **THE PEER-TO-PEER PARADOX**

No copyright phenomenon has generated as much attention, anxiety, and excitement as "peer-to-peer." The nature of peer-to-peer technology is widely misunderstood and the rhetoric surrounding it has been inflated and heated. Since the rise of Napster, a relatively centralized method of resolving information inquiries, popular accounts of the workings of peer-to-peer

functions have described them as being substantially new and profound. Yet at their most basic level, most common procedures on the Internet are already peer-to-peer. Every Web page search involves a resolution of an inquiry through an index, and then a link to a server on which the desired file sits. Searches through commercial services such as Google.com work in ways very much like the original Napster: a centralized index that links seekers to files held on third-party servers. The services we commonly call “peer-to-peer networks” (Napster, Kazaa, Gnutella, Grokster, etc.) are merely methods of resolving information queries laid over the network of networks we already use: the Internet. The rise of such resolution interfaces represents a return to the early state of the Internet, when individuals generated and distributed content as well as consuming it.<sup>38</sup>

However, recent moral panics about peer-to-peer distribution of copyrighted files have reached into the educational realm and disrupted reputable software engineering experiments that might yield better tools if allowed to flourish or fail outside the threat of civil judgments or state-imposed restrictions. Jesse Jordan, a student at Rensselaer Polytechnic Institute (RPI) in Troy, New York, settled a lawsuit in 2003 for \$12,000 after the Recording Industry Association of America filed suit against him for creating an indexed search engine for public folders on computers hooked up to the RPI computer network. Such a system would have been very helpful to those using the powerful university computer network. Often members of university communities host many volumes of reports, data sets, commentaries, reviews, teaching materials, and other libraries of data in remote corners of the network. Standard search engines only scan the indexed portions of the official sites and servers operated by university offices. But sometimes the best information sits on a connected computer on the edge of the network, virtually invisible to most researchers. Jordan’s system might have opened up many more interesting files to the RPI community. Jordan himself copied no files. He issued no encouragements to students or faculty to post copyrighted materials. Yet the very act of experimenting with creative media technologies resulted in a lawsuit and forced a settlement.<sup>39</sup> Educators and students have learned much from anecdotes such as Jordan’s. As a result, scholars hesitate to invent or deploy innovative peer-to-peer indexes and resolution processes that might spread data and processing power among a series of underused computers rather than centralizing such functions on one expensive computer.

As I write this, the U.S. Supreme Court is considering its ruling in a major case concerning peer-to-peer interfaces and the liability that companies that distribute such software might face. Media companies have asked the Court to reconsider and revise its decision in *Sony* and to create a notion of “inducement” as a cause for contributory infringement. In other words, if the court revises its standard from *Sony*, courts may hold software designers liable for the infringement that their work allows once it leaves their firms. Holding engineers responsible for the infringement others commit is a frightening prospect for many—not only technologists but artists and educators as well.<sup>40</sup>

Almost every act of teaching relies on the substantial replication and revision of others’ copyrighted works. Lectures, group projects, and assignments all rely on copying, distribution, and performance of copyrighted works. Teachers necessarily and consciously induce such copying. Many of the basic tools of teaching such as distributing photocopies, performing copyrighted works in class, and viewing film and video in class, would usually constitute copyright infringements. Yet Congress acknowledges that these functions are central to the mission of adequately educating students who live in an increasingly media-saturated society.

New technologies made media education and study more dynamic, effective, and accessible. For example, the proliferation of video cassette recorders (and such ancillary products as inexpensive video cameras and editing machines) truly unleashed the potential for media education. We copy and thus potentially infringe with video technology. But we have done so under the presumed protection of fair use. But such fair uses would have been impossible without the video recorder, the video camera, and without the confidence in technological experimentation set free by the U.S. Supreme Court’s ruling in *Sony Corp. vs. Universal City Studios, Inc.* Media education and scholarship never would have developed as an important field in college and university curricula and an increasingly important element of secondary education in the United States without such technology.<sup>41</sup>

Newer digital technologies are even more promising for educators and students. The costs of production and reproduction have fallen. Media studies are no longer unidirectional fields, with information flowing from the front of the classroom to the back. Digital technology has become democratized to such a degree that the walls among instructor, student, creator, and audience have



eroded. Every media student has the potential to build on the work of those who came before and comment critically on her media environments by answering in a multimedia, intertextual, dynamic manner, only because U.S. law has facilitated technological experimentation that has in turn generated a flurry of curricular initiatives.

One of the best examples of the creative use of the technology liberated by Sony comes from the Media Education Foundation, established in 1991 at the University of Massachusetts at Amherst. Under the direction of Professor Sut Jhally and with assistance from students and the public, the Foundation has been collecting video clips of copyrighted media messages and images and assembling them into annotated and narrative videos for classroom use. The videos produced by the Foundation have had a profound effect on media education at all levels. Without the strong and clear message sent by Sony, the Media Education Foundation would not have been able to produce videos examining the sexist images promoted by MTV or the troublesome relationship between musicians and the major recording companies. None of the concerned companies would have cleared their images for use in a critical educational video.<sup>42</sup> Sony made such productions—and many of the recent advances in higher education in general—possible.

The Fair Use provisions of the Copyright Act, as delineated by Sec.107, did not by themselves grant the confidence sufficient to spark technological experimentation and curricular initiatives such as the use of video cameras and editing in the classroom or teacher-produced media education videos. Only in the wake of Sony did such innovation emerge. In recent years, as digital technologies and powerful networks have granted remarkable creative tools to scholars, teachers, and students, the climate of panic and fear induced by the uncertainties of fair use in the new digital environment has generated a chilling effect. University and school administrators are cautious about or vehemently against experimenting with new methods of distribution, even for educational or research purposes.<sup>43</sup> For example, Professor Henry Jenkins at the Massachusetts Institute of Technology uses—as most media studies teachers do—clips and quotes from copyrighted works in his courses. On advice from MIT lawyers, the university has not allowed Jenkins to post the essential clips on its open courseware servers—only on server space closed to readers who are not registered MIT students. However, MIT

allows students from Harvard University to take courses at MIT. Such material is inaccessible to Jenkins' students from Harvard. This situation has frustrated Jenkins and prevented him from teaching his course as effectively as he might under a more relaxed and confident legal environment.<sup>44</sup>

Many scholars use peer-to-peer technology in their work. Some seek a song or video clip that is out of print and unavailable in their libraries, so they use the vast publicly generated library of files as an efficient index and virtual library. Others are curious about the function of such systems and their effects on culture and the culture industries. Still others are fascinated by the software itself and strive to understand and perhaps improve it. One of the most exciting scholarly proposals is "Edutella," an open-source project that builds upon metadata standards to generate similar standards for peer-to-peer applications. This project will make searching using peer-to-peer interfaces more precise and effective, thus unleashing the distributed nature of the Internet to store essential documents redundantly and dependably. Maintaining central servers is costly for educational institutions so many information experts see distributed information as way to make educational resources available to teachers and researchers who do not have access to large libraries or servers.<sup>45</sup> Other similar initiatives include "OAI-P2P," an effort to link all data in open archives via a peer-to-peer search interface that would link all the metadata attached to all the content in all the databases, and "Bibster," an effort to exchange bibliographic metadata across many institutions.<sup>46</sup>

Such scholarly peer-to-peer experiments are benign and potentially valuable. Yet the mere suggestion that researchers employ peer-to-peer technology invites scrutiny and suspicion. Henry Jenkins at MIT could solve his content distribution problem by deploying a search engine like the one Jordan developed at RPI. But without clear legal guidance that would enable Jenkins and MIT lawyers to allow such experimentation confidently, Jenkins will not even try. More interesting than what scholars do with peer-to-peer technology is what they might not do if the current mood of panic fails to ebb. Many other uses of distributed computing or peer-to-peer indexing and resolution have yet to be imagined in the educational context. Yet, like the democratization of video production twenty years ago, there is no way for anyone to predict the externalities (positive and negative) that might flow from granting confidence to scholars, teachers, and students.

## THE ANARCHY OF CULTURAL PRACTICE

While peer-to-peer has attracted the most attention, the most important element of the copyright rebellion comes from creative communities such as Free and Open Source Software (FOSS) advocates and digital music and video producers who have built impressive new works from the elements of culture and information that flow by them every day. Yochai Benkler calls this phenomenon “peer production.” In every one of these cases, people release their work to a wider audience of contributors and creators, who then add incrementally to the project, thus building large things out of many small pieces. The GNU-Linux operating system is the best known of such peer-produced projects. The Internet itself—or at least its core protocols—represents another.<sup>47</sup>

Inspired by the power of the GNU General Public License (GPL), legal language that travels with many Free and Open Source software projects, locking open further contributions so that the entire project remains out of proprietary hands, Lawrence Lessig, James Boyle and others developed a licensing system for other kinds of content. The Creative Commons project leverages the cultural power and political statements of the Free and Open Source community to demonstrate through its application to music, video, and text that many creators would prefer to have their work shared and altered as long as they retain credit and no one captures it. In this way, thousands of creators have enlisted in the copyright revolt by building something new rather than destroying something old.<sup>48</sup>

Another, more radically tinged element of the copyright revolt relies on twisting copyright horror stories into public lessons. The activist group Downhill Battle has taken on such projects as distributing video files of the civil rights documentary “Eyes on the Prize” after copyright clearance problems stifled its digital re-release. Downhill Battle’s most influential prank involved the distribution of an underground album called “The Grey Album.” Produced by an artist who calls himself DJ Danger Mouse, the “Grey Album” is a brilliant combination of the lyrical track from hip hop star Jay-Z’s “Black Album” with musical samples from the 1968 album, “The Beatles,” commonly known as “The White Album.” On the first Tuesday in February 2004, Downhill battle encouraged hundreds of website editors to distribute the illicit files of the “Grey Album” as a challenge to lawyers for EMI, the Beatles’ publisher. EMI had forced DJ Danger Mouse to cease distributing the album himself. As a result of Grey Tuesday, the

album was a major hit, and was even reviewed in *The New York Times*.<sup>49</sup>

## PROSPECTS

If the music and film industries continue to tighten the reins on use and access, they will strangle the public domain and the information commons. This trend presents a much greater threat to American culture than just a chilling effect on scholarship and creativity. Shrinking the information and cultural commons starves the public sphere of elements of discourse, the raw material for decision-making, imagination, and humor. In addition, these industries will fuel the growing outrage about these and other examples of copyright holders using their new legal powers to stifle criticism and undermine legitimate uses of their material. Loud protests have emerged from communities of software producers, artists, writers, librarians, and media activists. Activist organizations such as the Electronic Frontier Foundation and [publicknowledge.org](http://publicknowledge.org) are struggling to accurately define the “public interest” in copyright and debating how best to articulate the issues to a diverse public. At one point, Napster had 77 million registered users, more than twice the number of users that American Online enjoys. That means there were 77 million potential infringers walking our streets. And there are few Americans who have not wondered about the intrusive power of that video mattress tag—the FBI warning at the start of every rented videotape.

But we can’t have the conversation that would lead us to that best possible copyright system as long as we continue to work within the limited rhetorical frameworks that we have inherited. We make a grave mistake when we choose to engage in discussions of copyright along the terms of “property.” Copyright is not “property” as commonly understood. It is a specific state-granted monopoly issued for particular policy reasons. While technically, such terms describe real property as well, the public understanding of property is more fundamental, more exclusive, more natural, and precedes specific policy choices the state may make about its regulation and dispensation. When we engage in “property talk,” we can’t compete with the content industries. It’s impossible to have a clear and reasonable discussion about what sort of copyright system might be best for the United States and the world as long as those who hold inordinate interest in copyright maximalization can cry “theft” at any mention of fair use or users’ rights. This is the “property-talk trap”: You can’t argue for theft.

Two rhetorical strategies have emerged out of the concern about the “property talk trap.” Most prominent is “commons talk.” A growing number of activists and law professors are pushing for an appreciation of the “information commons.” Sparked by a brilliant paper by Duke law professor James Boyle titled “A Politics of Intellectual Property: Environmentalism For the Net?,” this movement toward preservation and expansion of an information commons resembles the environmental movement 40 years ago.<sup>50</sup> With some good luck and hard work, these activists hope to build a similar level of public concern and awareness about how information operates in society, and the need for it to be commonly owned and shared. The best defense of the information commons can be found in a new book by activist David Bollier called *Silent Theft: The Private Plunder of our Common Wealth*. In this sober and lucid book, Bollier considers issues as wide ranging as private exploitation of federal pharmaceutical research funds, the commercialization of public space, and the enclosure of the “academic commons.” It is essential reading for anyone concerned with the future of “the public” and its potential survival.<sup>51</sup> In addition, Lawrence Lessig argues persuasively for a commons on several “layers” of communication in his important book *The Future of Ideas*.<sup>52</sup>

The second rhetorical strategy involves focusing on uses and users of copyrighted material—everyone who reads, writes, watches, photographs, listens, and sings. This is a more pragmatic approach, intended to warn people that the harmless acts they have taken for granted for years, such as making a mixed tape or CD for a party or “time-shifting” television programs and skipping commercials, are threatened by these recent changes in law and technology.

In addition to promulgating a healthy vision of an information commons and emphasizing the practical ramifications of extreme copyright, Free Culture advocates must confront several other trends and issues. They must link their efforts to other democratic efforts such as the privacy and media reform movements. There is strong continuity among these areas of policy and practice. One of the reasons the digital environment has fostered such a strong level of vigilance and mania to restrict the use of copyrighted material in any form is that it also fosters an environment of surveillance.<sup>53</sup> Minor infringements that used to cause no concern in the analog world—sharing music among friends, for instance—now attract the attention of lawyers. And Free Culture advocates must

recognize that the massive consolidation among media firms in the 1990s increased their political power both in the United States and around the world, thus allowing them to dictate copyright laws globally.<sup>54</sup>

Six years after the U.S. Congress passed the 1998 Digital Millennium Copyright Act and the Sonny Bono Copyright Term Extension Act, it should be clear that they were both tremendous mistakes and failures. They have done much harm and no good. The Internet is ripe with unauthorized digital content of all kinds. Peer-to-peer systems are fulfilling the role of a disorganized global digital library. Street corners from Manhattan to Mexico City to Manila to Moscow to Mumbai are filled with pirated discs. And laws and technological locks have done little to change that. However, these laws have stifled legitimate and harmless users of digital materials, especially scholars, librarians, and researchers.

## ENDNOTES

1. 17 U.S.C. Section 512(c)(3) and 512(d)(3). Section 512(c)(3) sets out the elements for notification under the DMCA. Subsection A (17 U.S.C. 512(c)(3)(A)) states that to be effective a notification must include: 1) a physical/electronic signature of a person authorized to act on behalf of the owner of the infringed right; 2) identification of the copyrighted works claimed to have been infringed; 3) identification of the material that is claimed to be infringing or to be the subject of infringing activity and that is to be removed; 4) information reasonably sufficient to permit the service provider to contact the complaining party (e.g., the address, telephone number, or email address); 5) a statement that the complaining party has a good faith belief that use of the material is not authorized by the copyright owner; and 6) a statement that information in the complaint is accurate and that the complaining party is authorized to act on behalf of the copyright owner. Subsection B (17 U.S.C. 512(c)(3)(B)) states that if the complaining party does not substantially comply with these requirements the notice will not serve as actual notice for the purpose of Section 512.
2. Lawrence Lessig, *Free Culture: How Big Media Uses Technology and the Law to Lock Down Culture and Control Creativity* (New York: Penguin Press, 2004).
3. John Schwartz, "File Sharing Pits Copyright against Free Speech," *The New York Times*, November 3 2003.
4. *Online Policy Group, Et Al. Vs. Diebold Incorporated, Et Al.*, 337 F. Supp. 2d 1195 (2004). Also see Kim Zetter, Diebold Loses Key Copyright Case (*Wired*

News, 2004 [cited May 1 2005]); available from [http://www.wired.com/news/evote/0,2645,65173,00.html?tw=wn\\_tophead\\_2](http://www.wired.com/news/evote/0,2645,65173,00.html?tw=wn_tophead_2).

5. Sonia K. Katyal, "Privacy Vs. Piracy," *International Journal of Communication Law and Policy* 9 (2005).
6. John Dewey, *The Public and Its Problems* (New York: H. Holt and Company, 1927). Also see Walter Lippmann, *The Phantom Public* (New York: Harcourt, 1925), Walter Lippmann, *Public Opinion* (New York: Harcourt, 1922).
7. Peter Drahos, *A Philosophy of Intellectual Property* (Aldershot ; Brookfield, USA: Dartmouth, 1996).
8. Neil Weinstock Netanel, "Copyright and Democratic Civil Society," *Yale Law Journal* 106, no. 2 (1996).
9. Siva Vaidhyanathan, *Copyrights and Copywrongs: The Rise of Intellectual Property and How It Threatens Creativity* (New York: New York University Press, 2001).
10. Peter Lamarche Manuel, *Cassette Culture: Popular Music and Technology in North India, Chicago Studies in Ethnomusicology* (Chicago: University of Chicago Press, 1993). Also see Siva Vaidhyanathan, *The Anarchist in the Library: How the Clash between Freedom and Control Is Hacking the Real World and Crashing the System* (New York: Basic Books, 2004), Siva Vaidhyanathan, "Remote Control: The Rise of Electronic Cultural Policy," *The Annals of the American Academy of Political and Social Science* 597 (2005).
11. Jessica Litman, *Digital Copyright: Protecting Intellectual Property on the Internet* (Amherst, N.Y.: Prometheus Books, 2001).
12. *Eldred Vs. Ashcroft*, 239 F.3d 372 (2003). Also see Lessig, *Free Culture : How Big Media Uses Technology and the Law to Lock Down Culture and Control Creativity*.
13. *New York Times Co. Vs. Tasini*, 206 F.3d 161 (2001).
14. *Suntrust Vs. Houghton Mifflin*, 136 F. Supp. 2d 1357 (2001). Also see Lessig, *Free Culture: How Big Media Uses Technology and the Law to Lock Down Culture and Control Creativity*, Vaidhyanathan, *The Anarchist in the Library: How the Clash between Freedom and Control Is Hacking the Real World and Crashing the System*.
15. *The Martha Graham School and Dance Foundation, Inc. And Ronald Protas Vs. Martha Graham Center of Contemporary Dance, Inc. And Martha Graham School of Contemporary Dance*, (2001). Also see Anthony Lin, "Dances by Graham Held 'Work for Hire' in Ownership Case," *New York Law Journal*, August 23 2004; Vaidhyanathan, *Copyrights and Copywrongs: The Rise of Intellectual Property and How It Threatens Creativity*.
16. Vaidhyanathan, *The Anarchist in the Library: How the Clash between Freedom and Control Is Hacking the Real World and Crashing the System*.

17. William W. Fisher, *Promises to Keep: Technology, Law, and the Future of Entertainment* (Stanford, Calif.: Stanford Law and Politics, 2004), J. D. Lasica, *Darknet: Hollywood's War against the Digital Generation* (Hoboken, NJ: J. Wiley & Sons, 2005), Lessig, *Free Culture: How Big Media Uses Technology and the Law to Lock Down Culture and Control Creativity*, Kembrew McLeod, *Freedom of Expression: Overzealous Copyright Bozos and Other Enemies of Creativity*, 1st ed. (New York: Doubleday, 2005), Vaidhyanathan, *The Anarchist in the Library: How the Clash between Freedom and Control Is Hacking the Real World and Crashing the System*.
18. David Bollier, *Brand-Name Bullies: And Their Quest to Own and Control Culture* (Hoboken, N.J.: J. Wiley, 2005), McLeod, *Freedom of Expression: Overzealous Copyright Bozos and Other Enemies of Creativity*.
19. Robert S. Boynton, "The Tyranny of Copyright?," *The New York Times Magazine*, January 25 2004, Dan Hunter, "Culture War," *Texas Law Review* 83 (2005).
20. L. Ray Patterson, *Copyright in Historical Perspective* (Nashville: Vanderbilt University Press, 1968), L. Ray Patterson and Stanley W. Lindberg, *The Nature of Copyright: A Law of Users' Rights* (Athens: University of Georgia Press, 1991). Also see Paul Goldstein, *Copyright's Highway: From Gutenberg to the Celestial Jukebox* (New York: Hill and Wang, 1994), Benjamin Kaplan, *An Unhurried View of Copyright, James S. Carpentier Lectures, 1966* (New York: Columbia University Press, 1967), Vaidhyanathan, *Copyrights and Copywrongs: The Rise of Intellectual Property and How It Threatens Creativity*.
21. Goldstein, *Copyright's Highway: From Gutenberg to the Celestial Jukebox*, Bruce Lehman, "Intellectual Property and the National Information Infrastructure: The Report of the Working Group on Intellectual Property Rights," (Washington, D.C.: U.S. Patent and Trademark Office, 1995), Litman, *Digital Copyright: Protecting Intellectual Property on the Internet*, Vaidhyanathan, *Copyrights and Copywrongs: The Rise of Intellectual Property and How It Threatens Creativity*.
22. Lehman, "Intellectual Property and the National Information Infrastructure: The Report of the Working Group on Intellectual Property Rights." Also see Goldstein, *Copyright's Highway: From Gutenberg to the Celestial Jukebox*, Lawrence Lessig, *Code and Other Laws of Cyberspace* (New York: Basic Books, 1999).
23. Matt Loney, Google Pulls Anti-Scientology Links (*C-Net News*, 2002 [cited May 1 2005]); available from [http://news.com.com/2100-1023-865936.html?tag=cd\\_mh](http://news.com.com/2100-1023-865936.html?tag=cd_mh).
24. *Chilling Effects Clearinghouse* [Web page] (2005 [cited May 1 2005]); available from <http://www.chillingeffects.org/>.



25. Vaidhyanathan, *The Anarchist in the Library: How the Clash between Freedom and Control Is Hacking the Real World and Crashing the System*.
26. *Unintended Consequences: Five Years under the Dmca* [Web page] (Electronic Frontier Foundation, September 24, 2003 2003 [cited May 1 2005]); available from [http://www.eff.org/IP/DMCA/?f=unintended\\_consequences.html](http://www.eff.org/IP/DMCA/?f=unintended_consequences.html).
27. Cory Doctorow, *Microsoft Research Dm Talk* [Web page] (2004 [cited May 1 2005]); available from <http://www.craphound.com/msftdrm.txt>.
28. Vaidhyanathan, *The Anarchist in the Library: How the Clash between Freedom and Control Is Hacking the Real World and Crashing the System*.
29. Christopher May, "Digital Rights Management and the Breakdown of Social Norms," *First Monday* 8, no. 11 (2003). Also see Hunter, "Culture War."
30. Siva Vaidhyanathan, "Critical Information Studies: A Manifesto," (New York: 2005).
31. *Dmca: The Digital Millennium Copyright Act* [Web page] (American Library Association, 2005 [cited May 1 2005]); available from <http://www.ala.org/ala/washoff/WOissues/copyrightb/dmca/dmcaDIGITALMILLINIUM.htm>. Also see Nancy C. Kranich, *Libraries & Democracy: The Cornerstones of Liberty* (Chicago: American Library Association, 2001).
32. *Universal City Vs. Reimerdes*, 111 F. Supp. 325 (2001). Also see Dan L. Burk, "Anticircumvention Misuse," *UCLA Law Review* 50 (2003), Vaidhyanathan, "Remote Control: The Rise of Electronic Cultural Policy."
33. *Chamberlain Group, Inc. Vs. Skylink Technologies, Inc.*, 292 F. Supp. 2d 1040 (2003), *Lexmark International Inc. V. Static Control Components Inc.*, 387 F.3d 522 (2004). Also see Burk, "Anticircumvention Misuse."
34. Julie E. Cohen, "Lochner in Cyberspace: The New Economic Orthodoxy of "Rights Management", *University of Michigan Law Review* 97, no. 2 (1998).
35. *Eldred Vs. Ashcroft*.
36. *Campbell Vs. Acuff-Rose Music*, 510 U.S. 569 (1994), *Suntrust V. Houghton Mifflin*.
37. McLeod, *Freedom of Expression: Overzealous Copyright Bozos and Other Enemies of Creativity*.
38. Andrew Oram and Safari Books Online., *Peer-to-Peer: Harnessing the Power of Disruptive Technologies*, 1st ed. (Beijing; Cambridge [Mass.]: O'Reilly, 2001). Also see Siva Vaidhyanathan, *The Anarchist in the Library: How the Clash between Freedom and*

- Control Is Hacking the Real World and Crashing the System* (New York: Basic Books, 2004).
39. Tim Goral, "Recording Industry Goes after Campus P-2-P Networks: Suit Alleges \$97.8 Billion in Damages," *Professional Media Group LCC* 2003. Also see Lessig, *Free Culture: How Big Media Uses Technology and the Law to Lock Down Culture and Control Creativity*.
  40. *Metro-Goldwyn-Mayer Studios Inc. Vs. Grokster Ltd.*, 380 F3d 1154 (2004). Also see *In Re Aimster Copyright Litigation*, 334 F. 3d 643 (2003), *A&M Records, Inc. Vs. Napster, Inc.*, 239 F. 3d 1004 (2001).
  41. *Sony Corp. Vs. Universal City Studios, Inc.*, 464 U.S. 417 (1984).
  42. *Media Education Foundation: About* [World Wide Web page] (Media Education Foundation, [cited February 15 2005]); available from <http://www.mediaed.org/about>. Also see Kembrew McLeod, *Freedom of Expression<sup>a</sup>: Overzealous Copyright Bozos and Other Enemies of Creativity*, 1st ed. (New York: Doubleday, 2005).
  43. Andrea L. Foster, "Justice Department Wants Colleges to Do More to Stop File Sharing," *The Chronicle of Higher Education*, July 30 2004.
  44. Henry Jenkins, e-mail correspondence to Siva Vaidhyanathan, February 23 2005.
  45. Wolfgang Nejdl *et al.*, "Edutella: A P2p Networking Infrastructure Based on Rdf" (paper presented at the WWW2002, Honolulu, May 7-11 2002).
  46. Benjamin Ahlborn, Wolfgang Nejdl, and Wolf Siberski, "Oai-P2p: A Peer-to-Peer Network for Open Archives," (2002). Also see Peter Haase *et al.*, "Bibster-Semantics-Based Bibliographic Peer-to-Peer System," *Web Semantics: Science, Services and Agents on the World Wide Web* 2, no. 1 (2004).
  47. Yochai Benkler, "Coase's Penguin, or, Linux and the Nature of the Firm," *Yale Law Journal* 112, no. 3 (2002).
  48. *Creative Commons* (2005 [cited May 1 2005]); available from <http://creativecommons.org>.
  49. Sam Howard-Spink, "Grey Tuesday, Online Cultural Activism and the Mash-up of Music and Politics," *First Monday* 9, no. 10 (2004).
  50. James Boyle, "A Politics of Intellectual Property: Environmentalism for the Net?," *Duke Law Journal* 47 (1997).
  51. David Bollier, *Silent Theft: The Private Plunder of Our Common Wealth* (New York: Routledge, 2002).
  52. Lawrence Lessig, *The Future of Ideas: The Fate of the Commons in a Connected World*, 1st ed. (New York: Random House, 2001).
  53. Katyal, "Privacy Vs. Piracy."

54. Robert Waterman McChesney, *The Problem of the Media : U.S. Communication Politics in the Twenty-First Century* (New York: Monthly Review Press, 2004).

## REFERENCES CITED

- Ahlborn, Benjamin, Wolfgang Nejdil, and Wolf Siberski. "Oai-P2P: A Peer-to-Peer Network for Open Archives." 2002.
- Benkler, Yochai. "Coase's Penguin, or, Linux and the Nature of the Firm." *Yale Law Journal* 112, no. 3 (2002).
- Bollier, David. *Brand-Name Bullies: And Their Quest to Own and Control Culture*. Hoboken, N.J.: J. Wiley, 2005.
- . *Silent Theft: The Private Plunder of Our Common Wealth*. New York: Routledge, 2002.
- Boyle, James. "A Politics of Intellectual Property: Environmentalism for the Net?" *Duke Law Journal* 47 (1997): 87.
- Boynton, Robert S. "The Tyranny of Copyright?" *The New York Times Magazine*, January 25 2004.
- Burk, Dan L. "Anticircumvention Misuse." *UCLA Law Review* 50 (2003): 1095-135.
- Campbell Vs. Acuff-Rose Music*, 510 U.S. 569 (1994).
- Chamberlain Group, Inc. Vs. Skylink Technologies, Inc.*, 292 F. Supp. 2d 1040 (2003).
- Chilling Effects Clearinghouse* [Web page]. 2005 [cited May 1 2005]. Available from <http://www.chillingeffects.org/>.
- Cohen, Julie E. "Lochner in Cyberspace: The New Economic Orthodoxy of "Rights Management". *University of Michigan Law Review* 97, no. 2 (1998): 462.
- Creative Commons*, 2005 [cited May 1 2005]. Available from <http://creativecommons.org>.
- Dewey, John. *The Public and Its Problems*. New York,: H. Holt and Company, 1927.
- Dmca: The Digital Millennium Copyright Act* [Web page]. American Library Association, 2005 [cited May 1 2005]. Available from <http://www.ala.org/ala/washoff/WOissues/copyrightb/dmca/dmcardigitalmillennium.htm>.
- Doctorow, Cory. *Microsoft Research Dm Talk* [Web page]. 2004 [cited May 1 2005]. Available from <http://www.craphound.com/msftdrm.txt>.
- Drahos, Peter. *A Philosophy of Intellectual Property*. Aldershot; Brookfield, USA: Dartmouth, 1996.
- Eldred Vs. Ashcroft*, 239 F.3d 372 (2003).

- Fisher, William W. *Promises to Keep : Technology, Law, and the Future of Entertainment*. Stanford, Calif.: Stanford Law and Politics, 2004.
- Foster, Andrea L. "Justice Department Wants Colleges to Do More to Stop File Sharing." *The Chronicle of Higher Education*, July 30 2004, A22.
- Goldstein, Paul. *Copyright's Highway: From Gutenberg to the Celestial Jukebox*. New York: Hill and Wang, 1994.
- Goral, Tim. "Recording Industry Goes after Campus P-2-P Networks: Suit Alleges \$97.8 Billion in Damages." *Professional Media Group LCC* 2003.
- Haase, Peter, Bjorn Schnizler, Jeen Broekstra, Marc Ehrig, Frank van Harmelen, Maarten Menken, Peter Mika, Michal Plechawski, Pawel Pyszlak, and Ronny Siebes. "Bibster--a Semantics-Based Bibliographic Peer-to-Peer System." *Web Semantics: Science, Services and Agents on the World Wide Web* 2, no. 1 (2004): 99-103.
- Howard–Spink, Sam. "Grey Tuesday, Online Cultural Activism and the Mash–up of Music and Politics." *First Monday* 9, no. 10 (2004).
- Hunter, Dan. "Culture War." *Texas Law Review* 83 (2005): 1105.
- In Re Aimster Copyright Litigation*, 334 F. 3d 643 (2003).
- Jenkins, Henry. e-mail correspondence to Siva Vaidhyanathan, February 23 2005.
- Kaplan, Benjamin. *An Unhurried View of Copyright, James S. Carpentier Lectures, 1966*. New York: Columbia University Press, 1967.
- Katyal, Sonia K. "Privacy Vs. Piracy." *International Journal of Communication Law and Policy* 9 (2005): 7.
- Kranich, Nancy C. *Libraries & Democracy: The Cornerstones of Liberty*. Chicago: American Library Association, 2001.
- Lasica, J. D. *Darknet : Hollywood's War against the Digital Generation*. Hoboken, NJ: J. Wiley & Sons, 2005.
- Lehman, Bruce. "Intellectual Property and the National Information Infrastructure: The Report of the Working Group on Intellectual Property Rights." Washington, D.C.: U.S. Patent and Trademark Office, 1995.
- Lessig, Lawrence. *Code and Other Laws of Cyberspace*. New York: Basic Books, 1999.
- . *Free Culture: How Big Media Uses Technology and the Law to Lock Down Culture and Control Creativity*. New York: Penguin Press, 2004.
- . *The Future of Ideas: The Fate of the Commons in a Connected World*. 1st ed. New York: Random House, 2001.
- Lexmark International Inc. Vs. Static Control Components Inc.*, 387 F.3d 522 (2004).

- Lin, Anthony. "Dances by Graham Held 'Work for Hire' in Ownership Case." *New York Law Journal*, August 23 2004.
- Lippmann, Walter. *The Phantom Public*. New York,: Harcourt, 1925.
- . *Public Opinion*. New York: Harcourt, 1922.
- Litman, Jessica. *Digital Copyright: Protecting Intellectual Property on the Internet*. Amherst, N.Y.: Prometheus Books, 2001.
- Loney, Matt. *Google Pulls Anti-Scientology Links C-Net News*, 2002 [cited May 1 2005]. Available from [http://news.com.com/2100-1023-865936.html?tag=cd\\_mh](http://news.com.com/2100-1023-865936.html?tag=cd_mh).
- A&M Records, Inc. Vs. Napster, Inc.*, 239 F. 3d 1004 (2001).
- Manuel, Peter Lamarche. *Cassette Culture: Popular Music and Technology in North India, Chicago Studies in Ethnomusicology*. Chicago: University of Chicago Press, 1993.
- The Martha Graham School and Dance Foundation, Inc. And Ronald Protas Vs. Martha Graham Center of Contemporary Dance, Inc. And Martha Graham School of Contemporary Dance*, (2001).
- May, Christopher. "Digital Rights Management and the Breakdown of Social Norms." *First Monday* 8, no. 11 (2003).
- McChesney, Robert Waterman. *The Problem of the Media: U.S. Communication Politics in the Twenty-First Century*. New York: Monthly Review Press, 2004.
- McLeod, Kembrew. *Freedom of Expression: Overzealous Copyright Bozos and Other Enemies of Creativity*. 1st ed. New York: Doubleday, 2005.
- Media Education Foundation: About* [World Wide Web page]. Media Education Foundation, [cited February 15 2005]. Available from <http://www.mediaed.org/about>.
- Metro-Goldwyn-Mayer Studios Inc. Vs. Grokster Ltd.*, 380 F3d 1154 (2004).
- Nejdl, Wolfgang, Boris Wolf, Changtao Qu, Stefan Decker, Michael Naeve, Mikael Nilsson, Matthias Palmer, and Tore Risch. "Edutella: A P2p Networking Infrastructure Based on Rdf." Paper presented at the WWW2002, Honolulu, May 7-11 2002.
- Netanel, Neil Weinstock. "Copyright and Democratic Civil Society." *Yale Law Journal* 106, no. 2 (1996): 356-86.
- New York Times Co. Vs. Tasini*, 206 F.3d 161 (2001).
- Online Policy Group, Et Al. Vs. Diebold Incorporated, Et Al.*, 337 F. Supp. 2d 1195 (2004).
- Oram, Andrew, and Safari Books Online. *Peer-to-Peer: Harnessing the Power of Disruptive Technologies*. 1st ed. Beijing; Cambridge [Mass.]: O'Reilly, 2001.
- Patterson, L. Ray. *Copyright in Historical Perspective*. Nashville: Vanderbilt University Press, 1968.

- Patterson, L. Ray, and Stanley W. Lindberg. *The Nature of Copyright: A Law of Users' Rights*. Athens: University of Georgia Press, 1991.
- Schwartz, John. "File Sharing Pits Copyright against Free Speech." *The New York Times*, November 3 2003.
- Sony Corp. Vs. Universal City Studios, Inc.*, 464 U.S. 417 (1984).
- Suntrust Vs. Houghton Mifflin*, 136 F. Supp. 2d 1357 (2001).
- Unintended Consequences: Five Years under the Dmca* [Web page]. Electronic Frontier Foundation, September 24, 2003, 2003 [cited May 1 2005]. Available from [http://www.eff.org/IP/DMCA/?f=unintended\\_consequences.html](http://www.eff.org/IP/DMCA/?f=unintended_consequences.html).
- Universal City Vs. Reimerdes*, 111 F. Supp. 325 (2001).
- Vaidhyanathan, Siva. *The Anarchist in the Library: How the Clash between Freedom and Control Is Hacking the Real World and Crashing the System*. New York: Basic Books, 2004.
- . *Copyrights and Copywrongs: The Rise of Intellectual Property and How It Threatens Creativity*. New York: New York University Press, 2001.
- . "Critical Information Studies: A Manifesto." New York, 2005.
- . "Remote Control: The Rise of Electronic Cultural Policy." *The Annals of the American Academy of Political and Social Science* 597 (2005): 122-33.
- Zetter, Kim. "Diebold Loses Key Copyright Case," *Wired News*, 2004 [cited May 1 2005]. Available from [http://www.wired.com/news/evote/0,2645,65173,00.html?tw=wn\\_top\\_head\\_2](http://www.wired.com/news/evote/0,2645,65173,00.html?tw=wn_top_head_2).

# It Is Easy for Universities to Support Free Culture with Digital Libraries: The NDLTD Example

Edward A. Fox (Virginia Tech)


**Abstract:** Computer networks have made it easy and beneficial for students to upload electronic theses and dissertations (ETD) to digital libraries and repositories. This paper describes issues in the implementation of ETD repositories and recounts the development of a successful example, the Networked Digital Library of Theses and Dissertations (NDLTD). The paper argues that implementers may better understand digital libraries, and have greater impact on the movement toward free culture, if they build upon the 5S framework of Streams, Structures, Spaces, Scenarios, and Societies.

## INTRODUCTION

The Union Catalog of the Networked Digital Library of Theses and Dissertations, NDLTD ([www.ndltd.org](http://www.ndltd.org)), has over 200,000 electronic theses or dissertations (ETDs), from over 40 data providers supporting the Open Archives Initiative Protocol for Metadata Harvesting. NDLTD has over 200 members, including national libraries from countries such as Sudan and the UK. While a great deal of progress has been made to advance the ETD initiative worldwide, there is much yet to do, so as to revolutionize the future of the academy, so that the next generation of scholars and leaders will support and utilize digital libraries. Yearly, we could have over 100K ETDs submitted around the world and made available. This sharing would dwarf many initiatives that are being pursued at great expense, and which might have much less impact. Imagine the dramatic improvements to research and graduate studies if all theses and dissertations were prepared, submitted, and archived electronically—which could happen if graduate students were guided by their advisors, supported by the faculty, and energized by leadership in their universities as they prepared and submitted ETDs. Unfortunately, most students preparing a thesis or dissertation remain ignorant of the ETD

---

M. Halbert (Ed.): *Free Culture and the Digital Library Symposium Proceedings*. Atlanta, Georgia: MetaScholar Initiative at Emory University, 2005. pp. 43-55.

 This work is covered by the Creative Commons Attribution-NonCommercial-NoDerivs 2.5 License.

initiative, fear what might happen if they freely shared research results, and sense a void in academic leadership. Yet, it is quite easy for an institution to launch an ETD initiative. Universities around the globe have done so, and as a result have: saved money as a result of automation, increased their visibility in the global research community (which downloads ETDs, on average, hundreds or thousands of times each year), and shown their support for free sharing of knowledge. Whether one uses ETD-db, DSpace, Eprints, Digital Commons@, or another tool to help in this effort, digital library technology as well as software and services are key enablers of ETD activities.

### **FREE CULTURE AND DIGITAL LIBRARIES MADE EASY, WORLDWIDE**

One of the easiest and most effective ways to support free culture and digital libraries is to join a local initiative for electronic theses and dissertations (ETDs). If there is no such local initiative, you can help lead an effort to start one, or can support the Networked Digital Library of Theses and Dissertations, NDLTD (Fox 1997), established to help ETD activities worldwide and to encourage international collaboration in the academy (Fox et al. 1997). A key goal is to prepare the next generation of scholars, e.g., those who create a thesis or dissertation, to understand how to use and how to add to a digital library (Fox, Hall, and Kipp 1997). Since there are masters and doctoral programs in colleges and universities around the world, ETD activities can be in every continent, country, region, state, city, and town—a truly global initiative (Suleman et al. 2001a, 2001b) that is completely scalable.

To facilitate work on ETD projects, *The ETD Guide* (Moxley, Masiello, and Fox 2002) was prepared in multiple languages with support from UNESCO, by a team of authors and editors from around the globe (Fox 2001). An edited book was published as well, with broad coverage of all the key concerns, and detailed discussion of many effective solutions (Fox, Feizbadi et al. 2004). These works address perspectives of students, faculty, and administrators. They even supported the training of trainers who have aided the emergence of new national ETD programs, such as in a growing number of countries in Latin America. In addition, there are links from the NDLTD home page (Fox 1997) to a great deal of useful information, including reports from around the globe, as presented at the annual international conference, e.g., the one held in 2005 in Sydney, Australia (NDLTD 2005).



Regarding global needs, NDLTD builds its access infrastructure (Suleman and Fox 2002b) upon the gathering of metadata about ETDs (Suleman and Fox 2003) using the Protocol for Metadata Harvesting (Van de Sompel and Lagoze 2001; Lagoze et al. 2002; Suleman 2002) of the Open Archives Initiative (Van de Sompel and Lagoze 2000; Suleman and Fox 2001; Suleman and Fox 2002a). OCLC maintains a union catalog, collecting from sites that support the OAI-PMH (OCLC 2004). Search engines (Suleman and Fox 2003; VTLS 2004) work from that catalog, using various systems and approaches, covering a variety of languages, extended through mirroring (CALIS 2004), and, later in 2005, expanding to facilitate full-text searching through Scirus (Elsevier 2005). Ultimately the hope is that the metadata being collected will improve in quality and follow ETD-ms, the ETD metadata standard that evolved from several years of international discussion (Atkins et al. 2001).

ETD activities also have connections with local plans, policies, and practices (Fox, McMillan et al. 2004). At Virginia Tech, efforts (McMillan 1998) began in 1987, initially involving the Graduate School and campus computing (Eaton, Fox, and McMillan 1998). Then, with the help of the Library, a more robust treatment of practical and operational issues was developed (McMillan and Peters 1999). Concerns with access, cataloging, and preservation also were addressed (McMillan 1997, 1999a, 1999b). A software solution was developed at Virginia Tech in 1996, ETD-db (Kletnieks 2005); it has been refined, used, and adapted in locations around the world. Since that time, many other software packages have been adapted to the requirements of diverse campus ETD programs, including:

- DSpace (MIT 2003), with the ETD DSpace Implementers Group (Hemminger 2005)
- Eprints (EPrints.org 2002)
- VTLS's VITAL (VTLS 2005), built around Fedora (Staples and Wayland 2000)
- Digital Commons@ (bepress 2005) by Berkeley Electronic Press

Ultimately, however, the success of an ETD initiative depends upon students and faculty. While high level leadership paves the way for programs to be launched rapidly, e.g., within 6 months (Harrison et al. 2004), maximal benefit results from enthusiastic support by those most directly involved:

- Students can use online ETDs to facilitate their education and research, e.g., starting with NDLTD services, finding works related to their interest.
- Students can plan to describe their research through an electronic document, so as to facilitate rapid completion of an ETD when they finish their degree, learning the various skills needed for that process well in advance; those skills also can help them prepare electronic versions of proposals, submissions to conferences, and reports.
- Students can enhance their ability to communicate electronically beyond what is feasible with a document produced only to be printed; thus they can:
  - Include hyperlinks, both inside the ETD, and to other online resources;
  - Include digital images from digital cameras, from medical units, from special devices/sensors—or digitized after capture using analog equipment;
  - Include multimedia content to help explain time-based phenomena, such as animation, music, dance, surgery, and other processes and procedures; and
  - Include datasets, possibly with software to manage the data, to facilitate replication of experiments, and to popularize testbeds, tools, and methods.
- Students can allow worldwide access to their research, so that hundreds or thousands of others interested in their work can download copies and cite their results.
- Students can learn about free access as well as intellectual property issues, including about patents and copyright. If it is necessary, they can restrict access to some or all of their ETD for a limited period, e.g., a year, but can later benefit from increased visibility of their findings, while all along their home institution has a complete and current repository of works, with continually increasing levels of access.
- Faculty can rest assured that that they have prepared their advisees well for leading the next generation of scholars, can have key findings well documented and archived, and can benefit from globally expanding interest in their research.

Thus, digital library methods can enhance the free sharing of graduate research results, expanding immensely the growth of a free culture among scholars, who have had intense personal

involvement in this process (in connection with a valuable body of their own work, i.e., their thesis or dissertation, in which they have had long-term and serious interest). At Virginia Tech, where ETDs have been required since 1997, already by 2000, according to a survey, student views were overwhelmingly positive (Eaton, Fox, and McMillan 2000). Further, according to the data given in Table 1, based on Virginia Tech logs, access to Virginia Tech ETDs has grown rapidly and has been extensive.

	1997/98	1998/99	1999/00	2000/01	2001/02	2002/03
Requests for PDF files	221.6 K	481 K	578 K	2,173 K	4,497 K	7,320 K
Requests for HTML files	165.7 K	215.5 K	260.7 K	400 K	472 K	368 K
Requests for Multimedia	1.7 K	4.5 K	12.6 K	44 K	169.1 K	121 K
Distinct files requested	6.4 K	21 K	16.4 K	N/A	51 K	31.9 K
Distinct hosts served	29.8 K	57.9 K	87.8 K	N/A	425 K	681 K
Average transfer/day	156 Kb	219 Mb	382 Mb	945 Mb	2.15 Gb	3.49 Gb
Data transferred	55.6 Gb	78.1 Gb	137 Gb	332 Gb	780 Gb	1.2Tb

Table 1. Server log records of accesses to Virginia Tech ETDs.

### **A 5S PERSPECTIVE ON FREE CULTURE AND DIGITAL LIBRARIES, CONSIDERING NDLTD**

A more detailed analysis of the NDLTD example can shed additional light on that enterprise, while at the same time helping explicate the potential of free culture and digital libraries. We apply a relatively new approach to this analysis, based upon the formally defined 5S framework: Streams, Structures, Spaces, Scenarios, and Societies (Gonçalves et al. 2004). Since 1999 we have been developing this approach to explain digital libraries in particular and information systems in general (Fox 1999a, 1999b). Based on analysis of hundreds of papers in the digital library field, we discovered that Streams, Structures, Spaces, Scenarios, and Societies seem to be a necessary and sufficient set of well-defined concepts that can be used, individually or in concert, to precisely characterize an information system. Thus, we have applied 5S to prepare a minimalist explanation of “digital library” (Gonçalves et al. 2004), and, more recently, of a minimalist archaeological digital library (Shen et al. 2005). Accordingly, we consider NDLTD

below, from a 5S perspective (summarized in Table 2), arguing informally for the sake of brevity, emphasizing aspects related to free culture and digital libraries.

Ss	Examples	Objectives
Streams	Text; video; audio; image	Describes properties of DL content such as encoding and language for texts or particular forms of multimedia data
Structures	Collection; catalog; hypertext; document; metadata	Specifies organizational aspects of the DL content
Spaces	Measure; measurable; topological; vector; probabilistic	Defines logical and presentational views of the several DL components
Scenarios	Searching; browsing; recommending	Details the behavior of DL services
Societies	Service managers; learners; teachers; archaeologists; etc.	Defines managers, responsible for running DL services; actors that use the services; and relationships among them

Table 2. The 5S's, with examples and objectives.

## STREAMS

- The raw content (text, image, audio, video) of every document connected with NDLTD is a stream. So too is the raw stream of bits sent over networks, between computers, supporting all types of transmissions, including harvesting and downloading.
- Free culture encourages transmissions that are not constrained, e.g., cryptographic encodings are not necessary.

## STRUCTURES

- At a more descriptive level, ETDs are composed of structured streams; this is especially clear when one deals with XML, but also pertains to PDF files.
- ETD-ms is a metadata format, for descriptive metadata specifications. These in turn help define a metadata catalog, one of the key parts of a digital library.
- Another structure is the NDLTD Union Catalog; each OAI-PMH-enabled university or regional metadata catalog is another structure that feeds into the Union Catalog. These

structures refer to digital objects, which in turn are defined using Stream and Structure. Digital objects are in collections. Collections help define hypertexts, yet another type of structure, that is at the heart of a browsing service.

- Free cultures benefit from well-defined metadata formats, so that sharing can occur across conventional boundaries, building upon the common understanding that comes from wide-spread adoption of standards.
- Free cultures evolve through expanding social networks and patterns of communication, that can be described using graphs, and so have a particular structure.

## **SPACES**

- Spaces are crucial for indexing and searching services (e.g., vector or probability spaces).
- 2D spaces are the basis for common human-computer interaction that makes use of ever-advancing display technology.
- Free cultures benefit from use of such 2D spaces that follow guidelines for accessibility, and that are not constrained to only work for one language, culture, or nation.
- Free cultures involve people and systems in diverse locations across space.
- Free cultures that use digital libraries can go beyond constraints of time, as content can be accessed by store and get operations (see below). Thus, movement of knowledge takes place across time and space, intermediated by digital libraries.

## **SCENARIOS**

- As mentioned above, a digital library must support a variety of services, which implement scenarios such as indexing, searching, and browsing.
- Another key part of a digital library is its repository, which depends on its (content) collection. Repositories support at least the functions store, delete, and get.
- Regarding the construction of digital libraries, such as to better support NDLTD requirements, we have developed software that fits with automation scenarios according to the 5S framework (Shen et al. 2005; Raghavan et al. 2005), including:

- 5SL—a language for describing digital libraries (Gonçalves and Fox 2002)
- 5SGraph—a tool for digital library designers (Zhu et al. 2004)
- 5SGen—a tool to generate tailored digital libraries (Kelapure 2003)
- XML logging standard for digital libraries (Gonçalves et al. 2002; Gonçalves et al. 2003)

Deployment of these tools as part of digital library development can facilitate the creation of more advanced and useful services.

- One higher level scenario is the career of a graduate student, including education and research. That in turn may include work on a thesis or dissertation, further including preparation of an ETD.
- An even higher level scenario covers a university's graduate education and research activities, which overlap with the career scenarios of large numbers of students, as well as similar career scenarios of faculty, staff, and administrators.
- At an even higher level, above these scenarios, we have the worldwide federated ETD initiative, led by NDLTD, which integrates scenarios involving member institutions, through its harvesting as well as information access services. Ultimately it fits back in with the career scenarios of individual students and researchers, who use the services offered by NDLTD and its partners.
- The abovementioned is a key part of the set of scenarios that support free culture. However, many other scenarios must be supported as well for free culture to expand. While some scenarios that are common in the digital library world, such as authentication and authorization, may be less important, additional scenarios are needed to ensure collaboration and assessment of quality in free cultures.

## **SOCIETIES**

- Societies covered in NDLTD include students, faculty, staff, administrators, librarians, researchers, general patrons of NDLTD services, and those with particular roles related to ETDs: ETDreviewer, ETDcataloger, and ETDsearchManager. The latter is an automated actor; other such service managers support each of the key scenarios, like browsing.

- In a free culture, a key society is comprised of those who prepare and share content. Another key society is comprised of those who make free use of that content. Further, there are all the digital librarians and other helpers who support the infrastructure upon which the free culture is built, e.g., those who support the worldwide ETD initiative.

Clearly, 5S can be used to precisely characterize NDLTD and other digital libraries, as well as key aspects of free culture. It is hoped that such a clear understanding will reinforce the abovementioned explanation of how easy it is to work with ETDs to expand free culture.

## CONCLUSION

There is ample opportunity for students, faculty, staff, and administrators—at universities worldwide—to engage seriously in free culture, facilitated by digital libraries. A clear example, suitable for any location worldwide, is presented by the activities of the Networked Digital Library of Theses and Dissertations, NDLTD, further explained above using the 5S framework. Students seriously interested in research and scholarship can personally engage and contribute by preparing their theses/dissertations as electronic documents, submitting them to a local institutional repository, and indicating (possibly after a relatively short delay, e.g., a year) that access to their work can be freely allowed worldwide. They benefit from this process not only by learning important skills that will help prepare them for future use of digital libraries, and by saving money (as compared to paper submission), but also by having their work become more frequently cited and more highly regarded by larger numbers of their peers. They contribute to the visibility of their university, and help expand a free culture that can grow rapidly among a key sector of academia, as well as to the broader research community.

## Acknowledgements

The work reported in this paper has been supported in part by the US National Science Foundation (NSF), through grants DUE-0136690 and 0333531, as well as IIS-9986089, 0080748, and 0086227. Other support has been provided by the US Department of Education, through FIPSE Program P116B61190. Further support has been provided by Adobe, NDLTD, OCLC, UNESCO, and VTLS. Special thanks go to the Board of Directors and the members of NDLTD, Virginia Tech's Digital Library Research

Laboratory, and all the colleagues and students supporting these efforts.

## REFERENCES CITED

- Atkins, Anthony, Edward A. Fox, Robert K. France, and Hussein Suleman. *ETD-ms: An Interoperability Metadata Standard for Electronic Theses and Dissertations*. NDLTD 2001. [http://www.ndltd.org/standards/metadata/current.html]
- bepress. *bepress Repository Technology*. Berkeley Electronic Press 2005. [http://www.bepress.com/repositories.html]
- CALIS. *NDLTD Union Catalog Mirror Site in China* [Web site] 2004. [http://ndltd.calis.edu.cn]
- Eaton, John, Ed Fox, and Gail McMillan. 2000. Results of a Survey of Virginia Tech Graduates Whose Digital Theses and Dissertations Are Accessible Worldwide. *Council of Graduate Schools (CGS) Communicator* 33 (9):1, 7, 8. [http://www.cgsnet.org/PublicationsPolicyRes/Communicatorpdfs/2000/November00.pdf]
- Eaton, John L., Edward A. Fox, and Gail McMillan. 1998. The Role of Electronic Theses and Dissertations in Graduate Education. *The Council of Graduate Schools Communicator* 31 (1):1.
- Elsevier. *Scirus*. Elsevier 2005. [http://www.scirus.com/]
- EPrints.org. *E-Prints* [Web site]. University of Southampton 2002. [http://www.eprints.org/]
- Fox, E., G. McMillan, H. Suleman, M. Gonçalves, and M. Luo. 2004. Networked Digital Library of Theses and Dissertations. In *Digital Libraries: Policy, Planning, and Practice*, edited by J. Andrews and D. Law: Ashgate Publishing.
- Fox, Edward. 1997. NDLTD: Networked Digital Library of Theses and Dissertations. [http://www.ndltd.org]
- Fox, Edward A. 1999a. The 5S Framework for Digital Libraries and Two Case Studies: NDLTD and CSTC. In *Proceedings NIT'99, Taipei, Taiwan, Aug. 18-20*. [http://www.ndltd.org/talks/nit99.ppt]
- . 1999b. From Theory to Practice in Digital Libraries: 5S and Educational Applications (NDLTD, CSTC). In *NSF-CONACyT-ISTEC Workshop on Digital Libraries, Albuquerque, July 7-9*. [http://www.ndltd.org/talks/nm990708.ppt]
- . 2001. Overview of a Guide for Electronic Theses and Dissertations. In *Digital Libraries and Virtual Workplaces: Important Initiatives for Latin America in the Information Age*, edited by J.V. Reenen. Washington, DC: Organization of American States. [http://www.unm.edu/~jreenen/dlbook/chapter5.html]



- Fox, Edward A., Shahrooz Feizbadi, Joseph M. Moxley, and Christian R. Weisser. 2004. *Electronic Theses and Dissertations: A Sourcebook for Educators, Students, and Librarians*. New York: Marcel Dekker.
- Fox, Edward A., Robert Hall, and Neill Kipp. 1997. NDLTD: Preparing the Next Generation of Scholars for the Information Age. *The New Review of Information Networking (NRIN)* 3:59-76.
- Fox, Edward A., Robert Hall, Neill A. Kipp, John L. Eaton, Gail McMillan, and Paul Mather. 1997. NDLTD: Encouraging International Collaboration in the Academy. *Special Issue on Digital Libraries of DESIDOC Bulletin of Information Technology (DBIT)* 17 (6):45-56.
- Gonçalves, Marcos André, and Edward A. Fox. 2002. 5SL - A Language for Declarative Specification and Generation of Digital Libraries. In *Proc. JCDL'2002, Second ACM / IEEE-CS Joint Conference on Digital Libraries, July 14-18, Portland, Oregon*, edited by G. Marchionini: ACM.
- Gonçalves, Marcos André, Edward A. Fox, Layne T. Watson, and Neill A. Kipp. 2004. Streams, Structures, Spaces, Scenarios, Societies (5S): A Formal Model for Digital Libraries. *ACM Transactions on Information Systems* 22 (2):270-312.  
[<http://doi.acm.org/10.1145/984321.984325>]
- Gonçalves, Marcos André, Ming Luo, Rao Shen, Mir Farooq Ali, and Edward A. Fox. 2002. An XML Log Standard and Tool for Digital Library Logging Analysis. In *Research and Advanced Technology for Digital Libraries, Sixth European Conference, ECDL 2002, Sept. 16-18, 2002, Rome, Italy, Lecture Notes in Computer Science 2458*, edited by M. Agosti and C. Thanos: Springer.  
[<http://www.springerlink.com/index/6D5505P4WHKN4MVL>]
- Gonçalves, Marcos André, Ganesh Panchanathan, Unnikrishnan Ravindranathan, Aaron Krowne, Edward A. Fox, Filip Jagodzinski, and Lillian Cassel. 2003. The XML Log Standard for Digital Libraries: Analysis, Evolution, and Deployment. In *Proc. JCDL'2003, Third ACM / IEEE-CS Joint Conference on Digital Libraries, May 27-31, Houston, TX*: IEEE Computer Society.
- Harrison, Dianne, Judith Devine, Althea Jenkins, David Johnson, and Robert McDonald. 2004. ETD Implementation in Six Months: Five Principles. In *Proc. ETD 2004: Distributing knowledge worldwide through better scholarly communication, June 3-5, University of Kentucky, Lexington, KY, USA*: NDLTD.  
[<http://www.uky.edu/ETD/ETD2004/abstract2.html#devine>]
- Hemminger, Bradley M. *ETD DSpace Implementers Group*. UNC Chapel Hill 2005. [<http://etd.ils.unc.edu/etd-dspace/>]
- Kelapure, Rohit. 2003. Scenario-Based Generation of Digital Library Services. MS thesis, Computer Science, Virginia Tech, Blacksburg, VA. [<http://scholar.lib.vt.edu/theses/available/etd-06182003-055012/>]

- Kletnieks, Carolyn. *Resources for Developers of ETD databases: ETD-db*. Virginia Tech 2005. [<http://scholar.lib.vt.edu/ETD-db/developer/>]
- Lagoze, Carl, Herbert Van de Sompel, Michael Nelson, and Simeon Warner. 2004. *The Open Archives Initiative Protocol for Metadata Harvesting - Version 2.0*, Open Archives Initiative 2002. [<http://www.openarchives.org/OAI/2.0/openarchivesprotocol.htm>]
- McMillan, Gail. 1997. Electronic Theses and Dissertations: Merging Perspectives. *Cataloging and Classification Quarterly* 22 (3/4)
- . 1998. Digital Dissertations: The Virginia Tech Initiative. Paper read at Issues in Science and Technology Librarianship, American Library Association Midwinter, Jan. 10, at New Orleans
- . 1999a. What to Expect from ETDs: If you build it, they will use it. Paper read at Digital Library Symposium, Sept. 21, at Cleveland, OH. [<http://scholar.lib.vt.edu/staff/gailmac/presentations/CWRUGMc1.pdf>]
- . 1999b. What to Expect from ETDs: Library Issues and Responsibilities. Paper read at Digital Library Symposium, Sept. 22, at Cleveland, OH. [<http://scholar.lib.vt.edu/staff/gailmac/presentations/CWRUGMc2.pdf>]
- McMillan, Gail, and Len Peters. 1999. ETDs: Practical, Operational, and Technical Issues for Universities Implementing Electronic Theses and Dissertations. Paper read at Conference on Preservation and Access for Electronic College and University Records, Oct. 8, at Mesa, Arizona.
- MIT. 2004. *DSpace: Durable Digital Depository* [Web site]. MIT 2003. [<http://dspace.org>]
- Moxley, Joseph M., Diane Masiello, and E. Fox. *ETD Guide: The Guide for Electronic Theses and Dissertations*. UNESCO 2002. [[www.etdguide.org](http://www.etdguide.org)]
- NDLTD. *Past Conferences, Symposia and Meetings*. NDLTD 2005. [<http://www.ndltd.org/meetings/index.en.html>]
- OCLC. *XTCat NDLTD Union Catalog*. OCLC 2004. [<http://alcme.oclc.org/ndltd/index.html>]
- Raghavan, Ananth, Naga Srinivas Vemuri, Rao Shen, Marcos André Gonçalves, Weiguo Fan, and Edward A. Fox. 2005. Incremental, Semi-automatic, Mapping-Based Integration of Heterogeneous Collections into Archaeological Digital Libraries: Megiddo Case Study. In *Proc. European Conference on Digital Libraries, ECDL 2005, Vienna, Sept. 18-23*: Springer.
- Shen, Rao, Marcos André Gonçalves, Weiguo Fan, and Edward A. Fox. 2005. Requirements Gathering and Modeling of Domain-Specific Digital Libraries with the 5S Framework: An Archaeological Case Study with ETANA. In *Proc. European Conference on Digital Libraries, ECDL 2005, Vienna, Sept. 18-23*: Springer.

- Staples, Thornton, and Ross Wayland. 2000. Virginia Dons FEDORA: A Prototype for a Digital Object Repository. *D-Lib Magazine* 6 (7/8). [http://www.dlib.org/dlib/july00/staples/07staples.html]
- Suleman, H., and E. Fox. 2003. Leveraging OAI Harvesting to Disseminate Theses. *Library Hi Tech* 21 (2):219-227. [http://www.emeraldinsight.com/0737-8831.htm]
- Suleman, Hussein. *VTOAI OAI-PMH2 PERL Implementation* [website] 2002. [http://www.dlib.vt.edu/projects/OAI/software/vtoai/vtoai.html]
- Suleman, Hussein, Anthony Atkins, Marcos A. Gonçalves, Robert K. France, Edward A. Fox, Vinod Chachra, Murray Crowder, and Jeff Young. 2001a. Networked Digital Library of Theses and Dissertations: Bridging the Gaps for Global Access - Part 1. *D-Lib Magazine* 7 (8). [http://www.dlib.org/dlib/september01/suleman/09suleman-pt1.html]
- . 2001b. Networked Digital Library of Theses and Dissertations: Bridging the Gaps for Global Access - Part 2. *D-Lib Magazine* 7 (8). [http://www.dlib.org/dlib/september01/suleman/09suleman-pt2.html]
- Suleman, Hussein, and Edward Fox. 2001. The Open Archives Initiative: Realizing Simple and Effective Digital Library Interoperability. *Journal of Library Administration* 35 (1/2):125-145. [http://www.husseinspace.com/publications/jla\_2001\_article\_oai.pdf]
- Suleman, Hussein, and Edward A. Fox. 2002a. The Open Archives Initiative: Realizing Simple and Effective Digital Library Interoperability. *Journal of Library Automation, special issue on "Libraries and Electronic Resources: New Partnerships, New Practices, New Perspectives"* 35 (1/2).
- . 2002b. Towards Universal Accessibility of ETDs: Building the NDLTD Union Archive. In *Proc. ETD'2002, Provo, Utah: NDLTD*
- Van de Sompel, Herbert, and Carl Lagoze. 2001. The Open Archives Initiative Protocol for Metadata Harvesting: Protocol Version 1.0, Document Version 2001-01-21. Ithaca, NY: Cornell University. [http://www.openarchives.org/OAI/openarchivesprotocol.htm]
- Van de Sompel, Herbert, and Carl Lagoze. 2000. Open Archives Initiative. Ithaca, NY: Cornell University. [http://www.openarchives.org]
- VTLS. *Networked Digital Library of Theses and Dissertations Union Catalog*. VTLS 2004. [http://zippo.vtls.com/cgi-bin/ndltd/chameleon]
- . 2005. VITAL homepage. [http://www.vtls.com/Products/vital.shtml]
- Zhu, Qinwei, Marcos André Gonçalves, Rao Shen, Lillian Cassel, and Edward A. Fox. 2004. Visual Semantic Modeling of Digital Libraries. In *Proc. 7th European Conference on Digital Libraries (ECDL 2003), 17-22 August, Trondheim, Norway, Springer LNCS 2769*.

# The arXiv: Fourteen Years of Open Access Scientific Communication

Simeon Warner (Cornell University)

**Abstract:** The arXiv was started in 1991 as a way for high-energy physicists to share preprints fairly and efficiently. Since then it has evolved into an archive of more than 330,000 articles in physics, mathematics and computer science. Within certain disciplines, the arXiv is now the primary means of scholarly communication and has changed the way that scientists work. This paper charts the development and use of the arXiv e-print archive over the past 14 years in the context of changes in scholarly publishing. Lessons learned from this development include the importance of community and critical mass, and the difficulty of balancing openness with fairness and keeping submissions appropriate and relevant. I discuss how journal publishers have reacted to the arXiv, and ask what the arXiv reveals about the established system of journals and the importance of peer review. Finally, I consider the role the arXiv should play in the future scholarly communication landscape and ask how arXiv fits with emerging institutional repositories.


## INTRODUCTION

The arXiv was started by Paul Ginsparg in 1991 at Los Alamos National Laboratory (LANL) as a way for high-energy theoretical physicists to share preprints fairly and efficiently.<sup>1</sup> The initial user-base was an email list of 160 addresses assembled from existing pre-print distribution lists for the hep-th subject area. It has since evolved into an archive of more than 330,000 articles in physics, mathematics and computer science. The arXiv now accepts ~4,000 new articles each month, offers an alerting service, search facilities, and has 17 mirror sites around the world.

Figure 1 picks out some landmarks in the development of arXiv. Much of the history of arXiv is recorded in the logs of “What’s New” pages.<sup>2</sup> The first phase was marked by rapid development of new facilities and expansion in subject area coverage. Implementation of automatic TeX processing software in 1995 meant that readers no longer had to download TeX source files and

---

M. Halbert (Ed.): *Free Culture and the Digital Library Symposium Proceedings*. Atlanta, Georgia: MetaScholar Initiative at Emory University, 2005. pp. 56-68.

 This work is covered by the Creative Commons Attribution-NonCommercial-NoDerivs 2.5 License.

process or compile them to get a readable version of an article (a process akin to compiling a C++ or Java program; rather arcane to many word processor users). Instead, PostScript was available directly.

Aug 1991	Physics e-print archive started: hep-th archive with email interface.
1992	ftp interface added. hep-ph and hep-lat added locally; alg-geom, astro-ph and cond-mat added remotely.
Jul 1994	Web interface added.
Nov 1994	Data at some remote archives (using the same software) moved to main site, the remote sites become mirrors.
Jun 1995	Automatic PostScript generation from TeX source.
Apr 1996	PDF generation added.
Jun 1996	Web upload facility added.
Sep 1997	ATMP is first journal overlay on an e-print archive.
from 1996	Worldwide mirror network grows.
Jan 2001	OAI compliant coincident with the OAI protocol release.
Sep 2001	Administrative oversight transferred to Cornell.
Dec 2001	Cornell site becomes primary, LANL site now mirror and backup.
Jul 2003	email submission discontinued, new user registration.
Sep 2003	q-bio archive launched.
Dec 2003	holiday schedule announced for the first time.
Jan 2004	submitter endorsement system added.

Figure 1: Selected landmarks in the evolution of arXiv

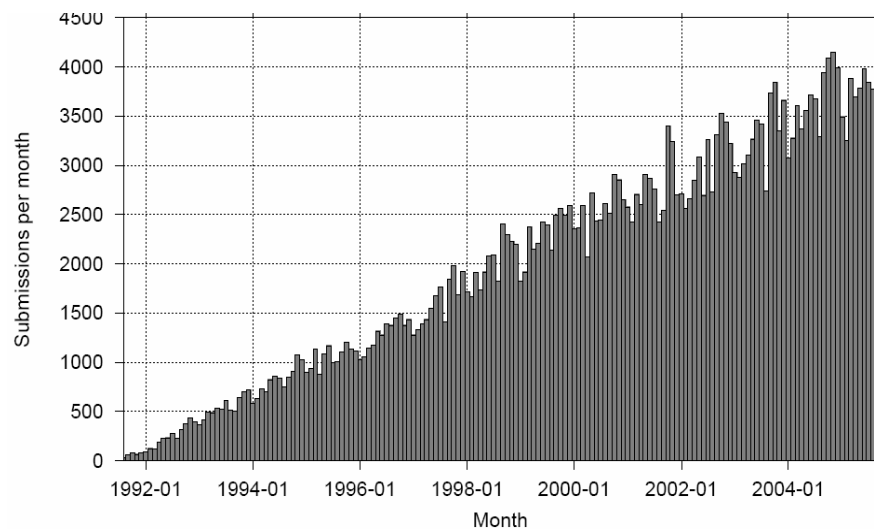


Figure 2: Monthly submission rate at arXiv.org. A up-to-date version of this graph is available from [http://arxiv.org/show\\_monthly\\_submissions](http://arxiv.org/show_monthly_submissions).

In 1996 the web submission interface was added. The facilities and scope of arXiv stabilized somewhat and arXiv offered most of

the features that it does today. Submission rates and readership continued to increase steadily, and the mirror network was enlarged.

The year 2001 marked the start of the most recent phase for arXiv. Metadata was made available for harvesting via an OAI<sup>3</sup> interface and could thus be added to other services (such as the NASA ADS<sup>4</sup>). The move to Cornell spurred a process of institutionalization which has included the development of a new user registration and authentication system, formalization of procedures and policies, and even scheduled holidays.

The overall submission rate to arXiv has increased approximately linearly since 1991, as shown in figure 2. Starting around 1995 the growth in the submission rate to the high-energy physics categories (hep-th, hep-lat, hep-ph, hep-ex) started to slow down. This did not mark any problem but instead a saturation in that almost all papers in these subject areas were being deposited on arXiv. This saturation is shown in figure 3.

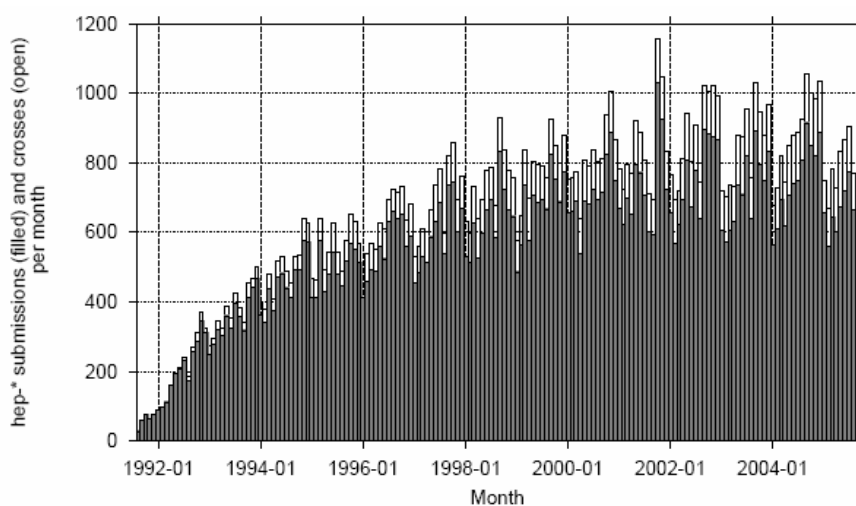


Figure 3: Monthly totals of new submissions to all the high-energy physics archives (hep-th, hep-lat, hep-ph and hep-ex) (filled bars) and cross-listings of papers from other archives (open bars). The submission rate started to flatten in 1995 and now shows just yearly fluctuations. This is an updated version of the graph available from <http://arxiv.org/Stats/hcamonthly.html>.

While initially envisaged as a self-contained preprint redistribution service, arXiv continues to evolve into part of an integrated global communication system. The creation of the journal *Advances in Theoretical and Mathematical Physics* in 1997, as an overlay on arXiv, demonstrated how conventional peer-review can be

implemented on top of an open access substrate. Such overlays continue to represent just a very small fraction of the literature but now include *Geometry and Topology*, *Geometry and Topology Monographs*, *Algebraic and Geometric Topology*, *Logical Methods in Computer Science*, *Theory and Practice of Logic Programming*, and all the journals of the Institute of Mathematical Statistics (IMS).

## COMMUNITY AND CRITICAL MASS

With some systems, it seems, one just has to “*build it and people will come.*” Of course, it helps to have built the right thing in the first place and maybe to have some prominent advocates suggesting a visit. This was the case with arXiv. Kling and McKim<sup>5</sup> argued that physics community was ready and arXiv fit its practices well. They cite the different route chosen by PubMed as an example of a different resource being developed to meet different community needs. A corollary to the argument that the arXiv was successful because of the particular user community is that this model might not be a good fit for other communities with different practices.

### Creation of the q-bio archive

A new top-level classification, *Quantitative Biology* or q-bio for short, was created in September 2003.<sup>6</sup> The creation of q-bio illustrates a number of elements of our strategy for expansion:

1. Logically, “Biology” would have been a better classification to sit alongside “Physics”, “Mathematics” and “Computer Science.” Quantitative Biology would then be a natural sub-field of Biology. However, there had been a number of requests for a separate subject area from key figures in the q-bio field, and they were already submitting papers to other parts of arXiv. There had not been similar request from others in the broader field of biology, so it was thought better to avoid a misleading Biology title without appropriate content, and to avoid creating additional subject areas that would likely be under-used and appear dead from the outset.
2. Those who had asked for a q-bio archive were charged with guiding its creation. This involved setting up a subject advisory board, deciding on sub-categories, and recruiting volunteer moderators for each sub-category.

3. The moderators for the q-bio archive identified a number of existing papers that should be classified in each sub-category of q-bio . These examples were used to train a machine learning system which was then used to find other papers that should be cross-listed to the new category.<sup>7</sup> These papers provided a seed for the category, and all the authors were emailed with an explanation of the proposed cross-list and an invitation to use the new category. This provided a set of articles going back to 1992 that were cross-listed to the new category.

Is q-bio a success? The first and obvious answer is yes: there is steady growth in the submission rate and informal feedback is positive. A second answer might be more reserved, as we see that there is no discontinuity in the characteristics submission rate graph associated with the creation of q-bio. It seems that q-bio is certainly successful in that it groups together submissions that were previously dispersed over other subject categories, but it has not perhaps attracted new users any faster than the underlying arXiv growth.

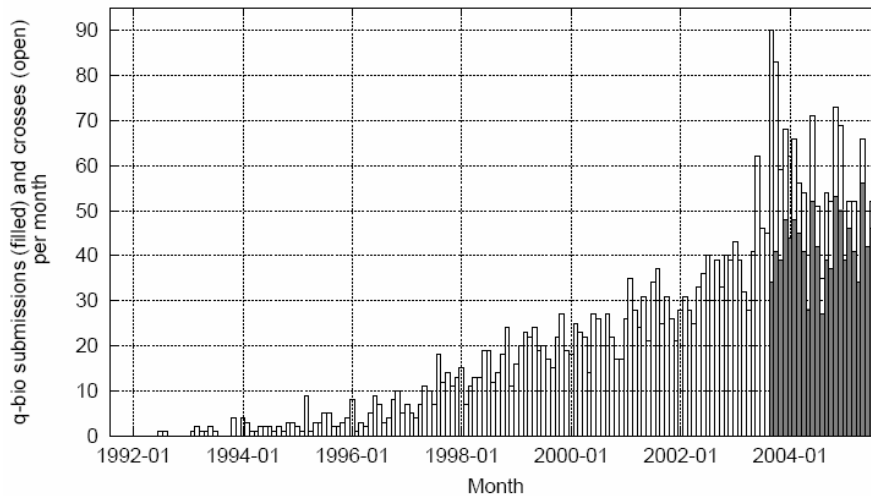


Figure 4: Monthly totals of new submissions to q-bio (filled bars) and cross-listings of papers from other archives (open bars). Ignoring the peak when the q-bio archive was started we see almost a straight line through September 2003 when the archive was started. This is an updated version of the graph available from <http://arxiv.org/Stats/remmonthly.html>.

### The price of popularity

Most practicing physicists occasionally receive emails or postcards from hobby scientists who believe they have unearthed errors in





Figure 5: Monthly download totals since the launch of the arXiv web interface in 1994, shown on a logarithmic scale. This data has been screened to remove robot downloads from crawlers and from internal processes, and to remove duplicate downloads from the same IP address within a month.

### Preferred download format

Figure 6 shows the fraction of downloads in each display format available from arXiv. We see three distinct phases: First, downloading the TeX source was most popular simply because there was no other option for papers submitted as TeX source rather than processed PostScript. PostScript generation was added in June 1995 and this immediately ushered in a second phase where PostScript downloads dominated and source package downloads became gradually less popular. PDF generation was added in April 1996 but the popularity of PDF as a download format grew only very slowly and over many years. Finally, in 2002 there was a rapid swing to the current phase, where PDF downloads dominate.

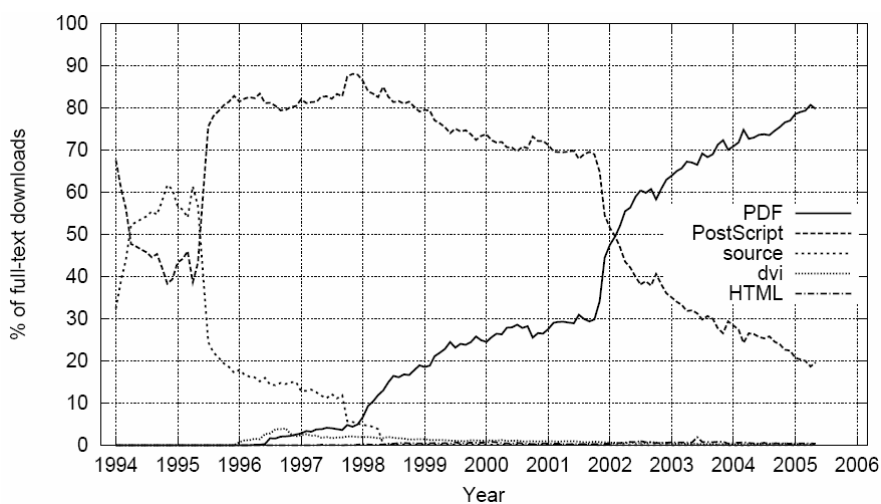


Figure 6: Graph showing change of preferred download format over time. PDF was first introduced by Adobe in 1993 (with the release of Adobe Acrobat 1.0) and from arXiv in 1996. In 2002 it replaced PostScript as the preferred download format and now accounts for 80 percent of downloads.

Here we have a common preservation scenario played out twice over: what to do when formats become obsolete? The first case is rather trivial, as users would likely have preferred to download PostScript all along and the source files were designed to produce PostScript output, however the facility wasn't available initially. The move to PDF is more interesting, as this format was not

known when early arXiv papers were submitted. The strategy employed was to process source files to produce PostScript much as usual (some differences in font use), and then to convert the PostScript to PDF on demand.

### The mirror system

Mirrors account for about 37 percent of downloads (42 percent including repeats), the distribution of downloads in June 2005 are shown in Figure 7. These data have been cleaned to remove mirroring and robot accesses as much as possible.

Site	Full-text downloads in June 2005	Abstract downloads in June 2005
(main site) arXiv.org	1027310	714162
lanl.arXiv.org	278572	151993
jp.arXiv.org	56302	27597
it.arXiv.org	50694	24622
de.arXiv.org	47393	30922
fr.arXiv.org	37038	18105
uk.arXiv.org	33702	21776
cn.arXiv.org	25297	10915
ru.arXiv.org	11784	8031
br.arXiv.org	10089	2034
tw.arXiv.org	8965	2864
il.arXiv.org	5461	3094
au.arXiv.org	5446	4464
kr.arXiv.org	4483	3494
aps.arXiv.org	4352	4363
es.arXiv.org	3241	1582
in.arXiv.org	3015	1888
za.arXiv.org	268	1153
Total	1613412	1033059

Figure 7: Numbers of full-text and abstract downloads from the main site (arXiv.org) and all mirror sites for June 2005. Mirror sites have country code prefixes except for “lanl” which is the LANL mirror and “aps” which is the APS mirror at Brookhaven. The total numbers of downloads have been counted as unique paper/IP address pairs to avoid over counting due to multiple downloads by the same user (plain counts are 30 percent higher). Considerable efforts have been made to remove robotic accesses which would otherwise inflate the counts.

### RIGHTS, LICENSES AND ACCESS

For a long time, arXiv operated without any explicit statements about rights. A non-exclusive license to distribute was assumed to

have been granted by the act of submission. A few years ago, this was made explicit in the submission process, which now involves two elements of click-through as shown in Figure 8. Without both boxes certifying submitter identity and agreements with terms checked, the submission will not be accepted.

**A. Verify Your Contact Information**

*...explanation omitted...*

First Name: Simeon  
 Last Name: Warner  
 Suffix: (‘Jr.’, ‘II’, etc; may be blank)  
 Affiliation: Cornell University  
 E-mail: simeon@cs.cornell.edu

I certify that the above contact information is correct.

**B. Legal Statement**

- I grant arXiv.org a license to distribute this article.
- I certify that I have the right to grant this license.
- I understand that submissions cannot be completely removed once accepted.
- I understand that arXiv.org reserves the right to reclassify or reject any submission.

I agree to the above terms.

Figure 8: License click-through during arXiv submission process.

We plan to offer the option of simply granting arXiv a license to distribute, or saying that a Creative Commons license applies which also gives us the permissions we need. Clarke<sup>8</sup> argues that the “Attribution/ NonCommercial/No Derivative Rights” (By-NC-ND) license is adequate for e-print use and would certainly give arXiv the necessary rights. However, one might want to encourage the use of the more permissive “Attribution” (By) license used by PLoS (for example).

## ARXIV AND THE CONVENTIONAL JOURNAL SYSTEM

Writing in 1994, Ginsparg said “The rapid acceptance of electronic communication or research information in my own community of high-energy theoretical physics was facilitated by a pre-existing ‘pre-print culture’, in which the irrelevance of refereed journals to ongoing research has long been recognized.”<sup>9</sup> To read this statement as an assertion that journals are irrelevant is to miss a disconnect between the practice of physics, for which peer review

is not considered very important (at least in the short term), and rewarding or professional progression for which the stamp of authority offered by journals is considered indispensable. Thus, physicists somewhat contradictorily argue that arXiv is essential for their work, and is how they communicate, and yet that the conventional journal system must remain as is.

It was recognized early-on that arXiv was not an informal means of communication,<sup>10</sup> even though it does not attempt to replicate the journal system. The format of articles is quite conventional and inappropriate submissions are rejected. Furthermore, all submissions are stamped with the submission date and time thus providing a record that can be used to settle disputes about priority. Finally, it has been a principle of arXiv that submissions cannot be removed or altered once announced. New versions may be submitted which update, correct or withdraw an article, but the original is retained for all to see.

The arXiv submission policy aims to screen submissions that are not of “reviewable quality.” Our experience with moderators has been that in the vast majority of cases it is trivial for a subject expert to determine whether a submission is acceptable or unacceptable. This ease is perhaps why physicists are happy to use pre-prints from arXiv.

## **INSTITUTIONALIZATION**

The arXiv is established as indispensable in some disciplines and of growing importance in others. The large number of submissions, almost 200 every working day, means that it is no longer feasible to run it with “a couple of postdocs” to both administer on a daily basis and develop it further. Since the move of arXiv to the Cornell University Library in summer 2001, efforts have been underway to transfer all daily operation and mainstream development efforts to library staff. This has necessitated a number of changes in operational and development strategies. The first was formalization of a number of previously ad-hoc procedures for daily operation. Another has been improvement in tools and practices to separate less skilled administration actions from those needing intervention by someone with detail knowledge of the system internals.

Benefits of the move to the Cornell University Library include a long-term institutional commitment to preserve and maintain access to the collection, and rationalization of policies that have been made both simpler and more uniform. There are, of course,

costs associated with larger management overhead and less development agility.

Significant steps have been made toward fairer and more sustainable governance through greater liaison with the overall arXiv advisory board and with the separate advisory committees for each of the main subject areas. In particular, the separate advisory committees are used to recruit moderators for each subject area and the physics advisory committee has recently reached consensus on reorganization of the physics subject categories.

### **THE FUTURE ROLE OF ARXIV**

Many authors have identified two roles fulfilled by scholarly publication: one being to communicate information necessary for continued research, and the other to provide certification necessary for professional rewarding and advancement. The arXiv has demonstrated a very efficient system for the former need, but has not addressed the latter.

One can think of the largely automated distribution system provided by arXiv as the “low hanging fruit” of the broader scholarly communication problem. Even the submission system for arXiv is extremely cheap, as most of the effort is offloaded to the author. Administration effort is less than two minutes per article on average (based on a single administrator being able to deal with problems relating to 250 submissions in a single work day, neglecting overhead of maintaining and developing the system). However, even this amount of time adds up to one full-time equivalent just for the daily administration.

Most of the expense of running arXiv is in handling new submissions. This has two positive results. First, the cost of maintaining the archive of old papers is negligible in the context of running the whole service so there is no incentive to reduce access or facilities for the archival collection. Second, if at some time new submissions were no longer accepted, it would not be expensive to maintain the archival collection alone.

Nascent institutional repositories may eventually replace arXiv. The distributed model is appealing although experience suggests that it is much more difficult to implement. In 1995 arXiv was distributed over several sites each dealing with separate subject areas but these were gradually brought back under central control for management convenience and stability. It may be that an

intermediate stage will be for arXiv to act as a slave subject-based publishing venue with institutional repositories serving as the primary archives, or vice-versa. We are already having these discussions with a few institutions where institutional repositories have been deployed.

There is a clear advantage in the funding model for institutional repositories in that if funded by the institution that runs them, puts material in them, and benefits from resulting publicity, then the ownership and benefit is clear. Contrast this with the current situation for arXiv where the Cornell University Library is putting significant funding into a resource where Cornell is only a minor benefactor. ArXiv is currently funded through the Cornell University Library and the NSF though other long term funding sources are being investigated.

## ACKNOWLEDGEMENTS

I am pleased to acknowledge that Paul Ginsparg, Thorsten Schwander, Mark Doyle and others in the past and present arXiv team have shared various historical facts and insights with me. They have helped shape my view of arXiv and scholarly publishing and thus the history presented here. The arXiv received some early support from DOE HEP and has been supported by the NSF since 1995 (previous awards were 9413208 and 0132355), and recently by Cornell University Library. This work is supported by NSF award number 0404553.

## ENDNOTES

1. Paul Ginsparg. First steps towards electronic research communication. *Computers in Physics*, 8(4):390–396, 1994. URL: <http://arxiv.org/blurb/blurb.ps.gz>.
2. What was new on the arxiv.org (xxx) *e-print archives*, 1991–2005. URLs: <http://arxiv.org/new/>. Downloaded 2005-09-07.
3. Carl Lagoze, Herbert Van de Sompel, Michael Nelson, and Simeon Warner. *The Open Archives Initiative Protocol for Metadata Harvesting, version 2.0*, June 2002. URL: <http://www.openarchives.org/OAI/2.0/openarchivesprotocol.htm>.
4. *The NASA Astrophysics Data System*. URL: <http://adswww.harvard.edu/>.

5. Rob Kling and Geoffrey McKim. "Not just a matter of time: Field differences and the shaping of electronic media in supporting scientific communication." *Journal of the American Society for Information Science*, 51:1306–1320, 2000. URL: <http://arXiv.org/abs/cs.CY/9909008>.
6. Terry Hwa and Michael Lässig. "Announcement of new quantitative biology (q-bio) archive," September 2003. URL: [http://arxiv.org/new/q-bio\\_announce.html](http://arxiv.org/new/q-bio_announce.html).
7. Paul Ginsparg, Paul Houle, Thorsten Joachims, and Jae-Hoon Sul. "Mapping subsets of scholarly information," 2003. URL: <http://arxiv.org/blurbs/sack03.pdf>.
8. Roger Clarke. "A proposal for an open content licence for research paper (Pr)ePrints." *First Monday*, 10, August 2005. URL: [http://firstmonday.org/issues/issue10\\_8/clarke/index.html](http://firstmonday.org/issues/issue10_8/clarke/index.html).
9. Ginsparg, 1994, op.cit.
10. *Ibid.*



# The Librarian Revolutionary: Creating Cultural Exchange and Preservation on the Internet


Debora Halbert (Otterbein College)

**Abstract:** This paper examines the role of librarians as information revolutionaries. While required to live within the world of copyright law, librarians by profession are dedicated to the exchange of information, a revolutionary act in a world overly burdened by copyright law and broad licensing agreements. A commitment to access to knowledge puts librarians at odds with copyright owners who seek to allow access only under restricted circumstances. This paper is part investigation and part call to action. First, I will investigate the state of access to knowledge as seen through the eyes of librarians. Second, I will argue that librarians in their capacity as gatekeepers should begin to shift the focus from for-profit databases towards free culture. It is in the tradition of the library, dedicated to serving a public, that access to knowledge be as widespread as possible. This makes librarians revolutionaries of the information age.

Academic and library communities have long operated under the assumption that the free flow of information should be supported and enhanced (Coyle 1995; Ferullo 2004: 24). With this assumption in mind, librarians have embraced the power of information technology to help their patrons access knowledge by providing digital materials. The digital environment is not an entirely free culture, however. Restrictive copyright legislation and digital rights management by copyright owners intent upon shoring up their “rights” against unauthorized use jeopardizes the role of librarians and libraries as sources of information. Libraries have been forced to become adept at understanding the implications of copyright law and the myriad licensing agreements associated with intellectual property, especially in relation to digital collections (Ferullo 38). In an effort to preserve the free flow of information, libraries have found themselves at the center of copyright-related controversies and have increasingly been pulled into political battles over access to knowledge.<sup>1</sup>

---

M. Halbert (Ed.): *Free Culture and the Digital Library Symposium Proceedings*. Atlanta, Georgia: MetaScholar Initiative at Emory University, 2005. pp. 69-81.

 This work is covered by the Creative Commons Attribution-NonCommercial-NoDerivs 2.5 License.

The ability to access knowledge has been transformed by the availability of the Internet, digital materials hosted in library databases, and electronic book resources. Where earlier research trips required a journey to the library, the use of a card catalogue, and photocopying journal articles (or reading them on location), it is now possible to conduct most basic library research from your office computer.<sup>2</sup> Researchers have access to on-line catalogues, peer-reviewed texts are available on-line, new avenues for on-line publishing exist, and the entire cacophony of the World Wide Web is a resource for those interested in locating information. With such widespread access to work, libraries are seeing both the costs and the benefits of the information age.

A primary benefit of digital materials and networked environments is the ease of access they provide, especially for small colleges and public libraries. Small institutions cannot purchase the scope of materials found at a large research library, and until the last decade, this made acquiring quality materials for anyone outside a research library time consuming and difficult. Furthermore, while some still mourn the old card catalogue system, the ease of access to knowledge made possible by information technology is truly astounding. In terms of content, the world of information networks has changed the process of research for the better, despite the fears of information overload and the overabundance of low quality materials.

However, there are costs associated with the information age as well. First, access to privately owned information databases is very expensive. As state funding for libraries is reduced, libraries face budgetary choices regarding how to manage their resources. Most cut increasingly expensive print journals, and will be forced to reduce access to on-line databases as costs continue to rise. Second, restrictive licensing agreements mean that at any point, access to an electronic journal can be denied, and in the absence of a hard copy, any individual library will be left without an archival copy. The World Wide Web can be even more intangible with websites flowing in and out of existence daily. Third, as the media becomes increasingly concentrated, copyright law and licensing agreements are used to centralize control (Bettig 1996; McLeod 2001) and bring new complexities to the world of public lending of texts. Librarians find themselves needing to become experts on copyright law and licensing agreements in order to continue to provide information to the public. Given these concerns, how we

seek to expand our access to knowledge in the future suggests some political and economic choices must be made.

The costs and benefits of the information age place libraries at the threshold of two parallel worlds of knowledge production and control. The first is the costly world of copyright, licensing agreements, and digital rights designed to restrict the flow of information to paying customers (Lessig 2004, p. 281). As Vaidhyanathan points out, this is the pay-per-view world that would have each reader pay for access (2004), although some would argue it is an unlikely model (Odlyzko 1999, 13). The second is the chaotic and anarchistic flow of information on the Internet—a world where information is free, but quality can be questionable.<sup>3</sup> Librarians are the gatekeepers to both worlds and in the future will play the role of helping the public negotiate the territory of digital information (Boucher, 2002 98). The way librarians guide us into the future has political implications. In other words, choices about what type of information will be available and how it will be accessed must be made, and librarians are key players in making these choices.

I wish to examine the role of librarians as information revolutionaries. While required to live within the world of copyright law, librarians by profession are dedicated to the exchange of information—a revolutionary act in a world overly burdened by copyright law and broad licensing agreements. A commitment to access to knowledge puts librarians at odds with copyright owners who seek to allow access only under restricted circumstances. However, there is a growing body of open access journals, open source materials, and publicly created peer-reviewed documents that can be accessed free on the Internet. Librarians can play an important role in directing patrons towards the latter and helping this body of knowledge grow while avoiding the former.

This paper is part investigation and part call to action. First, I will investigate the state of access to knowledge as seen through the eyes of librarians. Second, I will argue that librarians, in their capacity as gatekeepers, should begin to shift focus from for-profit databases towards free culture. It is in the tradition of the library, dedicated to serving a public, that access to knowledge be as widespread as possible. This makes librarians revolutionaries of the information age.

## A CRISIS IN ACCESS TO KNOWLEDGE

I live in Ohio and work at a small liberal arts college with a limited library collection. Access to quality, peer-reviewed information is acquired by joining OhioLink, a service that networks all major research universities, most of the liberal arts colleges and a growing number of public libraries into one vast interlibrary loan system for books, journals, and other materials. Full-text searchable journals are available through OhioLink and the Electronic Journal Center (EJC). Otterbein, for example, subscribes to OhioLink and other databases including Lexis-Nexis, Academic Search Premier, and JSTOR. These databases make it possible to access materials that would be both expensive and time consuming to find otherwise.

While OhioLink is a research paradise of sorts, it is in crisis, a problem that will impact all colleges and universities in the state. No single institution can afford to purchase a significant number of the scholarly journals, with the Ohio State University having the most purchasing power. The inability to pay the high purchasing costs has led to rationing and denial of access with the biggest impacts felt by small private colleges. For example, increasing costs have led Otterbein to substantially reduce the number of print-journals to which it subscribes and the availability of these journals on-line becomes increasingly crucial. Reliance on the on-line system means that when service is reduced there are not copies of the journals available.

Library budgets across the nation are in a position where they cannot keep up with increasing subscription costs. Some subscription prices increased 227 percent between 1986 and 2000 (Perzigian, Zimmerman, and Sanville 2004, 1). Odlyzko estimates that publisher revenues are around \$4,000 per article published (1999, 4). In some cases, the profit margins on these scholarly publications can reach 75 percent (Rosenzweig, 2001, 1). These publications have become increasingly privatized over the years, moving from professional associations to for-profit publishers who centralize control over academic scholarship.

Enormous increases to library subscription costs were initially justified by the pervasive use of the photocopy machine. While it is unclear that the photocopy machine substantially hurt journals, these costs have created such a disparity between individual subscription prices and institutional ones that most institutions cannot afford to retain the numbers of journals they once did

(Liebowitz 2002, 198-199). Thus, expanding costs require some libraries to cut expensive subscriptions leading journal owners to increase costs to make up the difference. When all things are considered, however, the high prices of most academic journals are difficult to justify when one understands that the bulk of labor put into the process is free (Willinsky 2002).

In response to the high and always increasing costs of print journals, OhioLink has negotiated electronic access for all its member libraries to commercial, non-profit, and society based publishing companies through its Electronic Journal Service (EJS). These negotiations allowed OhioLink members to retain access to most journals while controlling costs somewhat. However, group licensing is only a stop-gap measure. Journal prices continue to increase between 5 percent and 10 percent each year and library budgets continue to shrink (Rosenzweig, 2). The result is increased pressure to reduce costs with the most likely avenue being cutting subscriptions in both print and electronic form.

I offer only the briefest overview of the problems facing libraries. Problems associated with high costs are compounded by pressure to ensure that copyrights are not infringed. Thus, in addition to price negotiation and licensing agreements, libraries find themselves fighting battles over the expanding nature of copyright law. Libraries have lobbied to gain important exceptions to the law that will provide them with flexibility in sharing materials, but the crux of the problem stems from the fact that the mission of the library to share information is directly contradictory to the intent of copyright owners to control information they own. As a result, while being forced to license materials as a result of copyright laws, librarians are finding themselves politicized by copyright disputes.

Given the problems associated with scholarly publishing and copyright law for libraries, it is perhaps time to seek solutions that do not require libraries to fight a reactive war against copyright owners or spend endless hours negotiating licensing agreements that inevitably place libraries in a vulnerable position. In fact, there are more extreme measures that can be taken to rectify the high costs of access to scholarly knowledge—librarians and academics can become political actors who seek to change the type of materials made available. Information flows around barriers and it is clear that as copyright constructs barriers to some information, there are already enormous numbers of actors ready to facilitate the flow of information around these barriers. While

what is called piracy is part of these information flows, what I seek to discuss in the remainder of this paper is the development of a new paradigm for innovation, scholarship and the communication of ideas—a paradigm that needs the help of librarians to succeed. If one system of knowledge production has become too costly and refuses to change its method of pricing then a different system should be, and is being, constructed.

### **OPEN ACCESS AND CREATING CHANGE**

On New Year's Eve in 1940, American broadcasters announced an official boycott of all music licensed through the American Society for Composers, Authors, and Publishers (ASCAP). At the time, ASCAP had a monopoly on the music licensing business and could charge monopoly rates. In 1931, ASCAP raised its rates by 300 percent (Perlman 2002, 18). This process of increasing rates and restrictive licensing agreements continued until the first day of January in 1940. The Broadcast Music Incorporated (BMI) was born in resistance to ASCAP. On the first of January, radio listeners were treated to only music licensed by BMI or music found in the public domain. BMI licensed music ignored by ASCAP—blues, jazz, rock and roll, Latin music, and gospel. ASCAP's attempt to strictly control access to music via licensing and copyright resulted in both a resistance to restrictive copyright regimes and growing popularity for entirely new genres of music in the United States (Perlman 2002, 18). In order to stay relevant, ASCAP was forced to compromise with radio broadcasters and BMI. While BMI has in turn become a licensing giant that no longer acts as part of a resistance to copyright law, there is a lesson to be learned from the licensing wars of the 1930s and 1940s. Working to create alternatives to monopoly copyrights is not only an important form of resistance, but can bring valuable results—it is a lesson that librarians should listen to closely.

Given the state of licensing today, one possible future for libraries is that they will be disintermediated because electronic journals will remain protected by centralized copyright owners and information distribution will fall into the hands of journal owners (Odlyzko 1999, 10-11). In this scenario, the role librarians will play is primarily one of licensing negotiation and copyright policing (Odlyzko 1999, 10). Libraries may remain depositories for books and hard copy materials, but digital materials will become even more intangible with librarians serving as enforcers of property rights not guarantors of public rights. In this future scenario, a pay-per bite world where information is not free and

copyright control is monopolized by the publishing industry becomes conceivable. However, given the example of music, another possible scenario can emerge. In this scenario, the publishing industry is disintermediated and librarians become key players in an information future as gateways to the world of the open access movement and creative commons licensing agreements.

In response to the increasing costs of the for-profit publishing industry and the problems associated with copyrights in digital materials, librarians and scholars are developing an alternative system that retains the values of access and circulation of information. Librarians have been politicized by the threat they see related to the way copyrights and licensing are being used and the expanding nature of copyright law. Librarians have begun to seek alternatives.

One revolutionary idea already in development is to build and support a parallel system of scholarly production designed to ultimately replace (or diminish the importance of) the publishing industry paradigm where centralized copyright control allows only for expensive access to digital materials. Building a public alternative to private ownership of scholarly work is part of an open access movement that brings librarians and scholars together to confront the problems of costly journals and prohibitive copyrights. Librarians have been instrumental in the foundation of organizations that will allow knowledge to flow around copyright boundaries by teaming up with academics to put scholarly publishing back into the hands of the academic communities who produce the work and seek to have it disseminated as widely as possible. In this parallel system, librarians are important gatekeepers that can help create a digital future where access to materials is relatively cheap and easy. However, the choices made by librarians are political choices about the types of information networks they will choose to support. While not comprehensive, the examples I cite below are indicative of the work being done in the open access movement that will continue to grow as more librarians seek to make information freely available to the public.

First, Create Change is an advocacy website seeking to enlist librarians in the fight to provide access to academic work and bypass the expensive for-profit journal model that is currently pricing many libraries out of the leading research journals (Create Change, <http://www.arl.org/create/librarians/faq/scomm.html#q8>). The resistance is multi-layered. Create Change is co-sponsored by

the Association of Research Libraries, Association of College and Research Libraries, SPARC (Scholarly Publishing and Academic Resources Coalition). Their goals include:

1. To resist the cost increases in journal subscriptions.
2. To recognize the political economy of academic publications where authors lose their copyrights and receive no compensation for their work, but it is priced out of affordable range which harms the exchange of information.
3. To help develop an alternative journal system of on-line peer-reviewed journals that will provide free access to scholarly knowledge (*Ibid*).

The Create Change website includes a tool kit for librarians interested in working on these issues and developing partnerships with faculty at their colleges in an effort to change the system.

Another important resource for librarians and scholars interested in providing alternatives to the status quo can be found in the Scholarly Publishing and Academic Resources Coalition (SPARC) guide to placing science journals back into the hands of the communities who publish and use them (SPARC 2001). SPARC seeks to create alliances between libraries and scholars to publish on-line peer-reviewed materials that can be accessed free or for little cost and to use these new journals as a method for replacing the current for-profit paradigm. For example, a new journal called *Plagiary* that will deal with issues of plagiarism has opted to go with SPARC's model (<http://www.plagiary.org>). The advantages of the model are many including low start up costs and easy accessibility to a potentially large audience.

Besides these educational and informational guides, major projects with the intent of providing additional open access to knowledge are already underway. California librarians and scholars are developing the California Digital Library (CDL) to provide increased access to a range of quality peer-reviewed materials. (<http://www.cdlib.org/programs/escholarship.html>) This system is open to anyone, and serves as an example for how scholarship can be brought to the public relatively cheaply. Libraries and librarians are essential in this process, not only by helping develop these projects, but also by directing users to them once they exist.

There are numerous on-line projects that help support public access to knowledge that can also be supported by librarians as



inexpensive alternatives to expensive copyright controlled databases. In San Francisco, for example, a non-profit project called the Internet Archive is archiving on-line resources (Albanese 2005). Charles Bailey has compiled a detailed bibliography of open access resources that can also serve as a starting point for research utilizing materials from the open access world (<http://www.arl.org/pubscat/pubs/openaccess/>). These initiatives seek to bring knowledge to the public under affordable terms and help develop what Lawrence Lessig has termed “free culture” (Lessig, 2004).

Yet another dimension of the idea of open access found on the Internet is the phenomenon of the public construction of knowledge as a social act. Perhaps one of the most popular examples is the Wikipedia (<http://en.wikipedia.org>). This “free” encyclopedia allows all users to engage in collective editing and has rapidly become both a website for knowledge and a form of political conversation. This public access project is a potential resource for all users of the World Wide Web and one that may be useful to library patrons as well.

The search engine Google is in the midst of developing innovative ways to bring knowledge to the wider public. Recently, Google entered into partnerships with the University of Michigan, Oxford, Stanford, Harvard and the New York Public Library to digitize their materials and make them available on-line (Albanese 2005). Google has the lofty goal of making a full text searchable card catalogue of all books in all languages (Google Print Library Project, <http://print.google.com/googleprint/library.html>). While this project is running into copyright difficulties in the United States, in Europe it has sparked a counter-project by European libraries (except Oxford) to digitize European literature in an attempt to avoid what some consider cultural imperialism by the U.S. (Deutsche Welle 2005)

Librarians have been instrumental in Google’s plans, but not all librarians see this as a step in the right direction. In the minds of some, a massive search engine such as Google will call into question the role of libraries in the information age. While many agree the role of librarians will inevitably change as new mechanisms for information exchange develop (Sapp and Gilmour 2003), it is important to note that the concept of public access to information, through some form of library either virtual or tangible will remain a necessity. By extension, librarians have a role to

play in ensuring that information remains available and affordable to the public.

The Open Access movement is gaining speed and is not limited to a single nation. In fact, one of the more important implications of the open access movement is that it crosses national boundaries and opens up access to knowledge throughout the world (Suber 2005). With calls for access to knowledge being brought before the World Intellectual Property Organization (WIPO) by the developing world, shifting focus to open access instead of private ownership of knowledge can have enormous benefits for those in the global south who have even more limited resources with which to buy expensive journals and books.

A movement is growing to provide an alternative model for access to information. It is a movement that embraces the concept of free culture and the free flow of information and librarians are playing an important role in the production of this movement. Librarians have not only been political actors, but also play an important role in the type of information that will be available to access in the future.

### **LIBRARIANS AS REVOLUTIONARIES OF THE INFORMATION AGE**

Karen Coyle noted in 1995 that librarians must become advocates of information freedom and work to ensure that access to public knowledge remains a central focus of the information age (Coyle 1995). There is no doubt that the digital environment will provide new challenges and opportunities. It is my argument that librarians can be central to the shaping of the future digital environment. As gatekeepers, they can educate and direct scholars, students, and the general public towards open access projects.<sup>4</sup> As part of the educational process, they can inform academics about the opportunities to publish in open access journals and the mechanisms for seeking out publicly available information. While Google's quest to digitize everything and Yahoo's Creative Commons search engine may seem threatening to the future of brick and mortar libraries, I think the opposite is true. The most essential task in today's information saturated world is to help filter information—a job well suited to libraries. Librarians can direct patrons towards freely accessible materials of quality while controlling costs in new acquisitions and restrictive licensing agreements. Instead of becoming copyright police, librarians will be information networkers in the open access environment. In his

book *Information Liberation* (1998) Brian Martin suggests numerous options for helping to create the free flow of information. Among these are promoting non-owned work (Martin 1998: 54). In facilitating the promotion of open access sources, librarians are essential and can become central revolutionaries of the information age.

## ENDNOTES

1. Librarians have been publicly opposed to the Digital Millennium Copyright Act (DMCA), the Copyright Term Extension Act, and UCITA to name a few recent legal controversies related to copyright. These laws have convinced many librarians that they will need to become political activists in order to continue with their mission. For example, it was the librarians at our College who spread the word about UCITA and motivated the College to pass a resolution against UCITA that was then sent to our state representatives.
2. Jenny Levine discusses the vast shifts in how we acquire information in relation to libraries. She has coined the term “the shifted librarian” to describe the new portability of libraries and their content (Levine, 2004).
3. The issue of quality and the World Wide Web has generational qualities to it as well with younger generations tending to see all information on the Web as equally valid (Lombardi, 2000).
4. This argument has been made by several Mexican librarians (*e.g.*, Reyes 2002) who see the exchange of public information as a key aspect of a democratic system of governance. Note: I am citing to the English abstract to Reyes’ paper since I do not read Spanish.

## REFERENCES CITED

- Albanese, Andrew. January 15, 2005. “Google to Digitize 15 Million Books,” *Library Journal*.  
<http://www.libraryjournal.com/article/CA491156.html>. Site visited 6/07/05.
- Bailey, Charles. 2005. Open Access Bibliography: Liberating Scholarly Literature through E-Prints and Open Access Journals.  
<http://www.arl.org/pubscat/pubs/openaccess/>. Site visited 7/01/05.
- Bettig, Ronald V. 1996. *Copyrighting Culture: The Political Economy of Intellectual Property*. Boulder: Westview Press.
- Boucher, Rick. 2002. “The Future of Intellectual Property in the Information Age,” in *CopyFights: The Future of Intellectual Property in the Information Age*. Adam Thierer and Clyde Wayne Crews Jr., Eds. Washington D.C.: Cato Institute.

- Coyle, Karen. 1995. "Electronic Information—Some Implications for Libraries," <http://www.kcoyle.net/carlart.html>. Site visited 6/8/05.
- Create Change. <http://www.arl.org/create/librarians/faq/scomm.html#q8>. Site visited 5/20/05.
- Deutsche Welle*, "European Libraries Fight Google-ization," April 27, 2005. <http://www.dw-world.de/dw/article/0,1564,1566717,00.html>. Site visited 6/07/05.
- Ferullo Donna L. 2004. "Major Copyright Issues in Academic Libraries: Legal Implications of a Digital Environment." *Journal of Library Administration*, Vol 40, No. ½, pp. 23-40.
- "Google Print Library Project," 2005. <http://print.google.com/googleprint/library.html>. Site visited 6/07/05.
- Lessig, Lawrence. 2004. *Free Culture: How Big Media Uses Technology and the Law to Lock Down Culture and Control Creativity*, New York: Penguin Press. Available at: [http://www.ibiblio.org/ebooks/Lessig/Free\\_Culture/Free%20Culture.htm](http://www.ibiblio.org/ebooks/Lessig/Free_Culture/Free%20Culture.htm)
- Levine, Jenny. 2004. "What Is a Shifted Librarian." <http://theshiftedlibrarian.com/stories/2002/01/19/whatIsAShiftedLibrarian.html>. Site visited 6/8/05.
- Liebowitz, Stan. 2002. "Copyright in the Post-Napster World: Legal or Market Solutions?" in *CopyFights: The Future of Intellectual Property in the Information Age*. Adam Thierer and Clyde Wayne Crews Jr., Eds. Washington D.C.: Cato Institute.
- Lombardi, John. V. 2000. "Academic Libraries in the Digital Age." *D-Lib Magazine*. Vol 6: No 10, pp. 1-9.
- McLeod, Kembrew. 2001. *Owning Culture: Authorship, Ownership and Intellectual Property Law*. New York: Peter Lang.
- Odlyzko, Andrew. January 19, 1999. "Competition and Cooperation: Libraries and Publishers in the Transition to Electronic Scholarly Journals." ArXiv:cs.DL/9901009 v1. <http://arxiv.org/pdf/cs.DL/9901009>. Site visited 6/9/05.
- Perlman, Michael. 2002. *Steal this Idea: Intellectual Property Rights and the Corporate Confiscation of Creativity*. New York: Palgrave.
- Perzigan, Anthony, Zimmerman, Julia, Sanville, Tom. (January 28, 2004). The Challenge for Ohio Higher Education Providing Access to Scholarly Communication. Available at: <http://www.ohiolink.edu/journalcrisis/ChallengeOhio.pdf>. Site visited 5/20/05.
- Torres Reyes, José Antonio (2002) El Profesional de la Información como facilitador del libre acceso a la información. In *Proceedings Foros de Consulta Popular Sobre el Derecho al Acceso a la Información, Monterrey, Nuevo León (Mexico)*. English Abstract at: *E-Prints* in

*Library and Information Sciences.*

<http://eprints.rclis.org/archive/00003695/>. Site visited 6/8/05.

Rosenzweig, Michael L. (2001). "Dear Colleague." In *Declaring Independence: A Guide to Creating Community-Controlled Science Journals*. Washington D.C.: SPARC. Available at: <http://www.arl.org/sparc/di>. Site visited 5/20/05.

Sapp, Gregg and Gilmour, Ron. 2003. "A Brief History of the Future of Academic Libraries." *Portal: Libraries and the Academy* 3.1, pp. 13-24. [http://muse.jhu.edu/demo/portal\\_libraries\\_and\\_the\\_academy/v003/3.1gilmour.html](http://muse.jhu.edu/demo/portal_libraries_and_the_academy/v003/3.1gilmour.html). Site visited 6/8/05.

Scholarly Publishing and Academic Resources Coalition (2001). *Declaring Independence: A Guide to Creating Community-Controlled Science Journals*. Washington D.C. Available at: <http://www.arl.org/sparc/di>. Site visited 5/20/05

Suber, Peter. April 27, 2005. "Open Access Overview." <http://www.earlham.edu/~peters/fos/overview.htm>. Site visited 6/8/05.

Vaidhyanathan, Siva. (2004). *The Anarchist in the Library: How the Clash Between Freedom and Control is Hacking the Real World and Crashing the System*. New York: Basic Books.

Willinsky, John. November 2002, "Copyright Contradictions in Scholarly Publishing," *First Monday*, Vol. 7, No. 11, Available: [http://www.firstmonday.org/issues/issue7\\_11/willinsky/index.html](http://www.firstmonday.org/issues/issue7_11/willinsky/index.html) Site visited, March 22, 2005.

# Catalysts for Change: Librarians and the Open Access Movement

Meghan Miller Brawley (University of Tennessee - Knoxville)

**Abstract:** This paper explores the role librarians play in the Open Access (OA) movement, in both research institutions and secondary school and public libraries. By examining the reasons researchers and libraries do not use open access resources, the author looks at ways librarians can overcome resistance to open access and increase the number of people dedicated to advancing the movement. Although open access has significant momentum in the scientific community, the scholarly literature in the humanities and some social sciences has not caught up. This is where librarians need to focus their efforts in order to make an impact on open access as a whole, by bringing in new authors and researchers dedicated to OA publishing.


## INTRODUCTION

The open access (OA) movement has both enthusiastic proponents and detractors, and many scholars and publishers are fiercely pitted against one another over intellectual property rights and the economics of research and publishing. Libraries are chief consumers of scholarly and scientific publications and the librarians who determine libraries' holdings potentially wield considerable power. If they choose, librarians can add significant momentum and publicity to the OA movement by supporting and promoting OA journals alongside—or in place of—traditional fee-based publications.

The OA movement was born out of frustration with traditional publishing. The now much-discussed “crisis in scholarly publishing,” the epithet given to the growing disparity between rising subscription costs and falling budgets, was coupled with complicated licensing agreements publishers created to protect their distribution rights. As a result, many think information is locked away to all but those with the funding to access it. Additionally, electronic subscription licenses do not always allow for access to back issues once a subscription is canceled; this is

---

M. Halbert (Ed.): *Free Culture and the Digital Library Symposium Proceedings*. Atlanta, Georgia: MetaScholar Initiative at Emory University, 2005. pp. 82-92.

 This work is covered by the Creative Commons Attribution-NonCommercial-NoDerivs 2.5 License.

significantly different from the paper and microfilm environment, where a library can physically own the back issues as long as the material lasts (Burke 2001).

OA supporters believe information is for the public good; free access to information will give research greater visibility and encourage further research. The Budapest Open Access Initiative (BOAI), one of the leading OA organizations, states that

removing access barriers ... will accelerate research, enrich education, share the learning of the rich with the poor and the poor with the rich, make this literature as useful as it can be, and lay the foundation for uniting humanity in a common intellectual conversation and quest for knowledge.

(<http://www.soros.org/openaccess/index.shtml>)

This paper seeks to explore the place librarians and libraries hold in the OA debate and the changes open access will bring to libraries by examining reasons researchers do not choose to publish in OA resources and looking at the ways librarians can influence these choices. Additionally, the paper looks at gaps in secondary school and public libraries that open access can fill. Although open access has significant momentum in the scientific community, the scholarly literature in the humanities and some social sciences has not caught up. This is where librarians need to focus their efforts in order to make an impact on open access as a whole, by bringing in new authors and researchers dedicated to OA publishing.

A large amount of the current literature focuses on the economics of OA publishing and arguments for or against the model; while a great deal has been written about the impact of open access on research communities and academic libraries, not much has looked at the impact of libraries on the future of open access. Librarians are touted as the key players in open access; the Association of Research Libraries' Scholarly Publishing and Academic Resources Coalition (SPARC) is a shining example of an OA resource (Morrison 2004). Librarians are navigators and guides, helping to bring organization to the OA universe and adding value often seen from commercial publishers — things like indexing and abstracts, metadata and tools to help improve searches and item retrieval in an environment that currently relies on simple, inadequate Google searches to find one relevant and authoritative item among millions (Guédon, 2001). Others think librarians will be needed to convince academics to publish in OA journals and to self-archive, citing testimony given in front of the House of Commons Science

and Technology Committee by academics who are reluctant to pay to publish (Poynder 2004).

### **WHAT LIBRARIANS ARE ALREADY DOING**

Many librarians already support open access, especially in academic and research libraries. The ARL and Association of College and Research Libraries are prominent library organizations who have signed the BOAI. ARL's SPARC works diligently all over the world to support OA journals through advocacy groups, education campaigns and "incubation" of new OA resources through publisher partnerships (SPARC 2005).

Many university library websites now include OA resources and material designed to educate users about OA issues. The University of Tennessee - Knoxville's Hodges Library has preprint sources listed in the physics subject guide, and the Directory of Open Access Journals is listed with all databases, including many other public, free databases. UT also labels databases with icons describing the access limits, from library use only to global access to all users. All the DOAJ journals are integrated into the e-journals list. The library system also has a Scholarly Communications Issues Committee and weblog, where open access is a prominent discussion topic (University of Tennessee Libraries 2005). Vanderbilt University has a journal costs online exhibit, where costs of journals like *Brain Research* and *Surface Science* are compared to less-expensive popular consumer goods such as a Volkswagen Beetle and a plasma television. Visitors to the site can look at selected cost figures and learn that electronic journals do not always provide the perpetual access print subscriptions provide. The University of Maryland Health Sciences and Human Services Library, Cornell University Library and Georgetown University Medical Center Dahlgren Memorial Library all have similar sites (Journal Costs 2004).

Librarians also participate in activism. Recently, librarians and scholars staged a well-publicized revolt against Reed Elsevier, canceling subscriptions and refusing to renew "big deal" licenses, opting to purchase journals "a la carte" instead. The University of California system boycotted Cell Press until they were able to negotiate lower prices. Cornell University canceled its Elsevier titles—two percent of its serials collection—and freed 20 percent of its budget (Guterman, 2004; Medeiros, 2004).



## **OBSTACLES TO OPEN ACCESS**

While many librarians may admire and agree with the principles behind the OA movement, they often encounter obstacles to supporting it. They need to focus not on the group of authors already involved in open access but on those who are not. The authors who are already involved are able to affect the movement by being outspoken advocates and recruiting new contributors, but open access needs more authors, especially in the non-scientific community. While this may be obvious, it is important to look at who is not involved, why not and how to get them involved. A survey of authors by Key Perspectives, Ltd. on behalf of the Joint Information Systems Committee and the Open Society Institute indicates authors do not publish in OA journals for the following reasons:

- perceived low prestige and reputation
- perceived slow publication times
- perceived smaller readership
- perceived fewer citations
- unfamiliarity with OA journals

The author-pays funding model is “only of middling importance,” according to the survey (Swan and Brown 2004).

### **Vanity press**

Many publishers claim OA journals are not of as high a quality as traditional journals, some going so far as to imply OA journals do not perform peer review. In a 2003 article on open access, Elsevier Science’s director Pieter Bolman argues that the OA movement operates on the assumption “that peer review is the only essential service to be provided by journals” and OA sources consist only of parallel-published documents reviewed under the traditional system for publication in a fee-based journal, then simultaneously uploaded to the internet by the author for free online access. Bolman calls this “parasitic” (Bolman 2003), and if this scenario was an accurate representation of the totality of open access, he would be correct. However, peer-reviewed OA journals do exist, and they do not rely on articles reviewed first through traditional publishing. The Directory of Open Access Journals lists more than 400 journals that submit all articles to “quality control ... through an editor, editorial board and/or a peer-review system” (DOAJ 2005). A key tenet of the BOAI is peer review (BOAI 2005). Some journals review articles prior to publication; one of the

advantages of OA electronic publishing is the peer-review process can take place publicly and after publication (Henry 2003). Some OA journals, including BioMed Central's medical journals, post signed reviews with the articles so the review process is open. BioMed Central's peer review policies are also public for each journal (BioMed Central 2005). *BMJ*, while not fully open access—the journal keeps all original research free, but editorial content (reviews, letters and editorials) are only available for the first week after publication, then embargoed for 51 weeks—publishes reader responses with articles, although the original peer review comments are not public (bmj.com—Subscriptions 2005).

Librarians can combat misperceptions of the prestige of open access on many levels, including the individual level, the student level and the faculty and staff level. The librarian works on the individual level when assisting patrons who are looking for resources; these individuals can be students, faculty, staff, outside researchers, and members of the general public. Reference work is one of the primary ways a librarian—especially one in a university or special library—can reach the general public and other non-affiliated researchers who may not otherwise encounter OA resources or even know about the movement. During reference assistance, librarians have the opportunity to suggest OA resources as well as traditional ones, pointing out the peer-review policies and research by ISI suggesting OA journals are cited as often as non-OA journals, and possibly earlier. The ISI report also concluded OA journals “have the potential” to have a larger reader base than non-OA journals (Testa and McVeigh 2004). Bibliographic instruction, a significant component of reference work in university libraries, provides librarians with an excellent opportunity to reach out to budding researchers (and their professors). During presentations in which librarians point out the various resources available in the library, they can educate students about the OA resources available and about the OA movement. These students, already well-versed in the electronic environment and copyright issues by virtue of their activity in music and video sharing, are prime candidates for a new injection of support for open access. Institutional repositories provide an excellent first foray for students into the world of published research. Institutional repositories—especially if created on an academic department level—give young researchers a way to archive undergraduate attempts that may not be ready for the academic publishing world but could certainly benefit from more exposure and distribution than most class assignments get.

### **Open access? Is that like Linux?**

Even researchers who know open access is not the same thing as open source software are not always familiar with specific journals, especially ones in their fields. The BioMed Central and Public Library of Science journals may be fairly well known, but many authors are unaware of journals in the humanities, partly because there are fewer in the non-science areas. Authors can only publish in the journals they know about. Although almost two-thirds of the JISC/OSI surveyed authors were aware of open access, only a quarter of them found out about the movement through their institutions. Additionally, many authors who did not publish in OA journals said they were unfamiliar with OA journals in their fields, and only eight percent of OA authors have been aware of the movement for more than three years. Only six percent of OA authors identified journals by consulting a librarian for a recommendation; 47 percent had recommendations from colleagues (Swan and Brown, 2004). While this is not unusual information-seeking behavior, it does indicate a lack of involvement by librarians in publishing seeking. Although this is probably not because librarians are deliberately uninvolved, it does show there is room for librarians to actively point researchers to potential publications.

Again, reference interviews give librarians a chance to educate researchers one-on-one. As the researcher looks for information for an article, the librarian can take an active part in the future publication source by asking in what journals the researcher is wanting to publish, and suggesting equivalent OA journals.

### **RECOMMENDED FUTURE ACTIONS**

Research librarians are already active in the OA movement. ARL and ACRL are two associations that have put their weight behind the movement; many universities have education campaigns. However, more needs to be done to increase acceptance of OA journals as legitimate publications. In an e-mail interview with Rice University's Digital Library Initiative's Executive Director Geneva Henry, she discussed problems for tenure-seeking faculty members. "Junior faculty members," she said

once they understand the benefits of OA publishing, support it in theory. However, they are faced with the institutional restrictions that measure their contributions through traditional publications when considering tenure. Right now, this tension generally resolves in favor of the traditional publishers; getting tenure is the most important consideration for these junior faculty

(Henry 2005). Faculty librarians can lobby their institutions to change tenure rules to include OA journals rather than favoring traditional journals. Librarians on tenure committees obviously have significant opportunities to educate fellow committee members about OA resources and help give peer-reviewed OA journals the same recognition of traditional journals. Knowing that tenure committees will equally consider OA and traditional journals will help encourage faculty members to publish and use OA journals. Open access is not “vanity press;” while posting reflections on a weblog hardly qualifies as published work, the delivery vehicle of open access—electronic, on the internet—does not preclude it from being scholarly. Librarians with faculty status can serve on tenure committees and help weigh decisions in favor of—or at least not discriminating against—junior faculty whose minimum required annual publications include OA journals.

Although the author-pays model was not considered by authors in Swan and Brown’s survey to have much of an influence on their decision to publish in open access (55 percent have published without having to pay), authors may not be willing to pay publishing fees that can often be in excess of \$1,000. However, OA publications often offer memberships to offset publishing costs. Institutional, corporate and group memberships are available at BioMed Central where the membership fees (paid by the sponsoring organization) cover publishing fees, and member authors publish without paying directly (BioMed Central Membership 2005). Libraries can redirect subscription funds to cover these memberships, either entirely or in conjunction with departments and the larger institutions.

### **Non-research Libraries**

While ARL, ACRL and the American Association of Law Libraries are BOAI signatories, the American Library Association is not. ALA represents public and school libraries as well as research institutions and special libraries. Because many OA resources publish scientific and scholarly research, their place in public and school libraries is not readily apparent, but these libraries can use the OA movement to introduce young people to scholarship. A Peter J. Hart Research Associates/Public Opinion Strategies survey strongly indicated high school students feel they were not fully prepared for college. Forty percent of the students surveyed said they had “some gaps”—defined as “10 percent large gaps”—in their ability to do research, and professors had even harsher assessments of their students’ abilities. Fifty-nine percent

of instructors were dissatisfied with their students' research skills, and seventy percent of instructors reported dissatisfaction with students' abilities to understand complex reading material (Peter D. Hart Research Associates 2005).

OA journals can provide a tremendous benefit to school libraries. If universities and other research institutions are unable to purchase subscription journal access, public secondary schools are even less so. While many schools are able to purchase databases as part of consortia, OA journals can provide access to publications not typically included in these consortia. For instance, the Tennessee Electronic Library provides access to excellent resources for young children, as well as current events resources like InfoTrac OneFile, General Reference Center Gold and the National Newspaper Index. But much of the resources available cover mainly popular and non-peer-reviewed resources (Tennessee Electronic Library 2005). The federal government funds TEL through the Library Services and Technology Act, not through the schools, so OA resources would give school media centers resources they can access independently of the political climate towards library funding. With OA journals, teachers and school media specialists would be able to introduce complex research to students and help them learn to understand what they read. The research would also provide educators with examples of research methods and could introduce students to scholarly thinking skills, all at no cost to the school system beyond the electronic infrastructure necessary to access the internet and the time spent on the instruction.

In public libraries, many of the same arguments apply. Many public librarians help school students with research assignments; during reference interviews they can introduce students to OA journals as well as consortia resources. While the OA movement is concerned with scholarly and scientific publishing, and not geared toward the general public or popular material, open access will provide public librarians—especially reference librarians—with resources they would not otherwise have to offer to patrons who need more information than is provided in newspapers, magazines and other popular literature. Additionally, reference librarians can use OA resources to answer difficult reference questions.

## CONCLUSIONS

The OA movement may not eclipse traditional publishing, but it is likely the mutual competition will force open access to develop economically viable models, which would force traditional publishing prices down and publishers will adapt licensing agreements to allow for freer—if not free—access. Librarians will be instrumental in this evolution, both for their roles as navigators and organizers and as negotiators between publishers and scholars. In order to solve the “crisis” in scholarly publishing, librarians must be able to adapt to new technologies and maintain communication among all parties.

Librarians can use their role as educators and research assistants to reach out to faculty members and student researchers, introducing them to the OA movement and resources, as well as promoting and encouraging institutional repository use. Bibliographic instruction sessions and reference interviews provide two excellent opportunities for user education. Tenured librarians can also use their influence to change institutional attitudes toward the validity of OA publishing, explaining the OA peer review process and pointing to well-known and accepted OA resources.

While library websites include OA resources and information about the serials crisis, websites can be more “activist” by displaying more prominent links to information about open access, such as the Directory of Open Access Journals. The University of Tennessee Scholarly Communications Issues weblog was only found by searching the library site for “open access.” OA journals can be separated from regular journals on e-journal lists, websites could have an additional list for OA resources while keeping them on the main e-journal list, or they could simply be more prominently identified as open access. Promotion of institutional repositories is also important; links can be added to resource lists and subject guides.

There is considerable clout in the library community, both financial and political. ALA is considered to be largely responsible for the recent PATRIOT Act changes reinstating readers’ privacy rights. Institutional subscriptions outnumber personal ones and provide much of the financial income for traditional publishers; when the University of California boycotted Cell Press they were able to get the publisher to renegotiate prices. If they choose, librarians can add significant momentum and publicity to the OA movement by supporting and promoting OA

journals alongside—or in place of—traditional fee-based publications.

## REFERENCES CITED

- BioMed Central. 2005 [cited June 18, 2005]. Available from <http://www.biomedcentral.com> (accessed June 18, 2005).
- BioMed Central Membership. 2005 [cited June 18 2005]. Available from <http://www.biomedcentral.com/info/about/membership> (accessed June 18, 2005).
- bmj.com—subscriptions. in BMJ Publishing Group [database online]. London, 2005 [cited June 18 2005]. Available from <http://bmj.bmjournals.com/subscriptions> (accessed June 18, 2005).
- Bolman, Pieter. 2003. “Open access: Marginal or core phenomenon?: A commercial publisher's view.” *Information Services and Use* 23, (2-3): 93-98, <http://www.epnet.com> (accessed April 6, 2005).
- Budapest Open Access Initiative. 2005 [cited June 18 2005]. Available from <http://www.soros.org/openaccess/index.shtml> (accessed April 10, 2005).
- Burke, L. 2002. “The future role of librarians in the virtual library environment.” *Australian Library Journal* 51, (1) (Feb): 31-45, <http://www.csa.com> (accessed April 7, 2005).
- Directory of Open Access Journals. 2005 [cited June 18 2005]. Available from <http://www.doaj.org> (accessed June 18, 2005).
- Medeiros, Norm. 2004. “Of budgets and boycotts: The battle over open access publishing.” *OCLC Systems and Services* 20, (1): 7-10, <http://www.emeraldinsight.com/rpsv/~1183/v20n1/s1/p7> (accessed April 7, 2005).
- Morrison, Heather. 2004. “Professional library & information associations should rise to the challenge of promoting open access and lead by example.” *Library Hi Tech News* 21, (4): 8-10, <http://www.emeraldinsight.com/rpsv/~1147/v21n4/s4/p8> (accessed April 7, 2005).
- Guédon, Jean-Claude. In Oldenburg’s long shadow: Librarians, research scientists, the publishers, and the control of scientific publishing. Paper presented at 138th Membership Meeting of the Association of Research Libraries (ARL), Toronto, ON, Canada, <http://www.arl.org/arl/proceedings/138/guedon.html> (accessed April 10, 2005).
- Guterman, Lila. 2004. “The promise and peril of 'open access.'” *Chronicle of Higher Education* 50, (21) (Jan. 30): A10-A14, <http://www.epnet.com> (accessed April 6, 2005).
- Henry, Geneva. 2005. Personal e-mail.

- Henry, Geneva. 2003. "On-line Publishing in the 21st Century: Challenges and Opportunities." *D-Lib Magazine* 9, (10) (Oct), <http://www.dlib.org/dlib/october03/henry/10henry.html> (accessed June 15, 2005).
- Peter D. Hart Research Associates/Public Opinion Strategies. 2005. "Rising to the challenge: Are high school graduates prepared for college work?" Available from [http://www.achieve.org/dstore.nsf/Lookup/pollreport/\\$file/pollreport.pdf](http://www.achieve.org/dstore.nsf/Lookup/pollreport/$file/pollreport.pdf) (accessed June 18, 2005).
- Poynder, Richard. 2004. "U.K. academics and librarians disagree over open access publishing." *Infoday PB*: (May 3), <http://www.infoday.com/newsbreaks/nb040503-3.shtml> (accessed April 8, 2005).
- Scholarly Publishing and Academic Resources Coalition. 2005 [cited June 18, 2005]. Available from <http://www.arl.org/sparc/> (accessed June 18, 2005).
- Swan, A. & Brown, S.N. 2004 [cited June 18 2005]. *JISC/OSI Journal Authors Survey Report*. Available from [http://www.jisc.ac.uk/uploaded\\_documents/JISCOAreport1.pdf](http://www.jisc.ac.uk/uploaded_documents/JISCOAreport1.pdf) (accessed June 18, 2005).
- Tennessee Electronic Library. 2005 [cited June 18 2005]. Available from <http://www.tenn-share.org/tel.html> (accessed June 18, 2005).
- Testa, James and Marie McVeigh. 2004. "The impact of open access Journals: A citation study from Thomson ISI." Available from <http://www.isinet.com/oaj> (accessed June 18, 2005).
- University of Tennessee Libraries. 2005 [cited June 18 2005]. Available from <http://www.lib.utk.edu/> (accessed June 18, 2005).
- Vanderbilt University Jean and Alexander Heard Library. 2004 [cited June 18, 2005]. Available from <http://www.library.vanderbilt.edu/jcosts/> (accessed June 18, 2005).



# Millennial Net Value(s): Disconnects Between Libraries and the Information Age Mindset

Charles F. Thomas and Robert H. McDonald (Florida State University)<sup>1</sup>

**Abstract:** Libraries are facing a new generation of online users who are technologically savvy and integrate information access and use in all spheres of their lives to an unprecedented degree. They approach the traditional library with certain expectations that may conflict with the existing services, policies, and values of the library as information broker. This paper identifies the fundamental disconnects between current library values and this new generation of information user. In the process, the authors identify numerous opportunities for leadership in meeting the needs of the millennial generation.


## INTRODUCTION

Recently, the U.S. Supreme Court reached a decision in the MGM Studios vs. Grokster, Ltd. case. The published decision is naturally of interest to libraries because of the associated copyright and fair use issues. From the perspective of digital libraries, however, equally interesting is Justice Stephen Breyer's accompanying analysis of the case. In it, the Justice acknowledges the potential of peer-to-peer file sharing softwares for non-infringing sharing of content such as public-domain texts, freely downloadable music, media with accompanying Creative Commons licenses, and newscasts. Breyer's analysis also references evidence uncovered in this case suggesting that 75-90 percent of current peer-to-peer file sharing traffic consists of copyrighted works being exchanged illegally.

Justice Breyer's comments raise two important issues for digital libraries. The first is the fact that peer-to-peer file sharing network space is an unexplored frontier for disseminating digital library content. What would happen if digital libraries collectively began to push their free cultural materials into this space? How would file sharing software users react? What impact would this have on the

---

M. Halbert (Ed.): *Free Culture and the Digital Library Symposium Proceedings*. Atlanta, Georgia: MetaScholar Initiative at Emory University, 2005. pp. 93-105.

 This work is covered by the Creative Commons Attribution-NonCommercial-NoDerivs 2.5 License.

overall balance of unrestricted versus restricted content in this environment? Are some digital library materials already being circulated in this space without our knowledge? Undoubtedly, this space offers many untapped research opportunities and territory ripe for digital library entrepreneurs.

The second issue is a bit more abstract. Peer-to-peer file sharing software is just one type of tool among many employed by Internet users. Like many other technologies, it is used most extensively within particular communities. In this case, that community is most likely to be college students, who are nearly three times as likely to have used file sharing softwares as the general Internet population (Jones 2002). For libraries, the main concern should not be this particular type of tool, but understanding the underlying values and priorities of populations who employ such tools. What is most important to them? What co-existing values do they have that are balanced against one another? Can we even generalize in this way about online user groups? What are the service and policy implications of their values that we should try to accommodate? How do libraries' existing priorities and values mesh or conflict with the emerging community of online users?

This paper examines what we are learning about online user communities and their values. It relates them to the existing services, policies and values espoused by libraries, paying particular attention to the fundamental disconnects and strategies we might use to become more accommodating of this important group.

## **NET GENERATIONS**

Digital libraries have emerged during the past fifteen years, and as a whole have progressed through several identifiable stages of growth and maturity (Greenstein & Thorin 2002). In recent years, the digital library research agenda has progressed beyond standards and technical issues to focus on identifying and responding to the needs of online users and online communities. Since Howe & Strauss' *Millennials Rising* (2000), the "Millennials," also known as "Gen Y" or the "Net Generation," have claimed an increasing share of this attention. Kaimal (2003) notes the slight variations in the dates associated with this generation, ranging from Alch's (2000) definition as those born between 1976-1996, to alternative definitions that shift these dates slightly later. Despite the lack of precise agreement on who is part of this generation, numerous

studies and papers (Woodall, 2004; Lippincott, 2005; Kaimal, 2003) affirm some of their key characteristics, including:

- They are generally confident in large institutions;
- They are receptive of requests to be part of larger group efforts;
- They want to be independent problem solvers;
- They are accustomed to media-rich entertainment and computing;
- They are proficient in using many kinds of media;
- Multi-tasking is the accepted norm for their personal, social, and work activities;
- They are slow to build trusting relationships, but;
- They prefer to build a wide, sustained network of connections via technology.

DeRosa, Dempsey & Wilson (2004) synthesized earlier studies and their own work at OCLC to analyze the attitudes and impact of the leading edge of Gen-Y, those who are in the early years of college and the workforce. According to their report, this generation:

grew up with computers, multimedia, the Internet and a wired world. ... Their world is a seamless 'infosphere' where the boundaries between work, play and study are gone. Computers are not technology, and multitasking is a way of life...the lines between workplace and home are blurred.

Why is this group so important to librarians and digital library planners? As the name "Net Generation" implies, this generation is the first to have come of age in an environment where ubiquitous computing has existed as long as they can remember. Unlike previous generations such as the "Generation X" defined by Coupland (1991) or the Baby Boomers, this new generation was not forced to adapt in mid-adulthood to the changes wrought by the Internet.

Instead, their formative years coincided with the birth and growth of these new technologies and communication channels. As the first native generation of the Information Age, the Millennials are more likely to exhibit attitudes and behaviors that Frand (2000) characterized as the "Information-Age Mindset." This mindset, according to Frand, manifests itself as:

- Computers are not technology (too commonplace);
- Internet is better than TV;

- Reality is no longer real (cannot assume the truth or authenticity of anything online);
- Actions are more important than knowledge;
- Nintendo trial-and-error approach to improvement is a viable model for learning;
- Multitasking is a way of life;
- Typing is better than writing;
- Staying connected is essential;
- Zero tolerance for delay;
- Consumer and creator are blurring (copyright implications);

Now that the Millennials are entering early adulthood, notice this mindset and other generational differences are becoming apparent to educators, librarians and even employers. While in college, in particular, their behaviors are even more sharply distinct from those who came before them. They are more than twice as likely as the general population to rely upon their online peers and other Internet resources for entertainment (Jones 2002). Eighty-four percent of them own a computer, and twenty-five percent own more than one computer (Oblinger 2003). As noted earlier, they choose to remain near advanced computing and communications technologies, and they integrate those technologies into their social, personal and school activities. This generation is, more than any other, wired, and thus of great interest to those who create, implement and support the advanced “cyberinfrastructure” (Atkins 2003) that promises to revolutionize research, teaching and learning.

For all of their comfort and familiarity with technology, however, one cannot assume that this generation is any more intuitively skilled than other groups in finding or understanding information. Supporting all of the emerging cyberinfrastructure is a cyber-infrastructure that must be recognized and incorporated into plans for serving the Millennials and subsequent generations. This layer of information systems, standards, stakeholders, policies and data remains far from seamless and transparent. It must be learned and exploited through guided experience, which goes counter to the Gen-Y preference, noted earlier, for independent problem solving. Unfortunately, skills with technology do not necessarily translate into the skills needed to be successful information consumers, particularly in academic endeavors. To build upon their Information Age mindset, it seems that Millennials may need some

guidance to develop further knowledge of issues like privacy and intellectual property that are increasingly important in our Information Society (Lippincott 2005). Kaimal (2003) likewise suggests that they are very much like other young adults in their need for guidance and assistance to define goals, and to plan the exact paths and resources needed to achieve these goals.

With so much emphasis on the Millennials, it is sometimes easy to neglect the many other large and important online communities. Gen-Y is of such immediate interest because they are most likely to be exploiting fully the tools, communication possibilities, and resources that can be found online. In other words, they receive so much attention because they may be the most receptive to the new systems, resources and services that libraries are trying to provide to online users. Librarians and digital library planners must not forget, however, that earlier groups of online users have been providing helpful feedback and preferences since much earlier in the life of the public Internet. For example, Rayward & Miller (1998) noted almost a decade ago that early public Internet users wanted information providers to remove as many institution-specific rules as possible on access and use of information. This desire for a “functional consolidation” of information providers is very similar to Frand’s Information-Age mindset of no tolerance for delays, and meshes well with the attitudes described in numerous studies of the Net Generation.

In fact, this and other evidence suggest that libraries should expect more than one “Net Generation.” Bioarsky (2002) examined the aesthetic and functional preferences of the Millennials in cultural contexts such as the fine arts and in technical contexts such as web interfaces. Interestingly, when their preferences were compared with older others who had become accustomed to the conventions and norms of the same types of cultural, artistic and technical environments, few differences were perceived. Bioarsky hypothesized that Millennials soon will not be alone in the ways and extent to which they use computing and communication technologies. Others from previous generations are merely integrating these same preferences and behaviors into their lifestyle a bit more slowly. The Millennials are merely leading the way for a few years, and are a good forecaster of a mindset, behaviors and needs that will become more characteristic of all “Net Generations” in the near future. If so, traits such as a preference for group work among Millennials, may significantly alter how libraries serve them.

## LIBRARY VALUES

Library values stem from a desire to share knowledge and are based on principles of openness and accessibility. A good framework with which to measure library values stems from the five laws of library science developed by S.R. Ranganathan almost a century ago. Ranganathan states in his five laws (Ranganathan 1957):

1. Books are for use
2. Every reader his book
3. Every book its reader
4. Save the time of the reader
5. A library is a growing organism

If one can imagine these tenants as applied to libraries within the last 10 years, it is clear that Ranganathan's laws no longer apply to just the singular media format of the printed book or journal. In fact, in his tribute essay to Ranganathan, Garfield states, "Of course the term 'Books' here stands for all information items" (Garfield 1984). Within this context, much information has become a network-accessible resource that is provided by libraries via the publishing medium of the Internet. Law #5 is far reaching in its insight in that libraries have evolved over the last century as growing organisms. At some point, libraries may no longer be repositories themselves but purely access mechanisms to repositories of knowledge that are best stored in aggregate via collaborative partnerships either between content publisher and libraries or libraries as content publishers. DeRosa, Dempsey & Wilson (2004) predict substantial changes over the next decade related to libraries, information and ownership versus access. In such a scenario, libraries will be distinguished from one another primarily by ownership of sole copies of locally-published digital content not accessible elsewhere.

The last 10 years (1995-2005) bear particular significance when considering library values and online user needs. This period, starting in 1995 with the original IPO of Netscape and the ubiquitous availability of browser-based access to information (Kelly 2005), might be known as the *consumer net decade* and corresponds with the clear progress of libraries evolving from knowledge storehouses of print materials to service-based access nodes of online information (Bertot 2005). Prior to this period,

many academic and research libraries did offer access to electronic formats either through recorded media such as CD-ROM or data tapes, and to networked resources like gopher and university-based email systems.

However, it was the dawn of the World Wide Web and access via the Netscape browser that revolutionized thinking about information access. Easy access by the populace to a global web of information brought forth the creation and interchange of digital objects like electronic text, hypertext, images and rich multimedia. Early digital libraries focused on the WWW dissemination of images, namely image representations of rare documents and photographs (Cornell University Libraries 2005). During the consumer net decade, content in digital libraries has expanded to include valuable content such as class lectures and digital broadcast archives (Internet Archive 2005). This growth in the availability and range of materials and the ease with which these materials can be transformed has increased the access expectations of online users.

Additionally, the emergence of popular tools that enable “recombinant” mixtures of pre-existing digital content into new creations, have clouded areas of traditional importance to libraries, such as copyright and provenance (Dempsey 2003). It is clear that Millennials and others comfortable with a wide range of media and technologies will redefine the traditional manifestations of research and creative activity with these new mashed, cut and pasted creations. For them, the line between consumer and creator is blurred in a way that previously was not possible.

How have these new technical and access/use expectations influenced library values? Because of highly publicized legal actions by media companies and a general fear of being found guilty of non-compliance with copyright regulations, libraries generally have been reluctant to embrace or provide such capabilities for users. Even when dealing specifically with their own digital library collections and other materials over which a library might reasonably assert intellectual ownership, the library community has not embraced new technical capabilities, nor shown significant desire to enable online users’ preferences for personalization and recombination of information found online (Larsen 2003).

How have online users responded to libraries that do not demonstrate evolving values through changing policies and

services? Research shows that, to a large degree, libraries are being ignored by online users, particularly younger generations. For example, Jones (2002) learned that new students in college are seventy-three percent more likely to use the Internet for research than the traditional library. Library values, evidently, need to be considered carefully. Librarians have made significant strides over the past decade by expanding their collections into new formats. However, the values that support library services have been slower to grow. To a great extent, the services and policies that libraries have adopted for the net decade have been merely online equivalents of past policies and services, and are not really responsive to the new modes of learning, exploration, and collaboration that are characteristic of the Net Generation learner.

### **FUNDAMENTAL DISCONNECTS**

Clear rifts have emerged in the virtual terrain that is occupied by library policies, services and collections and is explored by online users. These rifts or disconnects can be grouped into three classifications for redress. These include technology (infrastructure and integration), policy (copyright, IT policy, liability), and unexploited opportunities.

### **TECHNOLOGY DISCONNECTS**

Libraries have done little to embed themselves and their resources into non-library tools and spaces that are important to Net Generation learners (Lippincott 2005). Furthermore, libraries have done little to make most of their systems and discovery tools customizable or easily integrated into the life of online users. Library technologies remain substantially limited by an enduring obsession with individual privacy and copyright. Other user desires that have emerged in recent years, such as capabilities for collaborative and group work, remain essentially unaddressed in library conceptions of technology development. Such oversights and lack of responsiveness are good examples of why Millennials may be disappointed when they find that the technology infrastructure they expect to find in libraries turns out to be more limited (Oblinger 2003).

Some of the other key technology disconnects between libraries and online communities include:

- Libraries typically do not provide tools, hardware or software, nor support for students to create new digital products (Lippincott, 2005);



- Dogmatic library protection of privacy inhibits library support for file-sharing, work-sharing and online trust-based transactions that are increasingly common online;
- Libraries are not using technology and standards like RSS to permit choice-driven alerts on new resources or services, or to enable more effective need-driven help from libraries;
- Libraries continue to segment systems and technology for designated types of activity and information, when online users want more seamless integration (Jones, 2002);
- Libraries are still focused on becoming “wired” while many users desire wireless access.

To give credit where it is due, initiatives such as the Research Library Group’s RedLightGreen (RLG 2003-2005) have contributed some useful analysis of the needs and preferences of online audiences, and produced test interfaces for evaluation. Such efforts are commendable, and provide good models for thinking outside traditional library perspectives. Meanwhile, the technology disconnects between libraries and the emerging net generations are apparent, widespread and continuing. Libraries that are interested in assuming leadership roles during the next decade should consider the numerous opportunities for innovation that remain in this space today.

### **POLICY DISCONNECTS**

Drawing a clear line between technology and policy can sometimes be difficult. For example, how many of the characteristics of current libraries, identified by Lippincott (2005), are driven purely by technology or by policy? These traits include:

- Mainly electronic text-based collections;
- Not enough multimedia content for online users;
- Constructed for individual use;
- Require users to learn from experts how to access, use information and services;
- Assume that work progresses in a linear fashion;
- Library presence is usually “outside” main online place for student activity.

Not many of these could be resolved simply by introducing new technology. Conversely, policies used consistently to guide changes in these areas would likely yield substantial changes.

Similarly, a policy solution might be required to address the following types of disconnects between libraries and online users:

- Library staff lacking sufficient technology and media competency;
- Deliberately pushing library search tools into other environments, and conversely;
- Integrating other popular external search tools into library systems;
- Libraries linking and pointing only to what they own;
- Re-structuring access to reflect use instead of library organizational structure (Lippincott 2005).

Imagine the changes in library services if libraries addressed these disconnects by long-term attention to effective remedial policies!

### **OPPORTUNITY DISCONNECTS**

What are we doing now to enable flexibility for new learners? As this paper shows, we know a lot about the emerging preferences and needs of the net generations. We are not yet exploiting their affinity for self-paced, independent trial-and-error method of learning. We are not taking advantage of opportunities to make our information look and behave like information that exists in the corporate world of the entertainment industry. We have not found ways to deliver information literacy skills to them in alternate channels and spaces (Lippincott 2005). We are not applying their desire for instant gratification to the ways we could be using technology for streamlined services. Because of administrative, security and policy restrictions, we are not permitting online users to use their skills and abilities fully (Oblinger 2003). In the process, we also are sacrificing lessons and feedback we could gain from them.

### **CONCLUSION**

Many of the most important disconnects between library priorities and net values are closely related to the way libraries conceive, create, and provide public computing infrastructure. Larsen & Wactlar (2003) emphasize that libraries need to be fitting technology-enabled opportunities into the social fabric, and matching system capabilities to user needs. (Larsen & Wactlar 2003). The promise of seamlessness that stems from ubiquitous computing access is, unfortunately, stifled significantly within libraries of today (Lippincott 2005). Certainly, accommodating

changing user preferences is not the only priority that drives library decisions. A basic philosophical issue for libraries is the extent to which we should move in the direction of the user, and how much we should expect them to move in our direction. Unfortunately, rigid adherence to old values clearly is forcing online users to find other paths to information, and often not even realize what we have to offer. Finding the right way to achieve balance between traditional values and the expectations and habits of the wired generations will determine whether libraries remain relevant in the social, educational and personal contexts of the Information Age.

## ENDNOTES

1. *Charles F. Thomas* is the Assistant Director of Libraries for Integrated Digital Services and Scholarly Communication at Florida State University. He can be reached at [cthomas@mailier.fsu.edu](mailto:cthomas@mailier.fsu.edu).  
*Robert H. McDonald* is the Associate Director of Libraries for Technology & Research at Florida State University. He can be reached at [rmcdonal@mailier.fsu.edu](mailto:rmcdonal@mailier.fsu.edu).

## REFERENCES CITED

- Alch, M. L. (2000). "The Echo-Boom Generation: A Growing Force in American Society". *The Futurist* 34:5, 42-46.
- Atkins, D. E., et. al. (2003). "Revolutionizing science and engineering through cyberinfrastructure: report of the National Science Foundation blue ribbon advisory panel on cyberinfrastructure." (Wash., DC: NSF): 30. Retrieved July 28, 2005 from <http://www.nsf.gov/cise/sci/reports/atkins.pdf>
- Bertot, J., McClure, C., & Jaeger, P. (2005). Public Libraries and the Internet 2004: Survey Results and Findings. Retrieved from <http://www.ii.fsu.edu/publications/2004.plinternet.study.pdf>.
- Boiarsky, C. (2002). "This is Not Our Fathers' Generation: Web Pages, the Chicago Lyric Opera, and the Philadelphia Orchestra." *Journal of Popular Culture*. 36(1): 14-24.
- Cornell University Libraries, (2005). Making of America Website. Retrieved on July 25, 2005 from <http://cdl.library.cornell.edu/moa/about.html>
- Coupland, D. (1991). *Generation X: Tales for an Accelerated Culture*. NY: St. Martin's Press.
- DeRosa, C., L. Dempsey, & A. Wilson. (2004). *The 2003 OCLC Environmental Scan: Pattern Recognition*. Dublin, Ohio: OCLC Online Computer Library Center.
- Dempsey, L. (2003). "The Recombinant Library: Portals and People." *Journal of Library Administration* 39(4): 103-36.

- Frاند, J. L. (2000). "The Information-Age Mindset: Changes in Students and Implication for Higher Education." *EDUCAUSE Review* 35:5 (Sept./Oct.), 14-24.
- Garfield, E. (1984). "A Tribute to S.R. Ranganathan, The Father of Indian Library Science. Part 1, Life and Works." *Current Contents* 6: 5-12.
- Greenstein, D. & S. Thorin. (2002). *The Digital Library: A Biography*. Wash., D.C.: Council on Library and Information Resources.
- Higher Education Special Report. (2004). "Digital Community Colleges and the Coming of the Millennials." *T.H.E. Journal*. (October): 14-15.
- Howe, N. & W. Strauss. (2000). *Millennials Rising*. New York: Vintage Books.
- Internet Archive. (2005). Internet Archive Movie Archive. Retrieved on July 25, 2005 from <http://www.archive.org/details/movies>.
- Jones, S. (2002). *The Internet Goes to College. How Students are Living in the Future with Today's Technology*. Retrieved July 28, 2005 from [http://www.pewinternet.org/pdfs/PIP\\_College\\_Report.pdf](http://www.pewinternet.org/pdfs/PIP_College_Report.pdf).
- Kaimal, G. (2003). *Gen-X Meets Gen-Y: Youth Perceptions and Concerns About the Future: A Review of the Literature*. Retrieved July 28, 2005 from <http://www.stonehill.edu/cs1/sa/Generation%20Y%20review%20of%20literature.pdf>.
- Kelly, K. (2005). "We Are the Web." *Wired* 13:08, 92-95.
- Larsen, R. L. & H. D. Wactlar. (2003). *Knowledge lost in information: report of the NSF Workshop on Research Directions for Digital Libraries*. Pittsburgh: University of Pittsburgh.
- Lessig, L. (1999). *Code and Other Laws of Cyberspace*. NY: Basic Books.
- Lessig, L. (2004). *Free culture: how big media uses technology and the law to lock down culture and control creativity*. NY: Penguin Books.
- Lippincott, J. (2005). "Net Generation Students & Libraries." *Educause Review*, (March/April): 56-66.
- Metro-Goldwyn-Mayer Studios, Inc. Vs. Grokster Ltd.*, 04-480 Supreme Court of the United States, 545 U.S. (2005). Retrieved October 4, 2005 from <http://www.supremecourtus.gov/opinions/sliplist/s545pt2.html>.
- Naughton, K. & Raymond, J. (2002). Hip to Be Square. *Newsweek*. 140(17): 50-1.
- Oblinger, D. (2003). "Boomers, Gen-Xers & Millennials: Understanding the New Students." *Educause Review*. (July/August 2003): 37-47.
- Oblinger, D. & Oblinger, J. eds. (2005). *Educating the Net Generation*. Retrieved July 28, 2005 from <http://www.educause.edu/educatingthenetgen>.
- Ranganathan, S.R. (1957). *The Five Laws of Library Science*. (London, Blunt), 456 p.

- Rayward, B. & G. Miller. (1998). "Electronic information and the functional integration of libraries, archives and museums." In E. Higgs, ed. (1998). *History and Electronic Artefacts*. (Oxford, Clarendon Press), 207-226.
- Research Libraries Group. (2003-2005). RedLightGreen: Articles & Updates [website]. Retrieved July 28, 2005 from [http://www.rlg.org/en/page.php?Page\\_ID=4661](http://www.rlg.org/en/page.php?Page_ID=4661)
- Strauss, W. & Howe, N. (1991). *Generations: The History of America's Future, 1584 to 2069*. New York: Quill.
- Tapscott, D. (1999). *Growing Up Digital The Rise of the Net Generation*. New York: McGraw-Hill.
- Weiler, A. (2005). "Information-Seeking Behavior in Generation Y Students: Motivation, Critical Thinking, and Learning Theory." *Journal of Academic Librarianship*. 31(1): 46-53.
- Woodall, D. (2004). "Enter the Millennials." *ASEE Prism*. 13(9): 59.

# Rights, Registries, and Remedies: An Analysis of Responses to the Copyright Office Notice of Inquiry Regarding Orphan Works

Denise Troll Covey (Carnegie Mellon University)

**Abstract:** The U.S. Copyright Office received hundreds of responses to the Notice of Inquiry regarding orphan works. The responses report encounters with orphan works in all types of media, and many propose solutions to the problem, ranging from the creation of support services to eliminate or alleviate the problem to new legislation that would provide exemptions or accommodations that allow unauthorized use of copyrighted works under certain conditions. A quantitative look at the responses shows their general contours. A qualitative examination of the pros and cons of different positions taken on the many issues that must be addressed to solve the problem of orphan works reveals the trade-offs and implications of different actions to address the problem and the different perspectives and agendas of the respondents. Following these objective analyses of the responses, the paper argues for multiple approaches to solving the problem aimed at balance, certainty, practicality, and progress.

## FRAMING THE PROBLEM


A free culture supports and protects creators and innovators. It does this directly by granting intellectual property rights. But it does so indirectly by limiting the reach of those rights, to guarantee that follow-on creators and innovators remain as free as possible from the control of the past. A free culture is not a culture without property, just as a free market is not a market in which everything is free. The opposite of a free culture is a ‘permission culture’—a culture in which creators get to create only with the permission of the powerful, or of creators from the past. *Lawrence Lessig, Free Culture, p. xiv*

The opportunity to create and transform becomes weakened in a world in which creation requires permission and creativity must check with a lawyer. *Lawrence Lessig, Free Culture, p. 173*

In the analog world, roles in the supply chain of information—from creation through consumption—were more clearly delineated

---

M. Halbert (Ed.): *Free Culture and the Digital Library Symposium Proceedings*. Atlanta, Georgia: MetaScholar Initiative at Emory University, 2005. pp. 106-140.

 This work is covered by the Creative Commons Attribution-NonCommercial-NoDerivs 2.5 License.

and rights more clearly constrained than they are in the digital world. The capabilities of digital technology challenge the practices and very definitions operative in the analog world. For example, “publication” in the analog world was more likely the result of a peer-review process that added value and assured the quality or authoritativeness of a work than currently occurs with many Web pages. Photocopying an entire book was not only illegal, but discouraged by the tedium, cost, and resulting hundreds of lower quality, loose-leaf pages. In the analog world, exercising the right of first sale was constrained by the physicality of the work, which also constrained the number of simultaneous users of the work. In contrast, copying a digital book occurs automatically upon viewing, and (unless constrained by technology) multiple identical, high quality copies can be created and distributed at the click of a button. Digital technology has changed or challenged the cultural practices of centuries, practices that turned on the physical rendering of intellectual property. It has yielded a paradigm shift in consumption, from purchased ownership to licensed access, and enabled a veritably unlimited number of simultaneous users of the same work at the same time. Digital technology has simplified and reduced the cost of all of the copyrights: reproduction, distribution, public display and performance, and the creation of derivative works. The ease with which these things can be done now has dramatically changed behavior and expectation. To paraphrase Lawrence Lessig, digital technology enables anyone with a computer to participate in building and cultivating culture. People using this power are changing the marketplace and these changes threaten content industries (Lessig 2004, 9). The upshot is vociferous debate among those who prefer to cling to the traditions of the analog world, to replicate and lock them down in the digital world at great expense, and those who prefer to adopt new policies and practices aligned with the capabilities and economics of the new technology.

The debate over the definition and scope of what should constitute an orphan work is discussed in this paper, but for the purpose of these introductory remarks, please allow the general understanding to be a work for which the copyright owner cannot be found—a diabolical problem in a permission culture. Orphan works no doubt existed before computers became popular consumer goods and before the invention of the Web. These technologies, however, have exacerbated the problem by increasing the demand for preservation and access to these works, especially as they are likely to be works of little commercial value but of great historical

value (*i.e.*, a treasure trove of knowledge about who we are, where we came from, and what we've done).

The issues surrounding orphan works are complex. The very topic puts a spotlight on the problems inherent in a permission culture, a culture created and sustained by a labyrinth of laws driven in large part by content industries that have “queered” (to use Lessig’s word) the marketplace and fundamental cultural values. The orphan works problem highlights just how far we have wandered from the free culture of our roots. We live in a world where the two legal options that enable innovations built on the past—permission and fair use—are so fraught with problems, risks, and costs that they discourage rather than encourage preserving and cultivating culture. Acquiring permission is difficult if not impossible and prohibitively expensive in many instances. Relying on fair use is too risky, even for wealthy content industries. “Just at the time digital technology could unleash an extraordinary range of commercial and noncommercial creativity, the law burdens this creativity with insanely complex and vague rules and with the threat of obscenely severe penalties” (Lessig 2004, 19). The opportunities digital technology provides to stir democracy and creativity are obstructed “in a world in which creation requires permission and creativity must check with a lawyer” (Lessig 2004, 173).

The age, physical format, and ephemeral nature of many orphan works threaten their very existence. Our cultural and intellectual heritage in film, music, photographs, art, books, archival documents, etc. can be preserved by converting these works to digital format or at least replicating them in print. The law allows, under certain conditions, the preservation of copyrighted works, but a *preservation* copy is not a *use* copy. It is a locked-up copy, at least for the copyright term of the work. Preservation is not enough if the goals are marketing and cultivating culture. Broad access and use are essential to achieve these ends. Providing online access to orphan works would be a first step, a significant step, but access without a right to use would create a “read only” culture. To truly encourage the creation of new works and enhance or advance scholarship, research, education, and lifelong learning, people must be able to use and “tinker” with these works. The Internet, more specifically the Web, enables for the first time in history a new kind of teaching and learning that respects different styles. Digital technology provides an opportunity for us to overcome limitations inherent in our linear, left-brain, analog



world, and to encourage curiosity and creativity. Requiring permission from copyright owners who cannot be found threatens loss of our heritage and harms our ability to teach, learn, create, and compete in a global marketplace. Those who share this view believe the government should do something to address the problem of orphan works. The opposing camp argues that allowing unauthorized use of copyrighted work would encourage copyright infringement and destroy our economy by eliminating the incentive to create. The government should strengthen protections, punish pirates and other infringers, and ensure that copyright owners are appropriately compensated. Granted, this is a simplistic view of the terrain. As will be seen in this paper, there are positions in between these polar opposites. But let this suffice for an introduction to the problem space.

Concerns about whether current copyright law “imposes inappropriate burdens on users, including subsequent creators,” of orphan works and whether these works “are being needlessly removed from public access and their dissemination inhibited” prompted Senators Orrin Hatch and Patrick Leahy to ask the Register of Copyrights to study the problem and report to the Senate Judiciary Committee by the end of the year (Notice of Inquiry 2005, 3). The result was the U.S. Copyright Office’s Notice of Inquiry regarding orphan works, posted to the Federal Register January 26, 2005. The Notice requested initial comments from interested parties by March 25, and reply comments by May 9, 2005.

The Copyright Office received hundreds of responses to their Notice of Inquiry, each of which shared some experience or expressed some concern about the problem of orphan works or its solution. The responses run the gamut from uninformed (or misinformed) to well informed, from unintelligible rants to thoughtful analyses. They report painful experiences and heartfelt concerns from positions both for and against any action to address the orphan works problem. They reflect naiveté, arrogance, ignorance, ingenuity, acuity, altruism, and self-interest. Taken as a whole, the responses provide a diversity of perspectives on U.S. copyright law from a self-selected cross section of citizens and for-profit and non-profit organizations. Indeed they are a rich read.

The problem of orphan works raises serious questions about the proper balance of private interest and public good inherent in copyright law. The burning questions are whether unauthorized use of copyrighted works, for example, use without the copyright

owner's permission, should be allowed in circumstances and if so, what those circumstances might be. Should we and can we devise a designation of "orphan" works that both protects the rights of copyright owners and enables preservation, access, and use of orphaned cultural artifacts? Understanding the scope of the problem and the harm it causes is critical to finding an appropriate solution.

### **THE AGENDA AND APPROACH OF THIS PAPER**

This paper provides a preliminary analysis and critique of the responses to the Copyright Office Notice of Inquiry, both the initial comments and the reply comments. The analysis includes a high level, quantitative look at all of the comments, and a qualitative, closer look at the objections to allowing unauthorized use of copyrighted works under any circumstances and the proposed solutions that would allow unauthorized use under certain circumstances. To the best of my ability, the analysis conveys an objective look at the responses, providing the pros and cons brought forth by the various respondents for each significant point under debate. Then taking a step back and looking at the debate through my personal lens as a professional librarian, as leader of the National Information Standards Organization (NISO) initiative on rights expression and management in the digital environment, and as a student of technologically driven cultural change, the paper provides my subjective, albeit preliminary, observations and recommendations for cutting a viable path through the maze. The work in this paper is preliminary in the sense that it has been constrained by the time available from the posting of the comments to the due date of this paper.

Unlike many of the reply comments to the Notice of Inquiry that dispensed with "outliers" and addressed what they claimed to be consensus in the initial comments, this paper acknowledges the outliers in the belief that the voices and opinions of a diverse citizenry should be heard in a deliberative democracy. Furthermore, absent a rigorous empirical study, there is no way to know if the outliers in these ad hoc comments represent the views of a significant segment of the population. Those who responded were self-selected, and it would be all too easy to dismiss as an outlier a view that was counter to our own or simply the view of a group less likely to self-select.

## INITIAL OBSERVATIONS

To get a handle on the general contours of the comments, I devised a simple coding scheme. The results of this scheme do not accurately indicate the popularity or weight of positions for or against action to address the orphan works problem. Some comments were submitted by single individuals. Others were submitted by one or more organizations with thousands of members. Furthermore, all comments were not created equal, so to speak. Some are very well informed, others are not. Nevertheless, some way to grapple with the volume of responses was necessary as a starting point. To begin my task of trying to understand how the populace responded to the Notice of Inquiry, I analyzed both the initial comments and the reply comments using the following categories and definitions:

- **Experience:** The comment reported first- or second-hand encounters with problems related to orphan works.
- **No:** The comment explicitly stated an objection to any action that would allow use of copyrighted works without the copyright owner's permission.
- **Yes:** The comment explicitly or implicitly stated approval of or requested action to address the problem of orphan works. Comments that described experience as a matter of fact, without requesting help or indicating harm caused by orphan work, were not coded as "Yes." I took this conservative approach as a precautionary measure to prevent my personal position from coloring my coding.
- **NIMBY ("Not In My Back Yard"):** The comment explicitly stated approval of or requested action to address the problem of orphan works, but requested that their content be exempt from any orphan works designation because there is no or only a minor problem in their domain or because there are other compelling reasons that warrant their exemption.
- **Solution:** The comment proposed some action that could help to alleviate the problem of orphan works.

A rare few comments received neither a "Yes" nor a "No" code, for example, the comment that simply asked who owned the copyright on a vacation photograph taken by a random passerby. A few comments were coded "No" and "Solution" because they objected to allowing use of orphan works without permission, but proposed some action to eliminate or alleviate the problem.

Table 1 shows the results of this analysis. Few respondents submitted both an initial and a reply comment, and few objected to action that would allow unauthorized use of copyrighted work under any conditions. Very few approved action to address the problem everywhere but in their domain. The overwhelming majority approve of allowing unauthorized use in some circumstances. Many respondents shared personal experience with orphan works and proposed something about the solution to the problem. Not surprisingly, the reply comments focused more on the solution to the problem than the experience of the problem.

	No	Yes	NIMBY	Experience	Solution
Initial Comments	8%	79%	1%	52%	54%
Reply Comments	5%	86%	3%	33%	62%

Table 1. Rudimentary content analysis of initial comments and reply comments.

The many comments that reported experience with orphan works reveal the broad scope of the problem in terms of users, uses, and media. From personal to professional use of photographs, graphic art, software, film, books, radio and television broadcasts, any media you can think of, works for which the copyright owner cannot be found have created problems for academic researchers, teachers, students, journalists, documentary filmmakers, radio producers, photo shops, authors, publishers, record producers, hobbyists, scientists, engineers, libraries, archives, and museums. Though a few respondents claimed that there is no problem or only a minimal problem in their area and therefore their domain should be exempt from any legislation that would allow unauthorized use, experiences reported and in some cases data provided by other respondents belie these claims. All of the “NIMBY” respondents, though seeking exemption from an orphan works solution for their content, proposed solutions for the orphan works problem in other domains.

To enable me to target “Solution” responses that warranted focused study, I also analyzed the initial and reply comments using the following categories and definitions:

- Simple solution: The comment proposed one to three actions that could help to alleviate the problem of orphan works. I

also noted whether the recommended action was to remove copyright protection from orphan works (make them public domain) immediately or upon meeting certain conditions.

- Detailed solution: The comment proposed more than three actions or solution criteria that could help to alleviate the problem of orphan works.
- Solution analysis: The comment articulated advantages or disadvantages of different definitions of orphan works or approaches to the problem.

My operating assumption was that comments containing “Detailed solutions” and “Solution analysis” were likely to contain the points of merit in “Simple solutions.” Note that these codes, like the previous ones, do not accurately indicate the popularity or weight of positions for or against action to address the orphan works problem. They simply provide a slightly more detailed view of the general contours of the comments.

The results of this analysis are shown in Table 2. Overall, most of the solution proposals were “Simple,” though the percentage of “Detailed solutions” and comments containing “Solution analysis” increased in the reply comments. Among the initial comments, over a third recommended that orphan works become public domain immediately or conditionally; significantly fewer reply comments proposed the public domain as the solution. In conducting this analysis, I observed that proposals for the public domain came from individuals, not organizations, and were typically quite brief. Responses from organizations were longer, more detailed, and more analytic, which is not to say that no individuals proposed detailed solutions or provided analyses. More importantly, I observed that reply comments that made claims about consensus in the initial comments simply ignored all the proposals that orphan works become public domain. Granted, the solution adopted for the problem of orphan works is not likely to be the public domain, but it is misleading at best and at worst irresponsible to not even acknowledge that more comments proposed the public domain solution than any other solution. The many public domain proposals reveal something of interest if not significance about our citizenry, and overlooking or dismissing their comments entirely reveals something important about those who claim to build on consensus.

	Simple solutions			Detailed solutions	Solution analysis
	Public Domain	Conditional public domain	Other		
Initial comments	26%	10%	42%	22%	19%
Reply comments	4.5%	4.5%	54%	37%	38%

Table 2. Analysis of solution proposals.

The remainder of this paper explores the responses to the Notice of Inquiry in detail, beginning with respondents' answers to the Copyright Office's questions about the definition of an orphan work and the scope of the designation as these frame the objections and approaches to allowing unauthorized use of copyrighted works. The analysis focuses on the initial and reply comments coded as "No" and those coded as "Detailed solution" with "Solution analysis." The interaction of perceptions, priorities, assessments of value, awareness of relevant international treaties, and concerns about abuse, bureaucracy, control, risk, and cost, along with the respondent's presumptions about the purpose of copyright protection and allowable unauthorized use color the responses and make it difficult to present the debate in a linear fashion. The same arguments are brought forth again and again to address different issues and are sometimes used to make different points. In some cases, the definition of an orphan work shapes the proposed solution. In other cases, criteria for an acceptable solution shape the definition of an orphan work. For example, those primarily concerned about a solution that will scale to meet the needs of libraries, archives, and other cultural heritage institutions take a significantly different approach to defining an orphan work from those focused on individual use. I will do my best to walk you through the quagmire. Following an initial exploration of the defining criteria and scope, our path in broad strokes covers:

- Objections to allowing unauthorized use
- Copyright registries to avoid or alleviate the problem
- Categorical approaches to solving the problem
  - Default licensing
  - Safe harbor exemptions
  - Registry of orphan works
- Case-by-case approaches to solving the problem

- Public domain
- Compulsory licensing
- Reasonable effort accommodation
- My recommendations and closing observations

Though the comments are posted on the Copyright Office website for public review, in the interest of objectivity and not biasing or influencing the reader's response, throughout this paper, the person or organization that submitted the comment is not named and—with rare exception—the frequency or popularity of the points made is not indicated. Instead, significant issues raised in the comments are briefly articulated and the pros and cons presented. Because the pros and cons often come from respondents with different priorities and perspectives, they do not always present a coherent whole.

## **DEFINING ORPHAN WORKS**

### **Copyright Owner Cannot Be Found**

Issue: Should an orphan work be defined as a work for which the copyright owner cannot be found?

Yes: “Unlocatable” copyright owners are the root of the orphan works problem and therefore should be the (or at least a) criteria for delineating what constitutes an orphan work. An agreed upon procedure for attempting to locate copyright owners is needed to substantiate the claim that a copyright owner is “unlocatable.”

No: Copyright owners might be locatable by more skilled or diligent seekers. Copyright owners should not have their copyrights infringed due to the laziness or incompetence of potential users. Furthermore, copyright owners might choose to be unlocatable. They should not have their copyrights infringed or be assumed to have neglected or abandoned their work because of their choice.

### **Copyright Owner Cannot Be Identified**

Issue: Should the definition of an orphan work include works for which the copyright owner cannot be identified?

Yes: The copyright owner of many copyrighted works is unknown because the work has no attribution. Clearly a potential user cannot locate an unknown entity.

No: Unscrupulous people often remove the attribution to plagiarize or steal the work.

### **Copyright Owner Does Not Respond**

Issue: Sometimes diligent efforts to identify and locate the copyright owner yield no response. Should some number of successful contacts (*e.g.*, three successfully delivered letters requesting copyright permission) be criteria for designating an orphan work?

Yes: If the copyright owner is not sufficiently interested in his or her work to respond to a request for permission to use the work, then the work has been orphaned. In the interest of the public good, “no response” should be treated as “permission granted.”

No: The copyright owner could choose not to respond or be unavailable or unable to respond when contacted due to illness, vacation, sabbatical, or other circumstances. Copyright owners should not have their copyrights infringed or be assumed to have neglected or abandoned their work because of their choices or circumstances. “The burden of having to be constantly available to those who may or may not really want to find the copyright holder is too much to put on copyright holders.” In the interest of copyright protection, “no response” should be treated as “permission denied.”

Discussion: Frequently the person or organization contacted is only the presumed copyright owner. Many comments detailed experiences where the presumed copyright owner responded only to say that the presumption was wrong. Sometimes multiple presumed copyright owners of the same work, contacted sequentially, responded that they did not own the copyright, sometimes referencing one another in a fruitless loop. According to legal counsel, no response from a presumed copyright owner could be “probative of the inability to find the owner.” The problem is that in many cases there is no way to know definitively who owns the copyright because the records of the Copyright Office are incomplete, out of date, and inaccessible. Even if more often than not it would be reasonable to claim that a request is being sent to the presumed copyright owner, there is no way to know for sure whether the presumed copyright owner is indeed the copyright owner.

### **Age**

Issue: Should the age of a work be considered in defining orphan work?



Yes: The age of a work is critical in categorical approaches to the orphan works problem. A categorical approach is necessary to provide certainty in designating orphan works. Without certainty, ambiguities will lead to self-censorship and gatekeeping and the problem will not be solved. (Categorical approaches are discussed later in this paper.)

No: The age of works likely to be orphans is often unknown. Excluding these works from acquiring orphan designation would prevent their preservation, access and use, and likely break up archival collections. Hinging the solution to the orphan works problem on the age of the work will introduce a new level of uncertainty. Any uncertainty in the definition will reduce the value of the solution and its impact on the problem. Furthermore, any setting of a minimum age to define an orphan work would necessarily be arbitrary. And if the designation of orphan works is restricted to older material, as is likely to be the case if age becomes part of the definition, more recent but ephemeral (endangered) materials on the Web will not be protected and preserved by the designation. The appropriate solution to the problem will apply to all works, regardless of age. The age of a work would be irrelevant if a “reasonable effort” accommodation or compulsory licensing were adopted to solve the orphan works problem. (The reasonable effort accommodation and compulsory licensing are discussed later in this paper.)

### **Publication Status**

Issue: Should the publication status of a work be a consideration in defining orphan work?

Yes: Copyright holders have the moral right of first publication, privacy rights, and the non-economic right to withhold a work from the marketplace, all of which would be thwarted by allowing unpublished works to be designated orphans. A compulsory licensing approach to the orphan works problem could not be applied to unpublished work.

No: The publication status of works is often unknown. Excluding these works or works known not to have been published from acquiring orphan designation would prevent preservation and use of valuable cultural and intellectual heritage material, access to which is in the public interest. Hinging the solution to the orphan works problem on publication status will introduce a new level of uncertainty. Any uncertainty in the definition will reduce the value of the solution and its impact on the problem. The 1992

amendment to Title 17 §107 expanded fair use to unpublished works (though courts and cautious publishers still discourage such use), so likewise an orphan designation should be applicable to unpublished works. The right of the copyright owner to first publication focuses on commercial exploitation, which likely does not apply to the bulk of the heritage materials that will be salvaged under an orphan works regime. Furthermore, international treaties do not provide for or require a moral right of first publication. The purpose of copyright law is not to protect privacy. Privacy laws will protect privacy concerns despite any orphan works regime that would otherwise enable public access to these materials. The publication status of a work would be irrelevant if a “reasonable effort” accommodation were adopted to solve the orphan works problem. (The reasonable effort accommodation is discussed later in this paper.)

### **Print Status**

Issue: Should the print status of a work be considered in defining orphan work? That is, if a work that had been commercially exploited (printed) is no longer commercially exploited (out of print), should this be a significant factor in designating an orphan work?

Yes: The primary purpose of copyright protection is the commercial exploitation of creative work. Works that are no longer being commercially exploited are likely to have been abandoned by the copyright owner. In a world where authors typically transfer exclusive rights to the publisher and the publisher ceases to disseminate the work, the rights may or may not revert to the author. In many cases authors would like to distribute their work, but either do not have the rights or the resources to do so. As libraries weed their collections, fewer and fewer copies of out-of-print books remain. As these books become worn and brittle, fewer and fewer copies circulate on interlibrary loan. Out-of-print materials are an endangered species in need of preservation. Access to these materials is in the public interest. Out-of-print materials are an easily identified corpus of works.

No: Works that are out of print now could be exploited later by the copyright owner. Designating these works as orphans and providing unauthorized public access to them could damage the future market for these works, and thus run afoul of international treaties by “unreasonably prejudicing the legitimate interests” of the copyright owner. Furthermore, the notion of print status is

becoming meaningless given print-on-demand services available in the digital environment.

### **Type of Work**

Issue: Should an orphan works designation be applicable to all types of copyrighted work?

Yes: All types of copyrighted work can be orphaned. Providing special exemptions or treatment for certain classes of works is inappropriate and would extend the existence of the orphan works problem associated with these classes. Special treatment for certain classes of works is unjustified and discriminatory.

No: There is no serious orphan works problem in some areas and in some cases licensing and collection agencies already exist that can handle the minor problems.

### **Application and Duration**

Issue: Should the orphan designation apply to the work that meets the defining criteria?

Yes: To be meaningful, the designation must apply to the work itself. Considerations of the identity and location of the copyright owner only make sense when applied to the work.

No: The designation should not apply to the work itself, but to a particular use made of the work, with each user having an independent duty to instantiate the definition prior to each use of a work.

Issue: Should an orphan designation endure in perpetuity?

Yes: Orphan designation should be an irrevocable status lasting for the remainder of the copyright term.

No: Orphan designation should endure only until the copyright owner comes forward to claim his or her work

### **SCOPE OF USERS AND USES OF ORPHAN WORKS**

Issue: Should the solution to the orphan works problem apply to all users and uses of designated orphan works?

Yes: Limiting users or uses would unnecessarily complicate the situation and fail to address the full scope and implications of the problem of orphan works. It can be difficult to distinguish commercial from non-commercial use. Any uncertainty in the definition will reduce the value of the solution and its impact on

the problem. The solution must apply to all uses on the basis of the assumption that permission for all uses could have been negotiated (at some price) if the copyright owner had been found. Providing special exemptions or treatment for certain users or uses is inappropriate. The solution should be uniform and equitable. Special treatment for certain user groups is unjustified and discriminatory. The solution should apply to all users and uses with legitimate access to a non-infringing copy of the orphan work.

No: Different users or uses should be treated differently. For example, preservation is one thing, online dissemination is another. For-profit uses should be prohibited because they could damage the future market for these works and therefore run afoul of international treaties by “unreasonably prejudicing the legitimate interests” of the copyright owner. The solution should take into consideration that there is no or only a minimal orphan works problem in some areas. Furthermore, not all rights should be granted for all media. In some cases the right to use orphan works should be limited to the rights of reproduction and distribution. (Proposals for specific exemptions are discussed later in this paper.)

### **OBJECTIONS TO ALLOWING UNAUTHORIZED USE**

A small percentage of the responses to the Notice of Inquiry objected to any action that would allow unauthorized use of copyrighted works. Most of the objections came from photographers and graphic artists. Reasons for objecting included:

- The Internet makes it easy to locate copyright owners. There is no orphaned art, only unscrupulous people who deliberately obscure signatures to pretend the creator is unknown. If publishers and the Copyright Clearance Center “took illustrators’ copyrights seriously, there would be no orphaned art.”
- The aim of copyright is to protect copyright owners. That protection should not depend on whether the copyright owner is locatable, available, or responsive. Allowing unauthorized use of copyrighted works would infringe the exclusive copyrights, including moral rights, of creators, take away their control of their work, and enable their work to be used to support organizations or causes to which they are opposed.
- Providing the incentive to create is more important than making life easier for users who do not have the time, skill, or resources to locate the copyright owner. Creativity and

innovation are the driving forces of a thriving economy. An orphan works regime would undermine the economy and threaten the livelihood of professional photographers and graphic illustrators.

- “The Internet has created a culture of appropriation, and immediate global access to artistic works has facilitated piracy, unintentional infringement and plagiarism.” Allowing unauthorized use of copyrighted works would encourage copyright infringement and favor corporate interests over individual creators. If potential users cannot locate copyright owners, they should use art created by copyright owners they can find or commission and pay for new art.
- Creativity is not chilled, free speech is not restricted, and culture is not endangered or impoverished by protecting orphan works. “The removal of copyright protection for orphaned work would reinforce the agenda of the ‘free culture’ movement to subvert existing copyright protection for other work.” The free culture movement is “using subtle language and deception to trick the masses.” Lawrence Lessig has convinced artists to give “their artwork away for free without them realizing the implications.”

The plea in these objections is for laws to be strengthened to further protect the rights of copyright owners.

### **COPYRIGHT REGISTRIES TO AVOID OR ALLEVIATE THE PROBLEM**

Some of those who objected to any unauthorized use of copyrighted work and some of those who approved of unauthorized use under certain conditions recommended a registry for current copyright owners to maintain up-to-date contact information. More aggressive respondents recommended not only the reinstatement of a copyright owner registry, but legislation to shorten the copyright term and to reinstate the renewal requirement. Some respondents blamed the orphan works problem on current copyright laws and practices.

Issue: Should copyright registration be required?

Yes: Requiring the registration of current copyright ownership and contact information would enable potential users to locate copyright owners and negotiate permission to use their work. Works not registered could with certainty be designated orphans. If registration is not a prerequisite for acquiring or maintaining

copyright protection, it does not run afoul of international treaty obligations prohibiting formalities as a condition for “the enjoyment and exercise of copyright.”

No: Copyright owners do not always know what copyrighted works they own. Requiring registration or renewal as a prerequisite to acquire or maintain full copyright protection would breach international treaty obligations prohibiting formalities as a condition for “the enjoyment and exercise of copyright.” Requiring registration would “perversely encourage publishers to abandon works” because of the burden entailed in registration. Reinstating registration requirements “would lead creators to manipulate the nationality of their work to sidestep U.S. formalities.” Periodic registration (renewal) would “diminish the market value of works.”

Issue: Should copyright registration be voluntary?

Yes: Voluntary registration, without which the copyright owner nevertheless acquires and maintains full copyright protection, does not breach international treaty obligations, but provides users with an indication of works to be excluded from orphan designation. Voluntary registration would signal a copyright owner’s intent to enforce copyrights on works not officially registered with the Copyright Office. Consulting such a voluntary registry would be a necessary, but not sufficient step in reasonable efforts to locate the copyright owner. Users could contribute information about their efforts to find copyright owners. Without filing formal proof of ownership or transfer documents such a registry would provide users with “precisely the information” on how to find rights owners and increase both the owner’s ability to assert his or her rights and the user’s ability to evaluate the risk of using a presumed orphan work.

No: Copyright owners do not always know what copyrighted works they own. If proof of ownership or transfer documents were not required, what prevents fraudulent claims of ownership? If registration were optional, how many copyright owners would register? Optional registration would only confuse copyright owners. “Freelancers and publishers alike often assume that registration is unnecessary because copyright is automatic. Frankly, when the complicated system of additional protections, deadlines, statutory damages, and so on, that result from registration are explained to them, their eyes go blank. The present system is one that only a lawyer could love. Duplicating it with

yet another complicated system would compound the problems that the system already fails to remedy.”

## **CATEGORICAL APPROACHES**

A categorical approach is an approach that provides certainty in designating orphan works and thus eliminates the risk in using orphan works. The rationale for supporting a categorical approach to solving the orphan works problem is that ambiguous definitions or criteria of “reasonableness” will go the same route as the “fair use” defense to copyright infringement: self-censorship by creators and gatekeeping by publishers. Ambiguity will yield to requiring permission because the risk of liability is too great.

### **Default Licenses**

The default license approach requires registration and renewal of published work for which copyright owners wish to retain the full copyright term and remedies for infringement provided by current copyright law. With the exception of software, online registration would be required within a 25-year period of publication. Software would be required to be registered within five years of publication. Renewal would be required 50 years into the copyright term. To facilitate use of registered works, copyright owners would be required to keep contact information up to date. Ideally the registry would include links to terms and conditions for licensing use of registered works.

Failure to register or renew would not remove copyright protection, but rather signal that the work is orphaned. A search of the registry would be sufficient to determine whether a work is an orphan. Orphan works could be used without permission for a nominal fee under a default license. No injunctions against use would be available under the default license. Copyright owners who do not register their work but later discover infringing uses may self-identify and claim the fees paid for use of their work.

A somewhat similar system is proposed for unpublished works. Registration is required within three years of the natural author’s death (if unknown to be set at 75 years after the creation of the work) or within ten years of the creation of unpublished works by corporate authors. Registration would signal that copyright owners wish to retain the full copyright term and remedies for infringement provided by current copyright law. Failure to register would not remove copyright protection, but signal that the work is

orphaned. Use of unregistered, unpublished work would be contingent on the user:

- Confirming the date of the natural author's death (or the date of creation of the work if the death date is unknown) or the date of creation of the work of corporate authors.
- Confirming the expiration of the appropriate registration period (three years for natural authors, ten years for corporate authors).
- Posting a notice of intent to use for a period of six months in a centrally administered Web accessible database. The notice of intent provides copyright holders of unpublished work with an opportunity to reclaim their work prior to its use.

Pro: Default licensing provides an efficient, effective way to balance private interest and public good. It provides certainty for copyright owners and users. It avoids the ambiguity and unpredictability of the "reasonable effort" approach and the accompanying chilling effects of the threat of litigation. It gives users a way to know whether a work has been orphaned and when, and gives copyright owners the power to signal that they have not abandoned their works. It also gives copyright owners the option, at an appropriate point in the life of a work, to decide whether the work warrants the high-cost route of infringement damages, injunctions and customized licenses under current copyright law, or whether it is better served through a lower-cost system of default licensing. Since failure to register or renew a work does not affect the term of copyright protection or forfeit any rights, the requirement does not run afoul of international treaty obligations prohibiting the institution of formalities or interference with the enjoyment and exercise of copyright. Default licensing would promote the enjoyment and exercise of copyright by creating a lower-cost market for works unable to be marketed in the high-cost environment of current copyright law. The U.S. Copyright Office's data on registration and renewals, when these were required by copyright law, indicates that most works were abandoned within 25 years of publication, that most material of continuing commercial value was renewed, and that the un-renewed material, while of little if any commercial value, was of great value to scholars and other specialists. These findings suggest that most copyright owners would be pleased with a low-cost system of default licensing (Notice of Inquiry 2005)—the upshot being that registration would be required of only those copyright owners who wish to pursue infringement damages,



injunctions and customized licenses some designated number of years after publication or creation.

Con: The claim that copyright owners who fail to register are maintaining control and asserting their choice to have their work covered under a default licensing is problematic and likely a trap for the unwary. The default license approach requires knowing the author and date of publication or creation. These will probably be known for new works or relatively recent works, but not necessarily for older works.

### **Safe Harbor Exemptions**

Exemption approaches would legislate safe harbors or exclude certain types of works or uses from the orphan works solution.

The safe harbor exemption for non-profit libraries, archives, and educational institutions would expand U.S.C. Title 17 §108 to enable the reproduction and dissemination of orphaned works. Driven by concerns about risk and scale, these institutions need a solution to the orphan works problem that is both low-cost and low-risk. Definitions of an orphan work that are ambiguous or cannot be applied cost-effectively—ideally by a computer—to identify large numbers of works will not solve the orphan works problem for this constituency. Cultural heritage institutions with missions to preserve and provide access to materials in the public interest have limited resources. Exorbitant transactional costs and the risks of uncertainty must be eliminated or greatly reduced.

The most detailed proposal for an expanded safe harbor recommended that it enable non-profit libraries, archives, and educational institutions to make and provide access to physical or digital copies of published written material for educational and scholarly purposes if the work was first published at least 30 years ago and is currently out of print and if the copyright owner has not registered the work to signal that it is to be excluded from orphan designation. Copyright owners that do not opt out of the exemption by registering their work could come forward later and require the institution to stop providing access to their work.

Pro: In the public interest, education and cultural heritage institutions should have a special exemption to encourage preservation, access and use of orphaned works. A legislated safe harbor for these institutions, for these works (published written materials that copyright owners have not signaled are to be excluded from orphan designation), and for these purposes

(reproduction and distribution for scholarly and educational use) would create an un rebuttable presumption of orphan work status and obviate the need for court intervention based on the nature of use and type of work. This approach is workable now. It avoids the unpredictability, costs and risks of the “reasonable effort” approach, and clearly identifies works that have not been abandoned. The registry requirement to avoid orphan works designation does not run afoul of international treaty obligations prohibiting the institution of formalities or interference with the enjoyment and exercise of copyright. An expanded exemption for non-profit libraries, archives, and educational institutions provides an efficient, effective way to support education and scholarship.

Con: The scope of the solution should address the scope of the problem. Limiting the solution to the orphan works problem to selected users, works, and purposes will not address the full scope of the problem. An appropriate solution to the problem will facilitate all creative users and uses. Special exemptions for educational and cultural heritage institutions should be prohibited because they could damage the future market for these works, which would breach international treaties by unreasonably prejudicing the legitimate interests of copyright owners. For example, complying with a take-down notice for an online copy of a work could be too late; multiple copies could have been made and distributed prior to the take down. Providing special exemptions or treatment for certain classes of works or user groups is inappropriate, unjustified and discriminatory. The solution to the problem should be uniform and equitable.

### **Registry of Orphan Works**

A few comments proposed a registry of works designated or certified as orphans. Potential users could presumably consult this registry to find works available for unauthorized use. Who would identify and register these works or maintain this registry is unclear. The only detailed approach to the orphan works problem that explores anything that even resembles such a registry is the “reasonable effort” accommodation described later in this paper.

## **CASE-BY-CASE APPROACHES**

### **Public Domain**

Comments that proposed the public domain as a solution to the orphan works problem typically did not indicate how an orphan work was to be identified, though it appeared to be a matter of

“unlocatable” copyright owners. The comments did not elaborate how this solution would be implemented. No comments that analyzed proposed solutions explored the pros and cons of the public domain solution, perhaps because the advantages and disadvantages are clear. Removing copyright protection from orphan works would obviously enable use of many works without the hassle of acquiring permission. However, removing copyright protection from orphan works would breach international treaties and take away copyright owners’ control of their work without remedy. The silence of other commentators speaks volumes about the lack of viability of the public domain solution.

### **Compulsory Licensing**

**Issue:** The Copyright Office Notice of Inquiry mentions the Canadian approach to orphan works, which is compulsory licensing. Potential users apply to the government for a license and pay a fee, which is reserved for the copyright owner who might later resurface. For each application, the government or other administrative body conducts an inquiry to determine whether efforts to locate the copyright owner were sufficiently reasonable and to determine the appropriate licensing fee for a particular use of a particular work. Though not designed for orphaned works, U.S. copyright law currently legislates compulsory licensing of recorded music through private agencies.

**Pro:** Compulsory licensing has worked well for the music industry in the United States and seems to work for well for orphan works in other countries.

**Con:** A compulsory licensing system modeled on Canadian law would be bureaucratic, inefficient, slow, expensive, “draconian,” and “inimical to the interest of both potential users and copyright owners.” There are no standards for what constitutes a reasonable fee for different uses of different media. The outcome for users would be uncertain and the licensing fees could be inadequate from the copyright owners’ perspective. Compulsory licensing could not apply to unpublished materials. Ambiguity regarding whether a use was fair or otherwise exempt would lead users to apply for a compulsory license for fear of liability under the licensing regime. Requiring navigation of a case-by-case adjudicatory system and up-front payment of licensing fees will seriously discourage if not prohibit use of orphan works. The cost would deter museums, archives, libraries and cultural heritage institutions from preserving and providing access to their materials. Requiring users to pay for

permissions that in many cases would be granted for free had the copyright owner been located is unfair. The money and time required of users is out of proportion to the scope of the problem given that many potential uses are personal or educational and non-commercial. Canada's experience with compulsory licensing of orphan works appears to be ineffective. The government has no right to claim ownership of copyrighted works and license rights to use them. Compulsory licensing could breach international treaties.

### **Reasonable Effort Accommodation**

This approach recommends amending U.S. Title 17 to include a "reasonable effort" defense and predictable limits or remedies for infringement. The accommodation hinges on the definition of an orphaned work as one for which the copyright owner cannot be located and clearly places the burden on the user, at least initially. Implementing a reasonable effort accommodation would require the development of guidelines and boundaries for what potential users must do to qualify for the accommodation and agreement on acceptable remedies should the copyright owner later come forward to claim infringement.

**Pro:** The reasonable effort accommodation would reward users who were diligent in their efforts to locate and acquire permission from copyright owners by allowing their unauthorized use of copyrighted works and limiting their liability for infringement should the copyright owner later come forward. Copyright owners would retain control of their work and have recourse in cases of infringement.

**Con:** The reasonable effort accommodation disadvantages both copyright owners and potential users. It disadvantages copyright owners by providing no way for them to signal that their work is not orphaned. It disadvantages users by providing only a defense in litigation; it aims to limit, but not eliminate the user's liability and therefore necessarily retains some uncertainty. Establishing objective criteria for reasonableness is impossible. Disputes and litigation are bound to happen. If potential users are left with uncertainty as to whether their effort meets the criteria of reasonableness and the available remedies are onerous, the reasonable effort accommodation will suffer from the same self-censorship and gatekeeping practices that currently constrain exercise of fair use rights. The solution will not scale to accommodate the needs of libraries, archives, and museums to

cost-effectively identify large numbers of orphan works. In short, it will not solve the orphan works problem. Allowing unauthorized use based on “reasonable efforts” will only encourage laziness and offer an excuse for infringement.

The reasonable effort accommodation engages more thorny issues that the other proposed solutions, including the level of specificity and flexibility that can be provided and will be acceptable, whether users should document their reasonable efforts or use of a work under the orphan accommodation, whether and where users should post a notice of intent and how this would operate, whether subsequent users can “piggyback” on the reasonable efforts of prior users, and whether or at what point in disputes the burden shifts to the claimant who later comes forward to declare copyright ownership and infringement of the designated orphan work.

### **Specificity and Flexibility**

Issue: What constitutes “reasonable effort?”

Some argue that “good faith” should be a defining characteristic, along with use of appropriate search tools and consideration of circumstances. Others believe the type of work, the nature of the proposed use, and the time, resources, and experience (expertise) of the user are essential criteria for assessing whether an effort is “reasonable.” For example, in the interest of preservation and access, perhaps there should be a lower threshold for ephemeral online works that would otherwise be lost (given the shelf-life of Web pages) or for personal uses like the reproduction of family photographs. Some argue that the user’s incompetence (lack of know-how or awareness of resources for locating copyright owners) is irrelevant and not an adequate defense for copyright infringement. Some prefer that a standard or test for “objective reasonableness” be established based on types of use and classes of works. Others want clearly delineated reasonable-effort practices that if followed would be recognized as reasonable per se, which would eliminate the possibility of litigation.

There does appear to be agreement on two points. First, Congress cannot prescribe safe harbor standards for what constitutes reasonable effort because what constitutes a reasonable search will vary with different media and over time. Second, guidelines and best practices developed by professional organizations could assist users in their search. These aids would need to be maintained and updated as new technologies and resources become available over time.

## **Documentation**

Issue: Should users be required to document their efforts to locate copyright owners and to retain the documentation as evidence of their claim of reasonableness in case the copyright owner later comes forward to claim infringement?

Common sense would support documenting the search for a copyright owner if unauthorized use of orphan works is a defense for infringement based on a reasonable effort to acquire permission. Key issues are whether such documentation should be required, and if so, how long it must be retained and the user's liability if it is lost. Some argue that loss of documentation should not subject the user to full infringement liability. Other issues are whether the documentation should be filed (registered) or certified. These issues are explored further in the discussion of registries below.

## **Notice of Intent**

Issue: Should users be required to post a "notice of intent" to use a work prior to using the work for which they could not locate the copyright owner through a reasonable effort?

Pro: Requiring the posting of a "notice of intent" to use a work is an essential and indispensable step in due diligence. Any inconvenience to the user is counter-balanced by bringing potential users and rights holders together and preventing works from being inappropriately designated orphans. A notice of intent would be a sign of good faith. A voluntary notice of intent would operate from the "false premise of symmetry between the situation of users and rights holders."

Con: Posting a notice of intent to use would be problematic in competitive contexts. Requiring a notice of intent prior to use would make planning difficult, delay preservation of and access to valuable resources, create the potential for illegitimate owners to corrupt the system, and add a step unlikely to connect potential users and rights owners. If copyright owners are required to check for notices of intent, does this run afoul of international treaties that prohibit formalities for copyright owners? If they are not required to check for notices of intent, how likely is it that they will check and what purpose would posting the notice serve other than to burden the user? Requiring formalities for potential users or copyright owners would be unfair and burdensome.

Issue: Where should notices of intent be posted and for how long prior to use of the work?

Those who support a notice of intent do not agree on how long a potential user must wait after posting the notice of intent before using the work. Suggestions include 90 days, six months, two years, and conducting a study to determine the appropriate time. There is also disagreement on where such a notice should be posted. Should a central database be created? If so, who should create and maintain it? How would it be funded? If notices should be advertised in major newspapers, as some suggest, in which newspapers and at what cost? High costs and long waiting periods will discourage preservation and use of orphan works.

### **Registries of Users and Uses of Orphan Works**

Respondents in favor of a “reasonable effort” accommodation proposed different registries that resemble a notice of intent with the exception that no waiting period would be required. Proposals included:

- Users are required to file voluntary sworn statements containing their search details with the Copyright office and pay a processing fee. The Copyright Office certifies the statement, but does not issue a license. The sworn statement provides prima facie evidence of reasonable effort. The burden is then on the copyright owner—within the statute of limitations—to prove either that the user did not do the search described or that the search was not reasonable under the circumstances.
- Users are required to register their use of orphan works with a designated licensing agency that would provide certification of the use via a limited license and renewal process. This proposal somewhat resembles compulsory licensing, but with the significant difference that use is allowed under an accommodation that leaves users at risk of the remedies for copyright infringement.
- Users can voluntarily register their uses of orphan works, presumably to assist subsequent users—which leads to the issue of piggybacking.

### **Piggybacking on Prior User Efforts**

Issue: Should potential users who want to use a work that a prior user’s effort designated as an orphan be able to rely on the prior

user's "reasonable effort" and orphan designation? The point is moot if in the interim the copyright owner came forward to claim infringement of the work, in which case the orphan designation no longer applies—though how the new user is to know that the copyright owner came forward is another question. The point is also moot if the orphan works designation applies to a specific use of a work. If, however, the designation applies to the work itself and the copyright owner has not come forward, the answer is open to debate.

On the one hand, in the case where potential users want to use the same work and are often working against deadlines, it would seem unreasonable to require redundant efforts. On the other hand, piggybacking on prior user efforts presents many problems. For example, what if the prior user's efforts did not meet the criteria for reasonableness? What if someone later comes forward to claim copyright infringement? What if new technologies or resources in the interim have enabled locating the copyright owner? Having each new user be responsible for the reasonableness of his or her effort avoids these issues. Consulting prior user efforts, if available, might be a reasonable start, but it seems reasonable to have each new user decide whether this suffices or whether repeating these steps or taking additional steps is warranted.

### **Liability of the User of an Orphan Work**

Issue: The reasonable effort accommodation requires remedies to handle cases when the copyright owner comes forward to claim infringement. Limiting remedies will enable potential users to manage the risk involved in using orphan works. Respondents who proposed the "reasonable effort" approach agree that remedies should be limited in cases where users have indeed exerted a reasonable effort in good faith to locate the copyright owner. Such users "qualify" for the reasonable effort accommodation. In cases where the effort was fraudulent or unreasonable, these users do not qualify for the accommodation and the full extent of the law should apply. Respondents disagree on whether the burden to prove reasonableness (qualification) remains the affirmative responsibility of the user, or whether the burden shifts to the copyright owner to prove unreasonableness (disqualification).

Pro: The remedies will limit only what copyright owners can do or recover in cases of infringement, not their exclusive copyrights. Limiting the remedies available to copyright owners is consistent



with international treaties and gives copyright owners some recourse in cases of infringement.

Con: While it is likely that legislative determinations of remedies for infringement do not impinge on the copyright owner's "exercise and enjoyment of copyright," it is less clear whether remedies that create de facto compulsory licenses for unauthorized use of orphan works would be compatible with international treaties. Who decides what users qualify and what remedies should be available?

The range of proposals for limited remedies runs the gamut from no forfeiture of any rights or remedies to complete immunity. Between these two end points, some would eliminate all statutory damages, criminal damages, and attorney fees. Others argue for a "reasonable royalty." Others argue for a cap on monetary damages. Still others for injunctive relief or a portion of the profits from any commercial use. Those who propose complete immunity typically would prohibit commercial use of orphan works. Detailed analyses in the responses focused on the options of capping monetary damages, requiring payment of a reasonable royalty, and depositing money in an escrow account.

### **Capping monetary damages**

Pro: Setting a cap is not price fixing because the payment would be within the range up to the cap. Having a set fee eliminates the problem of having to value different uses of different types of works on a case-by-case basis.

Con: Certain uses of certain types of works have greater market value than others, so setting a cap on all types of uses and works would be unfair. Furthermore, what cap would be appropriate, based on what criteria? On the one hand, the cap must be low enough not to discourage use. On the other hand, the cap must be high enough to provide an incentive for users to really try to locate the copyright owner and to make it worthwhile for a copyright owner to pursue cases of copyright infringement. A cap of a few hundred dollars would be so low that copyright owners would not likely pursue judicial redress for copyright infringement, which in turn would encourage users to exert less than reasonable efforts to locate the copyright owner or to refuse to pay the cap.

### **Requiring a reasonable royalty**

Pro: Requiring a reasonable royalty most closely resembles the market dynamic that would have been operative had the copyright

owner been located in the first place. Reasonable royalty fees can be predicted “within a reasonable range set by actual market practices.” If the user and copyright owner cannot agree on a reasonable royalty, the fee could be set by the court. Reasonable royalties should not discourage use because it is unlikely that the copyright owner of a truly orphaned work will come forward or file a law suit. The uncertainty of users should not be minimized at the expense of copyright owners’ rights. Use of an orphaned work could effectively preclude copyright owners from making profitable use of their work in the future.

Con: How is a user to know what a reasonable royalty might be for different uses of different media? Uncertainty will discourage use and fail to solve the problem of orphan works.

### **Depositing money in an escrow account**

Pro: Users can reclaim certainty by depositing into an escrow account a sum they believe in good faith constitutes a reasonable royalty fee.

Con: How is a user to know what a reasonable royalty might be for different uses of different media? Fees paid upfront in case copyright owners come forward later are likely not to end up in the pockets of copyright owners. Paying into an escrow account would be inefficient, ineffective, and involve third parties who have no interest in the transaction. It would require payment when in many cases the owner would grant permission with little or no fee. The music industry provides sufficient evidence of litigation between copyright owners and escrow administrators. Who would administer and pay to administer the escrow account? Who would pay for litigation? Any setting of the escrow amount would be arbitrary price fixing. Escrow entails bureaucracy and imposes an unnecessary tax that would be a hindrance to use of orphan works.

### **Ongoing and New Uses of Mistakenly Designated Orphan Works**

Issue: If a copyright owner comes forward to claim infringement, consensus appears to be that new uses of the mistakenly designated orphan work require permission from the now locatable copyright owner, but what happens to the new work a user created using the mistakenly designated orphan work before the copyright owner came forward?

There is some support for “ongoing uses” of new works created by qualified users of mistakenly designated orphan works, which

would allow the new works to continue unhindered in perpetuity or at least through some safe harbor period. “Successors-in-interest” (those who subsequently license or use the new work) would also have ongoing use without the approval of the copyright owner of the mistakenly designated orphan work. A different approach recommends an injunction against future sales of mistakenly designed orphan work, but no monetary damages for past use. If a mistakenly designated orphan work has been used in a derivative work such that the orphan work cannot be separated from the new work, there should be no injunctive relief going forward, but if it can be separated, then a reasonable license fee should be set for continued use.

**Pro:** If ongoing use of new works created with mistakenly designated orphan works is prohibited, many uses will be discouraged and the orphan works problem will not be solved.

**Con:** In the absence of payment of a license fee agreeable to the copyright owner, ongoing use might constitute a compulsory license that could breach international treaties.

## **RECOMMENDATIONS AND OBSERVATIONS**

Needless to say, the orphan works problem is profoundly complex. Clearly much is at stake and there are many stakeholders. Just as clearly, digital technology is implicated in the problem and its solution. Table 3 is an attempt to apply criteria for an acceptable solution articulated in the responses to the Notice of Inquiry to the proposed solutions. No proposal strikes me as a perfect match or conspicuous winner. Ideally, all of the cells in the Table for a given solution would be “Yes.” Part of the problem in applying the criteria is that many of the proposed solutions have more questions asked than answered. The Table also masks significant differences in the scope of application of the proposals.

The criteria reveal significant concerns about balance, certainty, and containing costs. The solution will require compromise and burden, the question is who gives and who endures. Under the current copyright regime, the balance is clearly tipped in favor of copyright owners, users are bewildered and threatened, and millions of valuable works apparently orphaned are not used. We need a practical solution and we need it now, a solution that is reasonable for creators, gatekeepers, and users of all stripes. Copyright owners are concerned primarily about compensation and loss of control. Users are concerned about costs, risks, preservation, access, and the right to use. Disenfranchised

creators, forced to transfer exclusive rights to publishers that no longer see a viable market for their work, are concerned about dissemination of their work. What can we make of this soup of concerns?

Solution criteria	Public domain	Compulsory license	Default license	Safe harbor exemption	Reasonable effort accommodation
Does it avoid harming copyright owners?	no	no	maybe	maybe	maybe
Does it lower risk to users?	yes	yes	yes	yes	maybe
Does it avoid unnecessary costs?	maybe	no	yes	yes	maybe
Does it avoid unnecessary bureaucracy?	maybe	no	yes	yes	maybe
Does it comply with international treaties?	no	maybe	yes	yes	maybe

Table 3. Solution Criteria

I believe solving the problem requires multiple solutions. We already have a copyright regime wherein one size does not fit all. There is no good reason to make that a requirement now.

I support the expanded exemption of U.S. Title 17 §108. This exemption, as proposed, is workable now with minimal effort. Current copyright law already grants exemptions and safe harbors for certain communities of interest and classes of works. It is not uniform and equitable now and those arguing for uniformity and equity in addressing orphan works do not make a case for reviewing and revising the entire multitude of copyright laws to make them uniform and equitable across the board. Their argument is disingenuous and defensive, prompted by fear of the capabilities of digital technology. An operational definition that can scale to identify large numbers of published written works at low cost is required to meet the urgent needs of libraries, archives, and educational institutions. In conjunction with a take-down option for copyright owners who fail to register their intent to exercise the full scope of copyright protection, expanding this exemption will encourage preservation and use of materials of

little commercial but great historical value. Allowing non-profit use of these works for scholarly and educational purposes is in the public interest. Those who argue against this exemption are likely those who would have outlawed the photocopier and used book stores. When a book goes out of print “it can be sold in used books stores without the copyright owner getting anything and stored in libraries, where many get to read the book, also for free. Used book stores and libraries are thus the second life of a book. That second life is extremely important to the spread and stability of culture” (Lessig 2004, 113). For the net generation, a work does not exist if it can’t be found online. Even those who prefer to use materials in print prefer to find them online. Digital libraries are essential to meet these needs, essential to democracy and the cultivation of culture in today’s world. Libraries are prepared to fund the digitization of these materials and provide equitable access to them. Their copyright owners, who see no market for these works, are not. They should not be allowed to deny access to them.

I acknowledge that expanding Title 17 §108 does not address the full scope of the orphan works problem. It’s a first step and a small step at that, but it would have a powerful impact on researching, teaching, and lifelong learning. Nevertheless, further steps are urgently required to address all users, all uses, and all orphan works. For the reasons noted in the respective sections of this paper, I strongly disapprove of making orphan works public domain and I disapprove of compulsory licensing schemes. I am not optimistic that the many issues swarming around the “reasonable effort” accommodation can be settled to the satisfaction of all interested parties or settled in a timeframe likely to enable salvaging valuable endangered works or to facilitate access and use in my lifetime. If working through the myriad issues inherent in a reasonable effort accommodation does not prove too expensive, unwieldy, or controversial to manage, such that the whole effort fizzles out like the attempt in 1994 to establish fair use guidelines for digital works, I predict that the power and self-interest of big media lobbies will push through the reasonable effort accommodation with the remedy of reasonable royalties, the uncertainty of which could yield the same results as the “fair use” defense (i.e., self-censorship and gatekeeping). Frankly, the whole notion of granting a legal right that is nothing more than a defense in litigation strikes me as nothing more than a taunt of the citizenry and a trap for the unwary. The reasonable effort accommodation is so fraught with problems that I hope it

collapses under its own weight. The burden it would place on users will do nothing to restore balance in our copyright system. The reasonable effort accommodation will likely do nothing of real value for copyright owners. It will not end or address the issue of piracy of commercially viable works. What it might do is make content industries reassess the value of a work on the spur of the moment and invent a “reasonable royalty” presumably designed to resemble actual market practice—but no actual market practice existed for this work prior to its use under the reasonable effort accommodation. The situation is analogous to the child who shows no interest in his toys until the neighbor kid starts playing with them, the difference being that the reasonable effort accommodation would make the neighbor kid guilty under the law. The group likely to benefit most from a reasonable effort accommodation is lawyers. Such a solution is not practical, preferable, or affordable.

I am most intrigued by the default licensing approach to solving the problem of orphan works. It is elegant in its simplicity, outward and forward looking in its thrust, commendable in reducing harm, burdens, and costs. I fully support but am not optimistic that default licensing will be adopted. I do believe that the time has come for radical change if we want to continue to have a free culture—not free as in free beer, but free as in not unnecessarily fettered by the past. But I sadly suspect that the default licensing proposal is ahead of its time. Significantly more grass roots work needs to be done. No comments with “Solution analysis” seriously considered the default licensing proposal, just as they dismissed the public domain as the solution to the orphan works problem. Those who objected to any action that would allow unauthorized use of copyrighted works attacked the free culture movement, though their comments reveal that they do not understand it.<sup>1</sup>

In my opinion, the ideal solution will not be framed to address the fears or protect the self-interests of content industries. Such a frame would only further burden users and cripple technological innovation. Instead the frame should harness the potential of the technology to create a future aligned with, but not controlled by, our past. Medieval monks controlled manuscript technology, censored what was copied, and were put out of business by print technology, which re-defined and democratized literacy itself. No one argues that this was a bad thing. Imagine our world today if the medieval Church had managed to lock-down or control the

printing press. Likely there would be many fewer readers and books, and Latin would probably have been the language of scholarship until Vatican II. Today those who rule in the analog world of print are at risk of losing their control in the digital realm. So be it. What we gain will far exceed what we lose. The default licensing proposal illumines and models a path that would both compensate copyright owners and encourage tinkering, creativity, and progress by embracing technology. What is needed is an easy, affordable process for registering all types of works. Granted, this will be a significant challenge with some media, but it is not an impossible task. Representative creators and professional associations could collaborate to prepare requirement specifications designed to meet the needs of each community of interest.

What's at stake is "Not *whether* creative property should be protected, but how. Not *whether* we will enforce the rights the law gives to creative-property owners, but what the particular mix of rights ought to be. Not *whether* artists should be paid, but whether institutions designed to assure that artists get paid need also control how culture develops" (Lessig 2004, 120). Once understood, what is there to legitimately resist in the default license proposal? It requires no unwieldy bureaucracy or exorbitant costs, entails no significant risks or sacrifices, and avoids creating jobs for lawyers. Furthermore, it exposes and leverages the mistaken assumption that the current copyright regime is in the best interest of all copyright owners and all copyrighted works throughout their copyright term. If all copyright owners approved of the current regime there would be no open source software, no open access movement, and no Yahoo! service to search only materials with Creative Commons licenses. There is a ground swell afoot that demonstrates strong dissatisfaction with current copyright law and practice. The problem is clearly bigger than orphan works. Nevertheless Congress should be commended for requesting an investigation and the Copyright Office commended for their public call for comments. I can't help hoping that this investigation opened Pandora's Box.

## ENDNOTES

1. Those who objected to any action to address the orphan works problem appear to be disenfranchised by the current copyright system. They are understandably frightened and angry. These communities, photographers and graphic illustrators, deserve special

attention in the inquiry into orphan works. The sheer number of photographs taken by a professional photographer and understandable practice of putting attribution information on the back of the work, where it is inconspicuous if not inaccessible, seems to me to warrant special handling in copyright law. The Copyright Clearance Office's payment of copyright royalties to primary copyright holders at the expense of third-party interests warrants investigation and redress.

### **REFERENCES CITED**

Lessig, Lawrence. 2004. *Free Culture: How Big Media Uses Technology and the Law to Lock Down Culture and Control Creativity*. New York: Penguin Press.

Notice of Inquiry, 70 Fed. Reg. 3741. January 26, 2005. Available at <http://www.copyright.gov/fedreg/2005/70fr3739.html>



# Heritage Under Lock, but No Key: The Troubled Status of Unpublished Works in Digital Archives Projects


William J. Maher (University of Illinois at Urbana-Champaign)

**Abstract:** Digital technology combined with new communications technologies have presented archivists with an unprecedented opportunity to make their unique research holdings widely available to audiences worldwide. While there are technical and resource hurdles to placing archival materials before a broader public, the most significant challenges are those in copyright law. This paper assesses the practical effect of the conjuncture of archival principles, copyright law, and archival collections, and examines current “digital archives” projects in light of these issues. While there is general consistency of objectives and policy at the consortia level, there is wide variation in the practices and content presented by individual repositories. All confront significant barriers when working with the vast array of non-public domain works in archives. To address these issues, the paper concludes with suggestions for project administration as well as a public action agenda.

Recent information technology developments have combined with increased attention to copyright law to create a deep tension in the lives of archivists and librarians. While clearly interested in doing all we can to disseminate information, we want to be good citizens and not tread on the rights of authors or publishers. No doubt, a system of creative commons licenses would allow greater dissemination of information, but these are of little value for those materials that fill archival repositories—pre-existing works locked in copyright but lacking any clear owners who might be approached for permissions. Indeed, perhaps the single most important intellectual property challenge facing archivists and their users, as well as the one most in need of a legislative solution, is that of these so-called “orphan works.” When neither scholars nor repositories can get permission to use these materials, public access to information is restricted, and both the archival mission and society suffer.

---

M. Halbert (Ed.): *Free Culture and the Digital Library Symposium Proceedings*. Atlanta, Georgia: MetaScholar Initiative at Emory University, 2005. pp. 141-153.

 This work is covered by the Creative Commons Attribution-NonCommercial-NoDerivs 2.5 License.

The 1976 copyright law was supposed to have clarified the murky territory in which archivists and their users had to work. The law ended the regime of unpublished works being under perpetual copyright according to myriad state common law rights, but it has failed in the new technological environment and, in regard to archival material, it fails to support the constitutional purpose of copyright. Meanwhile, the 1976 law has been of limited practical use in supporting scholarly, academic, and public use of archival material. For example, the fair use exemption is a weak, inadequate, confusing, and costly device for the support of scholarship, learning, and public education. As a result, archivists and manuscript curators receive frequent requests from authors whose publishers require written sign-offs for the use of a single quote or photograph. All too often, our response can only be that we do not own the copyright and have no way of knowing who does. Especially problematic are unpublished items—works created outside the economic assumptions of copyright, but which are incredibly valuable for research.

Such is the world of archival "orphaned works." These archival materials have so much value for society, yet they have been neglected not only by a copyright law written to satisfy large-scale commercial interests, they are also overlooked by reform efforts, as seen in some responses to the Copyright Office's recent call for comments on "orphan works" that address only published works.

So what is the diligent archivist to do when he or she receives the inevitable mandate to create a digital archive of those unique items that have never been seen before outside the archives' search room? Funding agencies and institutional legal counsels want clear-cut certification of copyright ownership or assignment, but what if, as is usually the case, the creators are long-since dead and their heirs virtually untraceable? What if, as is the case with archives, the documents were created not as self-conscious expressions for public dissemination but as accidents of some other administrative action, and that there was no thought of creating an effective record of the succession of rights?

The archivist may turn to copyright law in hope of finding some public-spirited exceptions that might allow the educational project to proceed. But how much guidance does copyright law really provide?

First of all, the center of copyright law is the recognition of exclusive rights, which define the limits on what archivists can

legally do with material they wish to digitize. These exclusive rights are summarized in Section 106, which says that no one but the author or his/her assigns can reproduce or make copies, prepare derivative works, distribute copies, perform the work, and display the work.

The law, however, provides two major limits on exclusive rights seemingly relevant to digital libraries and archives. Section 107 (Fair Use) was first added to the law in 1976, and it is based on the notion that the author's property rights should not hinder creative endeavors of society, and § 107 should enable copyright law to mediate between private and public interest. In practice, as one moves beyond these laudable principles, the archivist encounters enormous difficulties in relying on fair use to build robust digital archives. In strict terms, fair use is an affirmative defense against a claim of infringement, and it must be decided on a work-by-work and case-by-case basis in the course of a legal proceeding against the user. Fair use rules are not clear—they are overlapping and highly circumstantial—and certain kinds of transformative uses of archives have received little support in fair use decisions.

Then there is Section 108. One would hope that a section entitled “Limitations on Exclusive Rights: Reproduction by Libraries and Archives” might provide assistance to those building digital archives, but in fact Section 108 is narrowly construed to allow only limited copying for preservation purposes or one-time copying for end-users. Indeed, while the law now allows digital as well as analog copies, Section 108 does not allow such digital copies to be made available beyond the premises of the archives or library itself, a restriction inimical to the entire notion of the internet. Other absurdities include Section 108 (h), which allows preservation copying and distribution of works in the last 20 years of their term, but applies only to published works; and Section 108 (i), which eliminates most of the §108's exemptions for audiovisual, musical, and pictorial works, which are not only necessary for a complete record but which have historically been among the most sought-after of archival materials.

A final key issue in copyright law is the length of term of protection. Since 1978, all works, whether published or unpublished, are covered by copyright from the moment they are created. Second, thanks to entertainment industry lobbying, the term for all works is now life of the author plus 70 years. For archival material this presents several significant problems. First, few of the authors of works we hold are of such significance that

their date of death can be readily determined, and in fact, given the commonality of names among the millions and millions of document creators, it is very often difficult to establish the identity of many of the authors, let alone locate all their heirs. Second, because most of the documents so valuable to archives are created as accidents of some other action, few authors left means for the administration of their rights. Finally, the law's provision (§302 (e)) regarding presumption of death of the author does not open up the possibility of using the works of untraceable authors until the documents are at least 120 years old, and then only with a cumbersome process of checking with the Copyright Office. If one's efforts to create a digital archives are thus limited to only such works as are clearly past their copyright term, virtually all of the history of the twentieth century has been fenced off from use.

So why not just focus on pre-twentieth-century items and ignore the rest? For archivists, that would contradict our core mission which is to be purveyors of recorded knowledge and thereby to ensure that the knowledge created and accumulated by past generations is joined with that of the present, and in order to make it available for society to build a better future. Our role is to appraise, secure, arrange, describe, preserve, and make accessible an authentic record of the government, institutions, organizations, and peoples of our world. We preside over the past so that others may use it to form their own picture of the past. Because knowledge is cumulative, and because our work must result in an ultimate utility, we know that the content of archives must be copied, quoted, published, performed, broadcast, and otherwise disseminated using the latest technology. For this to happen, the copyright holder's exclusive rights are inevitably encroached upon.

All archives, whether in government, educational institutions, professional associations, businesses, or churches, share a common mission and a common methodological approach to the heritage in their charge. In fact, it is the nature of archival theory and methodology that makes the creation of a truly authentic digital "archive" in today's copyright world well-nigh impossible.

Why? First, the material that finds its way to archives is highly diverse not only in physical format, but also in terms of authorship. Any given correspondence file may contain anywhere from dozens to hundreds, thousands, or even millions of separate copyrighted works and an equal number of authors. Almost without exception, the authors of these works had little or no idea that their "works" had been deposited in a public repository, let alone might be

disseminated through some “digital archives” project. Second, the very characteristic that makes archives so extremely valuable as historical evidence—their spontaneous, almost accidental, creation—also means that few archival works have the commercial utility that is the underlying basis of American copyright law. Letters, photographs, sound-recordings, and other documents are, more often than not, created as the accidents of some other action, rather than as a conscious creative expression to provide testimony to the public. A third, and particularly important characteristic is that archives have a comprehensiveness that, while not absolute, is not compromised by artificial curatorial decisions. Indeed, the supposition that an archives is complete makes the spontaneously generated documents it contains so valuable for constructing an accurate historical record. Archives are valued precisely because they are fresh, unedited, and unselected. Such archives allow the readers to draw their own conclusions and make their own interpretations, while also leaving the document’s integrity intact so that the next person may draw different conclusions or challenge our interpretations.

To provide for these important archival values, a range of archival theories and practices have developed over the past two or three hundred years, the most important of which relate to the integrity of original order, provenance, and physical preservation and openness of access. In a conventional environment, creating repositories to ensure the key requirements of breadth of comprehensiveness, authorship, and spontaneity of creation only required large amounts of space, basic physical and intellectual control, and a survival-friendly storage environment.

Accomplishing this same objective in a digital environment is significantly more complex. New technology enables the delivery of archival content globally without the costs of building a distribution network, but unlike the conventional environment, the digital environment runs afoul of intellectual property law, even if all the other technical and resource issues can be resolved. That is because the items are no longer just held and examined under the “first sale” rights, but are copied and distributed by display on a network.

Given the diverse and complex nature of archival repositories and given copyright’s egregiously long term, it is no wonder that the number and scope of actual archival digitization projects with extensive unpublished materials less than 120 years old is so limited. Otherwise, there would be little hope of staying compliant

with U.S. Copyright law. This assessment may seem at odds with the appearance of several online historical projects with the word “archives” in their name, including those hosted by archives and manuscript repositories. In fact, on examination, one can see that the scope and depth of these efforts has been severely limited. While the materials that have been mounted are clearly of use by themselves, the fact that they cannot display a complete record means they do not constitute a genuine or significant archival presence.

To assess how various institutions and consortia have dealt with this dilemma, I examined thirteen digital archives sites.<sup>1</sup> Of course, since most of these projects function as portals to large consortia of institutions, and each institution generally has multiple collections and web products presenting “digital archives,” there is a dizzying number of collections and projects represented through these sites, and it should come as no surprise that it is virtually impossible to follow all links, quantify the results, or even apply uniform data collection tools. Thus, the following observations are more “strange attractors” of systematic chaos than fixed Ockhamite categories.

**Scope and Range of Projects:** There is no consistency among projects in terms of content, or at least in terms of the evenness of coverage. Some projects are heavily focused on providing consolidated, searchable EAD finding aids. Others put a primary emphasis on training, policy, and dissemination of information on best practices. Yet, others provide consortium-wide search engines for the subject content of collections, while still others settle for providing links to the home pages of the participating repositories, each with its own search tools.

**Content of Digital Collections:** Clearly the web presentation of archival finding aids is nothing new, dating back at least to the early days of Gopher and Mosaic, but what has developed more recently are the first glimmerings of a more robust exploitation of the possibilities of the Internet. However, when one looks at the actual archival content of the various sites, it is obvious that there is still a long way to go. The most common content seems to be individual photographs, generally selected from much larger collections. Another major component is clear public domain works—ones published before 1923 or authored by the U.S. government. This is particularly true of the *Making of America Digital Library*. In far too few instances, there are primary sources or unpublished textual documents, and these are almost always just

isolated documents selected from larger collections. With the exception of some largely pre-1923 items on LC's *American Memory* site, there is very little in the way of sound or audio-visual material.

**Range of Institutions Represented:** One striking characteristic is the extent to which public libraries and local historical societies have participated, a notable accomplishment since such repositories have traditionally not been party to large surveys of archives and manuscript collections. Although it varies by state or region, academic libraries and archives have also participated, although in some instances not to a very great extent (e.g., the Illinois Digital Archives Program). State archives and other governmental records repositories, on the other hand, are somewhat under-represented, and almost totally without representation are business and corporate archives.

**Project and Consortia Intellectual Property Policy:** When it comes to copyright policy, the greatest consistency among the sites is at the consortia level, where most all are emphatically aimed at providing the maximum legal protection against claims of having contributed to infringing activities. The projects generally state that the only materials to be digitized are ones in the public domain or those for which the participating institution has obtained written permission. They also normally contain a "kick-out" clause stating that any institution may be removed from the consortium if it violates copyright. Unfortunately, consortial guidelines seem to be generally silent on the problem of orphan works, although they often create the seeds of a justification for applying fair use when the owner cannot be located. Overall, the consortial guidelines on copyright seem more focused on instructing web visitors of the limits on what they may do with the material, with a near universal emphasis on the fact that the site is making material available only for non-commercial, instructional purposes and that any commercial or publishing use must have permission of the copyright owner. Overall, the policies provide clear evidence of having consulted educationally based, though risk-adverse legal counsel.

**Individual Institutions' Copyright Policy and Practice:** On a policy level, most of the sites I surveyed of individual institutions affirm adherence to the same copyright policies as the consortia. However, when one starts looking closely at the content—at what has been digitized—there is much variability. The following practices are representative of the lack of uniformity:

1. A number of institutions appear to want to play it safe and limit themselves to items published before 1923, items published by the U.S. government or by their parent institution (e.g., Illinois Wesleyan *Argus*), or to unpublished works written more than 120 years ago (e.g. *Valley of the Shadow*). This is fine if one's historical interests do not extend later than when there was a chance that the League of Nations could end conflicts among nations. A lot happened however in the last eight decades of the twentieth century, and given the need to connect history to a generation that soon will not be able to remember any president other than "W," it is clear that such digitization projects will be of limited value in providing a very complete historical account.
2. Some institutions present published and unpublished primary source items for which they have a plausible claim of ownership and present them in some depth. In a few instances, because the repository has negotiated a copyright transfer as part of a conventional deed of gift, it is able to make important unpublished or even recently published material available—for example, Colorado State University's extensive wild animal photographs or the University of California, Davis' mid-century commercial photographs of California. Where forward-looking curators obtained such rights, or when these rights can be secured with new accessions, this is a viable way to build meaningful digital archives. Unfortunately, given the vast quantities of valuable historical archives donated over the years without copyright transfers and given the extent to which third-party correspondence and other works make historical collections valuable, this approach has major limitations.
3. Some sites claim ownership, or act as if they do, even when the claim seems implausible according to information provided on the site. In some cases, the institutions may be claiming nothing more than copyright ownership over the digitized image rather than of the underlying work, or they may be seeking to limit what can be done with the image because they are the sole-source provider of the image. Clearly, this flies in the face of *Bridgeman vs. Corel's* finding that faithful copying of an underlying work does not by itself create a



work that can be claimed as a copyright by the digitizer. If participating institutions use such claims over their digitized images to discourage subsequent users from commercial uses, this may limit exposure to contributory infringement claims, but if this is used to limit public use of images that they truly do not own in the first place, it is a disservice to the educational purposes of digital archives.

4. Many institutions follow a “throw up their hands” approach. They indicate that they have made efforts to contact copyright owners without success but post the item(s) on the basis of fair use while also including a notice that should any copyright owner or other party have information about ownership, they should come forward. To show good faith, they sometimes include clauses indicating their desire to hear from any copyright owners, and in some further cases, they even promise they will remove material if a copyright owner does appear. A balanced statement is that from Cornell University: “The Kheel Center would like to learn more about these images and hear from any copyright owners who are not properly identified on this website so that we may make the necessary corrections.” They then provide a staff name and e-mail address.<sup>2</sup> Overall, fair use of this sort seems a reasonable approach, although one that ultimately places your institution at risk of having to invoke the always murky four-factor defense in an infringement case. If followed faithfully, it also requires considerable effort to track and document one’s efforts at pursuing owners.
5. On many sites, the content presented is less than complete, and thus far from archival. In some cases, such as sites that include pre-1923 published material from a collection but no correspondence from relevant individuals, the reason is probably copyright. A prime example of incompleteness is the Paul Eliot Green Papers at the University of North Carolina. In this case, all that has been digitized is about sixteen letters, comprising 111 pages of 1917-19 correspondence from a 192-linear foot, 110,000-item collection. There may have been sound intellectual reasons for the choices made with the Green Papers, but I suspect this is not always true. In a case I know more intimately, the James B. Reston Papers at the

University of Illinois, we have been able to digitize only about 2,000 pages of an estimated 146,000 pages in the collection, and the principal reason for the limit is copyright. When individual documents are “cherry-picked” out of an individual’s or an organization’s complete archives, the project may be a digital scrapbook, a digital exhibit, but hardly a digital archives.

6. Ultimately, we cannot say that copyright barriers are the only reason for such selection and narrowing of the digital content. In some cases, there seem to be understandable cost and project management reasons that the digitization and presentation is incomplete. For example, Colorado State University’s wildlife photos project included only 1,000 out of a total collection of 20,000 images. These are not indefensible editorial decisions, but nevertheless, the practice ends up short of the high hopes and promises of several digital library and archival projects.

Indeed this is quite at odds with the rhetoric by which such projects are promoted. For example, we are told that a driving concept for the Valley of the Shadow project was that it be “a research library in a box, enabling students at places without a large archive [sic] to do the same kind of research as a professional historian.” Yet, the Online Archive of California goal of providing “all” with “. . . access to information previously available only to scholars who traveled to collection sites” is clearly undermined by the occurrence, at least 85 times, of the following line in bibliographic records for collections included in the OAC database: “Items Online: None online. Must visit contributing institution.”<sup>3</sup>

As this review makes clear, creating a true digital archives will run afoul of copyright unless we can solve the orphan works problem for unpublished material. The issue at hand for this Symposium on Free Culture and the Digital Library is where to go from here in the present environment.

First, consortial projects and individual repositories should make clear the need for participants to examine copyright ownership before digitizing and mounting materials on the internet. While some of the consortia policy statements are a little draconian so that the consortia ends up functioning as an enforcing gatekeeper, as a general practice the emphasis on digitizing only that which you own or for which you have been granted permission is perhaps the only institutionally safe general rule. That said, however,

repositories are the ones in the best position to know the facts of a specific *fonds* and works therein to determine whether a hard line is needed or whether some flexibility is appropriate if investigation shows the works truly are orphaned.

Second, although fair use, as Lawrence Lessig has said, is not much more than a license to hire a lawyer, projects should look to fair use for establishing the context for digitizing and displaying material for which copyright owners cannot be readily located and which can otherwise be justified for their educational, non-commercial, cultural value. There is no assurance of protection from litigation, but if a repository's investigation shows that there is no existing market for the works, and if the site includes appropriate disclaimers, then fair use represents the only present basis for digitizing the inevitably orphaned copyright works, and these must be included if there are to be meaningful digital archives projects.<sup>4</sup>

Third, the library, archival, and internet community should make a focused effort to amend Section 108 (h) of the copyright law so it includes unpublished works in the scope of materials that libraries and archives can digitize and make accessible in the last 20 years of their copyright term. Better still would be adoption of a full-scale orphaned works exemption along the lines supported by the Society of American Archivists.<sup>5</sup> Archives, far more than published works, are very likely to be orphaned material, often for the very same reasons that it is valuable research material: it contains a multiplicity of authors, those authors are virtually anonymous, it is unclear those authors ever constructed their works for dissemination, and the works themselves are of research value but almost always of very limited or no commercial value.

Given that a few, potentially influential respondents to the Copyright Office's call for comments on orphan works categorically excluded unpublished works from any orphan works proposal, we need a fresh analysis of the very nature of the distinction between published and unpublished works. We also need a careful review of the case law on state court cases relating to suits involving appropriation of unpublished works, such as *First Trust Co. of St. Paul vs. Minn. Hist. Society* or *Hemingway vs. Random House*. Applying the old rules of state common law rights seems particularly dysfunctional and misplaced in the twenty-first century.

To ensure a solid basis for the kind of safe haven needed so digital library projects can include sufficient archival material to make them credible digital archives, public policy efforts should also focus on the international level. Specifically, we need to seek support within the Berne/WIPO treaties to allow for non-commercial, educational use of unpublished works. Efforts have been initiated in the International Federation of Library Associations (IFLA) and the International Council on Archives (ICA) following discussions at their separate 2004 congresses. However, funding is needed to bring together the librarians, archivists, and international copyright law specialists to craft language to be advanced to WIPO. Then, a concerted effort will be necessary to have this kind of change adopted, but in an environment where copyright policy is determined by large-scale commercial interests through the mechanism of the World Trade Organization, WIPO, and the European Union Commission, efforts to craft local solutions are unlikely to be persuasive without some grounding at the international level.

In conclusion, the many digitization projects to date have made a noble effort to expand the public's access to cultural research materials beyond those previously at hand through local libraries, but they can hardly be called "archives" in the full sense of the term because they have been unable to provide very deep or broad access to much truly archival material. In many instances, quite understandable cost and pragmatic hurdles have caused these efforts to be quite limited. However, the significance of the copyright law barrier is hard to overestimate. For any meaningful, robust, on-line digital archives to exist, the copyright issues must be addressed and the barriers they create reduced to manageable hurdles at the most. We need sophisticated keys that allow us to unlock these works and not become part of a gatekeeping for orphan works that serves neither the original author, the works, or the archives-using public.

## ENDNOTES

1. They were: Colorado Digitization Program, Northwest Digital Archives, Northwest Digital Archives, North Carolina Echo (Exploring Cultural Heritage Online), Making of America, Cornell University Library Digital Collections, New York University's "The Database of Recorded American Music, Library of Congress' American Memory Project, University of Virginia's Valley of the Shadow Project, American Museum of Natural History, Tufts

University's Perseus Project, the Illinois State Library's Illinois Digital Archives, and the University of Illinois' American Library Association Archives Digital Collections.

2. <http://www.laborphotos.cornell.edu/copyright.php?Kheel=7e23269dd4c420e8c06ea581a1f9e73e>
3. At more detail level, in one instance of a 43 linear foot photographic collection, 54 images have been digitized, and when one looks more closely, those images come from only 45 percent of the folders in one subseries of that collection.
4. For example, that from the American Museum of Natural History site reads: "While this Website is publicly-accessible, not all the materials are in the public domain—the majority of the images, texts and data are copyrighted to the American Museum of Natural History—and a number of other texts and images are still copyrighted to their original print publishers or digitizers and made available here with permission. We have put great effort and expense into producing this site, and we hope the results are useful to a broad audience."  
<http://library.amnh.org/diglib/conditions.html>
5. The SAA initial comment is at:  
<http://www.copyright.gov/orphan/comments/OW0620-SAA.pdf>, and the "reply comment" is at:  
<http://www.copyright.gov/orphan/comments/reply/OWR0088-SAA.pdf>

# Government Information in the Digital Era: Free Culture or Controlled Substance?

Karrie Peterson (North Carolina State University) and  
James A. Jacobs (University of California San Diego)

**Abstract:** Free public access to government information, to be anything more than a quaint tradition, must take advantage of new technologies, so that citizens can participate, intervene, and comment on government activities in a meaningful way. Technology must also be used to preserve the public record for study and scholarship as these are conducted in the digital era. Digital technologies make unprecedented sharing and innovative re-use of government information possible, but libraries and our nation's legally mandated program of disseminating public information play a key role in expanding and maintaining policies that enable citizens to use, distribute, share, repackage, and preserve government information.


## GOVERNMENT INFORMATION AS PART OF THE INFORMATION COMMONS

Government information—that is, information collected, compiled, and created by governments in their official capacity—is part of our information commons. This is not just by convention, but also by design. It is built into our laws. Government information is and must be free—and by free, we mean free as in free culture and free as in free beer. Whether it is information that the government collects, such as information about toxins in the groundwater; or information about the performance of government, such as reports by the Government Accountability Office on the effectiveness or legality of government policies; or Congressional deliberations as documented in the Congressional Record and committee hearings, government publications are the official public record of our democracy. They play an essential role in government accountability and government openness.

This is a value that predates the Constitution. Out of 30 articles of complaint leveled against King George in the Declaration of Independence, number four was that he made it difficult for us to

---

M. Halbert (Ed.): *Free Culture and the Digital Library Symposium Proceedings*.  
Atlanta, Georgia: MetaScholar Initiative at Emory University, 2005. pp. 154-178.

 This work is covered by the Creative Commons Attribution-NonCommercial-NoDerivs 2.5 License.

use our public records.<sup>1</sup> Open access to government information is rooted in our culture, our democracy, and our way of life.

Government information is created by us collectively through government agencies at all levels, acting under mandates of law that express the peoples' will. It is funded by our tax dollars and used by us for our common understanding and our common good. This information is, by law, in the public domain because it is too important to our collective welfare to be owned or controlled by any one person, group or entity. Government information is our information.

And government information is massive in quantity. For decades the US government has been the biggest single publisher in the US, if not the world—publishing more reports, hearings, pamphlets, journals, maps, and scientific documents than any private publisher. In some years, the government has published more than all private publishers combined (Schmeckebier and Eastin 1969).

If we need an example of a body of content that should by definition, tradition, and law be free of any limitations on public access, we should have to look no further than government information.

With the advent of digital publishing and the Internet and the World Wide Web, getting this kind of information has become even easier. Anyone with a computer and a fast network connection can get, it seems, almost anything.

The 9/11 report was downloaded over one million times (Russell 2004). Similarly, consumer pamphlets, the *Occupational Outlook Handbook*, the Energy Information Agency's *Monthly Energy Review*, the CIA's *World Factbook*, and many others are now more easily available on the web than they ever were before in print: easier to find, easier to browse, easier to get. And they are free—available without charge, because we've already paid for them with our tax dollars.

In addition, since government publications are mostly not copyrighted, private publishers are free to reprint and repackage government information. This has always made it easy to find copies of popular government publications—like the annual Budget of the President or the Pentagon Papers or excerpts from the Warren Commission Report—at your local bookstore.

And there's more good news. It used to be that, if you wanted to know the population of zip code 30322 where Emory University is

located, you'd have to find your local depository library, get to it, and dig through a bunch of volumes of census statistics. Now, you just go to the Census Bureau's "American Factfinder" website ([factfinder.census.gov](http://factfinder.census.gov)) and type in a zip code and get a nice report! Access is also enhanced because digital government information that is copyright-free can be re-used by anyone. Any group or person can put the same information online on their web sites in a context that makes sense for their audience.<sup>2</sup>

What could be a better example of the public domain and the information commons at work? What could be better for democracy than to be able to easily and freely find out what your government is doing and what it knows? What better way to leverage our tax dollars than to collect vital information and make it freely available to all? What better way to be sure our government officials remain accountable for their actions on our behalf?

Given all this, it would be understandable to think that the government's transition to the digital era is coming along pretty much the way it ought to. It would be easy to conclude that a government on the web is a more transparent government, that government information and services are more available to citizens than ever, and that online technologies are empowering the average citizen to participate in government more easily than ever. It would be easy to feel comfortable and to feel that at least one part of the digital information commons is in good hands and is well cared for.

Unfortunately, this is not true. Our goals today are to undermine your comfort level by sharing what we see as ominous emerging trends and to invite you to see that the struggle for free culture is deeply interconnected with the ongoing fight to keep government information accessible in the digital age. We don't have two single-issue battles, we have one big issue.

The trends that we will analyze point to government information becoming less available as governments shift to digital creation and dissemination of information. This shift is accelerating unnecessary and deleterious trends toward increased privatization of information, toward heightened government secrecy, and toward the alteration and loss of important public records and public knowledge.

We hope you will agree that, if we aren't assured of free and open access to public-domain government-produced information—



information that is supposed to be free, information that is essential to our democracy—then how will we solve the harder problem you all came into the room with—defending free culture for other kinds of intellectual content?

## **THE ROLE OF LIBRARIES**

Before we examine some of the issues that endanger free access, we want to briefly review the role of libraries in providing access to government information.

Libraries have helped ensure free access to government publications for more than one hundred years through laws that mandate deposit of government information in libraries across the country. Before the digital age, if you wanted to consult a federal government publication, you could go to any of over 1000 official depository libraries (44 USC 1901 et seq.)—there is at least one in every Congressional District—and find and read and borrow it. If your local library didn't have a copy, there were plenty of tools that libraries use to find it and get it to you.

Depository libraries receive copies of government publications without charge under the condition that they make them available to the public. Even libraries (such as law libraries at private universities) that normally don't let the public in are required by law to let you in if you want a government publication. All depository libraries follow strict rules for retention of government publications, some specially-designated depository libraries are obligated to keep all their federal documents forever.

In the print world, libraries bore the cost and responsibility for organizing and preserving government publications and providing access to them.

In the digital age, however, some people question whether there is a need for depository libraries anymore, citing the easy availability of government information on the web. The Government Printing Office (GPO) has been a frontrunner in getting government publications on the web, creating a service called "GPO Access" more than ten years ago.

GPO has continued to expand this new role and has begun the process of taking over responsibility of providing preservation of and access to government publications (U.S. Government Printing Office, Strategic Vision 2004). Some depository libraries have welcomed the fact that GPO is willing to assume the roles of preservation and access that they once provided. However, GPO's

vision of the digital future reduces the role of depository libraries to service centers without collections. In GPO's vision, the government, not libraries, will have collections and will decide what will be acquired and retained and who will have what level of access at what cost.

In contrast, we believe that libraries need to continue selecting, acquiring, organizing, preserving, and providing long-term free public access to digital government information. In fact, libraries—serving in their social role as institutions that share information—can help us solve or at least begin to address serious concerns in the realm of access to digital government information.

### **VERY STICKY WICKETS**

We categorize the problems that can interfere with public access to public information in three broad issues areas: technical, economic, and control.

#### **Technical issues**

We won't spend much time on this area because most of you are familiar with these problems. We do want to explain, though, why we believe that technical issues largely associated with copyright and so-called intellectual property have relevance to information that is, by definition, in the public domain.

First and most obvious is how free access to public domain information will be endangered if technical standards or legal rules and regulations are too stringent and limit fair use of public domain materials. For example, if peer-to-peer tools are made illegal or regulated in such a way as to make their use difficult or problematic, or if P2P technologies are undermined in a way that smothers innovation, then we will not be able to use such tools for dissemination and use of government information. Such tools are currently being used by libraries [LOCKSS—see <http://lockss.stanford.edu/>] and by citizens [see [outragedmoderates.org](http://outragedmoderates.org)] to provide easy access, distribution, and authentication of government information.

The Digital Millennium Copyright Act already worries librarians who wish to preserve copy-protected digital materials. The preposterous dilemma of a librarian or anyone being prevented from reverse-engineering public domain government information wrapped in a proprietary interface has been widely noted. The DMCA becomes even more worrisome when there is information which clearly ought to be in the public domain but is actually

copyrighted (such as state laws where printing has been contracted out to private publishers). If the government creates laws like the Induce Act or the Broadcast Flag regulation, these may limit how public domain materials can be used. If the hardware industry uses copy control technologies aimed at prohibiting unlawful copying, will hardware be able to make lawful copies (Doctorow 2005)? You know about these issues and we won't belabor them.

Of course, it is possible that more reasonable laws, regulations, and industry standards will be developed and we won't find ourselves in a world where a DVD of a presidential press conference is locked down the same way as a new Hollywood blockbuster. But there is still a large potential problem of governments using the 'wrong' tools or using tools in the 'wrong way.' What we mean by this is simply that governments as they create digital information will not necessarily go out of their way to find tools that allow for easy sharing of information. Governments are just like everyone else, they buy software that is readily available, meets industry standards, and so forth. As there is increasing commercial control of information delivery and information packaging, the tools for creating information will, by simple economics, conform to the needs of those creating commercial content. There may not be a big market for tools that create 'public domain' information. Governments will use the tools that are available and if those tools assume copy protection, digital rights management, and so forth, governments will create information that has those characteristics. In fact, government agencies have been directed by the Office of Management and Budget (OMB) to look for and use commercial products when they are available (McIntosh 1990). Since it is not the mission of each individual agency to consider the implications of these publishing decisions for future public access, this prodding by the OMB is almost guaranteed to lead to problems of public domain information wrapped in proprietary technologies.

We are already seeing the beginning of this. The Government Printing Office, for example, has expressed interest in using "Digital Object Identifiers" (DOI) for the reasonable purpose of better managing the pointers to online materials (U.S. Government Printing Office, *Managing* 2004.), but the intended purpose of DOIs includes far more than managing pointers. According to a working group of the information industry, the purpose and utility of DOIs include checking the authority of a person to access a document, to protect copyright, and to prevent "piracy" (Sidman 2001). How can we ensure that a technology designed to do these

things for commercial users won't subvert legitimate use of public domain materials?

We live in a time when powerful economic forces are dedicated to convincing lawmakers that technological lockdown is necessary. Cary L. Sherman, president of the Recording Industry Association of America, has said that Internet 2 "has been hijacked for illegal purposes" and further, that "We can't let Internet2 become a zone of lawlessness" (Read 2005). If reasoning like this persuades lawmakers and those who set standards, how difficult will it become to share public domain materials?

In short, the tools that make it possible to share all digital information freely are the same tools we need for government information: peer to peer technologies, non-proprietary formats, technologies that allow copying and other re-uses, open-source software, and so forth. And we don't need technologies that are built around government secrecy. Many of you will remember the Bush Administration's proposal in 2001 to sequester all the really important government information on a totally separate Internet known as Govnet. (Mitchell 2001)

## **ECONOMIC ISSUES**

The economics of digital information dissemination will, we believe, work against long-term preservation of and access to digital government information unless we endeavor together to create an economic environment for preservation that is sustainable. We will briefly examine three aspects of the economic problem. None of these is a new problem, but all are exacerbated by the shift to digital information with its capability to be easily copied and distributed.

### **The cost problem**

The first economic problem is being created by government and is avoidable. The context for this problem is that keeping digital information available requires ongoing expenses. Digital preservation, format and media migration, maintaining documents online—all these are expenses.

The problem is that these expenses that were once borne by multiple libraries in every state are being shifted by GPO's new policies so that the federal government will bear the complete cost. This will put the cost of information access and preservation in competition with other federal budget items.

The problem becomes evident by comparing the traditional methods of distribution for paper and ink documents with the way the government is now trying to manage dissemination of digital documents. In the paper and ink world, the government's cost for publishing ended at the time the document was published, but in the digital world the costs are ongoing. In the paper and ink world, an agency paid for publishing a book and its financial obligation ended. The economic costs of providing access to and services for the publication, and of preserving the publication for future access and use, were all borne by libraries—specifically by the more than one thousand libraries that are part of the Federal Depository Library Program (FDLP). But in the digital age, government agencies and the Government Printing Office (GPO) are depositing almost no documents in depository libraries—deposit of print documents has slowed to a trickle and no digital files are deposited.<sup>3</sup> Judith C. Russell, the Superintendent of Documents and head of the Depository Library Program reported recently that distribution to depository libraries has been reduced to only 14 percent of the documents that GPO handles (Russell, 2004). Instead of depositing digital files in FDLP libraries, GPO is providing pointers to files on government-controlled web servers and putting itself forward as the guarantor of permanent public access. This means that, when the government publishes anything, it is obligating itself not just to a cost in the current fiscal year, but to a perpetual cost to keep the document available. Of course, agencies can avoid this cost by removing information from the web, but that means that the information will no longer be accessible.

By relocating the expense of maintaining access to government information from a large number of separately funded libraries to a single government entity, the government is creating a new cost for itself without any guarantee of funding, thus endangering long-term access to this information.

Is this something to worry about? Just imagine Congress mulling over spending a few million dollars to maintain online access to employment data for women or minorities that is 10 or 20 years old, or an annual report from an agency that is now defunct. Imagine whether or not these expenses will get priority over national security, education, or social security.

### **The profit problem**

While making information available is a net cost to agencies and thus endangers information access, agencies do have an option—but this option also endangers public access. The option is to sell the information and turn a net cost into a profit. The information that governments hold is increasingly valuable to commercial resellers of information as well as to companies who can use government information to increase their profits. This can include everything from aggregate census information that allows marketers to identify neighborhoods for locating stores or zip codes for directing ads, to information about individuals who have bought or sold property, married or divorced, had a child or a death in the family, regularly travel a particular public road and are recorded as paying tolls, and so forth.

Though we are focusing our examination on federal government information, it is worth noting that the lure of selling government information exists at state and local levels of government as well (Newman 1998). For example, some states have removed criminal records data from their websites and now sell it to employee-screening companies, who in turn sell it to citizens (Fields 2005). The problem becomes more acute for government agencies when they have had to pay for data to be created to begin with and would like to recoup their costs: for example, local governments often pay for flyovers to create orthophoto base maps of their region, which they enhance with parcel-ownership or other data. This data has multiple uses for government and citizens, but real estate companies are willing to pay a high price for it. The Open Data Consortium Project is concerned about this loss of public information and offers advice on their web site about alternatives to selling data (Joffe 2003).

Governments selling public information was not a problem in the print era—people would be willing to pay for the convenience of their own copy if it saved them a trip to the library or a government office, where free copies were accessible. But with digital information, an agency cannot make information available to the public for free and sell the same information. We've seen dramatic evidence of this. An early attempt by the Government Printing Office to sell access to digital information it was simultaneously providing for free failed. Why would anyone pay for information they can get for free? Thus, when agencies want to sell information, they will find it necessary to avoid making the information freely available or risk a business failure.

While it is easy to understand how a cash-strapped agency faced with a net cost of keeping information online might jump at the opportunity of turning that liability into an asset by selling that information, it is also easy to see that such policies result in citizens having to pay for access to public information. When the government provides no free access and grants monopoly distribution rights to companies (as they often do), the result is government information being removed from the free-access commons and placed in the fee-access private sector.

### **The no-competition problem**

A third economic problem involves so-called competition between governments and the private sector. The publishing industry has argued for years that governments should not compete with the private sector, a slogan that became a policy in the Reagan years. In 1985, the Department of Agriculture granted exclusive distribution rights for its “AgNet” database of crop and livestock statistics, export sales reports, and other agricultural data from the Department to a private information company, which, in turn sold access to it for \$45 per hour (Gross 1991). A number of scholarly journals once published by the government and distributed free to depository libraries have been privatized and are now available only for sale. Journals falling into this category range from *Schizophrenia Bulletin* to *Public Health Reports*, the official journal of the U.S. Public Health Service.

The way business views this situation is that if they could make a profit offering some kind of government information, then the government should not be doing it, an argument reminiscent of the idea that any digital copy of music is lost revenue in the music industry. This has come up in legislative efforts this past April to prevent the National Weather Service from providing user-friendly weather forecasts, so that Accuweather and other private companies can re-package and sell NWS data and analyses (Johnson 2005).

In the digital age, the private sector argues that the government should have a very limited role in the dissemination of information and offer online services only under limited circumstances “even if private-sector firms are not providing them” and that governments “should generally not aim to maximize net revenues or take actions that would reduce competition” (Stiglitz 2000). The federal government has created policies, notably Office of Management and Budget Circulars A-76, which says that “The Government

should not compete with its citizens” and A-130, which warns federal agencies to avoid duplicating private sector information systems (OMB 1999; OMB 2000).

A recent controversy over USGS water quality data has emerged in committee reports accompanying the House of Representatives 2006 appropriations bills for the Department of the Interior. The report expresses concern that data collection efforts by the USGS are “in direct competition with the private sector.” What the USGS does is to conduct research on a cost-share basis with state and local partners for major regional projects that accomplish everything from removing pharmaceutical and organic waste matter from water supplies to providing new models for pathogen detection. Private sector firms say that the cost-sharing amounts to unfair competition. Many state and local officials say the benefits of working with USGS range from high quality, peer-reviewed, publicly-accessible data to common standards that enable data interoperability and comparability (Christen 2005). Are we really going to be better off if we prevent government agencies from working together? And end up with non-standardized data that may have licensing restrictions?

The potential for huge profits causes information industry moguls to wave the “no competition” banner in ongoing struggles over large and important indexing and abstracting databases. In 2002, despite widespread protest, the government shut down PubScience, an online database indexing peer-reviewed journal literature relating to work supported by the Department of Energy, after complaints that the database duplicated private-sector databases (Tenopir 2002). Now the American Chemical Society wants the National Institutes of Health to shut down its *PubChem* database of small organic molecules claiming that it duplicates the society's fee-based Chemical Abstracts Service (Kaiser 2005). There is concern among scientists and librarians that other scientific databases such as *Agricola* and *PubMed* may be the next to go (Jobe 2003).

Librarians and scholars alike have been watching the competition between two major re-packagers of government information in a race to digitize the Serial Set, a supremely important historical series of congressional documents. Some libraries have licensed these electronic collections. Libraries and the GPO have also expressed an interest in making a public domain version of these documents available. If at some point either of these companies



complain about “unfair competition,” might citizens be deprived of the advantages of digital access?

For many users of government information, these free resources help to level the playing field. Subscriptions to commercial indexing and abstracting databases are notoriously expensive. Individuals who are lucky enough to live near an academic library with public access may be able to use commercial databases, but for everyone else, including public library users, accessing free government-sponsored databases is the way they find out about new research in hazardous chemicals or cutting-edge therapies for diseases or any other life problem they are trying to solve.

### **Solution**

We believe that libraries can be a big part of the solutions to these economic problems. First, let’s not look to the GPO or other federal agencies as the sole solution for long-term preservation, where one bad funding cycle could spell disaster for public access. By insisting that the government deposit digital government publications in Federal Depository Libraries, we return to a system of locating the costs of long-term preservation and access of government information with locally-controlled libraries. Since it is the mission of libraries to preserve and share, we feel this is the best place to look for innovation, cooperation and solutions for the social costs of preserving digital government information (Jacobs, Jacobs, Yeo. 2005).

Some libraries may be reluctant to step up to the plate, but we think this is a tragic mistake. Any way you look at preservation costs for public information, the taxpayer ultimately foots the bill. Far better that decisions about what to keep easily and conveniently available to the public should be made as close as possible to user communities. Far better that libraries, with their long history of collaboration, continue searching for ways to accomplish for their users the same or even better services than they managed in the print world. This is innovation born of necessity, and will happen in an ongoing way over decades. The alternative is watching information disappear overnight when government agencies are hit with funding cuts and literally have no alternative to pulling the plug.

Second, libraries must continue to hold the government responsible for a strong and healthy program of information dissemination. Bruce James, who, as Public Printer, is leading the Government Printing Office’s strategic planning, has characterized

congressional funding for GPO as a “handout” (James 2003). Libraries must not succumb to that kind of cynicism. The cost of funding permanent public access to government information is the cost of democracy. Perhaps libraries should take up a demand for regular federal funding to libraries specifically for making government information available. While it is easy for an agency to quietly stop providing a service or take a few documents off-line without public notice, cutting funds that go to local libraries is more visible and more difficult politically.

Third, libraries and their allies must continue to oppose misguided efforts to sell public information. This is akin to the popular movement against government sales of other valuable assets such as public lands or radio spectrum. It doesn’t save the taxpayers money in the long run—it can only end up by removing public information from the commons. In fact, not a few veteran librarians can tell you stories about staff at government agencies calling them up in hopes of locating at the library their own agency information which they have lost track of. Will agencies who sell their data be able to get it back again free from commercial outfits? Will the commercial outfits even keep this data once they have made their immediate profit from it?

Finally, libraries must continue to expose demands by the information industry that their abstracting and indexing databases be granted monopoly status as just a pretty big grab for profits at the expense of the information commons. It isn’t right that citizens should be treated as “just one of the stakeholders” when it comes to public information—whether it’s about records of government activities or published results of taxpayer-funded research. It may be true that government should not compete with business in producing tomato soup or washing machines. But it is a calculated obfuscation to equate government’s protection of public information with those clearly different kinds of commercial activity.

There is no inconsistency between the government and libraries providing fully functional digital government information for free to the public and the private sector adding value to that government information and creating new information products. But we can take the controversy over National Weather Service data as a warning shot across the bow. Libraries must vigilantly reject any moves by the GPO to provide less than fully functional digital information for free to the public through depository libraries so that companies can profit by merely adding the value

that makes information usable. Perhaps by emulating academia's Open Access initiatives that aim to counter prohibitively expensive journal subscriptions, libraries might find a way to develop a sustainable model for these free abstracting and indexing databases. Rather than dismantling or dumbing-down these databases, perhaps the government should give the data to libraries to see if a sustainable partnership solution can be developed.

## **CONTROL ISSUES**

There is and will probably always be a tension between openness and secrecy, between government control of information and citizen access to and use of information. While publications originally intended for public consumption are happily disseminated far and wide, many documents related to the inner workings of policy and decision-making often get far less distribution. The reasons range very broadly, from concerns for national security to the more lowly short-term self-interest of officials trying to avoid political embarrassment.

Whistleblowers, librarians, scholars, and everyday citizens have at times had to deal with abuse of secrecy aggressively. It's a matter of checks and balances when citizens question or correct the decisions of those who represent them in government. If citizens are to be the highest decision-making body, then they must have the proper information at hand—again a philosophical underpinning that goes back to our founders.<sup>4</sup> Citizens will and should continue to question and correct government. We do not see this as an issue of Republicans vs. Democrats or liberals vs. conservatives. And, though many people have become more aware of government secrecy in the last few years, we don't see it as a post-9/11 phenomenon. Before we go further, we want to note explicitly that we are not proposing a conspiracy theory or asking you to believe that there is a diabolical plot to keep information from you. The control issues we want to examine are much more mundane. Digital information shifts the balance of control to users.

### **Public Access: Full Speed Ahead**

The miracle of digital information is that it can be easily copied, shared, and re-mixed. In fact, scholars, libraries, activists and just ordinary folks are doing that right and left with government information. Consider these projects as just a few of hundreds we could list:

- The National Security Archive at George Washington University, which started as a centralized repository for government information uncovered as a result of FOIA requests, and now is available as an online database.
- The Right-to-Know Network from OMBWatch, which provides for "multi-searching" across many databases having to do with hazardous wastes, chemicals, potential superfund cleanup sites, etc. Much easier for non-experts to use and more comprehensive than EPA's TRI Database [[www.rtknet.org](http://www.rtknet.org)];
- Websites that take advantage of RSS feeds and email newsletters featuring full-text Congressional Research Service reports—the extremely valuable think-tank reports written for members of Congress on currently important topics. These reports have been resolutely kept out of the depository program, but now several groups are preempting that policy by cooperating to gather, copy, and redistribute digital versions of these reports [[www.opencrs.com](http://www.opencrs.com)][<http://www.fas.org/sgp/crs/>][<http://digital.library.unt.edu/govdocs/crs/>];
- The TRACS Project at Syracuse University which compiles hard-to-obtain data from multiple sources and then analyzes it to evaluate the effectiveness and fairness of federal government enforcement of law and regulations [<http://trac.syr.edu>];
- The U.S. Congressional Bibliographies website at North Carolina State University, which also compiles data from multiple sources to provide the most complete listing available of congressional committee meetings and documents emanating from Senate committee meetings [<http://www.lib.ncsu.edu/stacks/senatebibs/>].

Some of you will remember the infamous John Ashcroft/Andrew Card memos which directed federal agencies to deny FOIA requests if there was any shred of legal standing for the denial, thus reversing existing Executive Branch policy. Don't you love it that those very memos that sought to raise the barriers for public access were immediately made accessible over the Internet to millions of people? (Ashcroft 2001) (Card 2002)

Where is all of this going in the future? Imagine government information presented in more usable contexts, like a federal or

municipal budget mounted on a library website, along with links to other relevant government documents, to annotations and explanations, to previous year's figures, to prior appropriations. Or imagine telecommunications laws and regulations presented in ordinary language so that people can actually understand and participate in the civic debates. Think of data mining technologies—not used to look for terrorists, but to find connections between collections of government documents and corporate profits. Advocacy groups will not only send out alerts on topical issues, but also present the full-text of relevant documents. Since these groups will be viewed as trusted third parties by their supporters, authenticity may not be an issue, but if it is, imagine links proxied through library servers where documents can be compared for authenticity and completeness. No more waiting for the GPO to collect, catalog and provide links to government information—we can get it for you this afternoon. Letters and memos and all those documents that were considered outside the scope of the depository program will appear everywhere on the web if even one copy is found by someone with an interest and access to the Internet. The documents associated with whistleblower court cases will be presented and generate comment long before the judges render their decisions. Libraries will capture and store policy documents from agency websites, providing thorough metadata for each version, so that political “spin-doctoring” of those documents is easily detectable and available for scholarly study. And blogging will allow communities to gather, educate themselves, debate the issues, formulate opinions, discover allies, and react—all in a timely fashion.

The possibilities for innovative and beneficial uses of government information are endless. And libraries belong right in the middle of it, experimenting and playing, expanding our collection policies to include documents that are released through FOIA and whistleblowers, creating metadata to make documents findable, hosting services to make documents usable. We are, hopefully, the ones who can organize and help provide access to the thousands of useful civic web projects, and we should also see ourselves as the potential preservers of these useful web-publishing efforts.

We do not see libraries as the only source of government information or librarians as intermediaries standing between users and information. We see libraries enabling more sharing and more use and re-use and making it easier for more people to find more

information more easily. We see libraries as places where government information, scholarly information and commercial information are available together rather than being separated into separate domains.

But oh my, are things getting “out of control?” Will there be a reaction to all this increased scrutiny of government activity? Of course there will be and again, we say, not necessarily due to a conspiracy. There’s an old saying about never ascribing nefarious planning and coordination to actions or events that can just be chalked up to bureaucratic short-sightedness or bumbling inefficiency.

And certainly government is not a monolith of secrecy-loving bureaucrats. Initiatives from government that also want to take advantage of digital technologies range from the National Institutes of Health’s original plans for requiring open access to publicly-funded scientific research to the Census Bureau’s digitization of historical census data.

But we have to talk about the darker side.

### **Control Mechanisms Threaten Public Access**

The classic way of trying to keep important information away from the public is to not release it. Internal documents, memos, minutes, and so forth, in the pre-digital world could be kept in offices and never published. In a kind of one-two punch, recent policies that heighten barriers for FOIA requests make it difficult for citizens to get the “first” copy, while regulations that threaten ISPs with liability for any information posted by their customers that is deemed “sensitive” create a chilling effect that will inhibit public sharing. In recent years the number of people who can classify documents as “secret” has increased drastically, now including Environmental Protection Agency, the Department of Health and Human Services, and even the Department of Agriculture. The President has expanded authority to reclassify declassified information. Categorizing information as “sensitive not classified” is increasing and at least one court case has found this power abused (Aftergood 2004). Even telephone books and non-government information such as almanacs have been labeled sensitive in this extreme climate. Even if each and every decision toward secrecy is made with the best of intentions vis a vis national security, all of this is going on in the absence of any meaningful citizen participation, despite concern that locking down information may harm public safety, environmental protection,

scientific progress, government accountability, electoral fairness, and civil liberties. Do we want such a radical change in our civic life without more widespread debate?

Government officials claim that increased secrecy is necessary not only to prevent individual “sensitive” documents from being seen on the web by the wrong kind of people, but also to prevent terrorists from making use of aggregate digital government information in ways that would have been impossible with print publications. Without a doubt, many citizens who feel uncomfortable with the increased secrecy also feel that they lack the technical expertise to evaluate the government’s claims or to take a strong stand in this political controversy.

In cases where the government will be politically unable to keep documents utterly and completely secret, there is still plenty of potential for maintaining control over how they can be shared and used.

While we have not yet seen widespread adoption of DRM technologies by government, we may see more soon. Some agencies refuse access to their web sites without passwords or registration; the Government Printing Office is examining techniques to assign “end user rights and privileges” (GPO 2005); private companies are marketing technologies to the government that will control access to government websites and “manage citizens’ online identity” (Gilbert 2005). Control can be asserted over digital information by agencies allowing documents to be read only on certain machines, or withdrawing permission to read after a document is downloaded, or releasing text documents only as images so that the text cannot be copied, and so forth. Librarians have repeatedly asked the Government Printing Office if it will refrain from using this kind of technology, but have yet to receive a reply.

Another control mechanism stems from the chilling effect of the government knowing who is reading what. From the USA Patriot Act, we already know that the government is interested in tracking use of libraries and bookstores. One recent editorial in support of the Patriot Act tellingly noted, “Books can also reveal what people are thinking” (Gelernter 2005). This kind of spying accompanied technological developments at least as early as 1987 when the FBI attempted to monitor who was using online commercial newspaper databases (Human Rights Watch 1992).

It will be doubly bad for democracy if the government uses technology to investigate the reading habits of citizens and develop programs like Total Information Awareness to mine public data about all of us, but prevents us from using the same technology to copy, share, and re-use government information and to mine data about government activities. When digital government information is available only from government-controlled web servers, it is very easy for the government to covertly collect information about readers of government documents. Citizens could feel a lot more comfortable accessing reports on intelligence failures, treatment of political prisoners, or other controversial topics if they knew they were accessing library copies, because libraries have traditionally been strong protectors of user privacy.

And in the category of “control due to inefficiency” we have to describe the problems citizens face when the government fails to live up to its own promise to deliver government information. “Fugitive documents,” those documents that are missed by the GPO for dissemination via the depository program, have grown in leaps and bounds since the days when government agencies were required to use the GPO for print publishing. Digital fugitive documents may be easy to discover on the web in the short term, but lacking bibliographic control and organized management, they are the most at risk to disappear from public view in the long run. Studies report that the information on government web pages is far from permanent, changing on average about every four months and information routinely disappears from agency pages (Cho and Garcia-Molina 1999. Lopresti and Gorin 2002).

High profile web sites like [whitehouse.gov](http://whitehouse.gov) or [science.gov](http://science.gov) are laudable efforts to help citizens get access to publications, but publications mounted there lack the long range endurance of publications gathered into the depository program. Libraries understand this, ordinary citizens may not.

### **WHAT IS TO BE DONE?**

Throughout this paper, we have mentioned libraries as part of the solution. In order to fulfill their original social function of providing access to government information, libraries need to take on new activities in the digital era. Here’s a summary of ways that libraries can help:

1. Support citizen innovations in making digital government information available, Help people find these great websites, make sure they’re preserved, explore ways to



build on them. These activities could help to prod government in the direction of more friendly user interfaces, more useful publications and who knows what else.

2. Investigate ways to capture and include documents in library digital collections that have traditionally been outside the scope of the depository program, but which are crucial to citizen engagement.
3. As always, watchdog policy changes masquerading as technological developments. Oppose DRM policies that could enable government agencies to restrict access or charge fees for public information, and continue to demand government publications free of proprietary constraints that will lead to access problems in the future.
4. Understand and explain to our user communities the consequences of selling digital government information, or of allowing government's dissemination of our information to be decried as unfairly competing with the private sector.
5. Continue to deepen relationships with other groups defending the public's right to know. In this vein, a group of librarians has created a website ([freegovinfo.org](http://freegovinfo.org)) to facilitate collaboration among libraries, government agencies, non-profit organizations, researchers, journalists, historians, economists and others who have a stake in the preservation of and perpetual free access to government information.
6. Perhaps most importantly, continue to insist on a strong and well-funded depository program that deposits digital documents with libraries. This protects reader privacy. It also allows libraries many opportunities for streamlining discovery and improving presentation of government information. It allows us to dig in and experiment with better ways to manage digital documents and spurs cooperation and innovation as efforts are made to cope with the cost of long-term preservation.

Holding government accountable for disseminating government information, and at the same time ensuring that government does not have exclusive, unnecessary, or abusive control of government information is a challenge in any era. However, if we don't stake

our claim to keeping government information in the public domain, the private sector's desire for profits and government agencies' varied agendas could drive the government into making the worst possible decisions. What should remain in the public domain is a question for citizens to decide—not to have it decided for them. Choosing how society will best benefit from digital technologies and free culture means getting our arms around all kinds of intellectual content, whether it's the latest publicly-funded scientific research or a decades-old folksong, because as we have seen, policies and attitudes that affect one type of content will also affect other content. New technologies create new opportunities, so let's not set our sites too low. Our struggle now should not be simply to maintain what we have enjoyed in the past, but to expand and enlarge meaningful access to our nation's public information.

## ENDNOTES

1. "He has called together legislative bodies at places unusual, uncomfortable, and distant from the depository of their public Records, for the sole purpose of fatiguing them into compliance with his measures."
2. We should mention, of course, that not all government information is copyright-free. For example, building codes for most states are privately published and demonstrate the outlandish consequences for citizens when the law itself is copyrighted (Balint 2001). This is a serious legacy problem, but not one that we will concentrate on here. Instead we'll look at the wealth of government information that is not regulated by copyright.
3. GPO does deposit so-called "tangible" digital materials, e.g., CD-ROMs and DVDs, perhaps for reasons of convenience.
4. Thomas Jefferson expressed himself on these ideas: "The information of the people at large can alone make them safe as they are the sole depository of our political and religious freedom." –Thomas Jefferson to William Duane, 1810. "The diffusion of information and the arraignment of all abuses at the bar of public reason, I deem [one of] the essential principles of our government, and consequently [one of] those which ought to shape its administration." –Thomas Jefferson: 1st Inaugural Address, 1801. Both quotations from *The Writings of Thomas Jefferson* (Memorial Edition) (Lipscomb 1903).

**REFERENCES CITED**

- Aftergood, Steven. 2004. "Court Blasts 'Frivolous' Withholding of Info By FBI, TSA." *Secrecy News*. 30 June 2004.  
<http://www.fas.org/sgp/news/secrecy/2004/06/063004.html>  
(Accessed June 29, 2005).
- American Library Association. 1988. *Less Access to Less Information by and about the U.S. Government: A 1981-1987 Chronology*. February 1988. Washington, DC: American Library Association.
- American Library Association. 1992. *Less Access to Less Information by and about the U.S. Government: A 1988-1991 Chronology*. February 1992. Washington, DC: American Library Association.
- Ashcroft, John. 2001. Memorandum for the heads of all federal departments and agencies. Subject: The Freedom of Information Act. <http://www.gwu.edu/~nsarchiv/NSAEBB/NSAEBB84/Ashcroft%20Memorandum.pdf> (Accessed June 28, 2005).
- Balint, Kathryn. 2001. "Public laws owned by the public? Think again, copyright rulings show." *San Diego Union-Tribune*, 13 May.
- Card, Andrew H. 2002. Memorandum for the heads of all federal departments and agencies. Subject: Action to Safeguard Information Regarding Weapons of Mass Destruction and Other Sensitive Documents Related to Homeland Security. <http://www.fas.org/sgp/bush/wh031902.html> (Accessed June 28, 2005).
- Cho, J., Garcia-Molina, H. 1999. *The Evolution of the Web And Implications For An Incremental Crawler*. December 1999. <http://rose.cs.ucla.edu/~cho/papers/cho-evol.pdf> (Accessed June 30, 2005).
- Christen, Kris. 2005. Congress seeks to privatize some USGS water activities. *Environmental Science and Technology*, 8 June 2005. [http://pubs.acs.org/subscribe/journals/esthag-w/2005/jun/policy/kc\\_congress.html](http://pubs.acs.org/subscribe/journals/esthag-w/2005/jun/policy/kc_congress.html)
- Doctorow, Cory. 2005. "Go Ask Hollywood - Why can't you back up your DVDs? Because entertainment execs don't want you to." *Popular Science* 266, no. 1: 40.
- Dowling, S.A. 1994. "Information Access: Public Goods or Private Goods." *Social Science Computer Review* 12 (Fall): 333-350.
- Fields, Reginald. 2005. "Public records not online are still available, for a fee." *Plain Dealer* (Cleveland), 22 January. Metro, p. B5.
- Gelernter, David. 2005. "You Can't Judge a Book by Its Cover, but We May Find a Terrorist by What He Reads." *Los Angeles Times*. 24 June.

- Gilbert, Alorie. 2005. "HP aims to help governments check IDs." *CNet News.com*. 26 May.  
[http://news.com.com/HP+aims+to+help+governments+check+IDs/2100-7348\\_3-5722206.html](http://news.com.com/HP+aims+to+help+governments+check+IDs/2100-7348_3-5722206.html) (Accessed June 29, 2005).
- Gross, Daniel. 1991. "Byting the hand that feeds them: information vendors are robbing the government blind." *Washington Monthly*. November 1991.
- Human Rights Watch. 1992. "Library Awareness Program. Electrifying Speech," *New Communications Technologies and Traditional Civil Liberties*. July 1992.  
[http://www.eff.org/legal/?f=electrifying\\_speech.paper.txt](http://www.eff.org/legal/?f=electrifying_speech.paper.txt) (Accessed June 29, 2005).
- Jacobs, James A., James R. Jacobs, and Shinjoung Yeo. 2005. "Government Information in the Digital Age: The Once and Future Federal Depository Library Program." *Journal of Academic Librarianship*. May 2005. <http://ssdc.ucsd.edu/jj/fdlp/> (Accessed June 30, 2005).
- James, Bruce. 2003. "Fall Council/Depository Conference Meeting October 19-22, 2003, Arlington, VA Remarks by Bruce R. James, Public Printer of the United States, Tuesday October 21, 2003," [http://www.access.gpo.gov/su\\_docs/fdlp/council/ney\\_james03conf.html](http://www.access.gpo.gov/su_docs/fdlp/council/ney_james03conf.html) (Accessed June 29, 2005).
- Jobe, Margaret M. 2003. "Another casualty." *Library Journal* 128: 9, p.54
- Joffe, Bruce. 2003. 10 ways to support GIS without selling data. Open Data Consortium Project.  
<http://www.opendataconsortium.org/documents/10Ways2SupportGIS-3.pdf> (Accessed June 28, 2005).
- Johnson, Neil. 2005. "Senate bill hobbles weather service." *The Tampa Tribune*, 28 May. Nation/World, p.1.
- Kaiser, Jocelyn. 2005. "Chemists want NIH to curtail database." *Science* 308 (6 May): 774.
- Lipscomb, Andrew and Albert Bergh, eds. 1903. *The Writings of Thomas Jefferson*. Memorial edition. Washington, DC: Thomas Jefferson Memorial Assn. of the U.S.
- Lopresti, Robert and Gorin, Marcia. 2002. "The Availability of U.S. government depository publications on the World Wide Web." *Journal of Government Information* January/February 2002.
- McIntosh, Toby. 1990. *Federal Information in the Electronic Age: Policy Issues for the 1990s*. Washington, DC: Bureau of National Affairs, Inc.

- McMullen, Susan. 2000. "U.S. government information: selected current issues in public access vs. private competition." *Journal of Government Information* 27, no. 5 (September/October): 581-593.
- Mitchell, Alison. 2001. A nation challenged: computer security. To forestall a 'Digital Pearl Harbor,' U.S. looks to system separate from internet. *New York Times*, 17 November, late edition-final, p.7.
- Newman, Nathan S. 1998. Net Loss: Government, Technology and the Political Economy of Community in the Age of the Internet. Ph.D. diss., University of California, Berkeley.  
<http://www.nathannewman.org/diss/chap7.html> (Accessed June 29, 2005).
- Office of Management and Budget. 1999. "Performance of Commercial Activities," *OMB Circular A-76* (August. 4, 1983, Revised 1999).  
<http://www.whitehouse.gov/omb/circulars/a076/a076.html> (Accessed June 21, 2005).
- Office of Management and Budget. 2000. "Management of Federal Information Resources," *OMB Circular A-130 Revised*, (Transmittal Memorandum No. 4, 11/28/2000) <http://www.whitehouse.gov/omb/circulars/a130/a130trans4.html> (Accessed June 21, 2005).
- Read, Brock. 2005. "Recording and Movie Industries Sue Internet2 File Swappers." *Chronicle of Higher Education* 51 (22 April): 37.
- Russell, Judith C. 2004. Remarks of Superintendent of Documents Judith C. Russell, Depository Library Conference, Washington, D.C. October 18, 2004. *Administrative Notes: Newsletter of the Federal Depository Library Program*. 15 October.  
<http://www.gpoaccess.gov/about/speeches/DLCFall2004.pdf> (Accessed June 28, 2005).
- Schmeckebier, Laurence F. and Roy B. Eastin. 1969. *Government Publications and their Use*. Washington, D.C.: The Brookings Institution, 1969
- Schrage, Michael. 1986. "U.S. limits access to information related to national security." *Washington Post*, 13 November.
- Sewell, Chris. 2002. "One network, under Gov." *Telephony* 242, no. 1 (7 January): 30-35.
- Sidman, David. 2005. The Digital Object Identifier (DOI): The Keystone for Digital Rights Management (DRM): Draft submitted to SIIA Digital Rights Management (DRM) Working Group, January 26, 2001. <http://www.contentdirections.com/materials/SIIA-DOIandDRM-DavidSidman.htm> (Accessed February 4, 2005)
- Stiglitz, Joseph., Peter R. Orszag, and Jonathan M. Orszag. 2000. The Role of Government in a Digital Age. Commissioned, Computer & Communications Industry Assn.  
<http://globalchange.gov/policies/role-gov.html> (Accessed June 20, 2005).

- Tenopir, Carol. 2002. "Disappearing databases." *Library Journal* 127, no. 20 (15 December): 38-39.
- United States Government Printing Office. 2004. Managing the FDLP Electronic Collection: A Policy and Planning Document, Council Discussion, Questions and Answers from the Fall Meeting 2004: Managing the FDLP Electronic Collection: A Policy and Planning Document. 19 April 2005. [http://www.access.gpo.gov/su\\_docs/fdlp/pubs/proceedings/04cbt/managingfdlp.html](http://www.access.gpo.gov/su_docs/fdlp/pubs/proceedings/04cbt/managingfdlp.html) (Accessed June 29, 2005).
- United States Government Printing Office. 2004. *A strategic vision for the 21st century*. Washington, D.C.: 1 December 2004. <http://www.gpo.gov/congressional/pdfs/04strategicplan.pdf> (Accessed December 13, 2004).
- United States Government Printing Office. 2005. Requirements Document (Rd V1.0) For The Future Digital System (FDsys) Final. 18 May 2005. [http://www.gpo.gov/projects/pdfs/FDsys\\_RD\\_v1.0.pdf](http://www.gpo.gov/projects/pdfs/FDsys_RD_v1.0.pdf) (Accessed June 29, 2005).
- Washington Internet Daily. 2002. White House holds cybersecurity comments from public view. *Washington Internet Daily*. 24 September 2002.

# Will Fair Use Survive the Digital Age?

Marjorie Heins<sup>1</sup> (NYU)


**Abstract:** Fair use is a critical free-expression safeguard in copyright and trademark law. As the Supreme Court recently recognized, fair use assures that the public can borrow “not only facts and ideas contained in a copyrighted work, but also expression itself in certain circumstances”—among them, criticism, comment, news reporting, and scholarship (Eldred vs. Ashcroft, 2003). But the coming of the digital age, combined with a tilt in law and public policy toward increased control by owners of “intellectual property,” now poses a major challenge to fair use as a vehicle for free expression and the growth of the digital library. “Cease and desist” letters from copyright and trademark owners, “take-down” letters under the Digital Millennium Copyright Act (the DMCA), and demands for “broadcast flags” and copyright filters on peer-to-peer software are just some of the developments that threaten the full exercise of fair use rights. After giving an overview of the dilemmas confronting fair use and free expression online, this paper describes empirical research conducted by the Free Expression Policy Project at the Brennan Center for Justice on the attitudes and experiences of artists, scholars, Web publishers, and others with copyright, fair use, and similar conflicts under trademark law. Stories from filmmakers, art historians, authors, and visual artists demonstrate the very real dilemmas that they face in trying to find out what fair use means, how to take advantage of it, and how to persuade publishers, distributors, insurers, and others to resist the sometimes overzealous industry enforcement of IP controls. The paper concludes that although many people are aware of their rights and do resist cease and desist letters or DMCA take-down notices, there is an urgent need for more information, better community support and legal backup, and statements of “best practices” by filmmakers and other artists that can be used to make fair use a reality in the digital age.

## INTRODUCTION

Fair use is probably the best-known of the free expression “safety valves” in copyright law. It allows anyone to copy, publish, and distribute parts of a copyrighted work without permission, for purposes such as commentary, news reporting, education, or

---

M. Halbert (Ed.): *Free Culture and the Digital Library Symposium Proceedings*. Atlanta, Georgia: MetaScholar Initiative at Emory University, 2005. pp. 179-217.

 This work is covered by the Creative Commons Attribution-NonCommercial-NoDerivs 2.5 License.

scholarship. Fair use not only encourages the creation of new, “transformative” works; it also allows criticism and parody of the myriad products of our past and present culture.

Fair use is thus critical for free expression and cultural life. If permission were required for every quote or new creation that borrows from works of the past, the costs and logistical difficulties of finding owners, seeking licenses, and paying their sometimes exorbitant costs would dramatically dampen political and artistic discourse. And equally important, copyright owners would be able to enforce orthodoxy simply by denying permission to quote a text or image to any author whose views they disapproved.

As the Supreme Court has recognized, every work of “literature, science and art borrows, and must necessarily borrow, and use much which was well known and used before.”<sup>2</sup> The fair use doctrine guarantees “breathing space within the confines of copyright,” and affords necessary “latitude for scholarship and comment.”<sup>3</sup>

But the digital age, bringing with it a tilt toward increased control by owners of “intellectual property” (or “IP”), now poses a major challenge to fair use, and to related doctrines that protect free expression under trademark law.<sup>4</sup> The employment of “digital rights management” to restrict access to and copying of cultural products; the Digital Millennium Copyright Act’s “take-down” procedure for alleged violations of copyright; and initiatives such as the “broadcast flag” all threaten the public’s ability to exercise fair use for material in digital form. Added to these hurdles are the inherent unpredictability of fair use, the high cost of defending it in court, and the crushing liability that may result if one guesses wrong. Industry practices also contribute to the problem—a “clearance culture” that assumes the necessity of licensing everything, and the common use of cease-and-desist letters to threaten artists, parodists and others who rely on fair use with dire punishments for copyright infringement.

In reaction to these developments, some free expression advocates have essentially given up on fair use, and have focused instead on alternative ways of combating an overly controlled culture. Initiatives such as Creative Commons,<sup>5</sup> which provides sample licenses that copyright owners can adopt to permit freer use of their works, are a wonderful innovation and show that alternatives to heavy copyright control are possible. But, of course, they depend on the willingness of the copyright owner. Fair use is an



irreplaceable element of our intellectual property system precisely because it *does not* depend on the owner's permission. It is particularly important today, when copyright seems to go on forever: life of the author plus 70 years, or 95 years for corporations. Although other approaches are welcome and desirable, fair use is a critical safeguard that needs to be strengthened, not abandoned.

This paper will first present an overview of the challenges confronting fair use as an essential element of free culture in the digital age. It will then summarize the results of empirical research conducted by the Free Expression Policy Project of the Brennan Center for Justice (FEPP) over the past year to gather information about the experiences of artists, scholars, and Web publishers when confronted with conflicts over copyright or trademark rights, on the one hand, and free expression or fair use on the other.<sup>6</sup> The paper will conclude with a discussion of policy changes that might strengthen and clarify fair use and free expression in the digital world.

### **DRM, THE DMCA, AND OTHER THREATS TO FAIR USE ONLINE**

The arrival of digital technology immediately created quandaries for copyright law. Because every online transfer is technically a new copy, some intellectual property owners argued that merely accessing a document on the Internet should be seen as a copyright violation. This was also the view of the Clinton Administration in a 1994 "Green Paper," produced largely in response to industry concerns about the potential for widespread copying and sharing of books, articles, movies, music, and virtually any other expression online. In her book *Digital Copyright*, Jessica Litman describes the dismay that the Green Paper caused among librarians, authors, online service providers, and makers of electronic devices and computer hardware. Public hearings were held, but in the end, the Administration did not fundamentally alter its views. The main difference between its initial Green Paper and its final White Paper on the subject was that the White Paper did not propose major changes in the substance of copyright law. Rather, says Litman, it interpreted existing law to assert that "most of the enhanced protection copyright owners might want was already available."<sup>7</sup>

This approach held serious dangers for fair use online. If one cannot even access a document without permission, then copying it for critical or other transformative purposes, or even making a

backup copy for home use, becomes unlawful. Yet these are well-recognized examples of fair use in the non-digital environment.

### **The Perils of DRM**

At the same time that public policy for the digital age was developing, private companies were introducing digital rights management, or “DRM,” techniques such as encryption to prevent the unauthorized sharing or copying of their movies, texts, software, and other products. DRM rests on an assumption of absolute control over both access and use, an assumption that is fundamentally at odds with the limited control built into copyright law. It undermines not only fair use, but open access to works that are in the public domain. And it threatens the first sale rule, which allows the purchaser of a book or other work to give, lend, or sell it to another.<sup>8</sup> Such common features of cultural life as bookstore, art gallery, and library browsing become difficult if not impossible in a DRM-dominated digital world.

Along with DRM came both “shrink-wrap” and “click-wrap” licenses that further undermine fair use. Shrink-wrap licenses are unsigned agreements that product manufacturers enclose inside the cellophane wrapping and that specify the conditions for use. Click-wrap licenses—those scrolls of legalese that one must accept in order to reach the desired content—are the online equivalent. These “take it or leave it” contracts not only impose restrictions that are inconsistent with the flexibility of copyright law; some of them flatly require viewers to relinquish their fair-use rights.<sup>9</sup>

### **The DMCA**

Having created DRM techniques, the media industry now looked for ways to secure them against the ingenuity of scholars, engineers, and Web hackers who might circumvent them. The next step in locking up content, therefore, was the 1998 Digital Millennium Copyright Act (the DMCA), a product of both the Administration’s White Paper and Congress’s desire to help the industry enforce digital rights management. The DMCA gives the force of law to DRM by criminalizing users’ efforts to circumvent any “technological measure that effectively controls access to a work.” It also criminalizes the creation and distribution of circumvention tools.<sup>10</sup> That is, it goes beyond penalizing copyright infringement to prohibiting research and communication that might be used for infringement—but also might be used for legitimate purposes such as fair use.<sup>11</sup>

Courts have recognized that the DMCA has the potential radically to shrink fair use in the digital era, but have upheld the law nonetheless, basically on the theory that Congress created the copyright system and can therefore alter it at will.<sup>12</sup> What these courts have ignored is that fair use is a constitutionally grounded principle—that, as the Supreme Court has said, if it did not exist, then serious First Amendment problems would arise because the intellectual property system would not be giving adequate breathing space to free expression.<sup>13</sup>

Another section of the DMCA is equally if not more troublesome for free expression and fair use. Section 512 of the law provides that in order to escape possible copyright liability, Internet Service Providers (ISPs)—including search engines—must “expeditiously” remove any material on their servers that a copyright owner tells them is infringing. The sender of the notice need only assert “a good faith belief;” no legal judgment is needed.<sup>14</sup> Although §512 does allow those targeted by take-down letters to send a counter-notice contesting the allegation of infringement, the legal requirements of the counter-notice are detailed and technical, and not every Web publisher or blogger is able to make use of the procedure.<sup>15</sup> Non-subscribers—for example, those contributing to ongoing discussions in newsgroups—may not even be aware that their words, images, or links have been removed.

Section 512 is misguided because ISPs should not be liable for copyright infringement in the first place, simply because someone is using their servers to infringe. ISPs are akin to telephone companies or highways: they provide a means of transport, whether for information or for tangible goods. Wrongs committed by users of telephones or drivers on highways are not attributed to those providing the conduits. Section 512 ignores this sound principle in the interests of helping copyright owners protect their rights, but it provides an insufficient check on overreaching, and creates an unacceptable shortcut around the procedures that are needed to decide whether speech is actually infringing.

The §512 procedure is thus a due process-free favor to corporate copyright holders whose squads of lawyers can churn out take-down notices that suppress the allegedly infringing content without any court having ruled on the matter. In some of these situations, the content will have a strong or at least reasonable likelihood of being fair use. Later in this paper, I give some examples, gleaned from FEPP’s fair use research.

### **The Broadcast Flag and P2P Filters**

Another development that threatens fair use is the entertainment industry's push for a "broadcast flag"—a mechanism enabling electronic devices that can receive digital broadcast signals—including TVs, radios, computers, DVD players, and video recorders—to recognize a code created by copyright owners. Once the flag recognizes the code, it can prevent recording (and potentially sharing) of the content. The Congressional Research Service has warned that the broadcast flag would hinder fair use—for example, when a consumer wants to record a program to watch at a later time or a different location, or when a student wants "to email herself a copy of a project with digital video content."<sup>16</sup>

In 2002, the industry persuaded the Federal Communications Commission to mandate a broadcast flag in all devices manufactured on or after July 1, 2005 that are capable of receiving digital TV signals. But in May 2005, the U.S. Court of Appeals voided the mandate, ruling that the FCC had waded too deeply into copyright matters that are beyond its authority.<sup>17</sup> The Motion Picture Association of America promptly turned to Congress to press for a mandatory broadcast flag.<sup>18</sup>

Next-of-kin to the broadcast flag are proposals to require that peer-to-peer (P2P) file-sharing software incorporate filters to prevent the duplication of copyright-protected works. File-sharing, especially of music, has been the leading complaint of the entertainment industry almost since the inception of the Internet. It led to a successful lawsuit shutting down Napster as a contributory copyright infringer, and a second suit which is likely to shut down Napster's successors—Grokster, KaZaa, and similar programs—which, in an effort to avoid liability, dispensed with the central website and indexing that doomed Napster.

In June 2005, the Supreme Court reversed a court of appeals decision that these second-generation products do not in themselves create liability for contributory copyright infringement. The justices ruled unanimously that the efforts of the defendants—Grokster and StreamCast—to attract Napster's former customers, and to profit from illegal file-sharing through advertising, probably made them guilty of inducing copyright infringement. One of the acts that seemed damning to the justices was the defendants' failure to make "an effort to filter copyrighted material from users' downloads."<sup>19</sup> The Court had received friend-of-the-court briefs from several filter manufacturers touting their wares and

suggesting that P2P creators should be required to install their products in order to minimize unlawful copying.<sup>20</sup>

But even putting aside the possible malfunction of these filters, they are likely to be overinclusive, blocking public domain along with copyright-protected materials, and preventing access for purposes of fair use. The filters function much like the broadcast flag—that is, they prevent copying of any work that is embedded by its maker with a copy-protection code. Although the Court did say that “in the absence of other evidence of intent, a court would be unable to find contributory infringement liability merely based on a failure to take affirmative steps to prevent infringement,”<sup>21</sup> there is a danger that creators of P2P software will start embedding filters in an effort to steer clear of possible copyright liability. Like the DMCA’s §512 take-down procedure, this development would distort the property rights/fair use balance by allowing IP owners to control all uses of their products, thus precluding or radically reducing opportunities for fair use.

### **Digital Libraries**

Another conflict between monopoly control and the free expression safety valves built into copyright law arose as a result of plans announced in December 2004 by Google, the New York Public Library, and four major university libraries to create an online archive, to include both public domain works and those still covered by copyright. The New York Public Library said it would allow the scanning and posting only of public domain material; Harvard said it would submit a relatively small sample of 40,000 books; but the University of Michigan and Stanford both announced that they would allow computerization of all their holdings.<sup>22</sup>

As one newspaper commented, creating digital archives and libraries is hardly a new idea; websites such as Project Gutenberg have been offering “virtual shelves of e-books” for years, but only for items in the public domain.<sup>23</sup> Conflicts between copyright owners and public libraries also date back at least to 2002, when librarians’ excitement at the potential of the Internet clashed with publishers’ desire to lock up access to copyrighted works.<sup>24</sup> What was new in 2004 was the ambitious scope of the Google project, combined with the money and technology to make it happen.

Publishers soon objected. A letter to Google from the Association of American University Presses in May 2005 asserted that “the

plan appears to involve systematic infringement of copyright on a massive scale.” Google, seeming to back off from its original announcement, replied that for books still in copyright, users would see only “bibliographic information and a few sentences of text.” Google wrote: “Although we believe there are many business advantages for publishers to participate in Google Print, they may opt out, and their books scanned in libraries will not be displayed to Google users.”<sup>25</sup>

Did Google, legally, have to make this concession? Thanks to a combination of fair use and first sale doctrines, books can be freely read—in their entirety—in offline libraries; and parts can even be copied. Now that the world has gone digital, does it make sense to prevent online libraries from fulfilling the same function and reflecting the same copyright balance? Even the New York Public Library, which is giving only public domain material to Google, announced in June 2005 that it was making 700 audio books—“from classics to current best-sellers”—available online for downloading onto computers, CDs, and portable listening devices. Users can borrow up to 10 digital books at a time from the library’s website. An official explained: “Library users today are much more technologically sophisticated than ever; our aim is to continue to provide our users with free access to materials in whichever format they prefer.”<sup>26</sup>

### **Cease and Desist Letters**

One final, and longstanding, fair use concern is the industry’s practice of sending threatening “cease and desist” letters to those making use of copyright or trademark-protected works. In many cases, the recipients of these letters might qualify for First Amendment protection or the fair use defense. Cease-and-desist letters, suffice it to say, do not mention this possibility.

To take one example from the cultural realm, the enthusiasm of movie, book, and music fans is often squelched by copyright owners through cease-and-desist letters. In the late 1990s, Fox Broadcasting succeeded in shutting down more than half of *The Simpsons* fan sites that were listed on one network.<sup>27</sup> Warner Brothers has suppressed sites containing irreverent parodies of such *Looney Tunes* favorites as Bugs Bunny, Daffy Duck, and Tweety, claiming that these “beloved characters” should not be maligned by lascivious humor.<sup>28</sup> Yet irreverent and unauthorized commentary is at the very heart of fair use.

None of this litany of fair use threats and woes is meant to deny or trivialize the fact that a great deal of genuine copyright and trademark infringement is going on, and that digital technology has exacerbated the problem. Digitization makes wholesale copying easier, more pervasive, and more visible than in the past. Copyright owners have the means of identifying and squelching much infringing activity online, but the existence of mirror sites and civil disobedience campaigns, and the relative ease of finding a new ISP and re-posting material that has been subject to a §512 take-down letter create formidable challenges to enforcing copyright online at the same time that they afford opportunities for circumventing overzealous enforcement that threatens free speech. The challenge for public policy is to identify means of copyright and trademark enforcement that are effective but not overbroad—the IP system was never intended as leakproof—and that truly protect fair use. It was toward this end that the Free Expression Policy Project began to conduct empirical research on how well fair use and related First Amendment defenses are actually working for artists, scholars, and other contributors to the worldwide digital library.

### **OVERVIEW OF THE FAIR USE RESEARCH PROJECT**

We used several methods to gather evidence of how well fair use is working. First, we examined all 332 letters in the “Chilling Effects” Clearinghouse database for one typical year, 2004. Chilling Effects began in 2002 in response to concerns that cease-and-desist and take-down letters were being used to squelch legitimate speech. Its founders created a database of such letters, submitted by members of the public or ISPs, and accompanied them with brief explanations of the relevant law.<sup>29</sup>

Of the 332 letters, 320 related to copyright, trademark, or related intellectual property issues.<sup>30</sup> Eighty-two percent, or 263 of them, were DMCA take-down notices received by ISPs or search engine companies.<sup>31</sup> The remaining 18 percent or 57 letters, were cease-and-desist demands sent directly by copyright or trademark owners to individuals or organizations that they claimed were violating their IP rights.

We analyzed our sample of 320 letters in several ways. First, we divided the letters into five categories: (1) those that seemed to state legitimate claims for copyright or trademark infringement; (2) those that, by contrast, seemed to state weak or nonexistent claims; (3) those that targeted expression with a strong claim to fair use, or

an analogous First Amendment defense under trademark law; (4) those targeting speech with less strong, but still reasonable claims to a fair use or First Amendment defense; and (5) those with possible fair use or First Amendment protection, but where more information was needed to make a judgment.

Whenever possible, we viewed the material that was targeted in order to make a decision about the legal strength of the copyright or trademark owners' assertions and the likelihood of a successful fair use or First Amendment defense. In cases where the targeted material was no longer accessible, we made our best judgment based on existing information, which included, in some instances, telephone interviews or responses to an online survey. Where the available information was insufficient to make a judgment, but we thought there was a possibility of a fair use defense, we used the fifth category: "possible fair use—more information needed."

In making our analysis, we adhered to the legal precedents governing fair use. Of course, it is notoriously difficult to make predictions about fair use based on these precedents because each case turns on its own specific facts, and the four factors that courts must consider in making a determination are open-ended and malleable.<sup>32</sup> Nevertheless, there were some clear differences among the materials targeted by the 320 letters. The major criterion that we used was whether the targeted material involved commentary, criticism, or other transformative use of the copyright or trademark-protected work. The other examples given in the fair use statute—scholarship, research, news reporting, and classroom use—were equally important, but less frequently encountered in our 320-letter sample. We also considered how much of the copyright-protected material was copied or quoted, and how much of it was commercial or noncommercial, where we could make this determination.

The ultimate legal result in any of these controversies, had they gone to court, would have turned on a much more detailed set of facts than we could assemble in evaluating 320 separate cease-and-desist or take-down letters. Our judgments are not definitive legal predictions. But it is not difficult, in many situations, to form a conclusion about the likely strength of a copyright claim or a fair use defense simply by looking at the material in question. Our evaluations are useful, therefore, in understanding the extent to which cease-and-desist and take-down letters may chill free expression.



In the first of our five categories, legitimate copyright or trademark claims, we placed many of the letters complaining about copying by commercial competitors. These seemingly legitimate claims ranged from a complaint by the Bhaktivedanta Book Trust that another site was displaying much of its artwork without permission<sup>33</sup> to Ovulation-Calendar.com's protest that a competing manufacturer had copied its "texts, descriptions, interfaces, images, and computer codes."<sup>34</sup> Other apparently legitimate claims involved texts describing architectural designs, National Health Scotland's "ReadySteadyBaby" website, PaddleAsia Company's description of its birdwatching tours, poetry texts used on posters, and Web links to unauthorized copies of photographs owned by the Brazilian version of *Playboy* magazine.<sup>35</sup>

Weak or nonexistent trademark and copyright claims, our second category, commonly involved assertions of control over nondistinctive phrases, or situations in which there was little or no likelihood of confusion in the use of trademarked words or images. Where we found weak or nonexistent claims, we did not move on to a further analysis of whether the targeted words or images might qualify for a fair use or First Amendment defense. We found 37 cease-and-desist or take-down letters in this "weak or nonexistent claim" category. Examples included the use of common terms like "penisimprovement" or "Pet Friendly" in a way that would not likely cause consumer confusion.<sup>36</sup>

Our third category—letters targeting speech with a strong element of a fair use or a strong First Amendment defense—generally consisted of political or cultural commentary. Examples included a parody of the *New York Times's* online corrections page, a site mocking American Express called "American Expressway," and an "Internet Infidels" site containing parodies of pro-Creationism religious cartoons.<sup>37</sup> We found 17 letters in this category, including four of 32 sent in 2004 by a "planetary enlightenment" group called Star's Edge International to Google, demanding the removal of newsgroup postings that contained copyright-protected course materials.

Our fourth category—a reasonable but not strong claim to fair use or analogous First Amendment protection—consisted of 13 items. Examples included an image of "The Wild Christmas Reindeer," which was arguably small enough to qualify as a fair-use "thumbnail" reproduction,<sup>38</sup> and seven postings of the *Grey*

*Album*, a musical amalgam created by Brian Burton, aka “DJ Danger Mouse,” from the rapper Jay Z’s then-recently released *Black Album* and the Beatles’s iconic *White Album*, owned by Capitol Records and EMI.<sup>39</sup>

Finally, we placed 86 items in the category of “possible claim to fair use, but not enough information to tell.” This category included a site offering a “recognizable portion” of the “Paris Hilton Sex Tape” (how much was copied would be relevant in determining fair use),<sup>40</sup> a site containing recipes for Cuban dishes (since it was taken down, we did not know whether substantial text was copied along with lists of ingredients; the latter are not copyright-protected),<sup>41</sup> and “harrypottergalleon.com” (this site can no longer be found on the Web, and we could not tell from the cease-and-desist letter whether it was simply a fan site using “Harry Potter” as part of its domain name or whether it was capitalizing on the name by selling mugs, T-shirts, and the like).<sup>42</sup>

Our total of strong, reasonable, or possible fair use or First Amendment defenses, combined with material that probably did not amount to copyright or trademark infringement in the first place, was 153 items out of 320. Thus, almost half of the cease-and-desist or take-down letters in our sample made claims that had the potential to chill free expression.

Our next step was to try to determine the outcome of the controversies in the 153 cases where we thought fair use and free expression might be at risk. Of the 17 instances where the targeted expression had a strong fair use or First Amendment defense, we found that five items were removed, 10 were not removed, one was probably removed (we could no longer find it on the Web, but we also did not find the usual Google disclaimer and link to Chilling Effects), and one was partially removed (the image in question was no longer in the Google image gallery, but the Web page still existed).

For the categories of reasonable and possible fair use, we found that almost all the material was removed or partially removed (or, in one case, the operator of the targeted website ended up paying the writer of the demand letter a fee for the use of his article). For the 37 targeted items that were likely not infringing in the first place, the number of removals was 18, with two additional likely removals, four partial removals, 10 non-removals, and three outcomes that we could not determine.

To supplement our analysis of the letters deposited with Chilling Effects, we created an online survey that we publicized through the listservs, websites, and newsletters of artists' and scholars' organizations, including the National Alliance for Media Arts and Culture, National Video Resources, Independent Feature Project, and College Art Association. Eventually, we received 290 completed surveys, describing attitudes about, and experiences with, copyright and fair use from both the owners' and the users' perspectives.

With the assistance of Chilling Effects administrators, we next invited all those who had submitted letters to the Clearinghouse and left contact information, to respond to the online survey or to be interviewed by telephone. We conducted 17 phone interviews, which gave us additional insight and more detail about a variety of conflict situations.

Finally, in collaboration with PEN American Center, Women Make Movies, the College Art Association, and the Location One art gallery, we conducted four focus group discussions, seeking the experiences and attitudes of writers, filmmakers, visual artists, and scholars on the issues of copyright, fair use, and free expression.

Our research was not strictly scientific. It would have been a daunting—perhaps impossible—task to create a truly random sample of artists or others affected by fair use, and then to design and implement a survey methodology yielding statistical results that accurately reflected the percentages of individuals who received cease-and-desist letters, acquiesced in their demands, and had reasonable fair use defenses. Nevertheless, we collected a huge amount of information through a variety of different methods, and the results are suggestive in thinking about threats to free culture in the digital age, and policy initiatives that might address them.

The examples that follow are gleaned from our research and are organized into four general—and somewhat overlapping—categories: Artistic Expression; Commentary on Religion; Political Speech; and Fan Sites and Free Culture.

## **EXAMPLES FROM THE FAIR USE RESEARCH PROJECT**

### **Artistic Expression**

**The Hero's Journey:** An artist who completed our online survey wrote that he had created a montage of Elvis Presley and Vietnam

War photos as part of a series of prints titled “The Hero’s Journey.” The works were also posted on the Web. He received a cease-and-desist letter from the archive that owned the original photographs, and replied “with a detailed fair use argument.”<sup>43</sup> He “discussed the matter with a lawyer, who showed little interest in taking on such a corporate entity.” Ultimately, he “agreed to take down the Elvis-in-Vietnam material and to never post it on the net again.” He did not agree to destroy the prints, which were later purchased by a museum in Finland. He reflected: “I made original artworks that commented on the myth of the hero in a contemporary context. Having shallow pockets and no outside support, I had no choice but to acquiesce.”

**Spoons Collective Deleuze and Guattari List Server:** Another politically engaged artist, who creates “deconstructive texts and artist statements,” wrote that for one major work, she “borrowed all over the place: literature, school textbooks, emails, letters, news media, philosophy and websites.” Among her sources of found texts were email posts from the “Spoons Collective Deleuze and Guattari List Server,” a discussion group dedicated to the authors Felix Guattari and Gilles Deleuze.<sup>44</sup> She did not receive a cease-and-desist letter, but “I did receive 100s, literally, of emails from professors, authors, artists and writers from all over the world,” protesting her quotations from the listserv postings.

“I replied to the actual authors of the posts. There were too many to remember. I deleted it all once it was over, about a year later. I continued to receive hate emails for months on the subject. I eventually replaced all controversial sections with quotes from Critical Art Ensemble’s essay on the ill effects of copyrighting.”

**DJ Danger Mouse:** A musical fair use controversy arose when hundreds of websites announced their intention to post DJ Danger Mouse’s *Grey Album* on “Grey Tuesday,” February 24, 2004, as a gesture of protest against a copyright system that fails to acknowledge the importance of mixing and sampling to musical creation. Danger Mouse’s mix, or “mash-up,” of rapper Jay Z’s *Black Album* and the Beatles’s 1968 *White Album* included such Beatles hits as “Julia,” “Happiness is Warm Gun,” “Dear Prudence,” and “Rocky Raccoon.” Capitol Records/EMI, claiming rights to the *White Album*, responded with lengthy cease-and-desist letters, seven of which we found in the Chilling Effects database. The letters not only warned against any distribution of the *Grey Album*, but demanded that each recipient: “identify the names and addresses of any third parties who have supplied you

with physical or digital copies of the *Grey Album* or who are otherwise involved in the *Grey Album*'s unauthorized distribution, reproduction, public performance, or other exploitation"; "provide Capitol with an accounting of all units of the *Grey Album* that have been distributed via your website"; and, if copies had already been made available, "make payment to Capitol in an amount to be discussed."<sup>45</sup>

Danger Mouse himself had agreed to Capitol's demands, and so did some cease-and-desist letter recipients. But DownhillBattle.org, coordinator of Grey Tuesday, reported that "for 24 hours, over 170 sites made the album available in protest, defying legal threats."<sup>46</sup> To the cease-and-desist letter that Downhill Battle itself received, the group responded:

Your efforts to suppress music stifle creativity and harm the public interest; we will not be intimidated into backing down. Downhill Battle has a fair-use right to post this music under current copyright law and the public has a fair-use right to hear it.<sup>47</sup>

In a legal memo, the Electronic Frontier Foundation (EFF) asserted that the Grey Tuesday protesters had "a credible fair use defense." Their postings were for a noncommercial purpose; downloads of the *Grey Album* would not "substitute for purchases of the *White Album* or other recordings of the Lennon-McCartney songs on the album;" "the *Grey Album* is a transformative use of the *White Album*, not a wholesale copy;" and the postings were intended as "a commentary on the use of copyright law to stymie new kinds of musical creativity."<sup>48</sup> Although some copyright experts would consider EFF's analysis to be a stretch for fair use, certainly the *Grey Album* was a creative and "transformative" use of the Beatles' material - a major factor in fair use cases.

**Joy Garnett:** At a focus group organized by the College Art Association in May 2005, painter Joy Garnett described her work as exploring "the relationship between found sources, such as photographs, and the transformative act of painting." In a recent project, she

started pulling images from the mass media, mostly from the Web. I mounted a show in January 2004 of about eleven paintings, of decontextualized figures pulled from news wires and all kinds of websites. The show was called "Riot," and the theme was basically people in extreme emotional states. It was all kinds of people—fighter pilots and guardian angels; skinheads rioting. The sources are forgotten; that's part of the process.<sup>49</sup>

One of the paintings, kind of the centerpiece for the show, which we decided was emblematic, was a picture of a guy in a beret with a cross around his

neck, and he's throwing a Molotov cocktail in a Pepsi bottle; he has a ponytail and he looks like a hippie. And he's really big; it's a big painting. So we made that the card, and we sent it out. That image went out on an email press release.

Halfway through my show, I get this email from an acquaintance, a photographer who lives in the Midwest, and the email was very polite, but the subject line was the name of a famous Magnum photographer, and the content of the email was, "She's coming to my studio next week. Your card is on the wall. I assume you asked her permission."

And I was like, "What are you talking about?" He sent me the URL to Magnum, and I found the original image that my fragment, which I had found somewhere on the Web, was from. It was a 25 year-old photograph by Susan Meiselas that was part of a famous book she had done in 1981, called *Nicaragua*. I was really interested in finding where that image came from, because I had found just a piece of it on some anarchist site. And I wrote back saying, "Well, I didn't even know. How could I ask her permission? I wouldn't have asked her permission anyway." I mean, implicit in that would be that for every painting that may or may not get shown, for every moment of my creative process, I would have to be concerned with finding the authors of these photographs, contacting them for permission and dealing with their attitudes about permission.<sup>50</sup>

I didn't contact her. A week after the show came down, my gallery and I—we have images on our websites; it's promotional—we each get a cease and desist from Susan Meiselas's lawyer saying that this painting "Molotov" is a derivative work based on her copyrighted photograph, that I had pirated the photograph, and that I had to sign a release form, a retroactive licensing agreement that would sign over all of my rights to the painting to Susan Meiselas in accordance with Magnum copyright derivative works formalities. And if I wanted to show or sell or reproduce the image again, it wouldn't just require a credit line to her, it would require that I got her prior written permission.

I was completely flipped out. So I did two things. I contacted a really good copyright lawyer. And I went on to my discussion group at Rhizome,<sup>51</sup> where only moments earlier we'd been discussing the *Grey Album* and sampling—this whole area of fair use remixing, licensing, and being sued. They were incensed on my behalf. In the meantime, I went to my lawyer who was also incensed. He thought that this was silly, that in terms of the four factors of fair use, it could be argued either way, but the main thing was I had used only a portion of the original. Whether or not an artwork is available commercially is not the whole issue. It's also creative. It's commentary. This comes under fair use. It didn't in any way change the market value of the original. And that's sort of the bottom line.

So we wrote a very brief letter saying that I was sorry; that I was not aware that I had chosen her photograph; that I was not interested in using seminal iconic famous photographs; that I was interested in photographs that had been lost. I suppose I was guilty of not recognizing her photograph. But it was famous in 1981, and I was little then. I wasn't around during the Sandinistas. I also mentioned that I was hurt by being accused of copyright infringement when in fact I had transformed the work, and that painting is all about quoting, referencing, copying. And I have my precedents in the contemporary art world, painters who use photography. So we said I would not sign the form

giving her rights to my work because I have something to protect, as well.<sup>52</sup> And in fact, that's when I began to get mad, instead of feeling like maybe I'd done something wrong. We sent out this letter, and discussion is still raging on Rhizome.

A few days later I get a 12-page very aggressive response from the lawyer saying that instead of waiving their licensing fee, now I have to pay \$2,000. And that there was all this caselaw, as though she'd cut and pasted from Copyright 101. It didn't make any sense, and I didn't think it related to what I had done. But it was incredibly intimidating and my lawyer agreed—it didn't seem like we were dealing with reasonable individuals. He was surprised that we got that kind of response because our letter was very polite and reasonable, and it was an apology.<sup>53</sup>

Now I believe that the whole thing was just a scare tactic to get me to take the stuff off the Web. And it worked. I called my lawyer and said I was taking my images off the website because I didn't want them to go to my Internet provider. I didn't want my website pulled. This is what I was really afraid of, because I use that site to send images to galleries, to writers, to critics. It's my inventory.

But here's the punch line. Seconds before I took the images off, people on Rhizome grabbed my images in solidarity. They copied the html and uploaded my pages to their websites; then one of these artists takes my painting, a JPEG of my painting, and flips it. He puts it up on his site, and now it's a derivative work based on my work. And everybody went for it. So everyone started making digital collage based on the Molotov image. It all turned into agitprop, protest art. They translated the story into Italian, Czech, Chinese, Spanish, French, Catalan.

It got me on a lecture tour, actually. But the weird thing is that the lawyer wrote to thank me for removing the images from my website. In the meantime, for the next five months, this image went global. And so there are a couple morals to this story. The idea that you can control what's going on right now in the digital realm with the older paradigm of copyright control is gone. The ideas of the new technology are part of our culture whether or not we're aware of it.

### **Commentary on Religion**

Religions, both mainstream and unusual, can be the subject of passionately conflicting arguments, and nowhere is this truer than on the Internet, where critics, true believers, and disillusioned followers engage in vociferous listserv discussions and create Web commentaries, some of them comic.

**Parody Site and Chick Publications, Inc.:** In February 2004, Chick Publications, Inc. wrote to a Web hosting company demanding that it take down a comedy site containing parodies of Chick's copyrighted artwork on religious themes.<sup>55</sup> The ISP notified the site owner, who complied. He wrote ruefully: "Taken down due to legal stuff. ... Sorry folks. The piece was done in fun, but you got to realize that the laws can censor you." A reader

responded that he thought the parody was “hilarious, and a nice spoof on the original. I am sorry that you had to take the page down. However, I am also surprised, as I thought that ‘fair use’ allowed the modification of copyrighted material for parody purposes.”<sup>56</sup> This reader was correct, but the law governing fair use is unpredictable; and most parodists cannot afford to defend a lawsuit.

**Roger Loomis:** Roger Loomis established his website as an “unauthorized investigator’s guide to The Church of Jesus Christ of Latter-day Saints” in order to share his thoughts on the Mormon Church. The site is not overtly hostile to the Church, but instead includes discussion and analysis, and accepts contributions from readers. It asks: “Is this really the most balanced Mormon Internet site on the web?” and replies: “the site will fairly present both sides and let the reader arrive at his own conclusions.”<sup>57</sup> Until 2002, it included excerpts from six of the Church’s official “missionary discussions.”

In July 2002, Loomis’s Internet Service Provider received a take-down letter from attorneys for Intellectual Reserve, Inc., the owner of the Church’s missionary discussions, demanding their immediate removal. The ISP forwarded the letter to Loomis, directing him to remove the items “ASAP.” Loomis replied directly to the Church’s attorney: “It is my personal belief that these pages were not copyright violations, but I will go ahead and remove them from the Internet as you requested.” Two months later, the attorney demanded that he remove three images as well, one of Joseph Smith, the Mormons’ founder; another of gold plates (important icons in Mormon doctrine), and the third of Gordon Hinckley, the current Church president.<sup>58</sup> Again, Loomis complied.

When we asked why he acquiesced, Loomis told us that it was much easier to remove the material than to get into a “big battle,” especially since he was worried about paying Intellectual Reserve’s legal fees if he received an unfavorable ruling. The risk of paying those fees was not worth the “emotional time commitment.” At the same time, he thought that he would have had a fair chance of prevailing if he had been able to afford a good attorney.

**Durango Bill’s Home Page:** Bill Butler operates “Durango Bill’s Home Page,” a collection of information on subjects ranging from paleography, the Grand Canyon, and number theory to “the



fallacies of Creationism and ‘Intelligent Design Evangelism.’” In July 2004, Yahoo, Butler’s ISP, received a take-down notice asserting that the site was using copyrighted material without authorization.<sup>59</sup> The material in question was a logo from Online Christ-Centered Ministries (OCCM), displayed on a Web page titled “Durango Bill’s Example of a Typical Young Earth Creationist.” The page critiqued one Jason Gastrich, a proponent of Creationism and the apparent proprietor of OCCM. Their disagreements ranged from Butler’s accusation that Gastrich faked his educational credentials to an exchange regarding Gastrich’s claim that the Colorado River flows uphill.<sup>60</sup>

In response to the take-down letter, Yahoo sent Durango Bill a “Notice of Infringement” instructing him to remove the “Jason Gastrich” page or else Yahoo would shut down his site. Over the following days, Butler repeatedly tried, without success, to contact Yahoo to determine which parts of the page could be deleted and still comply with the notice. On July 23, his site was shut down. Three days later, he was finally able to reach someone at Yahoo, but he had to delete the entire Gastrich page.

Meanwhile, Butler found a website with a sample DMCA counter-notice.<sup>61</sup> He used the sample to compose a response to Yahoo asserting that his website was well within fair use. Under the law, Yahoo now had to allow him to re-post his Web page, unless Gastrich filed a lawsuit against him. On August 4, 2004, Gastrich did sue, in California federal court, asking for \$2,000 in damages.

Butler retained a lawyer in California, who moved to dismiss Gastrich’s complaint on the ground that the court there had no jurisdiction, since Butler operates his website from his home in Colorado. The motion succeeded in December 2004. After nearly six months, Butler was able to restore his “Jason Gastrich” page.

**Star’s Edge:** Most of the take-down letters that Google deposited with Chilling Effects in 2004 concerned search-result links to websites, newsgroups, or blogs. The leading sender in this category was Star’s Edge International, a psychological improvement and “planetary enlightenment” organization that offers a course called Avatar.<sup>62</sup> Star’s Edge sent 32 take-down letters to Google in 2004, complaining that Avatar’s copyright-protected lectures and course materials were posted by participants in various newsgroups, including alt.clearing.avatar, nl.scientology, and alt.religion.scientology.

Several of the Star's Edge letters seemed to state legitimate copyright claims because, judging by the "subject" lines of the listserv postings—for example, "Here it is! The Avatar Course for FREE!"—the postings consisted of full texts of copyrighted course materials. Other postings, however, had possible claims to fair use—though more information was needed to make a legal judgment—because it appeared that the targeted messages included criticism and commentary.<sup>63</sup> Typical of this category were postings with such subject lines as "Look! The emperor is naked!" and "Harry Palmer talking about Feel its (clarified)" (referring to Avatar's founder).<sup>64</sup>

Ronald Cools, a former Avatar follower and resident of the Netherlands, was a regular contributor to these newsgroups. Cools also created his own website, in order, he told us, to "expose the policies of Star's Edge and to warn people about the Avatar program." Cools's website explains that he had been involved in this "psycho-cult for 8 years and became a licensed 'Avatar Master' in 1992. During this time I did not know that Avatar derived directly from Scientology nor that its creator Harry Palmer has been a dedicated Scientologist for many years." At the bottom of his site is a disclaimer: "This website is not affiliated with Avatar, Star's Edge Inc, Harry Palmer or/and Scientology. Its objective is to inform the public within the limits of Free Speech."<sup>65</sup>

Cools was not the only Avatar critic who posted course materials in newsgroups in 2004. Since Google presumably deleted all of the posts, it is difficult to determine which ones might have qualified as fair use. But whatever the mixture of commentary and unauthorized copying, the result was that throughout 2004, Star's Edge's take-down letters succeeded in suppressing criticism of its program. On the other hand, its critics continued to post new messages, so perhaps the conflict ended in a stalemate.

### **Political Speech**

**The National Debate:** One of the most striking examples of an ill-advised cease-and-desist letter came from *The New York Times* and sought to suppress a parody site. Robert Cox, operator of The National Debate blog site,<sup>66</sup> incurred the *Times's* wrath after he posted a parody of the *Times's* corrections page in order to critique the paper's lack of a correction policy for its op-ed writers.

Cox told us that when he first became concerned about the issue, in May 2003, all *Times* writers except the op-ed columnists were

subject to a strict policy requiring them to publish corrections. The op-ed columnists could decide on their own whether to publish formal corrections. To protest, Cox began chronicling errors in various op-ed columns and publishing them on a “Columnist Corrections” page.

Around this time, Cox told us, *Times* columnist Maureen Dowd “wrote a column quoting President Bush, and she manufactured the quote by taking something he actually said and putting in ellipsis dots, materially altering what he said. In fact, in the end, it was the opposite of what he said.”<sup>67</sup> Cox phoned the *Times* repeatedly to protest. “The column was published on Wednesday and was syndicated on Thursday, so there was time to correct before it got printed in other papers. But when I finally reached someone, I received questions like ‘Who are you with?’ and ‘Do you have an axe to grind against Maureen Dowd?’ I then spent the next year trying to use my blog and emailing other forums to get the message out—to pressure Dowd to issue a correction.”

A fellow journalist suggested that Cox tweak the *Times* by parodying its corrections page. He took both liberal and conservative editorials, and wrote fake corrections “in a *Times*-esque style. People in the blogosphere thought it was very funny.”

The *Times* did not. Its March 2004 take-down letter to the ISP Verio demanded that Cox’s “corrections” page be blocked because it infringed the *Times*’s copyright.<sup>68</sup> (Cox also received a cease-and-desist letter). Verio told Cox that he had to remove the offending page within 72 hours or his site would be shut down.

At that point, Cox took to the blogosphere, asking for help and advice. He soon received offers of *pro bono* legal representation. While his attorney, Ronald Coleman, negotiated additional time to respond to the *Times*’ demand, Cox took down his parody corrections page to avoid having his entire site shut down by Verio.

Coleman’s reply to the *Times* expressed amazement that the paper could have “a good faith belief that our client’s web page was not protected by the First Amendment as a parody.” He quoted the *Times*’s own impassioned editorial three years before, defending the novelist Alice Randall’s race-sensitive reworking of *Gone With the Wind* in her novel *The Wind Done Gone*.<sup>69</sup> Citing a 1994 Supreme Court decision indicating that an off-color rap version of Roy Orbison’s “Oh, Pretty Woman” probably qualified for fair

use,<sup>70</sup> the *Times* wrote that parody necessarily “requires some borrowing from the original,” that political expression generates “the highest level of First Amendment protection,” and that “in an era when media conglomerates control the rights to vast amounts of intellectual property, routine elevation of copyright to a right of censorship could easily squelch active debate and criticism of important ideas.”<sup>71</sup> Within six hours of receiving Coleman’s letter, Cox says, the *Times* announced that it would withdraw its complaint.

In the end, Cox agreed to state on his corrections page that it was a parody. The *Times* announced a new policy requiring corrections of op-ed columns. Cox later wrote: “This is an amazing David vs. Goliath story and one that may prove to be something of a tipping point in the battle between traditional media and new forms of media such as Blogs.”<sup>72</sup>

**Netfunny.com:** American Express sent a cease-and-desist letter to the political parody site netfunny.com, demanding that it stop using the terms “American Expressway” and “Membership Has Its Privileges” because they are “virtually identical to our client’s American Express mark” and are “likely to cause consumers to be confused, mistaken, or deceived as to the source of origin of your services.”<sup>73</sup> A visit to netfunny.com shows that its American Express parody is still posted, with a link to a site where “you can read about an effort by American Express to get this joke deleted.” On that page, blogger Brad Templeton writes that he received a “one of those bullying ‘cease and desist’ letters from American Express’s law firm,” and he posts another parody, this time of Amex’s lawyers. That parody concludes: “After all, Being Giant and Intimidating has its Privileges. ... American Express Lawyers: Don’t leave your home page without them.”<sup>74</sup>

**Attrition.org:** Another political commentary case involved Attrition.org, an acerbic website “dedicated to the collection, dissemination, and distribution of information” about the computer security industry. Among the humorous images on the site are several, contributed by readers and fans, that parody MasterCard International’s advertising campaign using the term “Priceless.”<sup>75</sup>

In June 2001, Attrition’s proprietor, Brian Martin, received a cease-and-desist letter from MasterCard claiming exclusive ownership of the “PRICELESS” trademark. Since 1998, the letter explained, MasterCard had aired TV and print ads for various products, all designated as “the MasterCard Priceless

Advertisements,” and concluding, after the word “priceless”: “there are some things money can’t buy; for everything else there’s MasterCard.”

The letter accused Attrition.org of “blatantly cop[ying] the sequential display of a series of items belonging to one or more individuals” in a manner similar to its advertising campaign, and thereby infringing the Priceless Mark not only “with impunity,” but with content that “is often obscene.” The letter demanded prompt removal of the allegedly infringing material; “otherwise, MasterCard will have no choice but to consider legal action.”<sup>76</sup>

Martin publicized the threat in a mood, as he told us, of “hey guys ... get a load of this ... what are they smoking?” A few of the lawyers on his email list told him that Attrition was “on solid ground,” but he did not feel the need to engage legal counsel. Instead he replied, based on his own research, that MasterCard’s trademark claim was groundless because there was no likelihood of confusion; and its copyright claim was equally bad, because parodies are a protected form of fair use under the Supreme Court’s “Oh, Pretty Woman” case.<sup>77</sup>

MasterCard took no further action, and the ad spoofs remained up for more than a year. They were eventually taken down for bandwidth reasons. Martin commented: “the biggest irony is that the legal pressure created negative publicity” and “has only spawned more parody,” yet MasterCard “still sends out those C&D letters.” He has received emails from other recipients to find out how he had responded. “I told them not to back down and they didn’t have to. I did say that I was not a lawyer, but based on experience, they had good cases.”

“This kind of issue needs exposure,” he concludes. Corporations are “trying to pressure people ... trying to use scare tactics and legal tricks.” In early 2005, one Priceless parody could still be found on the Attrition site. The image showed a youth thumbing his nose at an overweight police officer. The caption read: “Thumbing our nose at your pompous bullshit: PRICELESS. There are some things only hubris can buy. For everything else, there’s *Attrition*.”<sup>78</sup>

### **Fan Sites and Free Culture**

Fan sites for books, movies, and television shows have proliferated since the earliest days of the Internet. *Star Wars*, *Star Trek*, *The X-Files*, and *Buffy the Vampire Slayer* have been among the most

numerous; more than 3,000 sites were reportedly devoted to *The Simpsons* alone by the mid-1990s.<sup>79</sup> One commentator explains: “Fans created these websites for a variety of reasons, including a desire to celebrate television shows and movies they loved or to fill a void left by an unsatisfying official site.” Members of fandom communities “share exclusive, special knowledge that others do not have.” Fans are “able to forge symbolic communities created by shared interests in stories as opposed to physical proximity as in traditional communities.”<sup>80</sup> Naturally, most fan sites include greater or lesser amounts of text or imagery that is borrowed from copyright owners.

**Caroline in the City:** A site for the TV show *Caroline in the City* included transcripts from episodes along with “fan fiction” summaries. In January 2004, a lawyer at CBS television wrote: “in reviewing your website I see that you are a student, an artist and a big fan of the show. From what I can tell, it seems that you are operating the website for fun, and not for profit. This being the case, I thought I’d send you this informal e-mail first before taking any further action. CBS must insist that all transcripts and ‘fanfics’ immediately be removed from your website (and any other websites where they may be stored or accessed).” CBS closed by saying it was “quite happy to know that you are a fan of the show,” and that “it doesn’t bring us any pleasure to send legal letters to our fans.” Nevertheless, it threatened legal action if the material were not removed.<sup>81</sup>

The owner complied, explaining on the site that she had received a cease-and-desist letter. This resulted in an outcry from fans. A few months later, a new owner acquired the site and re-posted the plot summaries, though not the transcripts. She wrote: “woo, I got rid of that depressing letter Ann posted a few months ago. And if you look closely on the left, you’re in for a little surprise: The dearly missed section is back!!” She added a disclaimer explaining that the site is nonprofit and has no connection with the producers of the show.<sup>82</sup>

**Dee’s Dragonrider Art Gallery:** Artist Dee Dreslough was a devotee of science fiction novels by Anne McCaffrey—the *Dragonriders of Pern* series, *Renegades of Pern*, *Planet Pirates*, and many more. In the mid-1990s, Dreslough created a website and began taking commissions for original works of art. At the request of fellow McCaffrey fans, she would draw dragons and other characters. She told us: “I drew pictures of people and dragons inspired by her books, but not created by Anne

McCaffrey. These were original characters based in her book-worlds created by other fans.” She thought that there were “enough levels of filtration for the work to be allowed.”

Dreslough elaborates on her website: “my depictions of Pernese dragons differed from McCaffrey’s own definition of the dragons ... My dragons had ears, proto-horns rather than head-knobs, and non-bug eyes, as well as varied hues and stripes in the skin.” They were based on characters created by fans. “They would send me their character’s description, and I would do my best to draw it for them.”<sup>83</sup> Her charges were modest: she earned \$130 in six months from commissions. She posted her creations on a Web page called “Dee’s Dragonrider Art Gallery.”

McCaffrey’s attorney sent a cease-and-desist letter citing both copyright and trademark infringement in April 1997. The letter began by quoting Dreslough’s own Web statements—that her art was inspired by *Dragonriders of Pern*, and that “if mentioning my inspiration is something that can get me sued ... oh well. If I’ve goofed, let me know where to fix and I will.” The letter demanded that Dreslough “remove your art gallery from the Internet immediately,” and promised legal action if she did not comply. It said she was guilty of contributory copyright infringement for every other website where a fan had posted one of her Pern-inspired drawings; and closed by mentioning that in previous lawsuits, McCaffrey had won damages “in the middle and low six figures.”<sup>84</sup>

Dreslough told us that at first she “felt horrible;” she thought she had “done a bad thing” and offended her literary hero. But she was also distressed that the hero had resorted to legal threats. “If one of our friends-in-common had just told me, ‘Anne’s annoyed by your website; take it down,’ I would have done so in an instant.”

Dreslough was frightened by the threats—particularly the massive damages that might accrue based on contributory infringement. At the time, she told us, “\$15,000 was all that I made in a year. I couldn’t afford a lawyer to argue the case. I couldn’t even afford a lawyer to have the case researched.” She removed the art from her site, asked all of her fellow fans to do the same, and refunded the commissions.

“I folded up like a house of cards,” she reflected. “When I had to destroy my art, that really hurt. ... Did McCaffrey own my art, or did I? I’ll never know, and it doesn’t matter now.”

## CONCLUSION

Although FEPP's research methodology was not scientific, and those contributing to the Chilling Effects database or participating in our survey, interviews, and focus groups were to some extent self-selected, we can nevertheless draw some significant conclusions from the research.

First, whatever the precise percentage of meritless cease-and-desist letters may be, substantial numbers of them do state weak-to-nonexistent claims, or seek to suppress material that would qualify for a fair use or First Amendment defense. Recipients who do not acquiesce in the demands in these letters are often successful, and are not sued. Even without legal representation (and we found no strong correlation between legal representation and refusal to acquiesce in a copyright or trademark owner's demands), a non-acquiescent response often ended the controversy. But despite the fact that cease-and-desist letters are often empty threats, many recipients are cowed into acquiescing, and the validity of their fair use or First Amendment defense is never tested.

Section 512 letters have markedly different effects. By conditioning immunity from suit on compliance, the statute creates very strong incentives for ISPs and search engine companies to remove material merely on the "good faith" representation that it is infringing. As Google's experience with Scientology reflects, §512 has tremendous potential as a censorship tool for anyone seeking to suppress criticism. We saw other examples of this phenomenon in the take-down letters targeting the *New York Times* parody correction page, Durango Bill, Roger Loomis's Mormon discussion site, the Chick cartoons parodist, and newsgroup postings about Star's Edge International. The mere fact that in our typical year, take-down letters greatly outnumbered cease-and-desist letters on the Chilling Effects site suggests how pervasive their use has become.

What remedies could help restore the copyright balance? Section 512 is bad legislation, but it is not about to be repealed. One way to ameliorate its effects is through better use of the counter-notice procedure provided in the statute itself. Informational websites and advocacy groups could do a great service by explaining the counter-notice provision, and by working with ISPs to make sure that anyone whose website or other online speech is targeted by a take-down letter is given adequate information and help in preparing a counter-notice.



Two specific initiatives would be: (1) to mount a full-scale survey of ISPs to determine just what their §512 procedures are; and (2) to propose standardized notices for them to send to their users, which explain, in plain English, how to prepare a counter-notice, which provide sample language for a counter-notice, and which dispel some of the intimidating effect of the consent-to-suit requirement for the counter-notice by pointing out that any of us can be sued at any time, regardless of whether we consent to it or not.

Increased community support and pro-bono legal assistance are also needed. The experiences of painter Joy Garnett, the National Debate blog site, and fan sites such as *Caroline in the City* suggest that online communities can be critical in assisting those targeted by cease-and-desist or take-down letters. Although consulting legal counsel does not seem to correlate with non-acquiescence, this may be in part because many private attorneys take a conservative approach to fair use. Public interest lawyers at the Electronic Frontier Foundation and law student IP clinics have taken a less acquiescent approach. But there are not enough of them to fill the need. A national legal support backup center *cum* clearinghouse of legal pleadings, reply letters, and other materials, would be a worthy project. Bar associations can help by insisting that educational outreach campaigns of the type recently undertaken by the Copyright Society of the USA deal evenhandedly with fair use.<sup>85</sup>

Documentary film is probably the field of artistic endeavor where a pervasive “clearance culture” has been most effective in eviscerating fair use. The Center for Social Media at American University has studied the problem, and is now working on a proposed remedy: a statement of “best practices” by documentary filmmakers that could be used to persuade distributors, insurers, archival houses, and others involved in the transmission of film culture to accept the importance of fair use and make it a reality.<sup>86</sup> Unlike guidelines negotiated between copyright owners and users, which often contain numerical limits that have no basis in fair use law, a statement of best practices could advance, rather than narrow, fair use.<sup>87</sup>

In short, accurate knowledge of fair use; awareness of its importance; and determination not to let it wither despite its unpredictability are essential elements of free culture in the digital age.

**ENDNOTES**

1. Marjorie Heins is a Fellow at the Brennan Center for Justice at NYU School of Law, and is the founder and coordinator of the Free Expression Policy Project (FEPP), [www.fepproject.org](http://www.fepproject.org). Some of the material in this paper is excerpted from FEPP's forthcoming policy report, *Will Fair Use Survive? Free Expression in the Age of Copyright Control*. Both this paper and the policy report are covered by a Creative Commons "Attribution-No Derivs-Noncommercial" License. Thanks to Tricia Beckles for invaluable research assistance.
2. *Campbell v. Acuff-Rose Music, Inc.*, 510 U.S. 569, 575 (1994), quoting *Emerson v. Davies*, 8 F. Cas. 615, 619 (C.C.D. Mass. 1845).
3. *Id.* at 579; *Eldred v. Ashcroft*, 537 U.S. 186, 220 (2003), quoting *Harper & Row v. Nation Enterprises*, 471 U.S.539, 560 (1985).
4. The Supreme Court has said that under copyright law, the fair use defense is adequate to protect First Amendment rights; no additional constitutional protection is necessary. Under trademark law, by contrast, fair use is a narrower concept, and a number of courts have taken the First Amendment into consideration when ruling on trademark infringement claims. *E.g.*, *Mattel v. Walking Mountain Productions*, 353 F.3d 792,807 (9th Cir. 2003) (noting the importance of the First Amendment right to parody cultural icons in rejecting a trademark claim by makers of the Barbie Doll against an artist who created a series of comical "Food Chain Barbies"); *Rogers v. Grimaldi*, 875 F.3d 994, 999 (2d Cir. 1989) (rejecting trademark claim by actress Ginger Rogers against filmmaker Federico Fellini's use of the title *Ginger and Fred*, and holding that the Lanham Act [the trademark law] should "apply to artistic works only where the public interest in avoiding confusion outweighs the public interest in free expression"); *Mattel, Inc. v. MCA Records, Inc.* 296 F.3d 894, 904 (9th Cir. 2002) (holding that a rock song making fun of the Barbie Doll is a constitutionally protected parody).
5. See [www.creativecommons.org](http://www.creativecommons.org) (visited 6/13/05).
6. See <http://www.fepproject.org/policyreports/fairusesneakpreview.html> (visited 6/9/05).
7. Jessica Litman, *Digital Copyright* (Amherst, NY: Prometheus Books, 2001), 90-95. The White Paper is *Intellectual Property and the National Information Infrastructure: The Report of the Working Group on Intellectual Property Rights*, Bruce A. Lehman, Chair (U.S. Patent & Trademark Office, 1995), <http://www.uspto.gov/web/offices/com/doc/ipnii/> (visited 6/13/05).
8. See Marjorie Heins, "The Progress of Science and Useful Arts": *Why Copyright Today Threatens Intellectual Freedom* (Free Expression Policy Project, 2003), 6, 24; and, on the industry practice of encrypting both copyrighted and public domain material, Robert

Kunstadt, "Fair Use Should Not Die," *National Law Journal*, Nov. 11, 2002, A16.

9. See Brenda Sandburg, "Fair Use Fears Over Federal Circuit Ruling," *The Recorder*, Oct. 8, 2002, [www.law.com/jsp/article.jsp?id=1032128694823](http://www.law.com/jsp/article.jsp?id=1032128694823) (accessed 6/13/05) (describing a federal court's rejection of a challenge to a shrink-wrap license that prohibited reverse engineering, and a click-wrap license that requires viewers to waive the right to fair use).
10. 17 U.S.C. §1201(a)(1)(A), (b). The DMCA provides for the usual copyright law civil and criminal penalties – up to a \$500,000 fine or five years in prison for a first offense, and up to \$1 million or 10 years in prison for subsequent offenses. 17 U.S.C. §§1203, 1204.
11. The DMCA does direct the Librarian of Congress (after investigation by the Copyright Office) to decide every three years whether anybody seeking access to a "particular class of works" for legitimate aims such as fair use is likely to be "adversely affected" by the law's restrictions. 17 U.S.C. §1201(a)(1)(B). In 2000, after receiving hundreds of comments from educational and civil liberties groups on the importance of circumvention to the exercise of fair use, the Copyright Office recommended only two narrow exemptions to the access ban of the DMCA: lists of Web sites blocked by Internet filters, and works made inaccessible because of "malfunction, damage, or obsolescence" of encryption devices. In late 2003, the Copyright Office added two more narrow exemptions. See Heins, "The Progress of Science and Useful Arts," *supra* n. 8, 24-25.
12. *Universal City Studios v. Corley*, 273 F.3d 429, 459 (2d Cir. 2001) (conceding that the DMCA shrinks opportunities for fair use, but noting, "we know of no authority for the proposition that fair use, as protected by the Copyright Act, much less the Constitution, guarantees copying by the optimum method or in the identical format of the original"); *United States v. Elcom Ltd.*, 203 F. Supp.2d 1125, 1131 (N.D. Cal. 2002) (acknowledging that DRM embedded in Adobe's eBook Reader restrict first sale and fair use, but ruling that Congress can "sacrifice" these interests if it chooses; the law does not guarantee "the right to the most technologically convenient way to engage in fair use").
13. *Eldred v. Ashcroft*, 537 U.S. 186, 221 (2003) (where Congress alters such "traditional contours of copyright protection" as fair use, First Amendment scrutiny is needed). See also *Dr. Seuss Enterprises v. Penguin Books*, 109 F.3d 1394, 1399 (9th Cir. 1997) (fair use "permits courts to avoid rigid application of the copyright statute when, on occasion, it would stifle the very creativity which that statute is designed to foster").

The DMCA was defended as necessary to harmonize U.S. law with international treaties that oblige member countries to provide "effective legal remedies" for circumvention of electronic locks on

copyrighted works. WIPO Copyright Treaty, Art. 11, quoted in *The Digital Millennium Copyright Act of 1998 - U.S. Copyright Office Summary* (Dec. 1998), [www.loc.gov/copyright/legislation/dmca.pdf](http://www.loc.gov/copyright/legislation/dmca.pdf) (visited 7/29/03). But the head of the U.S. Patent and Trademark Office acknowledged that these treaties do not require the DMCA's "device-oriented" approach – as opposed to the more traditional "conduct-oriented" approach that targets copyright infringers and not the researchers who create new technologies. Testimony of Ass't Sec'y of Commerce and Commissioner of Patents & Trademarks Bruce Lehman, in *WIPO Copyright Treaties Implementation Act and Online Copyright Liability Limitation Act: Hearing on H.R. 2281 and H.R. 2280 Before the House Subcommittee on Courts and Intellectual Property*, 105th Cong., 1st Sess. (Sept. 16, 1997), 62.

14. 17 U.S.C. §512(c)-(g).
15. 17 U.S.C. §512(g) provides that an ISP can avoid possible liability to a subscriber for taking down content if it:
  - “(A) takes reasonable steps promptly to notify the subscriber that it has removed or disabled access to the material;
  - “(B) upon receipt of a counter notification described in paragraph (3), promptly provides the person who [sent the original take-down letter] with a copy of the counter notification, and informs that person that it will replace the removed material or cease disabling access to it in 10 business days; and
  - “(C) replaces the removed material and ceases disabling access to it not less than 10, nor more than 14, business days following receipt of the counter notice, unless its designated agent first receives notice from the [sender of the take-down letter] that such person has filed an action seeking a court order to restrain the subscriber from engaging in infringing activity relating to the material on the service provider's system or network.”

Paragraph (3) contains detailed requirements for a proper counter notification, including a statement that the subscriber consents to be sued by the sender of the take-down notice.
16. Carey Lening *et al.*, *Copyright Protection of Digital Television: The "Broadcast Flag,"* CRS Report for Congress, Congressional Research Service/Library of Congress, Apr. 5, 2005, 5. This report also notes that the flag could restrict the dissemination of “news or public-interest based content, or works that have already entered the public domain.” *Id.*, 6; see also Center for Democracy and Technology, *Implications of the Broadcast Flag: A Public Interest Primer* (Dec. 2003), <http://www.cdt.org/copyright/031216broadcastflag.pdf> (visited 6/9/05).
17. *American Library Ass'n v. FCC*, 406 F.3d 689 (D.C. Cir. 2005).
18. See “Letter Asking Congress to Oppose Attempts to Bring Back the Broadcast Flag,” June 17, 2005,

<http://www.publicknowledge.org/news/letters/bflag-opposition-letter-20050617> (visited 6/17/05).

19. *Metro-Goldwyn-Mayer Studios v. Grokster*, No. 04-480 (June 28, 2005), slip opinion, 8.
20. *E.g.*, Brief of *Amici Curiae* Snocap, Inc., in Support of Neither Party, *Metro-Goldwyn-Mayer Studios v. Grokster*, No. 04-480 (Jan. 24, 2005), 1 (maker of software that enables “digital licensing and copyright management services” wishes to “alert the Court to technological innovations” that may help “in resolving the issues presented”); Brief of *Amicus Curiae* Bridgemar Services in Support of Neither Party (Jan. 24, 2005), 6 (“effective technology exists to prevent infringing activity on a distributed P2P network”); Brief *Amici Curiae* of Audible Magic Corp. et al. (Jan. 24, 2005), 4 (describing “digital fingerprinting and filtering” and “digital watermarking” technologies); Brief *Amicus Curiae* of Video Software Dealers Association in Support of Neither Party (Jan. 24, 2005), 17 (suggesting that the Court should consider whether the defendants could modify their software to prevent infringement) .
21. *MGM v. Grokster*, *supra* n. 19, slip opinion, 22.; see Marjorie Heins, “Two Defeats – and a Silver Lining,” June 28, 2005, <http://www.fepproject.org/commentaries/grokster&brandx.html> (visited 6/29/05).
22. See Peter Grier & Amanda Paulson, “Google Plans Giant Online Library Stack,” *Christian Science Monitor*, Dec. 15, 2004, <http://www.csmonitor.com/2004/1215/p01s02-ussc.html> (visited 6/14/05). The fourth university library is Oxford’s Bodleian.
23. *Id.* Grier and Paulson note that even aside from copyright questions, the project will take years to complete. “At Michigan, for example, the library stacks contain about some 132 miles of books. Google hopes to get the digitization job at UM done in six years, according to John Wilkin, Michigan associate university librarian. ‘We feel this is part of the mission of a great public university – reaching out to the public with the resources that we have,’ he says.”
24. David Kirkpatrick, “Publishers and Libraries Square Off Over Free Online Access to Books,” *New York Times*, June 17, 2002, C7 (reporting that “the Web sites of more than 7,300 libraries, including the New York Public Library, provide patrons 24-hour remote access to the texts of a few hundred to several thousands electronic books”; the Queens Public Library in New York “provides unlimited simultaneous access to 3,500 Chinese texts,” which more than 400 Chinese speakers used in a single month).
25. Associated Press, “Publishers Protest Google’s Online Library,” *Int’l Herald Tribune*, May 24, 2005, <http://www.iht.com/articles/2005/05/24/technology/web.0524google.php> (visited 6/14/05). In September 2005, the Authors Guild filed

- suit challenging Google's plans to post even these brief excerpts from copyrighted books. One columnist commented: "I'm sorry to see authors buy into the old-school protectionism of the Authors Guild, not realizing they're acting against their own self-interest" in making their works, many of them out-of-print, more widely known. Tim O'Reilly, "Search and Rescue," *New York Times*, Sept. 28, 2005, p. A27.
26. "N.Y. Public Library Starts Digital Library," *Yahoo News*, June 13, 2005, [http://news.yahoo.com/s/ap/20050613/ap\\_on\\_en\\_ot/digital\\_books](http://news.yahoo.com/s/ap/20050613/ap_on_en_ot/digital_books) (visited 6/16/05).
  27. Cecilia Ogbu, "Note: I Put Up a Website About My Favorite Show and All I Got Was This Lousy Cease-and-Desist Letter: The Intersection of Fan Sites, Internet Culture, and Copyright Owners," 12 *S. Cal. Interdisc. L.J.* 279, 303-07 (2003).
  28. See Julie Keller, "Buffy Cyberfans Slayed By Fox," *Eonline.com*, Dec. 23, 1999, [www.eonline.com/News/Items/Pf/0,1527,5782.00.html](http://www.eonline.com/News/Items/Pf/0,1527,5782.00.html) (visited 9/8/03); Letter from Warner Brothers to Jonathon Woodward, Nov. 8 1995, [www.io.com/~woodward/@cme/served.txt](http://www.io.com/~woodward/@cme/served.txt) (visited 9/8/03). The status of fan sites and fan fiction under copyright law depends on the application of the usual fair-use factors (see n. 32). A court ruled in 1993 that a book consisting largely of detailed summaries of episodes of the *Twin Peaks* TV show did not qualify as fair use; but a more recent decision held that a novel using many of the same characters and plot elements as *Gone With the Wind* for purposes of parody and political critique did amount to fair use. Compare *Twin Peaks Productions v. Publications Int'l*, 996 F.2d 1366 (2d Cir. 1993) with *SunTrust Bank v. Houghton Mifflin*, 268 F.3d 1257 (11th Cir. 2001).
  29. "Chilling Effects Clearinghouse," <http://www.chillingeffects.org/> (visited 5/10/05). Actually, there were 339 documents in the Chilling Effects database for 2004, but we did not count seven of them because one was no longer available on the site and the other six were not cease-and-desist or take-down letters.
  30. The other 12, which we eliminated from our analysis, solely concerned defamation or child pornography. Defamation is a legal wrong unrelated to intellectual property; but the Federal Trademark Dilution Act (15 U.S.C. §1125(a)) does impose a kind of liability for defamation where a trademark is "tarnished" or "disparaged." We excluded defamation complaint letters from our analysis unless they also specifically mentioned trademark dilution.
  31. The great majority of the 2004 take-down notices (245) were submitted by Google, which began contributing them to Chilling Effects soon after the Clearinghouse began. Google had been getting complaints from the Church of Scientology about a site called

“Operation Clambake,” which the Church accused of violating its copyrights. See “Operation Clambake,” <http://www.Xenu.net> (visited 5/11/05), which includes links to such pages as “Scientology Illustrated” and “Baloney Detection Kit.” Google removed links to the allegedly offending pages, but was concerned that in consequence, when Web surfers searched for “Scientology,” criticism did not show up. According to Chilling Effects founder Wendy Seltzer, Google “realized that didn’t serve its public very well, but they were looking for something to help minimize their legal risk, and Chilling Effects was there to help.” Quoted in Mark Thompson, “Overzealous Lawyers Beware: Today’s Sites Are Fighting Back,” *USC Annenberg Online Journalism Review*, Sept. 15, 2004, <http://ojr.org/ojr/law/1095284771.php> (visited 4/12/05) Google began forwarding its take-down notices to Chilling Effects even as it removed the contested listings and, in their place, inserted links to the notice in the Chilling Effects database.

32. 17 U.S.C. §107 provides that “the fair use of a copyrighted work ... for purposes such as criticism, comment, news reporting, teaching (including multiple copies for classroom use), scholarship, or research, is not an infringement of copyright,” and then enumerates four factors “to be considered” in evaluating a claim of fair use:
  - “(1) the purpose and character of the use, including whether such use is of a commercial nature or is for nonprofit educational purposes;
  - “(2) the nature of the copyrighted work;
  - “(3) the amount and substantiality of the portion used in relation to the copyrighted work as a whole; and
  - “(4) the effect of the use upon the potential market for or value of the copyrighted work.”
33. <http://www.chillingeffects.org/dmca512/notice.cgi?NoticeID=1204> (visited 5/11/05).
34. <http://www.chillingeffects.org/dmca512/notice.cgi?NoticeID=1315> (visited 5/11/05).
35. <http://www.chillingeffects.org/dmca512/notice.cgi?NoticeID=1269> (Haiku houses); 1359 and 1259 (National Health Scotland); 1581 (Paddle Asia); <http://www.chillingeffects.org/dmca512/notice.cgi?NoticeID=1531> (poetry); <http://www.chillingeffects.org/dmca512/notice.cgi?NoticeID=1221> (Brazilian *Playboy*) (all visited 5/11/05).

We also categorized as legitimate all but one letter alleging online file-sharing or circumvention of software locks. Of our 320-letter sample, 23 letters targeted file-sharing and another 14 concerned circumvention. Some might question this judgment as overly conservative because not all file-sharing constitutes infringement.

36. <http://www.chillingeffects.org/dmca512/notice.cgi?NoticeID=1329> (“penisimprovement”); <http://www.chillingeffects.org/domain/notice.cgi?NoticeID=1065> (“Pet Friendly”) (both visited 5/11/05).
37. The *New York Times* and “American Expressway” controversies are described later in this paper. The letter to Internet Infidels is at <http://www.chillingeffects.org/protest/notice.cgi?NoticeID=1069> (visited 6/1/05).
38. See <http://msophelia.blogspot.com/prancer.jpg> (visited 5/11/05); the take-down letter is at <http://www.chillingeffects.org/dmca512/notice.cgi?NoticeID=1071> (visited 4/7/05).
39. See the section of this article marked *DJ Danger Mouse* for more on the *Grey Album*.
40. <http://www.chillingeffects.org/dmca512/notice.cgi?NoticeID=1330> (visited 3/16/05).
41. “Recipes,” FL-122 (U.S. Copyright Office, Jan. 2004) (“mere listings of ingredients as in recipes, formulas, compounds or prescriptions are not subject to copyright protection. However, where a recipe or formula is accompanied by substantial literary expression in the form of an explanation or directions, ... there may be a basis for copyright protection”). The take-down letter is at <http://www.chillingeffects.org/dmca512/notice.cgi?NoticeID=1089> (visited 4/7/05).
42. <http://www.chillingeffects.org/dmca512/notice.cgi?NoticeID=1225> (visited 3/16/05).
43. Quotations in this section are from the online surveys, written notes of telephone interviews, transcripts of focus group discussions, or follow-up emails, all on file at the Brennan Center.
44. <http://lists.driftline.org/listinfo.cgi/deleuze-guattari-driftline.org> (visited 6/6/05).
45. <http://www.chillingeffects.org/dmca512/notice.cgi?NoticeID=1123>, 1124, 1129, 1131, 1132, 1133, 1142 (all visited 5/16/05).
46. <http://www.downhillbattle.org> (visited 4/17/05). For more on the Grey Tuesday event, see <http://www.greytuesday.org/> (visited 4/17/05).
47. “Response to EMI’s Cease and Desist Letter,” Feb. 23, 2004, <http://www.downhillbattle.org/index.php?p=68> (visited 6/24/04).
48. [http://www.eff.org/IP/grey\\_tuesday.php](http://www.eff.org/IP/grey_tuesday.php) (visited 5/16/05). EFF also noted that EMI/Capitol did not have a statutory copyright claim because sound recordings weren’t covered by the federal law until 1972; it might have state law remedies, which would vary from state to state. Sony/ATV, which evidently owned the rights to the Lennon-McCartney compositions, would have had a federal



copyright claim – subject to the Grey Tuesday protesters’ fair use defense.

49. I have edited the focus group transcript for readability. Also in the interests of readability, I have not included brackets and ellipsis dots. The full transcript is on file at the Brennan Center.
50. The letter is at <http://firstpulseprojects.com/1stletter-from-bhoffman2.26.04.html> (visited 6/15/05).
51. “Rhizome.org is a nonprofit organization that was founded in 1996 to provide an online platform for the global new media art community. ... A rhizome is a horizontal, root-like stem that extends underground and sends out shoots to the surface. ... Rhizome is also a figurative term used by Gilles Deleuze and Felix Guattari to describe non-hierarchical networks of all kinds. “About Us,” <http://rhizome.org/info/index.php> (visited 6/10/06).
52. The reply letter is at <http://firstpulseprojects.com/letter-to-bhoffman2.27.04.html> (visited 6/15/05). It agrees to credit Meiselas and Magnum on the title of the painting, but not to seek Meiselas’s approval for any reproduction.
53. The second demand letter is at <http://firstpulseprojects.com/2ndletter-from-bhoffman3.03.04.html> (visited 6/15/05).
54. Garnett’s story, with copies of the correspondence with Meiselas’s attorney, and links to the many variations on “Molotov,” can be found at [http://www.firstpulseprojects.net/riot\\_2003/joywar](http://www.firstpulseprojects.net/riot_2003/joywar) (visited 6/15/05).
55. <http://www.chillingeffects.org/dmca512/notice.cgi?NoticeID=1226> (visited 5/16/05).
56. <http://www.howardhallis.com/bis/cthulhuchick> (visited 4/17/05).
57. <http://www.lds4u.com>; <http://www.lds4u.com/balanced.htm> (both visited 5/3/05).
58. The original take-down letter is at <http://www.chillingeffects.org/notice.cgi?NoticeID=352> (visited 5/4/05); the later correspondence is on file at the Brennan Center.
59. <http://www.chillingeffects.org/dmca512/notice.cgi?NoticeID=1340> (visited 4/22/05).
60. See <http://www.durangobill.com/JasonGastrich.html> (visited 4/22/05) for the complete story.
61. “Do It Yourself Counter Notification Letter,” <http://www-2.cs.cmu.edu/~dst/Terrorism/form-letter.html> (visited 4/22/05). The proprietor of this site explains: “One of the favorite tools of both cults and corporations seeking to take embarrassing information off the Internet is to falsely claim violation of a copyright or trademark. The Digital Millennium Copyright Act ... set out a notification procedure that can be used to request an ISP to remove allegedly infringing material from a web page. However, there is a defense

against this attack: it's called a counter notification letter. Most people don't know how to write such a letter, which is why I've put together this helpful form."

62. "What is Avatar?", <http://www.avatarepc.com/html/whatis.html> (visited 4/24/05).
63. See, for example, *Religious Technology Center v. Netcom On-Line Communication Services*, 923 F. Supp. 1231 (N.D. Cal. 1995) (finding a "strong probability of success" for Scientology because the defendants did not include much commentary along with the copyrighted materials); *Religious Technology Center v. F.A.C.T. Net*, 901 F. Supp. 1519 (D. Colo. 1995) (denying a preliminary injunction to Scientology because of a likely fair use defense).
64. <http://www.chillingeffects.org/dmca512/notice.cgi?NoticeID=1500,1137> (both visited 5/19/05). We categorized four additional postings with criticism or commentary subject headings as having a strong claim to fair use: "Look Who's Talking," "Avatar is an Offshoot from Scientology," "Harry Palmer the NUT or the CLONE of Hubbard," and "The Curse of Avatar: Harry Shows His True Sleazy Colors."
65. <http://www.AvatarScam.com/> (visited 4/17/2005).
66. [www.thenationaldebate.com/blog](http://www.thenationaldebate.com/blog) (visited 6/14/05).
67. It is open to question whether Dowd changed the meaning. On May 5, 2003, President George W. Bush remarked: "That group of terrorists who attacked our country is slowly, but surely being decimated. Right now, about half of all the top al Qaeda operatives are either jailed or dead. In either case, they're not a problem anymore."  
<http://www.whitehouse.gov/news/releases/2003/05/20030505-4.html> (visited May 11, 2004). In her May 14, 2003 column, Dowd quoted Bush as saying: "That group of terrorists who attacked our country is slowly but surely being decimated. ... They're not a problem anymore." "Osama's Offspring," NY Times, May 14, 2003, A25.
68. <http://www.chillingeffects.org/dmca512/notice.cgi?NoticeID=1178> (visited 5/19/05).
69. See n. 28, *supra*.
70. *Campbell v. Acuff-Rose*, 510 U.S. 569 (1994).
71. "Gone With the First Amendment," *New York Times*, May 1, 2001, A22. Coleman's letter is at <http://www.chillingeffects.org/responses/notice.cgi?NoticeID=1186> (visited 5/19/05).
72. "BLOGOSPHERE 2 NY TIMES 0," Mar. 28, 2004, [http://www.thenationaldebate.com/blogger/archive/2004\\_03\\_01\\_TND-ARCHIVE.htm](http://www.thenationaldebate.com/blogger/archive/2004_03_01_TND-ARCHIVE.htm) (visited 4/21/05).
73. <http://www.chillingeffects.org/dmca512/notice.cgi?NoticeID=1218> (visited 5/16/05).

74. <http://www.netfunny.com/rhf/jokes/91q2/amexprew.html>;  
<http://ideas.4brad.com/archives/000064.html> (both visited 5/16/05).
75. “What is Attrition?”, <http://attrition.org/attrition/about.html>;  
“Mastercard Threatens to Sue Attrition,” July 1, 2003,  
<http://attrition.org/mews/mc/> (both visited 4/24/05).
76. <http://www.chillingeffects.org/protest/notice.cgi?NoticeID=153>  
(visited 5/3/05). This is one of three cease-and-desist letters from MasterCard on the Chilling Effects site. A near-contemporaneous one complains of defamatory trademark infringement “under the heading ‘Priceless MasterCard Adverts you will never see,’” and the third, sent in April 2001 to Brad Templeton, proprietor of [www.netfunny.com/rhf/jokes/Apr/columbine.html](http://www.netfunny.com/rhf/jokes/Apr/columbine.html), complains of a tasteless takeoff on the MasterCard ads that “denigrates the tragedy at Columbine High School and holds our client out as sponsoring this highly distasteful material.”  
<http://www.chillingeffects.org/protest/notice.cgi?NoticeID=473>;  
<http://www.chillingeffects.org/trademark/notice.cgi?NoticeID=17>  
(both visited 5/3/05).
77. Martin’s reply is at <http://attrition.org/news/mc/mc010.txt>; a description of the entire controversy is at <http://attrition.org/news/mc/> (both visited 5/3/05).
78. “Mastercard Threatens to Sue Attrition,” July 1, 2003,  
<http://attrition.org/mews/mc/> (visited 4/24/05). At the time Mastercard threatened Attrition, it had already sued Ralph Nader’s 2000 Presidential Primary Committee for using takeoffs on the “Priceless” campaign in its televised campaign ads. In March 2004, a federal court dismissed the suit, ruling that there was no likelihood of confusion and therefore no basis for trademark infringement.  
*Mastercard Int’l. v. Nader 2000 Primary Committee*, 70 U.S.P.Q.2d (BNA) 1046 (S.D.N.Y. 2004).
79. Ogbu, *supra* n. 27, 286-87.
80. *Id.*
81. <http://www.chillingeffects.org/fanfic/notice.cgi?NoticeID=1067>  
(visited 6/15/05).
82. See “Fanfic and Transcript Removal,”  
<http://www.sincereamore.com/board/viewtopic.php?t=462>;  
“Disclaimer,”  
[http://www.sincereamore.com/main\\_content/disclaimer.html](http://www.sincereamore.com/main_content/disclaimer.html) (visited 4/17/05).
83. “Important Notice,” <http://www.dreslough.com/main/important.htm>  
(visited 4/25/05).
84. <http://www.chillingeffects.org/domain/notice.cgi?NoticeID=143>  
(visited 5/3/05).

85. The “Copyright Kids” section of the Copyright Society’s Web site states: “Unless you are absolutely sure, relying on the doctrine of ‘Fair Use’ to avoid seeking Permission to copy a work is risky ” “Copyright Basics,” <http://www.copyrightkids.org/cbasicsframes.htm> (visited 6/17/05).
86. See Pat Aufderheide & Peter Jaszi, “Untold Stories: Creative Consequences of the Rights Clearance Culture for Documentary Filmmakers,” <http://www.centerforsocialmedia.org/rock/index.htm>, and “Recommendations,” <http://www.centerforsocialmedia.org/rock/recommendations.htm> (both visited 6/17/05).
87. At the same time that Congress was preparing what eventually became the Copyright Act of 1976, in which it would for the first time write the fair use doctrine into the law, affected groups were negotiating guidelines to give specificity to the broad terms of the new statute. One set of guidelines that became part of the official legislative history of the 1976 Act governed multiple copies for classroom use and set a series of numerical limits, combined with requirements of “brevity,” “spontaneity,” and lack of “cumulative effect,” which bore little relationship to the law – for example, no more than 250 words from a poem or 2,500 words from a story or article, *if* the copying is for only one course, using not more than one story, poem, or article, or two excerpts, by the same author, *and* it is done at the inspiration of the teacher and so close in time that it would be unreasonable to expect a timely request for permission. Ad-Hoc Committee of Educational Institutions and Organizations on Copyright Law Revision Guidelines, “Agreement on Guidelines for Classroom Copying in Not-For-Profit Educational Institutions,” in House Report No. 94-1476, 94th Cong., 2nd Sess. 65-74 (1976), reprinted in William F. Patry, *The Fair Use Privilege in Copyright Law* 626 (2nd ed. 1995). Although these guidelines state that they represent only agreed-upon *minimums* and do not purport to define the legal limits on fair use, as a practical matter they are often viewed as setting the rules, and they create strong incentives not to exceed their specifications.

In the 1990s, the government convened a Working Group on Intellectual Property Rights that proposed guidelines for educational fair use of digital images; these incorporated some of the numerical limits of the 1976 guidelines, and added new ones for multimedia, *e.g.*, “up to 10 percent, or 3 minutes, whichever is less, in the aggregate of a copyrighted motion media work.” “Proposal for Educational Fair Use Guidelines for Digital Images,” in Information Infrastructure Task Force, Working Group on Intellectual Property Rights, Conference on Fair Use (CONFU): *Final Report to the Commissioner on the Conclusion of the Conference on Fair Use*, Nov. 1998, <http://www.uspto.gov/web/offices/dcom/olia/confu/confurep.pdf>

(visited 6/17/05). One participant in this process noted that these classroom limits bear little if any relationship to the law of fair use, and reported that “among the groups opposing some or all of the CONFU Guidelines were: American Association of State Colleges and Universities (AASCU), American Council on Education (ACE), American Historical Society (AHS), American Library Association (ALA), Association of American Universities (AAU), Association of Research Libraries (ARL), Medical Library Association (MLA), National Association of State Universities and Land-Grant Colleges, and the American Association of Law Libraries (AALL).” Kenneth Crews, “The Law of Fair Use and the Illusion of Fair Use Guidelines,” 62 *Ohio State L.J.* 599, 610 n. 36 (2001).

# Greasing the Wheels of Regulation: the Google Print Library Project

Barrie Howard (Digital Library Federation)

**Abstract:** The Google Print Library Project, a mass digitization project undertaken by Google Inc., in partnership with five major research libraries, has caused quite a clamor throughout the library and publishing domains since it was announced on December 14, 2004. Some have lauded it, others condemned it, but there's no ignoring it. Like it or not, this project will have profound repercussions on the library profession. Many pundits are criticizing the project, threatening litigation over copyright issues. Others are claiming it to be the largest heist of public domain cultural heritage in recorded history. Regardless of the argument, it is symptomatic of a much larger issue: *communities of practice* are using the Internet to drive commerce, communication, and creativity, which is forcing libraries to reinvent themselves. The opportunities are many, viable solutions few. How will libraries meet the challenge of the digital revolution?

## THE REVOLUTION WILL BE DIGITIZED


The longer-term implications for universities of the paired forces of digitization and commercialization are poorly understood. This should not surprise us. - William Bowen

Any library that can be replaced by Google, should be. - David Seaman

Digital technology and the Internet have revolutionized the library profession, at once challenging and enabling libraries to become more efficacious at fulfilling the needs of their mission-aligned communities. There has been a paradigm shift from collecting, circulating, and storing inscribed artifacts to aggregating, linking, and pooling bundles of open-source and rented digital content across wired and wireless networks for desktop delivery. While many libraries still maintain their traditional functions, some have developed and implemented sophisticated Web-based services to connect their users to the information resources they seek.

---

M. Halbert (Ed.): *Free Culture and the Digital Library Symposium Proceedings*. Atlanta, Georgia: MetaScholar Initiative at Emory University, 2005. pp. 218-229.

 This work is covered by the Creative Commons Attribution-NonCommercial-NoDerivs 2.5 License.

Statistical analysis by the Association of Research Libraries (ARL) shows that “nonpurchased serial subscriptions make up a growing amount of the content that libraries offer, increasing by an annual average of 6 percent since 1986” (Kyrillidou and Young, 2004). This pattern is further documented in a case study conducted at Drexel University by Hansen and King. “The transition from print to electronic publications will cause fundamental changes in scholarly communications, and the impact on libraries, their funders and their users will be great” (2002).

Research libraries have assumed a leadership role in the information age, becoming the vanguard for developing new products and services through their *digital library* programs. As the use of electronic resources continues to rise, so does the importance of deciding how best to manage hybrid collections of analog and digital formats, and deliver products and services across local user networks and beyond. In a culture of accountability, this is a matter of survival rather than choice.

### **WHAT IS A DIGITAL LIBRARY?**

In 1998, the Digital Library Federation (DLF) described digital libraries as:

organizations that provide the resources, including the specialized staff, to select, structure, offer intellectual access to, interpret, distribute, preserve the integrity of, and ensure the persistence over time of collections of digital works so that they are readily and economically available for use by a defined community or set of communities. (Waters, 1998)

This definition maintains a great deal of currency today, but as Jeng has recently pointed out, there is still no consensus (2005). Given that there is no shared perception of what digital libraries are, DLF recognized the need for a common understanding of what they do. The DLF Abstract Service Framework Working Group was established at the end of last year to lay the foundation for a reference model of digital library services (2005).

The operations of many digital libraries are similar to their brick-and-mortar counterparts: access, budgeting, cataloging (metadata control), collection development, preservation, resource management, rights management, and strategic planning. However, the business model for a digital library may differ significantly from that of a traditional library. Some digital libraries don't even own collections and are built around a bundle of services and tools. This model of a library is based on access to content, rather than ownership of artifacts.

A study of ARL libraries reveals a steady decline of in-house use of library materials from 1996–2003 (Kyriallidou and Young 2004). Meanwhile, the use of online resources continues to rise. Research libraries have responded by turning the stacks inside-out; digitizing analog collections and providing access over LANs, WANs, and the Internet. Encoding protocols such as EAD (Encoded Archival Description) and TEI (Text Encoding Initiative) have been employed by numerous special collections units to provide access to their collections on and offline. At the same time, born digital content is being amassed in institutional repositories, using the same infrastructure that facilitates the delivery of digital surrogates.

The Internet has catalyzed communication and scholarship in an unprecedented way. Web-based library services are enabling scholarly communication to flow at the speed of thought. New tools, *e.g.*, University of California, Berkeley’s Scholar’s Box and Michigan State University’s MATRIX, are empowering scholars to capture, manipulate, and share content in ways previously impossible. This is largely due to the fact that content in digital formats behaves differently than content fixed in analog formats.

Digital content is fluid, malleable, and wants to be free to flow from one application to another, through the air on wireless networks, or across the world over the Internet. Analog content is brittle and static, encumbered by the technological shell that binds it. The ease and speed at which bits, bytes, and batches of information are transmitted across the Internet with no direct costs to users has paved the way for a virtual *free culture*. Physical and political barriers are less potent obstacles to free thought in the digital age. The new frontier is the mind.

### **ARTIFACTS, CONTENT, AND COPYRIGHT**

In *The Book on the Bookshelf*, Henry Petroski gives a wonderful account of the history and development of book collection and management through the ages. He dwells on the limitations and barriers to access that were accompanied by each system, which gave rise to new technologies. As the bank of human knowledge has grown, so has the issue of how to house this immense collection. Petroski introduces a number of innovators and how they proposed to resolve the storage and maintenance of information in print, and how, as times changed, the utility of their solutions became inadequate for each new age. This is associated to what Anderson calls “the tyranny of physical space” (2004).



Books have been, and continue to be, *the* tried-and-true technology for fixing human thought in a tangible form. The anatomy of the book and the process of publication spawned industries and legal instruments to regulate the creation, control, and use of the intellectual property embodied within a cultural work and its manifest formats. New technologies are changing the rules, developing new methods for both liberating and regulating content. Copyright owners, armed with the deep pockets and influence of the entertainment industry, recognized a potential slippage in their control and have put pressure on Congress to update copyright law to keep pace with the changing infoscape.

At one point in history, reading text was so integrally linked to the printed page that one was thought hardly possible without the other. People were accustomed to this symbiosis, and the rules of the game were determined by it. Digital content is very different from printed content. Its behavior, while dependent on software and hardware for creating, managing, and rendering, is not necessarily beholden to a unique container. Its nature is ephemeral. The rate at which ideas can be shared in a digital environment is one step behind spoken word and two behind thought. These qualities are as much a blessing as a curse. Attitudes regarding digital content are divided between the revolutionaries, who tout the promise of a free culture on the Internet, and information age Bourbons that are digging in their heels to maintain control of their intellectual property. With the Digital Millennium Act of 1998, the balance seems to have tipped in favor of the Bourbons.

The reign of print and the tyranny of physical space are steadfast and stubborn. Many publishers have clung to the traditional ways and means of business. They are fully aware of the change in the wind, but are still struggling to retool for the digital age. A recent report posted on the Web site of the Association of American University Presses (AAUP) illustrates the changing tides:

Electronic book technologies, laws, and practices continue to evolve and will play a significant role in scholarly publishing in the future. However, printed books will continue to dominate the market for the foreseeable future. In general, scholarly publishers view e-books as a complementary means of distributing their works to libraries and readers, not as a substitute for printed books. Once the rights issues are solved, electronic access to scholarly works may bring more readers to these titles and enhance sales of paper editions (Nitterhouse 2005).

The Constitution of the United States, Article I, Section 8, states: “The Congress shall have Power . . . To promote the Progress of Science and useful Arts, by securing for limited Times to Authors

and Inventors the exclusive Right to their respective Writings and Discoveries” (U.S. National Archives & Records Administration 2005). Time and time again, this noble mission to balance the progress of arts and sciences for the benefit of society with the rights of copyright owners to profit from their creativity has been upheld by judges and politicians. This hasn’t been an easy task, especially when new technologies have enabled the circumvention of current laws, and on occasion it has been necessary to overhaul the system, *e.g.*, the Copyright Act of 1976.

As Lawrence Lessig points out in his book *Free Culture*, “the Internet should at least force us to rethink the conditions under which the law of copyright automatically applies, because it is clear that the current reach of copyright was never contemplated, much less chosen, by the legislators who enacted copyright law” (Lessig 2004). Every time content is accessed over the Internet, a copy is made and copyright assignment is automatic for all content in fixed form. Copyright law grants exclusive rights to copyright owners for control over the reproduction, distribution, and public display of their content, all of which fall within the scope of downloading a Web page. Therefore, the simple act of reading online is technically an infringement of copyright. However, the *fair use doctrine* would limit these exclusive rights in many cases. The key issue is that no one had imagined the behavior and nature of digital content and the Internet during the reign of the print paradigm. Copyright law has become very complicated and cumbersome. It may be time for another overhaul.

### **GREASING THE WHEELS OF REGULATION**

All things being equal, a free culture is an economic, political, psychological, and social environment in which intellectual enterprises are permitted and promoted, such that invention and innovation flourish and manifest in tangible products and services in an open marketplace. We are not operating in a free culture, and the first phrase of the aforementioned definition illuminates the reason why; all things are not equal. The current state of copyright law has progressed to such a complex state that it is more likely to impede the progress of invention and innovation rather than promote it. The good news is that there are many highly qualified information and library professionals working to rectify this problem. For example, The Library of Congress has spearheaded two initiatives to address the issues; the Section 108 Study Group and the National Information Infrastructure and Preservation Program (NDIIPP). While NDIIPP’s mission extends beyond

intellectual property issues, both groups are working diligently on recommendations for recovering the balance between the interests of copyright owners and the public good. On another front, the Google Print project promises to provide “universal access” to a critical mass of public domain works on the Web.

### **SEARCH AND DISCOVERY V. CEASE AND DESIST**

Some people are aghast that Harvard University, The New York Public Library, Stanford University, and the universities of Michigan and Oxford (hereafter called the Google Five) have teamed up with Google, Inc. in the search service giant’s more recent endeavors to “organize the world’s information” (Google 2005). The controversy surrounding the Google Print Library Project appears to have polarized as a defense for intellectual property rights on one hand and the public domain on the other. Publishers are concerned about the protection of their investments and librarians are worried that the Google Five have given away the farm.

AAUP and the Association of American Publishers (AAP) have independently issued letters of concern that the enterprise will infringe upon the copyrights of their constituencies. The later demanded a project moratorium until Google could ensure that the group’s economic interests were not at stake (Carnevale and Young 2005). Google maintains that its mass-digitization project is within the bounds of fair use, however have postponed scanning copyrighted works until November until this issue can be further discussed with concerned publishers (Noguchi 2005).

Publishers are not the only ones keeping a skeptical eye on Google and its partners. A statement posted on Google Watch admonishes Michigan’s partnership with Google lamenting, “The library at the University of Michigan has betrayed the trust that we placed in them, as a public institution that acts as a custodian of our public-domain heritage” (Brandt 2005). Michael Gorman, president-elect of the American Library Association, says, “If you’re going to spend millions of dollars and you’re interested in getting research materials to scholars, wouldn’t it be better for Google to spend money improving their own business” (Young 2005).

On the other end of the spectrum, some remain optimistic of the promise of universal access. In a press release from Stanford University, University Librarian Michael Keller stated, “This is a great leap forward” (Palmer 2005). The University of Michigan President Mary Sue Coleman said, “We are exhilarated to join a

partnership with Google that perfectly advances our mission as a great public university to share knowledge within the academic community and far beyond it” (Connell 2004). At the Coalition of Networked Information Spring 2005 Task Force Meeting, Executive Director Clifford Lynch mused, “This may be one of the first mass digitization projects of the 21st century . . . the first of truly large-scale enterprises” (Lynch 2005).

Lynch’s observation articulates an important output of the project; establishing a critical mass of digital content available on the Web. However, the ultimate outcome is that Google may then become *the* online public access catalog. Call it GooPAC, the Google open public access catalog, on the open Web. Log on, look up, and link to.

Whether or not librarians like the idea of GooPAC, scholars and lay people alike are turning to Google as their first-choice search tool. In the United States, a nationwide telephone survey by the Pew Internet & American Life Project and comScore placed Google as the most often used search engine (Fallows and Rainie 2005). In the United Kingdom, the Contextual Resource Evaluation Environment (CREE) project, funded by the Joint Information Systems Committee (JISC), conducted a national online survey of higher education institutions in the United Kingdom, observing that “Google is used by everyone” (Awre 2005). Given this data, the CREE project has been developing a Google portlet prototype to be used in course or learning management systems and library portals.

## **OUTPUTS AND OUTCOMES**

What are the benefits of Google Print to communities of practice, be they the general public, libraries, or publishers?

- multiple access points to resources
- interoperability with any Web browser
- no technical support required
- no direct costs to users
- full-text and natural language searching enabled
- on-demand availability
- platform-neutral search tool
- increased visibility of print resources and libraries
- marketing tool to drive book sales, new and used

One outcome of Google Print is exposure to foreign language, hard-to-find, and out-of-print books to Internet users worldwide. This impact will receive its share of praise from the general public. More specifically, providing Web access to these materials for children and disabled, elderly, and infirmed adults, who may be conditionally confined to their homes, will undoubtedly be met with applause. K-12 students and teachers will undoubtedly flock to Google Print for learning and teaching. In a recent study funded by the Institute of Museum and Library Services (IMLS), Lankes interviewed teachers and discovered they had maybe 30 minutes a day for information seeking and retrieval. Google Print could facilitate lesson planning for teachers in certain subject areas, *e.g.*, American history and literature, by expanding their access to resources. In addition to providing the *K to gray* slice of the general public access to public domain works online, Google Print stands to increase awareness of copyright law, intellectual property issues, and the concept of the public domain. Increasing awareness may increase civic involvement in copyright legislation and public policy.

Where are libraries in the Google Print Library Project? First of all, libraries are providing the content that is the foundation for the enterprise. Finally, in the case of public domain works, libraries will be the destination for users looking to borrow print copies of the books they discover using the search service. OCLC has teamed up with Google and provided a link to the Open WorldCat database. The Find in a Library feature, included on the browsing page for public domain items, searches WorldCat to connect users to libraries that hold physical copies of the digital files they view online. OCLC began exposing WorldCat records to Google during a pilot program in 2003 (Quint 2003). This promises to draw users back into libraries and increase the circulation of books.

The Google Print Library Project will build a critical mass of digital content, surpassing the best efforts of government- and foundation-subsidized projects. With Google financing and executing the rote task of scanning, their library partners can concentrate on building products and services around the digital files they receive. This is where the value of a library lies; adding value to content by enriching it and developing services to facilitate resource discovery, retrieval, and use. Google hasn't released any details about the digitizing workflow, and one can imagine there will need to be additional work done to ensure the integrity and preservation of the digital files and the accompanying

metadata. Google's mission is "to organize the world's information and make it useful and universally accessible" (Google 2005). Libraries share this mission, and have a lot experience in achieving it.

### **THE CHALLENGE FOR DIGITAL LIBRARIES**

Overcoming the legal hurdles to building digital library services around digital surrogates of copyright-protected works is not trivial. Libraries are permitted to lawfully make limited copies of copyrighted materials upon request from users and for preservation purposes. Preservation reformatting can even be outsourced to commercial vendors. However, systematic copying of such works is prohibited. Google claims its digitization program is exempt from copyright infringement pursuant of the fair use doctrine. Some publishers sorely disagree.

It's not clear how the Google Five intend to allow their users to access the digital files attained from Google. Current copyright law only permits libraries to publicly display digital surrogates of copyright-protected work on the premises of the library. The digital revolution has left the walls of the library porous and library users are becoming very accustomed to remote access to and network-enabled document delivery of resources. Will the Google partners confine all the digital files they receive from Google to single library workstations? Intranets or LANs? VPNs? WANs?

The University of Michigan has posted its cooperative agreement with Google on the Web. The contract reveals that Michigan has reserved the right to share their files with consortial partners, *e.g.*, members of DLF. Since Michigan is a member of DLF, it may choose to arrange for document delivery or interlibrary loan agreements with other research libraries in the DLF membership (University of Michigan 2005). Stanford's terms of use are non-exclusive, and Oxford has an agreement to share their files with other universities in the UK. This could pave the way for cooperative collection building of digital content on a very large scale.

There are many challenges and opportunities for the Google Five in the future. Long-term preservation of digital assets is still an unanswered problem. Google's Web-based, full-text index doesn't answer all of the organizational solutions that libraries regularly build into their products and services, *i.e.*, controlled vocabularies, dictionaries, taxonomic indexing, thesauri, subject heading classification. One thing's for sure, Google chose the right partners

to work with. Michigan alone is providing half of the items to be digitized. It has been a pioneer in digital library development, establishing the Digital Library Production Service (DLPS) in 1996. Michigan's experience with digital projects as varied as Early English Books Online (EEBO), JSTOR, and the Making of America (MOA) project, attest to why it is an excellent choice for collaboration, governance, and oversight of the Google mass-digitization project.

The time has come for more libraries to step up to the plate to leverage their collections and services to better serve their constituencies. Libraries no longer just passively collect, lend, organize, and preserve relics of cultural production. There has been a shift in the universe. More and more libraries are assuming the creative and proactive roles of content providers. This is risky business, since it breaks with the culture of traditional librarianship, and it may become essential for the survival of libraries as online communication, learning, and teaching increases.

As the former Council on Library and Information Resources (CLIR) Director of Programs Abby Smith points out, "We have little understanding of the online environment at present and no understanding of how it and its users might evolve over time. Hence the need for projects that take calculated risks and try something new."

## REFERENCES CITED

- Anderson, Chris. "The Long Tail," *Wired Magazine* 12.10, (October 2004). <http://www.wired.com/wired/archive/12.10/tail.html> (accessed on June 29, 2005).
- Association of Research Libraries Collections and Access Issues Task Force. "Collections & Access for the 21st-Century Scholar: Changing Roles of Research Libraries," *ARL Bimonthly Report* 225 (December 2002). <http://www.arl.org/newsltr/225/main.html> (accessed on July 7, 2005).
- Awre, Chris. *Project CREE: Examining User Behavior*. Presented at the CNI Spring 2005 Task Force Meeting, Loews L'Enfant Plaza Hotel, Washington, D.C., April 4-5, 2005.
- Bowen, William G. *At a Slight Angle to the Universe: The University in a Digitized, Commercialized Age*. Princeton: Princeton University Press, 2001.

- Brandt, Daniel. "Google-eyed U.Michigan Gives Away Its Library," *The U.Michigan-Google Caper*, (June 15, 2005). <http://www.google-watch.org/umich.html> (accessed on June 29, 2005).
- Carlson, Scott. "The Revolution Will Be Digitized," *The Chronicle of Higher Education* 51, Issue 34 (April 29, 2005). <http://chronicle.com/weekly/v51/i34/34a03001.htm> (accessed on June 30, 2005).
- Carnevale, Dan, Jeffrey R. Young. "Publishers' Group Asks Google to Stop Scanning Copyrighted Works for 6 Months," *The Chronicle of Higher Education* 51, Issue 43 (July 1, 2005). <http://chronicle.com/weekly/v51/i43/43a02901.htm> (accessed on June 30, 2005).
- Connell, Nancy. "Google/U-M Project Opens the Way to Universal Access to Information," *University of Michigan News Service* (December 14, 2004). <http://www.umich.edu/news/?Releases/2004/Dec04/library/index> (accessed on June 29, 2005).
- Dempsey, Lorcan, Brian Lavoie, et al. *DLF Service Framework for Digital Libraries*. Washington, D.C.: Digital Library Federation, 2005. <http://www.diglib.org/architectures/dlfserviceframe.htm> (accessed on July 11, 2005).
- Fallows, Deborah, Lee Rainie. *Data Memo*. Washington, D.C.: Pew Internet & American Life Project, comScore, 2004. Distributed by James Farmer, June 30, 2005.
- Google, Inc. "About Google Print," Google Print. <http://print.google.com/googleprint/about.html> (accessed on July 7, 2005).
- Hansen, Carol, Donald W. King. "Comparing Library and User Related Costs of Print and Electronic Journal Collections," *D-Lib Magazine* 8, no. 10 (October 2002). <http://www.dlib.org/dlib/october02/Montgomery/10montgomery.html> (accessed on June 29, 2005).
- Jeng, Judy. "What is Usability in the Context of the Digital Library and How Can It Be Measured?" *Information Technology and Libraries* 24, no. 2 (June 2005), pp. 47-56.
- Kyrillidou, Martha, Mark Young. *ARL Statistics 2002-03*. Washington, D.C.: Association of Research Libraries, 2004.
- Lamolinaro, Guy. "Section 108 Study Group Convenes to Discuss Exceptions to Copyright Law for Libraries and Archives," *News from The Library of Congress*, (May 13, 2005). <http://www.loc.gov/today/pr/2005/05-121.html> (accessed on June 30, 2005).
- Lankes, David. *Breaking the K-12 Crust: The Realities of Digital Libraries for Education*. Presented at Web Wise 2005: Teaching and



Learning with Digital Resources, February 16-18, 2005, Washington, D.C.

- Lynch, Clifford. "Opening Plenary Session," CNI Task Force Meeting, Loews L'Enfant Plaza Hotel, Washington, D.C., April 4-5, 2005.
- Nitterhouse, Denise. *Digital Production Strategies for Scholarly Publishers*. Chicago: The University of Chicago, 2005.  
[http://www.bibliovault.org/digital\\_prod\\_strategies.pdf](http://www.bibliovault.org/digital_prod_strategies.pdf) (accessed on June 29, 2005).
- Noguchi, Yuki. "Google Delays Book Scanning," *The Washington Post*, (August 13, 2005), p. D1.
- Palmer, Barbara. "Deals with Google to Accelerate Library Digitization Projects for Stanford, Others," *Stanford Report*, (January 12, 2005).  
<http://news-service.stanford.edu/news/2005/january12/google-0112.html> (accessed on June 29, 2005).
- Pope, Barbara Kline, P. K. Kannan. "An Evaluation Study of the National Academies Press's E-Publishing Initiative: Final Report," *Public Version* (January 31, 2003).  
<http://aaupnet.org/resources/mellon/nap/index.html> (accessed on June 29, 2005).
- Quint, Barbara. "OCLC Project Opens WorldCat Records to Google," *NewsBreaks* (October 27, 2003).  
<http://www.infoday.com/newsbreaks/nb031027-2.shtml> (accessed on June 30, 2005).
- Smith, Abby. Issue in Sustainability: Creating Value for Online Users. *first monday* 8, no. 5 (May 5, 2003). Available from  
[http://www.firstmonday.org/issues/issue8\\_5/smith/index.html](http://www.firstmonday.org/issues/issue8_5/smith/index.html) (accessed February 16, 2005).
- United States. *The Constitution of the United States: A Transcription*. Washington, D.C.: U.S. National Archives & Records Administration, 2005.  
[http://www.archives.gov/national\\_archives\\_experience/charters/constitution.html](http://www.archives.gov/national_archives_experience/charters/constitution.html) (accessed on July 7, 2005).
- University of Michigan Digital Library Production Service. *About the Digital Library Production Service*. Website.  
<http://www.umdl.umich.edu/about.html>.
- Waters, Donald J. *The Digital Library Federation: Program Agenda*. Washington, D.C.: Digital Library Federation, 1998.
- Young, Jeffrey R. "From Gutenberg to Google," *The Chronicle of Higher Education* 51, no. 39 (June 3, 2005), pp. A24-27.
- Young, Mark, Martha Kyrillidou. *ARL Supplementary Statistics 2002-03*. Washington, D.C.: Association of Research Libraries, 2004.

# How Do We Sustain Digital Scholarship?

Bradley J. Daigle (University of Virginia)

**Abstract:** Digital scholarship can be viewed as a method of scholarly communication, research, and exchanging ideas that employs modern forms of technology (*i.e.* a collated digital edition of a text). Digital scholarship exponentially increases the services and support expected from librarians, pushing libraries to revisit their strategies for meeting researchers' needs and introducing the new role of librarian *cum* technologist. With the integration of library services and digital scholarship, libraries now find themselves playing a leading role in how faculty research is developed and disseminated. A major challenge for libraries in the face of these changes lies in developing, forecasting, and even imagining a consistent approach to digital scholarship. This paper will explore the development of a flexible model for sustaining digital scholarship that can be applied to any level of research.


## **DIGITAL SCHOLARSHIP: I'LL KNOW IT WHEN I SEE IT.**

Within the last few years, research libraries have been transforming themselves into centers for emerging technologies. This recent development is largely due to the nature of what can be described as “digital scholarship.” But what does this term convey? The nature of “digital scholarship” seems to elude easy categorization. If this is the case, then why bother? Often the most interesting aspect regarding the variety of definitions is the entities that are trying to make them. For example, one definition states the following:

Digital scholarship is publication that (1) exists in digital format, *i.e.* as an electronic file or set of files that can be stored, transported, and displayed on general-purpose computers or other devices that manipulate digital files; (2) is incapable of being translated without loss of information or value into a non-digital format, such as that of a printed book, because it makes use of media, tools, structuring, or other features of computer presentation that cannot be conveyed in any other medium; and (3) is subject in all other respects to the demands of traditional print scholarship for originality, value, and selection via a process of peer review.<sup>1</sup>

---

M. Halbert (Ed.): *Free Culture and the Digital Library Symposium Proceedings*. Atlanta, Georgia: MetaScholar Initiative at Emory University, 2005. pp. 230-240.

 This work is covered by the Creative Commons Attribution-NonCommercial-NoDerivs 2.5 License.

Another definition—this one from a university faculty member states: “Digital Scholarship can be defined as any element of knowledge or art that is created, produced, analyzed, distributed, published, and/or displayed in a digital medium, for the purpose of research or teaching.”<sup>2</sup> And finally, by way of example, I offer my own as a librarian at a public university: it can be viewed as a method of scholarly communication, research, and interplay of ideas that employs modern forms of technology. Of these definitions, the first clearly draws the connection to the print industry; the second, is more wide open and emphasizes the medium. Lastly, my definition (which is not necessarily the same as the University of Virginia Library’s) focuses on the tradition of the communication of ideas though now by using modern technology.

The two elements I believe that most proponents of these ranging views would agree on is that the definitions are all “working” definitions and that it is a far simpler task to state what digital scholarship *is not*. In looking at these definitions, I see the struggle to define a term that we are all attempting to support. If something is easily defined, we can adjust our missions and our strategies to support it. This support will certainly be different for each institution. In this essay, I will be speaking primarily to the role of academic libraries—specifically the University of Virginia—and our relationship to digital scholarship. Clearly, in terms of free culture, our mission is to make this scholarly output freely available to the public. This is not the case for every stakeholder in supporting and sustaining digital scholarship. Thus, for those faculty members who want to keep their work freely available, they turn to the libraries. This is a new role that some academic libraries have taken on.

Why would this be the case? Certainly libraries have been involved in new forms of technology since their inception. They have also always supported faculty research and instruction. In fact, librarians pride themselves on their roles as purveyors of information—information in all media and mediums. Have the parameters changed? Have the missions of research libraries really changed?<sup>3</sup>

With the advent of digital scholarship, one could posit that the libraries’ missions have not really changed tremendously. Libraries and librarians still strive to support the teaching and research of

their users—this will always be true. What has changed is the *how*. Digital scholarship exponentially increases the services and support expected from librarians. These services involve the digitization of materials (books, manuscripts, motion media), sophisticated data manipulation (statistical data sets, GIS), complex database management (Perl, SQL), and textual markup (XML, TEI)—to name but a few. Conventional scholarship required faculty and students to use library materials in traditional methods (publishing in journals or books, using reference materials, again to name a few). Based on the discussions that we have held here at the University of Virginia, scholarship itself (at least in the humanities and social sciences) has not changed dramatically with the advent of digital scholarship—once again it is *how* it is employed. What has changed for libraries is the bifocal nature of their existence: on one hand, “traditional” services still form the backbone of all library services—digital and otherwise; and now there is an entirely new layer of services actively engaged in research libraries—most of which involve digital technology in some form. As most would agree, digital libraries and librarianship will not replace brick and mortar libraries or for that matter, supplant “traditional” librarianship. The fact that digital and traditional aspects of librarianship will never be mutually exclusive remains a daunting one for all libraries. Digital scholarship acts as the nexus for that interrelation, pushing libraries to revisit their strategies for meeting researchers’ needs. It has added new roles for librarians—technologist and intellectual property rights manager to name a few.

### **Research Libraries’ Renaissance of Self-Fashioning**

The fact that a concrete definition of digital scholarship remains elusive does nothing to belie the need for rapidly changing library services to support it. How do research libraries identify their role within digital scholarship? Libraries are now expected to employ experts in emerging technologies, particularly with respect to the services they offer both in hardware and software. Libraries in general, and research libraries specifically, have begun this process of self-fashioning—adapting their support models to the digital demands of their users. However, as the term “self-fashioning” suggests, this is really an artificial distinction: underneath the surface, the mission of the library remains constant. Thus, in order to support what they have always excelled at, namely organizing and preserving information, libraries now have research and development units, digital specialists, and legal counsel.

This is not surprising in the least. The synergetic nature of digital scholarship has the potential to harmonize the efforts of units who historically saw themselves as separate. Thus one major change in the role of research libraries is the need to bring in support partners. With the integration of library services and digital scholarship, libraries now find themselves playing a leading role in how faculty research is developed and disseminated. This makes it unlikely that a single unit within a higher education institution can be entirely successful in offering a comprehensive approach to digital scholarship. Libraries certainly cannot do all this by themselves. Collaboration among university units that are conducting digital scholarship seems to be the best solution. It becomes absolutely necessary that libraries work closely with other units within the university sphere such as technology units, communication units, faculty-driven centers and institutes, and academic departments to explore the changing relationship among libraries and the university environment. As Clifford Lynch points out: “[o]nly an institutionally based approach to managing these data resources, which operates in alignment with what the faculty at each individual institution are actually doing, can provide a comprehensive dissemination and preservation mechanism for the data that supports the new scholarship in the digital world” (Lynch 2003). All these elements comprise what is now called a “trustworthy” repository.<sup>4</sup> Libraries cannot do it alone but they can certainly lead the way.

A major challenge in this overall interdisciplinarity in the higher education environment is the need for flexible and scalable support structures to support and sustain digital scholarship. While there are many established processes that handle traditional print scholarship, it is clear that institutions do not yet have the necessary policies and infrastructure to handle issues related to digital scholarship: core library issues such as collection, preservation, and access. The rules of engagement on these fronts have largely changed. How do we begin to approach the questions of what to collect, how to preserve collections for posterity, and how to provide the appropriate access and rights management to the materials? Libraries need to have a definable support structure for digital scholarship. These structures should be tied directly to the institution’s own repository. With respect to the goals of digital scholarship, the role of the library and its institutional repository they should function in a complementary manner, and as a “supplement, rather than a substitute, for traditional scholarly publication values” (Lynch 2003).

It is with these issues in mind that the University of Virginia library has been working for the last five years with a variety of partners to develop an approach to supporting digital scholarship. The University of Virginia Library partnered with the Institute for Advanced Technology in the Humanities (IATH) in 2000 in the Mellon sponsored Supporting Digital Scholarship (SDS) grant. The goals of this project were to “propose guidelines and document methods for libraries and related technical centers to support the creation and long-term maintenance of digital scholarly projects.”<sup>5</sup> Both ambitious in its scope and detailed in its documentation, the original SDS grant forms much of the underpinnings of our current approach. It analyzed digital scholarship from both a technical and a policy perspective. The current goal of what is now termed as “sustaining digital scholarship” can be stated as follows: to develop a socially and technologically sustainable and scalable model for the support of digital scholarship. The operative words in the statement are *sustainable* and *scalable*. Sustainable gestures to the “trustworthy” nature of the institution (both technologically and conceptually) to continue to support faculty research and scalable to grow those research support models as needed. A scalable model is absolutely required to continue to grow and support shifting users and technologies.

As stated earlier, current methodologies for production, delivery, access, and preservation of digital content need to extend beyond local solutions and move toward a more collaborative model to become a viable, sustainable, and scalable solution. Thus in order to fully articulate the impact of libraries on digital scholarship and the impact of digital scholarship on libraries we need look more closely at the model we are developing at the University of Virginia.

### **Digital Scholarship: The Process**

A major challenge for libraries in the face of both the nature of digital scholarship and its transformative effect on the roles of library support lies in developing, forecasting, and even imagining a consistent approach to digital scholarship. As mentioned earlier, the goal of cross-unit collaboration is to develop a flexible model for sustaining digital scholarship that can be applied to any level of research. In order to sustain digital research one must first define its scope. The process that the University of Virginia Library is currently testing will form a model that will look like this:

Stage 1: Determine the scope of the research as defined by project, class, or some other delineation;

Stage 2: Collection and selection;

Stage 3: Assessment or analysis of the digital scholarship components;

Stage 4: Develop and formalize agreements between parties;

Stage 5: Implement service and procedural methods to formally “collect” the digital scholarship;

Stage 6: Deliver via agreed upon method.

Buried in these stages are multiple layers of policies and technologies. They form the base that can function at a high level to provide a general structure where an institution can develop critical strategies and employ partners more efficiently. There are clearly stages that can be owned by one institutional unit and others that have shared responsibilities.

Determining the scope is often the most difficult aspect of digital scholarship. Where does scholarship begin? Where does it end? This is also the most critical step in the support model. It is the element that defines the digital scholarship without having to define digital scholarship overall. At the University of Virginia library, we began this phase by selecting mature, highly complicated digital endeavors and found almost immediately that the process stalled out by Stage 3. It became too difficult to analyze the projects because we had inadequately defined their scopes. Thus, each instance of digital scholarship was pared down to a smaller, discrete element we termed the “pilot.” This is a critical juncture and one which was agreed upon by both the faculty member and the library. Selecting smaller but representative elements of the overall project provided the proving grounds for the pilots.

Once the scope was agreed upon the library could decide whether or not the digital scholarship was something it was willing to collect.<sup>6</sup> Assuming the selection and collection phase closely resembles that of traditional library content selection. The need to involve both content specialists in the library as well as peer-review documentation of the digital scholarship will ensure the longevity of the materials. The fact that digital collections should

mirror physical collections and that they both reflect the overall institutions research mission should be underscored.

After determining the scope and whether or not the project would be collected, the next phase (Stage 3) deals with the assessment and analysis of the scholarly work. It involves a detailed analysis of the technical components, access and delivery issues, and preservation strategy. Each of these steps involves several policy decisions. These are classic library-related issues that arise frequently. What to do if the materials fall squarely within the library's collecting parameters but we do not have any current method to support its specific technology? Should we reject the material or accept and add the technology to the library's existing support model? The response to this question will vary with each collection we approach. However, if the decision is to support the new model in order to collect this scholarship then it clearly forms the basis of an argument for why we would need the additional resources.

Secondly, issues relating to access/delivery with the digital scholarship need to be clearly defined prior to collecting. This is another element of the collaborative nature of digital scholarship and is directly related to the concept of free culture. These are often closely tied to the exact needs of the researcher and end-user though they are not always the same. For the majority of all digital scholarship that falls within the purview of our broad definitions the question becomes one of intellectual property (IP) rights. These IP issues, perhaps most importantly, need to be clearly delineated from the outset. This is why we include it on our initial analysis of the content. Part of the analytical process involves breaking the actual technology up into its component parts as best as we are able (images, text, data sets, motion media, software) in order to ascertain the specific rights associated with each piece. Each project must submit a rights inventory assessment form. This is meant to begin the process of identifying the rights associated with each part—often the most muddled element of the entire process. As a public institution, one of the main objectives for the University of Virginia Library is to support the free access of its materials to the widest possible audience. Hence, a project with significant rights entanglements will not be the first to be prioritized for collecting. It is imperative that librarians “maintain [a] current awareness of these issues [IP and access] to understand emerging technologies and assist in its development, advocate for users' rights as legislation is drafted and revised, and ensure both



easy and enduring access to information resources available in, or through, the library” (Von Elm 2001). At a minimum, the University of Virginia Library will aggressively work to provide access to these materials for people within the University of Virginia network. The type of allowable access will also have an impact on what is delivered. For example, while on the University of Virginia campus a user may have access to images and text as part of a faculty project but once off campus, only thumbnails are available. In some cases, this will be documented at the item level before any agreements are put in place. The University of Virginia Library is currently investigating five separate access levels from tight restrictions to none: restricted (*i.e.* closed to all but the project members); classroom (for specific courses only); University of Virginia only (physical presence required on campus); Virginia consortium membership use (for participating institutions); and unrestricted.

One of the last elements of Stage 3 is to develop a preservation strategy that the library and faculty member can agree upon. This can only really be done once all the pieces that go into the particular project have been identified and analyzed. The original SDS grant proposes several preservation levels—each of these specific levels is based on select criteria and collecting strategies.<sup>7</sup> We are using these today as the basis for the collection level agreement. Each of these levels contains criteria against which all the above elements are tested. This part deals primarily with both “look and feel” as well as format compliance. Obviously, the closer digital scholarship is to the library’s accepted standards the easier it is to provide sustainable services. On the other hand, we do not want to apply standards as an all or nothing scenario. This is the primary reason we engage in the process of collaborative decision making. No one unit sets all the terms by which these agreements are made. Rather each element of the “pilot” is weighed in context of the technology and policies that surround its use. Chances are high that the work faculty and students are doing will recreate standards and push open new dialogues. For this reason we make each policy flexible to recognize when it must bend to accommodate emergent thinking and technology. In this manner, libraries can respond to digital scholarship in the most synergistic manner possible.

The remaining steps follow much more quickly once the above work is accomplished. Stage 4’s main objective is to formally document three elements: what will be the nature of the

collaboration (scope) down to the most minute detail in or order to formalize the reciprocal expectations of what exactly is being collected; then, establish the workflow that needs to be executed; and lastly a timeline that can be mapped to that process. These parts form the essential elements to the agreement or memorandum of understanding. The majority of the work for Stage 4 should have been completed in the prior analysis and assessment stages.

Stage 5 is where the actual work is done and if any technology or workflow problems arise it may result in the digital scholarship's return to an earlier stage for revisions. If all proceeds according to the agreement then the final delivery in Stage 6 is all that remains. These last stages (4-6) are where we are currently in the process. There are layers of agreement (including tandem delivery of local project and those ingested into the institutional repository) that will still need more attention and piloting.

## CONCLUSION

Free cultures are cultures that leave a great deal open for others to build upon; unfree, or permission, cultures leave much less (Lessig 2004).

When one considers digital scholarship, it is easy to cast it in many different terms—specifically as a form of cultural output. Even within a single institution, faculty and students have different options the dissemination of their scholarship. They can adopt profit models such as technology services (paysites etc.), bookstores, university presses. However, I find it significant that it is to *libraries* that users frequently turn for the support, delivery, and preservation of the materials that are being delivered “free” to the world. The new services and roles that librarians are taking on reflect this change. Though delivery can mean publishing to many, it is the long term support and preservation that speaks to the real value of the service. This is why academic libraries today are trying to balance the need for standards-based support models for longevity while at the same time creating digital workbenches where creativity, not form is foremost. As a true form of cultural expression, digital scholarship must be adequately supported to ensure its availability to the public for generations. Libraries are not doing this work alone—that is clear. Yet the historical role of libraries as an institution hopefully will highlight the need to keep that access available to as many people as possible and as long as we continue to have libraries.

## ENDNOTES

1. This definition comes from the University of Virginia Press's Electronic Imprint—a commercial entity (<http://www.ei.virginia.edu/digitalscholarship.html>).
2. Kirsten Foot, Assistant Professor, Department of Communication, University of Washington (<http://www.lib.washington.edu/digitalscholar/>).
3. For the purposes of this argument, the specific institutions I discuss are research libraries in general and the University of Virginia specifically. Whereas it is true that digital scholarship happens in many levels in many different types of libraries, I believe that each institution has to develop its own approach according to its established mission. One would hope that despite this variety we could agree on a general level how digital scholarship can interoperate among us all.
4. The concept of a “trustworthy repository” comes from the commercial world. A quick internet search for the term returned 97 hits, mostly dealing with commercial companies assuring depositors that their materials will be stewarded appropriately.
5. See the overview of the SDS Final Report, for a more detailed outline of the project's goals. The SDS project laid much of the foundation for the University of Virginia Library's current approach to what we now term as sustaining digital scholarship. The work of the original SDS grant is well documented (<http://www.iath.virginia.edu/sds/>).
6. One of the remaining questions that is still under development remains—namely, how to relate the pilot project back to the full project if that larger project has not been collected. This can be specific to an institution or to a specific faculty member. If the faculty member leaves the institution and continues the research elsewhere the question of scholarly continuity needs to be addressed.
7. The original grant suggests five separate levels: Level 1: Collecting metadata; Level 2: Saving the project as a set of binary files and metadata only; Level 3: The content can still be delivered as in the original; Level 4: Look and feel; Level 5: The project is completely documented. See the SDS final report for more specifics for each level ([http://www.iath.virginia.edu/sds/SDS\\_AR\\_2003.html](http://www.iath.virginia.edu/sds/SDS_AR_2003.html)—Section 4).

## REFERENCES CITED

- Andersen, Deborah Lines, ed. 2004. *Digital Scholarship in the Tenure, Promotion, and Review Process*. New York: M.E Sharpe.
- Buckland, Michael K. 2003. “Five Challenges for Library Research.” *Library Trends* 51. 4: 675-686.

- Lessig, Lawrence. 2004. *Free Culture: How Big Media Uses Technology and the Law to Lock Down Culture and Control Creativity*. New York: Penguin Press.
- Liu, Lewis. G. 2003. "The Economic Behavior of Academic Research Libraries: Toward a Theory." *Library Trends* 51.3: 227-292.
- Lougee, Wendy Pradt. 2002. *Diffuse Libraries: Emergent Roles for the Research Library in the Digital Age*. Washington, D.C.: Council on Library and Information Resources.
- Lynch, Clifford A. 2003. "Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age." *ARL* 226: 1-7. <<http://www.arl.org/newsltr/226/ir.html>>.
- Mann, Thomas. 2001. "The Importance of Books, Free Access, and Libraries as Places—and the Dangerous Inadequacy of the Information Science Paradigm." *Journal of Academic Librarianship* 27. 4: 268-282.
- Owusu-Ansah, Edward K. 2001. "The Academic Library in the Enterprise of Colleges and Universities: Toward a New Paradigm." *Journal of Academic Librarianship* 27. 4: 282-294.
- Quandt, Richard E. 2003. "Scholarly Materials: Paper or Digital?" *Library Trends* 51.3: 349-375.
- Von Elm, Catherine and Judith F. Trump. 2001. "Maintaining the Mission in the Hybrid Library." *Journal of Academic Librarianship* 27. 1: 33-36.

# A Scholia-based Document Model for Commons-based Peer Production

Joseph Corneli and Aaron Krowne

**Abstract:** Commons-based peer production is a term that describes authorship of shared information resources. In this article we examine the technical aspects of writing-in-common. We begin with a simple model: that of text and commentary. This scholia-based model emphasizes *ownership of speech* and *freedom of speech*. We then consider what happens when the *freedom to create derivative versions* is added to the mix. The resulting model proves to be quite sophisticated, and flexible enough to describe many different commons-based peer production systems. We provide an overview of our implementation of this model, and suggest some ideas for subsequent work. We conclude by discussing the implications of our model for distributed authorship and writing.

## INTRODUCTION

The simplest model of a document is a list of characters; a string, file, or buffer. This model is not diachronic; a diachronic model keeps track of editing operations, or summarizes them as a sequence of differences between document versions.


Many texts have features which can not be modeled adequately without still more information: markup or metadata. Compilations, hypertext, and collaboratively written documents are examples. In this essay, we advance the idea that it can be useful to treat such documents and document ensembles as collections of *scholia*.

### The Digital Library and the Document

For our purposes, the document and the library are essentially the same. In other words, the traditional library-document dichotomy can be viewed as a smooth spectrum, which we consider as a whole. Towards one end of the spectrum, the number of authors decreases and the topics under discussion become more integrated, and the information artifacts look more document-like. Towards the other end, the number of authors grows and the semantic gaps

---

M. Halbert (Ed.): *Free Culture and the Digital Library Symposium Proceedings*. Atlanta, Georgia: MetaScholar Initiative at Emory University, 2005. pp. 241-254.

 This work is covered by the Creative Commons Attribution-NonCommercial-NoDerivs 2.5 License.

between topics increase, and the information artifacts become more library-like. To be clear, when we wish to refer to elements of a given information artifact in a generic way, we will call them *articles*. We will use the term “scholium” to describe an article that is about another article. While a given article need not be about another article in general, our view is that it is about *something*, or perhaps about many things. An article that does not express these relationships explicitly is a degenerate scholium.

### **Scholia**

According to Webster,<sup>1</sup> a *scholium* (pluralized *scholia*) is,

1. A marginal annotation; an explanatory remark or comment; specifically, an explanatory comment on the text of a classic author by an early grammarian;
2. A remark or observation subjoined to a demonstration or a train of reasoning.

The Talmud is an excellent example of a document that is comprised of scholia. This document is a collection of interconnected commentaries and reflections on Jewish law (the *Torah*), composed by generations of Jewish religious scholars. There is an obvious and oftnoted comparison between the internet and the Talmud: we readily see ancient scholia-based documents as primitive hypertext. However, a deeper comparison is not simply technological but psycho-social.<sup>2</sup> There are similarities between our experiences of “culturally comprehensive” documents, whether secular or religious in nature. Of course, if you subscribe to the McLuhan view of media (“the medium is the message”), you would expect to see similarities (Federman 2004; McLuhan 1964).

Scholia also appear in the writings of mathematical scientists (*e.g.* Euclid, Galileo, and Newton) or philosophers writing in a similar style (*e.g.* Spinoza). Today, mathematicians typically call peripheral observations (frequently, of secondary importance) “corollaries” or “remarks.” Mathematicians also mark up parts of their texts as “axioms,” “definitions,” “propositions,” “theorems,” “lemmas,” or “proofs.” Markup, in general, can be thought of as scholia-based, *i.e.*, as commentary that instructs the typesetter or the reader to treat a given piece of content in a certain way.

## Free culture and the scholia-based document

This paper, and the system it describes, are a response to certain concerns that face us when we work with and think about shared information resources. These concerns have to do with the issues of *intersubjectivity*, *ownership*, and *freedom*. Content in a shared document is typically of non-trivial intersubjective importance. However, it is more than likely that different parties will have some different ideas about the information that is presented in a given text (it is no coincidence that the root of *intersubjective* is *subjective*)!

Free content (*i.e.* free as in freedom, or *libre* content) is one way to nurture difference. Free content can be modified and redistributed (and, in particular, *forked*) without permission or apology. Nevertheless, free content typically manifests aspects of a common resource as well as an open access resource; while anyone can do essentially whatever they wish with the content offline, in its online life, the content is managed in a socially-mediated way. In particular, rights to *in situ* modification tend to be strictly controlled. The details differ from ownership model to ownership model (*e.g.*, PlanetMath has article owners and access control lists, software projects like Emacs have a limited number of developers with commit privileges, members of the Wikipedia community enforce rules about what kinds of content are allowed, and so forth). In these instances and generally, unwillingness to cooperate comes at a cost of valuable support from the community, which must be balanced against the limitations the community imposes on the individual.

By finding new ways to support freedom of speech within CBPP documents, we embrace subjectivity as a way to enhance the content of an intersubjectively valued corpus. In the context of “hackable” media and maintenance protocols, the semantics with which scholia are handled can be improved upon indefinitely on a user-by-user basis and a resource-wide basis. This is free culture in action.

## Hyperreal Texts

Our interest in scholia-based documents largely derives from an interest in helping assemble a human-friendly, AI-enriched mathematics learning and communication interface. The immediate goal is a system in which text and code are both well-supported, as two sides of the same coin. Qualitatively, we want to be able to represent and work with complex ontological

relationships between entities that are encoded in the system. The system should be useful for both humans and programs—balance along the human/artificial axis is as important toward this end as balance along the freedom/ownership axis or the intersubjective/subjective axis. In short, we find it natural to use a scholia-based platform as the foundation of our *Hyperreal Dictionary of Mathematics* project.<sup>3</sup> Applications in other areas of intellectual inquiry (including digital libraries), as well as in business, government, and day-to-day life also seem to hold great promise.

### Inspiration

We have already seen glimpses of the ideology informing this paper (in particular, we've touched on free software, commons-based peer production, the Talmud, and mathematics). We'll come to more of this later on. Here, we would like to describe certain key technological inspirations. These should serve to illustrate the naturality of a scholiabased document model, as well as to contextualize our implementation of the model. Usenet, Slashdot, the World Wide Web and WikiWikiWeb are all inspirations. In the first two, the most obvious type of scholium is the *followup*. In the latter two, it is the *link*.

PlanetMath<sup>4</sup> uses several types of scholia. Discussion fora are attached to pretty much every “substantial” object in the system; there are auto-generated invocation links between articles; and attachment relationships can be asserted to apply to between objects. Objects can themselves be distinguished as being one of several different types and belonging to one of a number of different subject categories. Also, and most importantly for the current discussion, PlanetMath employs an explicit article ownership model.

Emacs text properties provide a facility for unlimited markup of strings, buffers, and files.<sup>5</sup> However, text properties are hierarchical (treelike), whereas we are looking for something web-like. Of course, locally, text properties are great (and, indeed, we use them in our implementation).

The semantic web project is exploring ways in which to make metadata about web pages available to computers with reasoning capabilities.<sup>6</sup> Our scholia-based documents are small “semantic webs” with a particular flavor, though it may be more appropriate to use the older and more generic term “semantic network” (Quillian 1968) to describe them. As a data structure, a scholia-



based document is a network with arbitrary text (and metadata) on its nodes and arbitrary annotations on the connections between nodes. Graphically, this closely resembles the notion of a *concept map* (Novak 1998), albeit with significant text content attached to each node.

Semantic networks and concept maps are just two of the many different strategies for knowledge representation. Cyc<sup>7</sup> and KM<sup>8</sup> are two other systems we've studied. Our goal is a system with similar expressive power, but which is more human-friendly (with *reasoning* capabilities to come eventually).

The first author's earlier hypertext experiments were heavily inspired by LISP, and the results ended up resembling Gopher (Karger 2000). The current system is meant to be much richer, but it is still LISP-inspired, particularly in terms of support for self-modification. Indeed, one of the key inspirations for the current system are the text-based "game" environments (MUDs,<sup>9</sup> etc.), in which people are able to "hack" the text-based virtual worlds that they are interacting in and with, in real time.<sup>10</sup> Scholia can include actionable features, and with appropriate support, contributing authors can modify the medium in which the scholia reside, not just its content.

## Survey

The scholium-based document is not a new idea. Here, we are referring not to classical works, but to Ted Nelson's project and document model entitled *Xanadu*, which is described in his book *Literary Machines* (Nelson 1990), originally published in 1982. His development strategy was quite different from the one taken here, however—in particular, while he and his cohorts were focused on developing powerful server-side technology, we focus on implementing user-ware with a simple p2p extension. Importantly, while at least some of Nelson's ideas have been implemented and released under a free software license, the code didn't compile for us.<sup>11</sup> A biography of sorts together with a history of the Xanadu project has been published in *Wired*.<sup>12</sup> Nelson is admired by FSF General Counsel, Eben Moglen, for "identifying the predicament of information ownership in the digital age" (Williams 2002).

One rather important difference between Nelson's vision (as expressed in *Literary Machines*) and the present one is that he deprioritizes artificial intelligence in favor of human intelligence. As remarked above, actionable features and artificial intelligence

occupy an important place in the scholium system. Our situation with regard to artificial intelligence is similar to the one described by Minsky in *The Society of Mind* (Minsky 1985). The connection between ideas and agents is not so dissimilar; only the translation to actionable form is missing to make an idea into an agent. Meaning clusters translate to complex agencies. As time goes by, we expect to find actionable and non-actionable features paralleling one another along some dimensions, intertwining along some, and diverging on others.

“Superimposed information” is a subject of current research in the field of digital libraries.<sup>13</sup> Note that while this model is *locally* similar to the scholia-based document model, this branch of research focuses on one superimposed layer. This makes a certain amount of sense for traditional libraries, which hold a specific, administratively-controlled collection of information. The superimposed information model emphasizes making this primary artifact more useful via value-added “attachments” (annotations and so forth). However, for us, neither library nor document is static, and annotations are an integral part of both. Thus, scholia-based documents are as much “community” as they are “collection.”

We find it compelling that, in the context of a digital library, marginal conversations within a text provide a chance for readers to interact with primary authors and with each other, and to become primary authors themselves, all at once. While marginalia are considered to be *vandalism* in physical library books, in a digital library, there is no reason to fear them—they can easily be hidden away. The scholia-based document model reflects the standard postmodern pun which says that writing on a text or subject (*i.e.* criticism, discourse) is writing *on* the text (making ones mark). We expect meaning to accumulate in the “margins” of texts, and for meaning complexes to grow by stitching documents together along their margins.

## IMPLEMENTATION OVERVIEW

In this section we give a tour of our implementation of the scholium system. The critical elements of the system are described in general terms here. Details, including code (and literate markup of the code) may be found at (Corneli 2004).

## Articles

The foundation of the scholium system is a catalog of articles (list or hash table). Adding or updating the reference for an article to the catalog is the fundamental operation in the system. This operation stores

- the article’s name;
- its text or a pointer to its text;
- a record of what it is about—nothing if it is degenerate, otherwise, some article(s) or passage(s) of articles in the collection;
- the designation of a type (link, followup, forum, action to take if a certain event happens at the parent, etc.);
- and bookkeeping information to keep track of ownership information and editing history.

In Figure 1, we illustrate the conceptual model of the scholium system with a hypothetical content instance. Shown are the key entities of *people*, *articles*, and *references*.

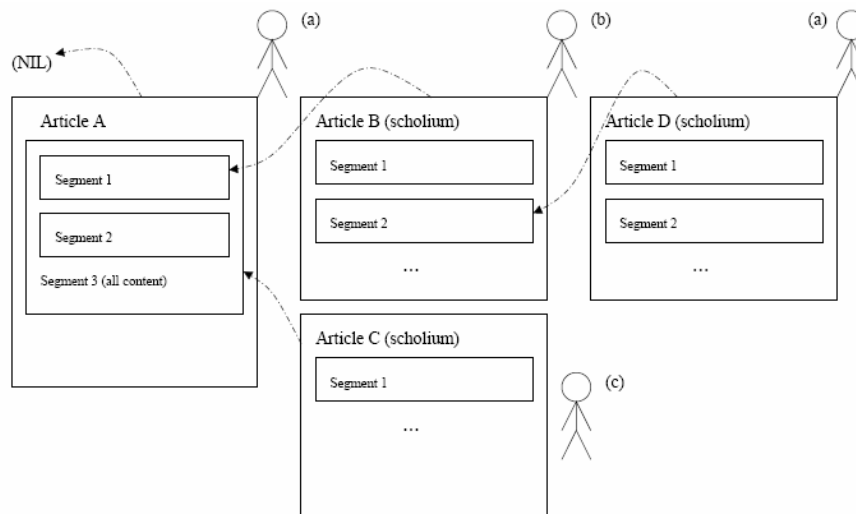


Figure 1: The key elements of the scholium system, for a hypothetical small document/library. There are three people (a, b, and c), and four articles (A, B, C, and D). All articles are scholia (i.e. refer to another article), except A. The contents of each article have been broken into “segments” for clarity. Links are shown as dashed arrows from articles to article segments.

## Environment

The current prototype system runs under GNU Emacs. We try to conform to the Emacs Way to the greatest extent possible. Typically, scholia are written about files and buffers, and are displayed alongside them. The default display uses color to associate displayed scholia with regions that are being commented on, but delimiters can be used for this purpose as well (for the benefit of those working without fontlock). Color can also be used to distinguish between different types of scholia (*e.g.* comments by different authors).

Our ability to render articles is limited to things Emacs can display: plain text, code, and pictures all work, but special proprietary formats are not supported in this implementation, nor do we have support for making scholia about specific pieces of a rendered diagram, for example.

The user can navigate the display in various ways, *e.g.*, by scrolling between marked-up regions of the main article, or finding the region(s) associated with a given scholium, or the scholia associated with a given region. Additional browsing methods are described below.

## Displaying Scholia

When an article is displayed, the system finds attached scholia and displays them too. Selective displays are possible (*e.g.* just show all the *links* attached to a given article). Finding scholia in the catalog requires search; sometimes we can limit the search to make it faster (this will be described further below). In Figure 2, we show a number of ways in which the underlying content from Figure 1 can be displayed.

## Adding to the library

We support various convenient ways to add articles. Different kinds of articles need different treatment. For example, the content of a buffer is lost if the buffer is killed, so a backup of the buffer's contents should be made immediately. In addition to specifying the article's text and stating what (if anything) it is about, one can specify the article's type. We provide built-in functions for creating scholia about the current buffer, and for creating scholia about mixed collections of other articles and passages.

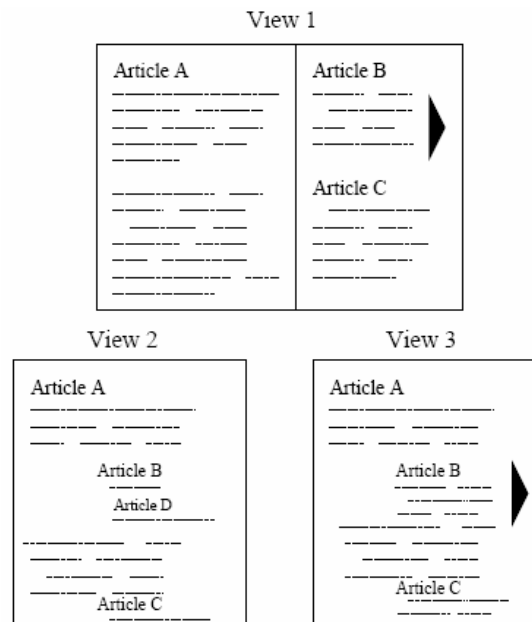


Figure 2: Different possible views of the articles from Figure 1. Three views are shown. View 1 is a two-pane style, with attached scholia on the right and the main article on the left. View 2 includes only title and linkage information for attached scholia, unfolding recursively to a user-specified depth. View 3 is a single-frame view, with one level of attached scholia inclusion (deeper levels are indicated with an arrow).

## Browsing

We provide several different browsing mechanisms. Simple local navigation features (scrolling etc.) were mentioned above. “Following” a scholium to make it into the new current article is supported. Browsing by catalog (*i.e.*, by list) is supported; we use a generic menu mechanism that make it easy to select different collections of articles matching different criteria, and perform various actions on them. Finally, web-browser-like history-based navigation is provided.

## Editing, deleting, saving, restoring

Articles are versioned and entire document versions or, alternatively, a collection of patches sufficient to move between versions, is maintained for each article. When articles change, it is typically necessary to propagate data to attached scholia in order to maintain coherence of references. Propagated changes need not be accepted, in which case the attachment relationship is fixed at the previous version.

Similarly, permission is typically requested whenever a portion of an article that has been *transcluded* changes. (“Transclusion” is Nelson’s term for “included by reference”; both transclusion and inclusion of pieces of other articles are supported.)

Changes that take place to articles outside of the system (*e.g.* moving a file from the shell) are of course nearly impossible to deal with. However, scholia can be saved in explicit, invariant forms. (CVS is natively supported).<sup>14</sup> Saved articles can be read back in selectively.

Currently, name conflicts pose a bit of a problem; if every scholium has a unique name, then the problem goes away. In general, we can approximate a solution by asking the user to uniquify names when conflicts are encountered.

### **Namespaces**

Another solution to the problem of name conflicts is provided by namespaces. For example, if article *A* has type “sublibrary” and we read in a scholium of type “member of sublibrary *A*,” then we don’t have to worry about name conflicts with articles not in sublibrary *A*. Namespaces can be used to make search convenient. For example, we can store all assertions of type *Z* in a given namespace; then if we are building a display that only relates to objects of type *Z*. In other words, we need only search one namespace instead of the whole library.

### **DISTRIBUTED AUTHORSHIP**

Distributed authorship is actually very easy using this system. Each contributing author posts her or his contributions to the document at their own chosen location. Then the other authors download the articles stored in all of the locations they want to use. (All authors don’t necessarily need to have all of the articles.) In the case of document revision, changed versions are simply posted, and other authors learn of the changes whenever they sync.

### **Derivative versions**

In general, the author may be the only person with permission to create modified versions, but we can also distribute this permission over a wider group of people. As we’ve already mentioned, derivative versions can be put together using two principles, inclusion and transclusion. Inclusion tends to reduce search but increase storage.

Here we illustrate *tracking* of derivative versions with a scenario. Suppose that a series of definitions was quoted in a textbook-style entry. The author of the textbook might receive a question from a reader and then adjust one of these definitions to include more expository text. If the original author was tracking derivative versions, the new expository text could be added as a scholium attached directly to the original, or the original could be modified.

These sorts of exchanges should still be possible even if one of the agents is working outside of the scholia-based system. In order to make *bi-directional updating* work in this case, both content-sharing parties need to be able to read a stream of diffs generated by edits taking place in another system and decide how to incorporate the modifications.

These issues should be of interest to anyone maintaining a collaborative digital library; information-sharing between such entities typically needs to support content that can change on both ends of the pipeline.

## **FUTURE WORK**

Some ideas for explorations to undertake with the scholium system and related concepts follow. Note that some of these could be considered feasibility or proof-of-utility experiments:

- Import a wiki and build a wiki-like interface to the scholium system;
- Use the scholium system to write a synchronous or asynchronous multiplayer game;
- Use the scholium system to maintain a text-based forum or set of fora, as found on PlanetMath or Slashdot;
- Implement a Slashdot-like scoring system for quality control;
- Use the scholium system to manage an evolving codebase (*i.e.*, take advantage of the functionality which subsumes a system like CVS);
- Use the scholium system to manage TODO lists;
- Port the GNU Collaborative Dictionary of English (GCIDE) to Emacs and give instant access to definitions as scholia;
- Implement code to make an index or do autolinking as you type;
- Port WordNet15 to the scholium system and use the system to collaboratively improve the database;

- Implement semi-automated content sharing between two collaborative digital libraries, for instance, PlanetMath, and Wikipedia.

This list should give some idea of the range of capabilities the scholium system, in theory, encompasses. Long term investigations under conditions of fairly wide uptake would shed light on broader social implications of the system. We imagine that social institutions—from peer review to popular science, and from online shopping to participatory government—would tend to be transformed by widespread use of these systems. In fact, some of this social transformation is already apparent—in specialized implementations of the scholia concept (*e.g.* weblogs, wikis, online forums, PlanetMath). Currently we can only speculate as to how the general system outlined here would interact with these trends.

One thing we can say at this point is that the model seems to provide a useful basis from which to explore the design and implications of social contracts in online communities. We hope future work will take up this issue.

## CONCLUSION

In this article, we have described a scholia-based document model and outlined an implementation of a system that supports this model. We have discussed some ramifications of scholia-based documents and libraries. We have also shown how the model can be used to facilitate powerful collaboration dynamics in a wide array of scenarios and social settings—by fostering and managing alternative perspectives, encouraging responsible maintainership, and enabling readers to routinely make useful contributions.

The scholia-based document model corresponds to a culture with empowering conceptualizations of freedom and ownership. Accordingly, this paper has been a description of a model as well as something of a social manifesto. We hope to see the ideas presented here take off, as we and others work to push the limits of the model.

## Acknowledgments

Thanks to Ray Puzio for helpful comments and encouragement as the scholium system developed. Thanks also to Thien-Thi Nguyen and Sacha Chua for their comments on free software and hypertext, including comments on an earlier prototype of the system by the first author.



## ENDNOTES

1. <http://www.ibiblio.org/webster/>.
2. See “The Talmud and the Internet” (Rosen, 2000). You can also find the Talmud on the internet, at <http://www.sacred-texts.com/jud/talmud.htm>.
3. See <http://planetx.cc.vt.edu/AsteroidMeta/HDM>.
4. <http://planetmath.org/>.
5. (info "elisp(Text Properties)")
6. <http://www.w3.org/DesignIssues/Semantic.html>.
7. <http://www.cyc.com/cyc/technology/whatisyc>.
8. <http://www.cs.utexas.edu/users/mfkb/RKF/km.html>.
9. Marshal McLuhan’s theory ties in very nicely with MUD-like systems. Those with a user-hackable infrastructure embody a distinctly free “message.” For cultural materialists, the upshot is hackable superstructure. The fact that the MUD is virtual to begin with adds an interesting twist in this analysis.
10. It is interesting to compare the experience of these immersive worlds to the experience of internet or Talmudic scholars, mentioned previously. See (Rheingold 1993).
11. Xanadu development seems to be going slowly at present; see <http://udanax.com/> and the mailing list at <http://xanadu.com.au/mail/>.
12. See [http://www.wired.com/wired/archive/3.06/xanadu\\_pr.html](http://www.wired.com/wired/archive/3.06/xanadu_pr.html), but note that this work is not endorsed by Nelson, who writes at <http://ted.hyperland.com/whatsay/> “I believe the piece is a study in cunning and deliberate dishonesty, the most dastardly piece of dirty journalism I have ever seen.”
13. For example, see (Maier and Delcambre 1999), [http://nsdl.org/community/projects.php?this\\_sort=start\\_date&keyword=&project\\_id=0435496](http://nsdl.org/community/projects.php?this_sort=start_date&keyword=&project_id=0435496), or <http://datalab.cs.pdx.edu/sparce/>.
14. Code Versioning System, an extremely popular free software program to collaborative manage software codebases. See <http://www.gnu.org/software/cvs/>.
15. WordNet is a “lexical database for the English language”—essentially a semantic network of words and relationships, which can be used as a dictionary or thesaurus. See <http://wordnet.princeton.edu/>.

## REFERENCES CITED

- Joe A. Cornelli. A scholium-based document model, 2004. URL [http://planetx.cc.vt.edu/AsteroidMeta/A\\_scholium-based\\_document\\_model](http://planetx.cc.vt.edu/AsteroidMeta/A_scholium-based_document_model).

- Mark Federman. What is the meaning of the medium is the message?, 2004. URL [http://www.mcluhan.utoronto.ca/article\\_mediumisthemessage.htm](http://www.mcluhan.utoronto.ca/article_mediumisthemessage.htm).
- Bjorn Karger. The gopher:// manifesto, 2000. URL <gopher://gopher.quux.org/0/Software/Gopher/whygopher/gopher-manifesto.txt>.
- David Maier and Lois M. L. Delcambre. "Superimposed information for the internet." In Sophie Cluet and Tova Milo, editors, *ACM SIGMOD Workshop on The Web and Databases*, 1999. URL <http://www-rocq.inria.fr/~cluet/WEBDB/maier.pdf>.
- Marshall McLuhan. *Understanding Media: The Extensions of Man*. McGraw Hill, 1964.
- Marvin Minsky. *The Society of Mind*. Simon and Schuster, 1985.
- Theodore Holm Nelson. *Literary Machines*. Mindful Press, 1990.
- Joseph D. Novak. *Learning, Creating, and Using Knowledge: Concept Maps as Facilitative Tools in Schools and Corporations*. Lawrence Erlbaum, Mahwah, NJ, 1998.
- M.R. Quillian. "Semantic memory." In M. Minsky, editor, *Semantic Information Processing*, pages 227–270. MIT Press, Cambridge, MA, 1968.
- Howard Rheingold. *The Virtual Community*, chapter 5. Perseus Books, 1993. URL <http://www.rheingold.com/vc/book/>.
- Jonathan Rosen. *The Talmud and the Internet*. Farrar Straus Giroux, 2000.
- Sam Williams. *Free as in Freedom: Richard Stallman's Crusade for Free Software*, chapter 13. O'Reilly, 2002. URL <http://www.faifzilla.org/>.

# Adapting CBPP Platforms for Instructional Use

Robert Milson (Dalhousie University) and  
Aaron Krowne (Emory University)

**Abstract:** Commons based peer-production (CBPP) is the decentralized, net-based approach to the creation and dissemination of information resources. Underlying every CBPP system is a virtual community brought together by an internet tool (such as a website) and structured by a specific collaboration protocol. In this paper we will argue that the value of such platforms can be leveraged by adapting them for pedagogical purposes. We report on one such recent adaptation. The Noösphere system is a web-based collaboration environment that underlies the popular Planetmath website, a collaboratively written encyclopedia of mathematics licensed under the GNU Free Documentation License (FDL). We argue that a “proof of concept” experience also strongly suggests that every successful CBPP platform possesses latent pedagogical value.

## INTRODUCTION


### Background and motivation

The capacity of communications networks to create value is well recognized (Metcalf 1995). There is a theoretical argument that internet value creation is an even more dramatic process, because it is dominated by exponential rather than polynomial scaling effects (Reed 1999). To put it another way, the internet engenders powerful emergent phenomena, because every potential group with a shared interest can interact, collaborate, and create intellectual value through internet (and especially WWW) software applications.

Thus, with the advent of powerful search and indexing technologies, the World Wide Web is evolving into a ubiquitous reference resource (Berners-Lee *et al.* 2001). The network transforms the disconnected efforts of millions of web page

---

M. Halbert (Ed.): *Free Culture and the Digital Library Symposium Proceedings*.  
Atlanta, Georgia: MetaScholar Initiative at Emory University, 2005. pp. 255-272.

 This work is covered by the Creative Commons Attribution-NonCommercial-NoDerivs 2.5 License.

authors into something of practical value. Another noteworthy project is Wikipedia (Wales and Sanger 2001), a knowledge-oriented virtual community that successfully employs the wiki collaboration protocol (Leuf and Cunningham 2001) to unite the efforts of thousands of volunteers around the scholarly goal of a public domain encyclopedia (Kantor 2004). In both of the above examples, the underlying process lacks explicit organization and is non-hierarchical. In both cases the value is governed by an emergent phenomenon: the value of the whole is significantly greater than the sum of the individual parts. A recent economics-based theory attempts to explain such emergent value phenomena as instances of *commons-based peer production* (CBPP), an idealized mode of production that is complementary to firms and markets, and one that manifests naturally on the internet (Benkler 2002). However, economic theory is insufficient to fully understand and exploit the complex, emergent phenomena that underly internet value creation (Iannacci and Mitleton-Kelly 2005). The study of the internet is inherently cross-disciplinary; no one discipline, or even a blend of two will suffice.

In the present article we report on and discuss a recent adaptation of Noösphere (Krowne 2003a), a web platform for mathematics collaboration, for the purpose of teaching a graduate course in mathematics. A convenient categorizing label for our project is computer supported collaborative learning (CSCL), a field that brings together perspectives from cognitive science, computer and information science, education, and philosophy (Stahl 2006). Our thesis is inherently cross-disciplinary. We argue that CBPP, the phenomenon of internet value creation, crosses over naturally into the world of CSCL. We argue that the infrastructure of collaborative, knowledge-related projects, like Wikipedia and Noösphere, can be leveraged to yield concrete educational assets.

This value stems in large part from the inherent unity and collaborative nature of the scholarly enterprise. A context that fosters the formation of communities which acquire, organize, generate, synthesize, and transmit knowledge will also be a context where learning and pedagogy are of central importance. These qualities naturally lead us to the concept of a *digital library*. Traditionally, libraries have been the cornerstone of scholarship, providing a space for both research and learning, and other, more intangible benefits. It would therefore be surprising if emergent collaboration phenomena and educational scenarios did not play a

role in the evolution of the digital library (Robertson and Reese, 1999).

### **Re-conceptualizing the digital library**

The concept of a digital library is a natural outgrowth of the development of modern, network-oriented information technology. Information, once digitally encoded, can be stored electronically and distributed over the internet. Physical and geographical barriers disappear. There are no limits to the size of the library. It's contents are potentially available to everyone, everywhere, all the time.

The word *library* carries with it connotations of a nearly static archive, one where the primary information-related activity are storage, classification and retrieval. The shift of information content from the physical to the digital realm undermines this traditional conceptualization (Levy and Marshall 1995). Various recent internet-focused developments—powerful and ubiquitous search engines, virtual communities and the free culture movement, to name just a few—challenge us to move beyond the simple notion of an “electronic traditional library,” and to embrace benefits beyond the elimination of space and scarcity concerns.

Older information technologies, such as paper, foster a dichotomy between *information* and *knowledge*. The latter is the more dynamic concept; knowledge implies research, dissemination, debate, synthesis, activation, history and evolution. As well, knowledge cannot be conceived as something separate from people; knowledge implies a community of scholars, teachers, learners, and practitioners (Ehrlich and Cash 1994).

Therefore, the digital library concept needs to evolve to more fully realize the potential of the underlying network technology and software technology. New library tools and modalities that address collaboration, superimposed information, knowledge creation, and education will have to be developed (Delcambre *et al.* 2001; Frumkin 2005; Krowne 2003b; McRobbie 2003).

### **CBPP**

In this regard, *commons based peer production* (CBPP) shapes up to become a key phenomenon in the digitally mediated transition from *information* to *knowledge*. Internet-based CBPP has its origins in the open-source software movement, a collaborative, extra-commercial process of software creation.<sup>1</sup> The existence of numerous successful internet projects, Wikipedia and Project

Gutenberg/Distributed Proofreaders (Lebert 2004), to cite just two examples, indicate that the phenomenon of collaborative internet value creation has pertinence well beyond generating software programs.

With peer production on the Internet, distributed ensembles of people share open production of complex products and services—generally for no financial compensation. While the idea of non-market, non-corporate production is not new (science has traditionally worked this way), large-scale, decentralized, sustained, open production by diverse groups of peers is a new phenomenon: a development that has been enabled and encouraged by the confluence of computers, networking and the information economy. This form of non-market, internet-based peer production has been applied to create a wide variety of significant knowledge assets (Galiel 2004).

The impact of a knowledge-centric community like Wikipedia on the digital library landscape cannot be ignored. Neither should the enormous productive leverage of a project like Distributed Proofreaders. Therefore, it makes good sense (for both practical and idealistic reasons) to expand the “digital library” concept to incorporate an internet-based CBPP aspect. PlanetMath (Krowne and Egge 2001) is another CBPP project, of special connection to our study.

Planetmath is a collaboratively written encyclopedia of mathematics licensed under the GNU Free Documentation License (FDL), and implemented using the Noösphere system. The PlanetMath project is an instance of CBPP; the aim is to create a community-oriented, web-based repository for mathematical knowledge. The project attracts a diverse and international body of participants. These people are students and members of the wider public with an interest in mathematics, graduate students pursuing advanced mathematics degrees, professional mathematicians who make their living by practicing or teaching mathematics classes and by conducting mathematics research. Planetmath and Noösphere also have an extended role as a testbed for research and development in semantic extraction, digital information exchange, and collaborative authority models (Krowne and Bazaz 2004).

### **Academia, instruction, and engagement**

Academic communities are concerned with knowledge in all its manifestations; both the information and community-related

aspects are important. Certainly, instruction and the teacher-learner relationship are central academic concerns.

Instruction can be conceptualized as a structured interaction between senior and junior members of a knowledge community. The instructor is more than just a particular medium for the storage and transmission of information. Rather, for the student, the lecture hall is a portal to the community of knowledge (Clancey 1995). Let us use the term *engagement* to describe the process of active student participation and scholarly development (Stahl 2005).

In addition to the immediate goals of any particular course of academic instruction, there is, in the teacher-student relationship, an implicit invitation to “do as we do;” to join the community, and to become involved in knowledge-related activities. Pedagogical structures: exercises, discussions, individual and group projects, examinations and other assessment modalities, are the devices of guided scholarship. Engagement, rather than skill-set and information “download” is the deeper goal of academic instruction. The ultimate measure of success is the metamorphosis of the *student*, an individual at the outset capable and interested only in passive, assisted knowledge activities, into the scholar, an individual engaged in independent knowledge activities.

It is worth briefly examining the critical elements of scholarship. Of paramount importance is that for scholar, no “oracle” exists to provide the answer to a research question. Peers can provide critique but not guaranteed answers. The scholar also lacks a roadmap towards a solution, and must prioritize his/her efforts, evaluate the intellectual contributions of others, and act upon their own judgments. This is the universal situation of the scholar, and it is utterly different from the environment of the formal student. While attempts are made to deliberately teach students many of the tools upon which scholars rely, the aims and trajectory of classroom activities are by definition preset. Thus, the characteristics of the true scholarly environment induce a sharp division of students who have meaningfully become scholars from those who have merely learned to regurgitate information with relative success.

CBPP projects like Wikipedia and Noösphere possess a remarkable capacity for fostering engagement in scholarly activity. We suggest that it is reasonable to tap such free-culture phenomena for the purposes of academic instruction. Indeed, nothing could be more natural, because of the inherent compatibility between academic

and free-culture goals and values.<sup>2</sup> Let us make a sketch of how such an evolution can take place.

A re-conceptualized, more dynamic and community-oriented digital library is a natural context for both public domain knowledge activity and for pedagogical efforts that involve students in online knowledge activities. Such activities should include not just information retrieval, but collaborative knowledge creation and organization (Brown 1999). The physical community of the classroom can be extended to the network. The same community and collaboration tools and technologies that enable CBPP projects can be used to create a virtual space in which the participating students can carry out knowledge-related activities, albeit in an assisted and structured fashion.

We hypothesize that such an approach can lead to a heightened level engagement, because of the subtle but important shift of emphasis from “I will teach, you will learn” to “let us collaborate on a knowledge project”. The change of attitude is natural and desirable from an academic point of view, but is difficult to implement using traditional classroom methods and technologies.

Our hypothesis is that adoption of CBPP technologies into an instructional setting will facilitate just such a shift of emphasis. The student goal-set and motivations will be enriched by incorporating a network-based, collaborative aspect into the classroom experience. At one level, the instruction process can proceed in the traditional manner: the teacher guides the students through a fixed syllabus, assigns tasks, and performs evaluation. However, since the setting is now a “research library” as well as the classroom, since the medium of interaction includes a virtual collaboration environment, and since the goal-set includes the incorporation of individual efforts into a digitally encoded body of knowledge, the end result will manifest as a collaboration between all involved. Such a process should lead to heightened levels of student engagement.

## **A TRIAL OF NOÖSPHERE AS A PLATFORM FOR COLLABORATIVE INSTRUCTION**

### **Test scenario and goals**

In the Winter of 2003, the Noösphere system was used to host *Math 5190: Ordinary Differential Equations*, a graduate mathematics course at Dalhousie University, in Halifax, Canada. One of the current authors served as course instructor. A “tabula



rasa” Noösphere system was set up on a dedicated server. The primary course goal was the collaborative creation of a set of course notes, including a number of worked-out exercises to illustrate the key concepts. Assessment criteria included the quantity and quality of the online participation, as well as a more conventional final project.

The course attracted three graduate students and an auditor, who in the coming semester created and organized an online body of knowledge on the topic of differential equations. The end result was a 70-page document containing definitions, theorems, proofs, and examples. When taken together, these constitute a mini-treatise on certain aspects of the theory of ordinary differential equations. The trial addressed the following research goals:

1. Our main hypothesis was that CBPP platforms are suitable for advanced mathematics instruction, and that a course structured around collaborative principles and online tools can serve and advance conventional academic goals.
2. We evaluated the feasibility of deploying Noösphere as a CSCL environment. Experiences with CoWeb (Guzdial *et al.* 2001), show that CSCL-type mathematics courses present special challenges related to specialized notation and division of labor issues. Noösphere’s LaTeX based design incorporates the full range of advanced mathematical notation. As well, Noösphere possesses a unique authority model and groupware capabilities. The trial examined the capacity of these designs to address the above challenges. In particular, we wanted to compare the patterns of student activity in a collaborative, online environment with those in a traditional mathematics courses, and to consider the impact on student engagement. Our secondary hypothesis is that student engagement benefits from the introduction of CBPP elements.
3. We also considered the impact of a collaborative, online course environment on the students’ scholarly development.

### **Methodology**

Math 5190 is a one-semester course at Dalhousie University on the theory and methods of ordinary differential equations. Such

courses, typically aimed at beginning graduate students and advanced undergraduates, are offered, with certain variations, by most mathematics departments in North American universities.

In the Winter of 2003 this course served as a proof-of-concept study of the Noösphere system in an educational setting. The course included a number of conventional instructional components: three hours/week of lectures, a reading list, regular meetings with of the instructor with individual students, a final project, and student presentations. The core component, however, was a dedicated website set up as a “tabula rasa” Noösphere environment.

The basic unit of content in Noösphere is the *entry*, which any registered user can create. The entries comprise the main section of the system, which is called the “encyclopedia”. This reflects the general orientation and pedagogical style of the system.

Noösphere entries consist of title, content, and various metadata. The entries are interlinked, which means that the text of each entry contains hyperlinks pointing to other entries where appropriate. The general intent of this is to provide definitions for each concept utilized, in an easily navigable fashion. Entries are written in LaTeX (Lamport 1986), which serves as the basis for Noösphere’s mathematics support in addition to allowing for the expression of general document formatting. Displayed in rendered form, the mathematical portions of each entry “look right” with a standard browser (with no plug-ins), a considerable improvement over most other attempts to publish mathematics to the web to date. This mathematics support makes Noösphere a good candidate for use in all of the mathematical sciences.

A key feature of Noösphere is the *corrections* system. If any registered user determines there is a problem with an entry, he or she can voice concern by filing a correction to that entry. Until addressed, this correction is displayed when the entry is shown, ensuring that the critique is “out in the open.”

Finally, each entry in Noösphere has an owner, who is initially the person who created the entry. An owner has the option of orphaning an entry, or transferring ownership to another user. Orphaned entries are flagged by the system and may be adopted by any interested user.

Noösphere has a number of other services that provide direct community support:

1. The requests service, which functions as a global “to-do” list of content addition for the Noösphere site. Users can fulfill particular requests, rendering them inactive, by creating an appropriate entry.
2. The discussion service provides threaded, asynchronous messaging. A discussion can be attached to most of the core objects of Noösphere. This includes encyclopedia entries, corrections, and requests.
3. Noösphere’s notification system keeps members of the community aware of activity relevant to them through e-mail and a Noösphere system “inbox”. Corrections to an entry result in a notice to the entry’s owner. A resolved correction results in a notice to the filer, indicating what action was taken and why. Similarly, replies to a message posted result in a notice that makes the initial poster aware of the reply. An important part of the notification system is the ability to create configurable watches. Watches placed on any object by any user result in (e-mail or web) notices about events to that object being sent to the user.

At the outset, the students were informed that the main course objective was the collaborative creation of a set of lecture notes using the online environment. The instructor’s role was to facilitate and to structure this effort. As such, the instructor mirrored lecture topics and contents with Noösphere *request* objects that enumerated the key definitions, theorems, proofs, and techniques covered in the lectures. The students were responsible for filling these requests by creating the requisite *entries* and subsequently evolving and improving them based on *corrections* received from the instructor and fellow classmates. The students had to cooperate to decide how to divide the requests and to share the corresponding workload.

It is well recognized that mathematics instruction is greatly facilitated by supplementary problems and exercises. In place of the conventional system of regular assignments with specific deadlines, course exercises were presented to the students as illustrative examples to be included in the collaborative notes. The instructor, on a regular basis, created and *orphaned* exercise-type entries. The students were then responsible for *adopting* the entries and furnishing solutions. Again, students were given the opportunity to evolve and improve their solutions through

interactions with instructor and classmates. *As such, an incorrect solution did not necessarily result in a poorer evaluation, but rather served as an additional learning opportunity* in the context of Noösphere's system of corrections. Students had the opportunity to continuously improve their entries up to the course termination deadline. The collaborative, online aspect of student progress was assessed according to the number of owned entries, and according to the extent the entries were developed. At the termination of the course, a score of 1, 2, or 3 was assigned to each student entry according to the following criteria:

- Degree of participation was measured by the number of filled requests, and adopted exercises. An adopted entry with even a minimal amount of content was assigned a score of 1.
- A reasonably well developed entry with unresolved corrections was assigned a score of 2.
- A correct, well written entry with no outstanding corrections was assigned a score of 3.

The instructor issued corrections in response to student errors, and to suggest improvements to the mathematical content and presentation format.

Course assessment did not include an examination component. Rather, an assessment of scholarly development was based on a final project, which was implemented conventionally, and involved both an oral presentation and a written report. With input from the instructor, students selected a relevant topic,<sup>3</sup> delivered a classroom presentation, and submitted a written report. The project component played a particularly important role in the trial, providing a measure of student progress independent of the online activity.

### **Results data**

By the end of the course, the three registered participants, all first year MSc students, had created a total of 122 entries. The entry totals and the corresponding scores (see above) are displayed in Table 1. A score of 0 indicates an entry with non-existent or negligible content. At the conclusion of the course there were a total 12 unfilled requests and unadopted exercise problems. A total of 78 corrections were issued.<sup>4</sup>

Subsequently, the website contents were converted into document form and redistributed to the students. The resulting document spans 74 typeset pages.

The Noösphere collaboration protocol proved to be very suitable for student-instructor interactions. The entry ownership system and email updates allowed the instructor to easily follow student progress, and to issue timely feedback in the form of corrections. With minor adjustments, the Noösphere scoring system proved valuable as a highly visible indicator of individual participation levels.

Student	Entry score				Total
	0	1	2	3	
1	0	1	10	26	37
2	1	2	10	27	39
3	3	6	10	16	32

Table 1: Student entries and assessment scores.

Student-instructor interactions stabilized around the following cyclical pattern. The instructor delivered lectures and suggested deadlines for the fulfillment of requests and the adoption of exercise entries. This was followed by posted corrections and occasional email “nags” and feedback.

As is often the case in conventional courses, the students functioned as largely passive knowledge agents. There was no evidence of direct online collaboration among the students. Students did not give each other corrections, nor did they use the online forums to discuss mathematical content. Rather, students reported collaborating in more conventional ways. They held study group meetings to discuss course material, and to decide on the division of labor for their online tasks.

Student behavior and outlook in the trial was typical for courses at the beginning graduate level. Students at this level still require explicit goal structure and assessment criteria, and are often passive in their approach to the material. Students in the trial displayed typical procrastination behaviors, and regarded their participation as “necessary duty” to be balanced against time requirements from other courses and from outside jobs. As such, their online efforts tended to occur in bursts of concentrated activity. An example of this behavior pattern is visible in Figure 1, which shows the temporal distribution of student responses to corrections.

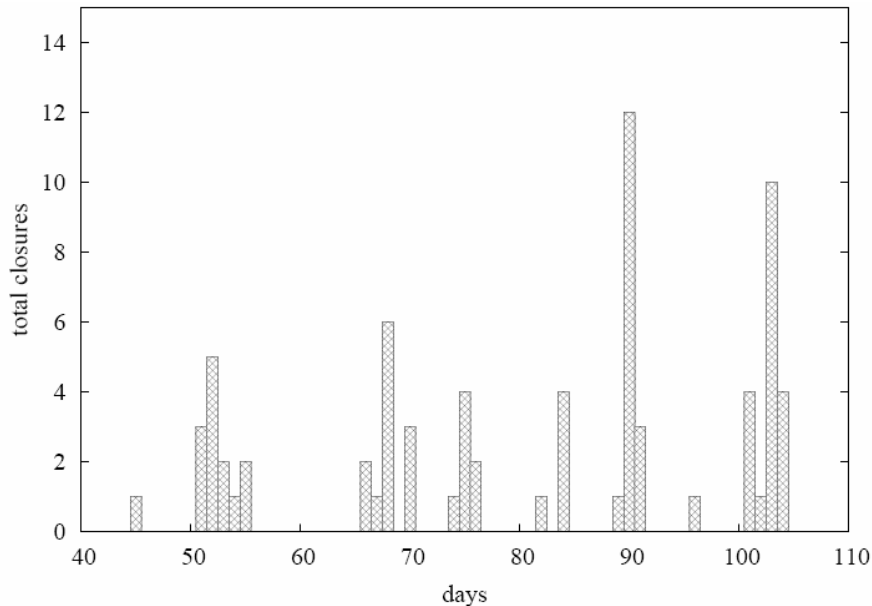


Figure 1: Chart of correction closures by students (with each bar representing a day), revealing the “bunching” effect.

The conventional educational objectives of the course were fulfilled. The content of the final projects and the website entries, especially the exercises, provided clear and substantial evidence of progress toward mastery of the subject matter, and progress in scholarly development. Relative to these metrics (exercise solutions and final projects), progress of the students in the trial was directly comparable to the progress of students in the same course taught by the same instructor conventionally in other years.

### Findings

Given the limited enrollments and the advanced nature of the material characteristic of graduate courses, and keeping in mind the natural variation of student backgrounds and abilities, it is not feasible to render a judgment on the relative merit of conventional pedagogy versus collaborative, online learning. However, our observations allow us to make the following points.

1. Our experience with Math 5190 and Noösphere provides strong support for the hypothesis that conventional educational objectives can be met by a course based on online learning and CBPP principles. Importantly, we found no evidence that the inclusion of a CBPP component diminished or disrupted traditional classroom learning. Our outcomes should be reproducible by groups of advanced students at other institutions, and with other

courses in the mathematics curriculum. To make sense of this claim, however, one must incorporate assessment components that can provide an objective measure of student progress.

2. The students in the trial readily accepted the mechanics of Noösphere and expressed appreciation at being able to do their work in an online setting. Nowadays knowledge of LaTeX is a near-universal prerequisite for the scholarly development of mathematics students. The LaTeX component of Noösphere provided our students with a useful opportunity to develop their typesetting skills.

Based on the instructor's observations and communication with the students, Noösphere's protocol of entry adoption and ownership allowed the students to exercise control over their participation, and thereby facilitated engagement. The fulfillment of requests and the adoption of exercises manifested as an act of commitment on the part of a student. Thus, the authority model allowed the students to pursue a division of labor, but in a transparent and principled fashion that is usually lacking in conventional courses. A potential weakness of this approach is the possibility that an overly selective focus on the part of some students may lead to a spotty coverage of essential topics. The instructor has an important role to play here, and must encourage students to contribute to a variety of course topics. Such difficulties did not visibly manifest in the trial under discussion. However, without a comprehensive final examination it is difficult to discount the possibility that some of the students received inadequate exposure to some of the topics.

3. The collaboratively compiled course notes are a valuable asset that is not readily available in the context of conventional instruction. From the point of view of the students, the document is far more than a transcription of the instructor's lectures. In a very real sense, the students are the authors of the document. As such, the notes concretely encapsulate their learning experience.

There are a number of benefits to producing such a document. The notes can serve as a source of reference

for future work in the subject. Perhaps, more importantly, the very existence of the notes embodies a latent, but powerful message about the students' capacity for scholarship, and about the nature of the academic enterprise. In an important sense, the creation of the notes transforms the asymmetrical relationship between instructor and the students into something more closely resembling scholarly collaboration. There is also the intriguing possibility that collaboratively produced course notes can serve as contributions to public domain knowledge repositories.<sup>5</sup> The primary responsibility of the course instructor centers around the student learning experience. As such, it would not be appropriate to make full scholarly use of the course notes without addressing issues of consent and attribution. Still, it is important to provide students with opportunities for independent scholarly activity. If nothing else, the format of the trial made the students aware of ongoing CBPP efforts, and served as an invitation to contribute to them.

## DISCUSSION

The Noösphere/Math5190 trial constitutes a limited, proof-of-concept experiment regarding the application of CBPP tools in an educational, academic setting. Though our experiment was a success, the small scale of the trial limits the inferences we can draw in support for our hypothesis regarding CBPP and education. It will be necessary to subject the hypothesis to further testing: one needs to organize more CBPP-based courses, involve more students and instructors, employ control and experimental groups, and to consider diverse academic subject material.

One also has to come to grips with the limitations revealed by our experience. Collaborative learning methods are not a panacea for improving student engagement (Guzdial et al. 2002). Indeed, it would be useful to undertake a systematic examination of the effects of CBPP on academic engagement. Methodologically, the undergraduate curriculum, with its larger enrollments, may be a more appropriate setting for such studies.

Wiki-based courses in the humanities and the social sciences are the subject of ongoing research and discussion (Boyd and Lohnes 2005). Wiki software is widely available, notational demands are lower, and the wiki interface is easier to learn than the LaTeX-based Noösphere. The ostensible aim of such a course should be a



well-developed body of “wikified” content that encapsulates a subject of interest, and that provides a concrete record of individual students’ participation. An initial study on this topic (Scharff 2002) supports the conclusions of our own trial. It would also be interesting to study to the effect of such an experience on scholarly evolution. To what extent does student exposure to wikis as an instructional medium encourage contributions to sites like Wikipedia, or the pursuit of more conventional scholarly publications?

## CONCLUSION

The joining together of the themes of collaborative education, the internet, and digital libraries is not a new idea (Roschelle and Pea 1999). Rather, the relatively recent emergence of successful CBPP knowledge projects should be viewed as a timely and complementary development (Tomek 2003). Much of the infrastructure, interface, and design issues are the same for both contexts. There is strong common focus on extraction of semantics, collaboration interfaces, and educational applications. We believe the potential for mutual benefit and a convergence of interests is evident.

In the context of a symposium on Digital Libraries and Free Culture, it is also appropriate to note the relevance of our hypotheses to the continuing debate about intellectual property and the public domain. Pragmatism and utility are strong arguments for commons-based knowledge activity. The impact of the open source and the free software movements on development of information technology is, at this point, beyond question. Likewise, projects like Wikipedia, PlanetMath, and Distributed Proofreaders are beginning to make a significant contribution to the intellectual commons.

As is the case with emergent internet value phenomena, the potential value of such projects is unconstrained and will manifest in unforeseen ways. But, this is just one instantiation of the general argument in support of public domain knowledge and culture (Lessig 2004). Synergy and flexibility is the point here, and a *libre* free project like PlanetMath is good example. This project began as a mathematics encyclopedia, then evolved into a groupware platform and test-bed for digital library research (Noösphere), and is now being used as an educational delivery system.

Academic involvement in CBPP projects allows researchers, librarians, and educators to exploit the kind of internet value that

IT companies enjoy when they employ open-source software. Conversely, free-culture projects benefit from academic attention and investment. Successful adaptation of CBPP technologies for academic instruction is a powerful argument in support of free culture. However, much work remains to be done in the crossdisciplinary exploration of CBPP, CSCL, and digital libraries.

## ENDNOTES

1. This is not to imply that open source software is without commercial value. Rather, the process of creation is governed by something other than a simple exchange of money for software end products.
2. The open access movement illustrates this nicely (Suber 2004).
3. The three registered students chose the following topics: convergence of iterative integral solutions, predator-prey models, differential equation modeling of guerrilla vs. conventional warfare.
4. All but one of these corrections originated with the instructor.
5. The students in the trial were encouraged to convert their course contributions into PlanetMath entries— though none of them chose to pursue such activity.

## REFERENCES CITED

- Y. Benkler. "Coase's penguin, or, linux and the nature of the firm." *Yale Law Journal*, 112:369–446, 2002.
- T. Berners-Lee, J. Hendler, and O. Lassila. "The semantic web." *Scientific American*, May 2001.
- D. Boyd and S. Lohnes, editors. *Proceedings of the Social Software in the Academy Workshop, 2005*. URL <http://www.annenberg.edu/symposia/ssaw/program.html>.
- J. S. Brown. Sustaining the ecology of knowledge. *Leader to Leader*, 12: 31–36, 1999.
- W. J. Clancey. "A tutorial on situated learning." In J. Sel, editor, *Proceedings of the International Conference on Computers and Education*, pages 49–70, Charlottesville, VA, 1995. AACE.
- L. Delcambre, D. Maier, S. Bowers, M. Weaver, L. Deng, P. Gorman, J. Ash, M. Lavelle, and J. Lyman. "Bundles in captivity: An application of superimposed information." In *Proceedings of ICDE 2001*, pages 111–120. IEEE Computer Society, 2001.
- K. Ehrlich and D. Cash. "Turning information into knowledge: Information finding as a collaborative activity." In J.L. Schnase, J.J. Leggett, R.K. Furuta, and T. Metcalfe, editors, *Proceedings of Digital Libraries '94*, pages 119–125. Texas A&M University: Texas Hypermedia Research Laboratory, 1994.

- J. Frumkin. "The wiki and the digital library." *OCLC Systems and Services*, 21:18–22, 2005.
- D. Galiel. "A lever long enough: Value driven enterprise in the networked information economy." In Sherrie Bolin, editor, *The Standards Edge: Open Season*. Sheridan Books, Ann Arbor, Michigan, 2004. URL <http://www.galiel.com/pie/aleverlongenough.pdf>.
- M. Guzdial, P. Ludovice, M. Realff, T. Morley, and K. Carroll. "The challenge of collaborative learning in engineering and math." In *Proceedings of IEEE/ASEE Frontiers in Education (FIE) 2001 Conference*. IEEE, 2001.
- M. Guzdial, P. Ludovice, M. Realff, T. Morley, and K. Carroll. "When collaboration doesn't work." In P. Bell and R. Stevens, editors, *Proceedings of the International Conference of the Learning Sciences*, pages 125–130. International Society of the Learning Sciences, Lawrence Erlbaum Associates, 2002.
- F. Iannacci and E. Mitleton-Kelly. "Beyond markets and firms." *First Monday*, 10(5), May 2005. URL [http://www.firstmonday.org/issues/issue10\\_5/iannacci/index.html](http://www.firstmonday.org/issues/issue10_5/iannacci/index.html).
- A. Kantor. "Today's best encyclopedia might surprise you." *U.S.A. Today*, pages 2004–11, Nov. 5 2004. URL [http://www.usatoday.com/tech/columnist/andrewkantor/2004-11-05-fifthofnovember\\_x.htm](http://www.usatoday.com/tech/columnist/andrewkantor/2004-11-05-fifthofnovember_x.htm).
- A. Krowne. "An architecture for collaborative math and science digital libraries." Master's thesis, Virginia Tech Department of Computer Science, 2003a.
- A. Krowne. "Building a digital library the commons-based peer production way." *D-Lib magazine*, 9(1082), 2003b. URL <http://www.dlib.org/dlib/october03/krowne/10krowne.html>.
- A. Krowne and A. Bazaz. "Authority models for collaborative authoring." In *Proceedings of the 37th Hawaii International Conference on System Sciences - 2004*, pages 18–24, 2004.
- A. Krowne and N. Egge. Planetmath: a collaborative, web-based mathematical encyclopedia, 2001. URL <http://www.planetmath.org>.
- L. Lamport. *LaTeX: A Document Preparation System*. Addison-Wesley, Reading, MA, 1986.
- M. Lebert. Michael Hart: Changing the world through e-books, Mar 2004. URL [http://www.gutenberg.org/about/marie\\_lebert](http://www.gutenberg.org/about/marie_lebert).
- L. Lessig. *Free Culture: How big media uses technology and the law to lock down culture and control creativity*. Penguin Press, New York, 2004.
- B. Leuf and W. Cunningham. *The wiki way: Quick collaboration on the web*. Addison-Wesley, 2001.
- D. Levy and C.C. Marshall. "Going digital: A look at assumptions underlying digital libraries." *Commun. ACM*, (38):77–84, Mar 1995.

- M. McRobbie. "The library and education: Integrating information landscapes." In *Council on Library and Information Resources (CLIR) 2003. Emerging Visions for Access in the Twenty-first Century Library*, Washington, DC, 2003. CLIR.
- R. Metcalfe. "Metcalfe's law: A network becomes more valuable as it reaches more users." *Infoworld*, 17:53, 1995.
- D. P. Reed. "That sneaky exponential— beyond metcalfe's law to the power of community building." *Context Magazine*, Spring 1999.
- S. Robertson and K. Reese. "A virtual library for building community and sharing knowledge." *International Journal of Human-Computer Studies*, 51:663–685, 1999.
- J. Roschelle and R. Pea. "Trajectories from today's www to a powerful educational infrastructure." *Educational Researcher*, 43:22–25, 1999.
- E. D. Scharff. "Applying open source principles to collaborative learning environments." In *Proceedings of the Conference on Computer Supported Collaborative Learning (CSCL 2002)*, pages 499–500, 2002.
- G. Stahl. "Engagement with learning (foreward)." In D. Hung and M. S. Khine, editors, *Engaged learning with emerging technologies*, pages i–v. Springer, Boston, 2005.
- G. Stahl. *Group Cognition: Computer Support for Building Collaborative Knowledge*. MIT Press, 2006. URL <http://www.cis.drexel.edu/faculty/gerry/mit/>.
- P. Suber. "The primacy of authors in achieving open access." *Nature*, June 10 2004.
- I. Tomek. "Collaboration via internet and web." In R. Klein, H.W. Six, and L. Wegner, editors, *Computer Science in Perspective: Essays Dedicated to Thomas Ottmann. Lecture Notes in Computer Science*, volume 2598, pages 318–329, New York, 2003. Springer-Verlag.
- J. Wales and L. Sanger. Wikipedia: a free-content encyclopedia, 2001. URL <http://www.wikipedia.org>.

# On Free Math and Copyright Bottlenecks

Raymond S. Puzio (University of Memphis)

**Abstract:** Modern computing and storage technology makes it possible to manipulate mathematical information in novel, unprecedented ways. As the state of the programming art advances, it will become possible to have the computer do such things as check statements in different mathematical works against each other for consistency and arrive at new conclusions based upon known results in the literature. For this to be possible, one needs to enter mathematical information into a computer. In attempting this, one runs into a somewhat paradoxical situation: although mathematical knowledge is free, it cannot be conveyed apart from some form of expression—but this expression is subject to the strictest legal protection under copyright law. The Free Math movement seeks to address this disparity by making available mathematical works which are as free as the concepts they embody. Unfortunately, in the process of doing this, we run into all sorts of bottlenecks having to do with copyright issues. This paper describes these bottlenecks and the dangers to mathematical progress.


## INTRODUCTION

It is a well-established principle of copyright law that ideas and concepts are in the public domain, exempt from legal protection and free for all to use. It is an equally well-established principle that the expression of these ideas is entitled to the strictest protection under law and that authors are entitled to claim recompense for their creativity and control the distribution of their works.<sup>1</sup>

Because ideas and concepts cannot be transmitted without somehow expressing them, there is a dynamic tension between these two principles and the aim of copyright law is to establish a balance between these two conflicting goals which respects both the general public's right to access knowledge and the author's right to be compensated for creative work. In recent years, this equilibrium has been disturbed by the emergence of computer and

---

M. Halbert (Ed.): *Free Culture and the Digital Library Symposium Proceedings*. Atlanta, Georgia: MetaScholar Initiative at Emory University, 2005. pp. 273-299.

 This work is covered by the Creative Commons Attribution-NonCommercial-NoDerivs 2.5 License.

network technology which radically changes the nature of the creation, transmission, and storage of information.

The purpose of this essay is to consider some of the effects of this situation in the field of mathematics. In particular, the author will concentrate his attention on several ways in which provisions of copyright law can impede the progress of mathematics in the age of computers. It is clear that some restriction of scope was necessary in order to contain the essay within its bounds; the reasons for this particular choice are twofold: First, the author, being a mathematician and having encountered issues of intellectual property law primarily within the context of projects having to do with the use of computers in mathematics, has first-hand personal experience of these issues and of their effects on the progress of the subject on which to draw. In particular, he has been involved with two projects, PlanetMath and the Hyperreal Dictionary of Mathematics, both of which seek to make mathematical knowledge available in digital form under free license terms and provide mathematicians with tools for making use of this knowledge. In connection with work on these projects he has encountered various situations which form the basis for the scenarios described in this essay. Second, mathematics is primarily concerned with ideas and concepts, so these issues manifest themselves in a peculiar form which may not occur in subjects whose primary focus is observable facts or artistic expression.

### **HISTORICO-TECHNOLOGICAL BACKGROUND**

Since the material cause of the difficulties to be discussed here is the change in the technology for disseminating texts, we shall begin this essay with an account of the relevant aspects of this technology and of the legal protections that regulate its use.

From the late 15th century to the late 20th century, the primary means of disseminating texts was the printing press, which may be characterized as centralized and static. On the one hand, a printing press being a rather large piece of machinery, it is not plausible that most individual authors would own or operate one. To use a press, one requires the services of several different types of craftsmen, such as typesetters and typesetters. Thus, this technology is conducive to a centralized mode of distribution in which many authors submit their works to a publisher who arranges to have them printed, and then distributes the results to booksellers for sale to the general public. On the other hand, the

process of typesetting a book is laborious and, once a work has been typeset and is ready to be printed from, there is no practical way to alter its contents. The best one can do is to add sheets of errata and come out with a new edition every few years. Thus, this technology encourages a static conception of works of literature in which the goal of writing is to produce a carefully edited final version which, if not “fixed in stone,” will at least be “fixed in lead,” and is conducive to a sharp demarcation between authors, editors, publishers, and readers.

By contrast, the computer networks which are now replacing the printing press as the primary means of disseminating works of literature may be characterized as decentralized and dynamic. Unlike printing presses, computers are now small and affordable enough that most authors own and operate computers. Since a computer can perform the operations of typesetting and distribution, authors in the digital age find themselves taking over the role of publisher. In particular, this means that authors now find themselves having to assume legal responsibilities which publishers would ordinarily discharge for their authors. Just as was the case in the days of oral and handwritten literature, so too with computers it is not much harder to redistribute a modified copy of a work than to redistribute the work verbatim. This stands in stark contrast to the static viewpoint described in the last paragraph—in the digital age, one can again conceive of a work of literature as a living, growing entity. When a reader can add comments or otherwise redistribute a modified version, the distinction between authors, editors, and readers can become blurred. This also leads to the revival of certain old forms of literature and the creation of new forms which are based on the ability of readers to interact with a work and add to it.<sup>2</sup> Finally, computers not only can store and transmit information, they can process it as well. As the state of the art advances, machines are increasingly able to perform operations on texts which were once only possible by humans. This too comes with legal ramifications which will be discussed here.

Copyright law is very much a result of the invention of printing.<sup>3</sup> Before printing, there were no legal restrictions on the copying of books—anyone with access to a book and writing materials was free to make a copy.<sup>4</sup> To be sure, one can find instances in which a text was considered the property of an individual or of a group, but the legal principle involved resembles trade secrets more than copyright since members of the general public were forbidden

access to these texts. Soon after the invention of printing, the need to regulate the printing trade was felt. At first, this took the form of ad hoc licenses and privileges granted to printers, but soon the practice became codified and standardized to serve three ends: to check the spread of heretical, seditious and otherwise undesirable works; to regulate the printing industry and protect publishers; and to uphold authors' right to control dissemination of their works and obtain recompense for their labors.

While the United States' copyright law is part of this historical evolution and resembles previous (specifically British) copyright law in many aspects, it also differs radically in outlook—if not in outward form—from earlier conceptions of copyright. In the new conception, the primary purpose of intellectual property law is to benefit the public by promoting the production and dissemination of intellectual content (U.S. Constitution, Article I, Section 8, Clause 8; Jefferson 1813). Censorship cannot play a role because it runs counter to the basic freedoms of speech and religion. While the notions of regulating publishers and rewarding authors are still present, they are no longer regarded as the *primary* purpose of copyright but as *secondary* goals which are legitimate only insofar as they are consonant with the primary goal.

Due to the fact that modern technology offers alternative means of mass dissemination which do not require a centralized publisher, these ends have come into conflict. This conflict and its effects on the advance of knowledge form the subject of this essay.

### **TENSIONS BETWEEN AUTHORS AND PUBLISHERS**

The relation between author and publisher has traditionally been an amiable symbiosis with the publisher as agent and patron of the author. Both shared a common economic interest in prosecuting infringements of copyright—not only would a pirated edition represent a loss of income for the publisher, it would likewise represent a loss of royalties for the author.<sup>5</sup>

However, this is no longer always the case nowadays, especially in scientific publishing, where royalties have typically not amounted to a significant source of income for the author and the economic value of printing was and is primarily as a means for disseminating information to a wide audience.<sup>6</sup> Therefore, as soon as other technologies for dissemination became available, mathematicians and other scientists have shown little compunction in supplementing the distribution of information through the printing press with alternative means, even when this might violate



copyright law. This has turned the relation of partnership into one of competition, putting scientists and their publishers on a collision course with occasionally disastrous results, as the following two examples illustrate.

As soon as the copy machine became readily available, scientists made use of it to produce reprints of research articles (and even books) for personal use, since it is much more convenient to copy an article in a journal than to send a letter to the publisher or the author requesting a reprint.<sup>7</sup> In 1994, the American Geophysical Union sued Texaco because researchers in the company's labs photocopied articles from the society's journal without permission (*American Geophysical Union vs. Texaco, Inc.*; Harper 2001). The court ruled that the scientists' copying of journal articles was not covered under the fair use provision of copyright law and that Texaco would have to pay the publisher back royalties.

The other example involves Eric Weisstein's "Treasure Trove of Mathematics." Originally, this was a website containing mathematical definitions and facts. Mr. Weisstein decided that this resource would be of greater value to the mathematical community if it were also available in the form of a book, so he engaged a scientific publisher, CRC Press, to issue a printed version, the "CRC Concise Encyclopedia of Mathematics." Soon thereafter, the publisher ordered the author to take down the website and was sued for infringement of copyright when he did not comply (*CRC Press vs. Wolfram et al.*; Weisstein 2001). The author was ordered by the court to close the website until the matter was resolved. A year later, the matter was settled out of court when the publisher relented and allowed the website to reappear, but on terms which make it clear that the publisher, not the author, has the upper hand and that the author only has those rights to the work which the publisher deigns to bestow—if the online version causes a perceived decrease in sales of the printed book, the author is to reimburse the publisher for the claimed difference.

### **ACCESS TO TEXTS IN THE PUBLIC DOMAIN**

Unfortunate as it might be that a sizable portion of the literature is "off limits" due to restrictive copyright restrictions, it is even more unfortunate that copyright property law is even being used to restrict access to material which has already passed into the public domain. Not only does this undermine the Constitutional provision that copyright is only available for a limited time, it also restricts the progress of science directly by making it difficult for scholars

to base new works on these existing works or to add value to digital distributions of public domain works.

Several institutions of higher learning have made digital images of old books in their collections and made these images and/or transcriptions based on these images available online in various file formats. Even though the copyrights on these books have clearly expired, the libraries nevertheless issue them under restrictive license agreements which typically only allow use as personal reading material, explicitly prohibiting redistribution of these files.<sup>8</sup> Such a practice effectively removes these items from the public domain.

When queried about this practice, a librarian replied that digitized versions of books in the public domain may nevertheless be covered by a “thin copyright” on account of the human decisions and quality control present involved in the production of a digitized version.<sup>9</sup> This claim that the effort involved in preparing an edition is sufficient grounds for copyright protection sounds rather questionable. At least as much human effort and expertise is involved in the typesetting and proofreading of a text in traditional printing, yet no publisher claims a copyright interest in modern editions of ancient texts for such a reason. In the case of *Corel vs. Bridgeman*, the court rejected such a claim saying “the point of the exercise was to reproduce the underlying works with absolute fidelity. Copyright is not available in these circumstances.”

Nevertheless, as long as digital libraries stick by their licensing claims, they will remain a menace. Even if these contracts ultimately prove unenforceable, the libraries can still sue those they consider infringers. Given that many of the entities involved in making older texts available online freely are small, not-for-profit organizations which rely on volunteer effort and have neither the time nor the money to mount a challenge in court (and may not even be aware of the fact that such license agreements stand on shaky legal ground), cease and desist letters from digital libraries are likely to go unchallenged and public domain texts will be treated as if they were still—or newly—copyrighted.

### **PERMISSION TO ADAPT**

In the course of writing, one often finds it expedient to quote or adapt an exposition of some subject from an already-existing source. To do so legally, one needs to obtain the permission of the copyright owner. If the work has not been released under a license which grants these rights, one needs to obtain explicit permission.

Traditionally, checking that such permissions are in order is one of the services which publishers have offered their authors. Nowadays, however, as authors are increasingly becoming their own publishers online, they are forced to take this responsibility onto themselves. At times, the result is that well-meaning but naive authors run afoul of copyright law despite their best efforts to the contrary. This can be illustrated with two scenarios based on the author's personal experience. In the first scenario, someone bases a number of entries for an online encyclopedia on a book which is in print. Before posting the entries, this person took care to obtain the permission of the author of the book on which they were based. Later on, this leads to some consternation when it is realized that these entries may nevertheless infringe on the book's copyright. The problem arose because the publisher of the book has a policy of obtaining all rights to books which they publish and asks that requests for permission to adapt works be made through their permissions department. Apparently, the author of the original work did not fully understand the terms of the contract which he had signed with the publisher and the author of the derived work saw no reason to suspect that this might be the case.

The other scenario involves the posting of an electronic version of a book to a digital library. Someone sees the an electronic edition is available for free through its author's homepage and concludes that it would be permissible to also make copies available through the digital library. This line of reasoning is faulty because, while the author had indeed made a special arrangement with the publisher to distribute the book through his website, this contract did not permit anyone else to distribute the work electronically.

As these two examples illustrate, the complexities and sometimes counterintuitive requirements of copyright law can appear as a confusing "fog of copyright" to the average author.<sup>10</sup> Not only does this lead to situations in which authors unintentionally infringe on copyrights, it can also lead to situations in which authors may not use texts to the extent allowed by law because they are unsure of what the law allows or do not know the correct legal procedures and prefer to err on the side of caution.

### **THE FOG OF COPYLEFT**

In order to make works available to the public and to permit the creation of derivative works which will also be freely available, authors have released works under so-called free licenses. The legal basis for this practice lies in the author's right to control

distribution and creation of derivative works. However, to use this right in a way which prevents others from exercising it in the future in order to keep a work free requires a certain amount of legal maneuvering; this invariably gives rise to undesirable side effects which can impede the free flow of information.<sup>11</sup>

As an example of a free license, we can take the GNU Free Documentation License (Free Software Foundation 2002; Stallman 1985; Moglen 2001). This is a legal instrument in ten sections, most of which are devoted to ensuring that no one could circumvent the requirement that future versions of the work being licensed be made available on similar terms. To achieve this goal, it makes use of a system which requires seven definitions and 15 conditions which govern modifications to the document and makes use of such devices as history lists and invariant sections. While this may not be overly complex as far as contracts go, it can be overwhelming to the average author who has no legal experience and consequently authors may not comply with all the terms of the license out of ignorance. This can have devastating results since one of the provisions of the license states that it is not permitted to disseminate copies of the work except as expressly provided by the license and that any attempt to do so will result in termination of the license and, hence, make it impossible to disseminate the work at all (by section 9).

In an attempt to avoid these Byzantine complexities, the Creative Commons license (Creative Commons 2005) takes a much starker approach—it states that derivative works can only be distributed under exactly the same license terms. While this is easily enough understood, it creates a different problem, “license lock”—once a work has been licensed under such terms, it and all derivative works are locked into the same license. This can lead to some unfortunate situations, as the following example shows.<sup>12</sup>

Suppose that an author wishes to make use of two articles in the course of preparing a book and that this use would be in excess of what is permitted by the doctrine of fair use. Suppose further, that one of these articles was released under a GNU license and the other was released under a Creative Commons license. Under such circumstances, the author would be no better off than if the articles had appeared in a journal whose publisher plays copyright agreements close to his vest because the conflicting demands of the GNU license and the Creative Commons license would render the work undistributable.<sup>13</sup> This situation can appear confusing and paradoxical to a newcomer. Naively, one would expect that, if an

author declares that his work is free, then anyone can do anything which one desires with that work. However, as we have seen, this is not quite true—the very mechanism by which free licenses are enforced legally places sometimes confusing, sometimes counterintuitive restrictions on authors, creating a “fog of copyleft” through which only those who have made a careful study of free licenses and copyright law or who have access to expert help can safely navigate. The net effect is to discourage those who are not in such a position (the vast majority) from taking full advantage of the possibilities of free license agreements.

Especially hard hit by this state of affairs are projects like Wikipedia (Wales and Sanger 2001) which combines contributions by thousands of authors and is released under the terms of the GNU Free Documentation License. Since it is safe to say that most contributors will not be familiar with the legal intricacies of free licenses, there is a good chance that users will violate the terms of these licenses unintentionally on a regular basis. This puts an extra burden on the facilitators of these sites. In addition to their usual duties, they must educate users on these issues and monitor content to ensure compliance with license terms.<sup>14</sup> Given that such operations are typically volunteer efforts funded by contributions, this places a hardship on people who have already assumed a fair load of work on their shoulders.

### **OPEN-ENDED COLLABORATIVE WORKS**

While collaborative multi-author works such as encyclopedias, journals, handbooks, and newspapers have long been common, computer technology has made possible new modes of production of collaborative content which can pose special challenges with respect to copyright. In the print world, the only feasible way to produce such a work was to first gather the contributions then arrange, edit, typeset, and print the whole. In the virtual world, however, one can instead produce a collaborative work by running a computer program which allows users to add their contributions to a website interactively.

As an illustrative example of such a collaborative work, one may consider the PlanetMath encyclopedia of mathematics (Krowne and Egge 2001). This work is maintained by an online community of mathematicians, membership in which is open to the general public. The articles are written by members of this community. As soon as an entry is submitted, it immediately becomes available to the general public. The editorial function is assumed by the

same community by means of a procedure whereby members informally review each other's articles and submit errata and editorial suggestions. All this happens in the open; the entries and the corrections and comments posted to them are all accessible to anyone with access to the internet.

Such an arrangement has several advantages over the traditional procedure for writing an encyclopedia such as more rapid dissemination of knowledge and lower overhead costs.<sup>15</sup> At the same time, it is open to the possibility of various sorts of legal difficulties. To a large extent, this is due to the fact that copyright law has evolved in an environment where such means of collaboration were not available, and hence it is no surprise that existing law may not suit their needs well. The major problem has to do with the all-too-real possibility that a malicious, ignorant, negligent or misguided user may submit material which infringes on an existing copyright. This can put the entire project in jeopardy, should the owner of the material's copyright press suit. At best, it might mean that the website, or significant portions thereof, is taken offline while its legality is investigated and contested. At worst, it might mean that the website goes offline permanently, if the entity which maintains it does not have the means to go through with a lawsuit.

Also, the question arises of who can be held liable. To be sure, the author of the offending entry is responsible. However, other parties may also be held liable for their role in allowing such a violation to occur. The organization who hosts the website may be held liable for disseminating the work. If the community of users takes upon itself the responsibility of policing the website for violations of copyright law, users may found responsible for failing to exercise due diligence. Until precedents are established or laws are enacted, one cannot be sure what will be the extent of liability for these different parties and, in the meanwhile, people who would otherwise be willing to participate in such undertakings may be reticent to do so because of the possibility of having to pay for the indiscretions of careless users.

The role of an entity such as PlanetMath.org Ltd. which maintains such a website is not so clear. At first sight, it may be likened to a publisher, but the analogy is seen to be of limited validity on closer inspection. In the world of print, a publisher decides which works they will print and supplies or finances the labor which is required to produce books such as layout, editing, typesetting, printing, storage, and distribution. Here, however, the analogous operations

are either automated or provided by the community of users. The role of PlanetMath.org Ltd. is merely to provide the infrastructure needed by the community and the formal organization takes a hands-off policy when it comes to content. In this respect, the role more closely resembles that of a common carrier such as a telephone company but this analogy too has its limitations—such an organization does not fit into any of the categories recognized by the FCC.

In the current political climate, it is highly unlikely that a claim to such a status would be accepted. Even as this essay is being written, the Supreme Court has ruled that programmers whose software which induces users to violate copyright can be held liable for these violations (*M.G.M. vs. Grokster 2005*; Electronic Frontier Foundation 2005; Consumer Electronics Association et al. 2005). Should mathematical publishers feel that Noösphere, the software package underlying PlanetMath is somehow undermining their business and encouraging illegal copying from their books, they could use this precedent to argue that PlanetMath.org Ltd. is liable for acts of infringement performed by its users.

Even if they would not be able to force the site to close, it is still plausible that they might be able to force major changes to be made to the Noösphere program and the policies of PlanetMath. For instance, while the court stated that technologies could not be outlawed if they were capable of substantial non-infringing uses, it also accepted the fact that Grokster did not incorporate filtering tools into its software as evidence of intentional facilitation of infringement. Therefore, a publisher might insist that Noösphere incorporate such filtering as a sign of good faith. This would represent a great hardship in more than one sense—not only would filtering require much work to implement,<sup>16</sup> but it might require the use of proprietary components.<sup>17</sup> In addition to putting an extra burden on the already strained budget of PlanetMath.org Ltd., this would conflict with the free, open source nature of the organization.

While PlanetMath should get by relatively unscathed, this ruling is a serious threat to preprint servers like Paul Ginsparg's ArXiv (Ginsparg 1996) because they are set up on a basis which could be legitimately be construed as encouraging users to violate copyright law. The purpose of such websites is for scientists to distribute copies of articles which they intend to publish in journals. It is well known that these preprints will stay available online long after they are published in the journals and that this is in violation of the

terms spelled out in the transfer-of-copyright agreement which most journals require. Furthermore, far from filtering the content for copyright violation, the maintainers of the ArXiv site acknowledge infringing uses by adding metadata which states in which journals papers were and even encouraging authors to update their submissions to reflect changes made in the printed version. They try to save themselves by a statement in their disclaimer that they do not represent that the use of their product would not infringe property rights. In light of the Supreme Court ruling, this disclaimer is worthless; the authors of the underlying software and the maintainers of the website can be held liable for each and every article in their archive whose copyright is owned by a publisher who distributes it under terms which prohibit dissemination other than through the journal. Looked at objectively, they seem to be in the same untenable position as Napster or Grokster—the contention that the website could or should only be used to distribute articles which are released under a free license would fall flat on its face as soon as is pointed out that, in actuality, there are relatively few such articles to be found on the website. The fact that the primary purpose of this website is the promotion of scientific research rather than entertainment sounds rather irrelevant as a defense in light of *American Geophysical Union vs. Texaco*.

The loss of preprint servers would have a serious negative impact on the scientific community. In some fields of physics, preprint servers have become the primary means of communicating ideas between researchers, and journals only serve secondary purposes such as peer review and establishing reputations. These servers have accelerated progress in these fields by allowing rapid dissemination of results (Jackson 2002). Were they to disappear, progress in these fields would grind to a halt until a suitable replacement would be found.

### **NOVEL USES OF TEXTS**

Advances in computer technology make it possible for machines to perform operations on texts which hitherto could only be performed by humans, and one can confidently predict that, as time goes on, they will be able to perform more such operations. For instance, it is relatively easy to use a computer to search for occurrences of a phrase in a text, to find the differences between different editions of the same text, and even to automatically generate concordances and frequency tables. One can also do such things as copy mathematical expressions found in an article into a



computer algebra program to use them for a computation or translate between notations. In the foreseeable future, it should be possible to perform such tasks as verifying the soundness of logical arguments and having computers draw simple conclusions based information found in texts.

While the use of machines to perform such tasks is certainly desirable because it eliminates drudge-work and human error, copyright can restrict one's ability to make use of such labor-saving devices. To have a computer perform such operations on texts, it is first necessary to store a digital representation of the text in the computer's memory. Since memory circuits are now considered a tangible medium of expression, this is not legal without the permission of the copyright owner.<sup>18</sup>

This can lead to some curious situations in which the same use of a text may or may not be legal depending on whether it was done by hand or by machine. For instance, suppose one were to copy a large number of mathematical formulae from a mathematical handbook into a mathematical encyclopedia or a textbook or even a more comprehensive handbook. The precedent of *Feist Publications vs. Rural Telephone Service Co.* makes it clear in no uncertain terms that this would not infringe on the copyright of the handbook. However, using a computer to prepare such a text may not be permissible if it means that one needs to prepare an electronic version of the text or, if an electronic version is available, but the licensing conditions do not permit such use.

In particular, such a scenario is especially worrisome for such projects as the *Hyperreal Dictionary of Mathematics* (Corneli 2005), which features a database of mathematical knowledge in a suitable format and tools for manipulating this knowledge. In order to produce a comprehensive database, one will need to scour thousands of mathematical books for content. The only practical way to accomplish this is to have a machine automatically extract the data from the texts. However, one cannot do so without first obtaining permissions and likely paying licensing fees to do so. In such a situation, the costs in labor and money involved in obtaining permissions can easily dwarf the costs actually involved in the production of the database.

### **POSSIBLE REMEDIES**

Having expounded a catalogue of woes, the author would like to end this essay on a more positive note by describing possible remedies to this situation lest the reader be abandoned in a

quagmire of despair. This is by no means intended as an exhaustive list or as a selection of the best proposals, but rather as a demonstration that the situations described in the preceding sections can be ameliorated if we, as a society (or a subset thereof), are willing to make an effort to change them and as a starting point for a discussion on plans of action which will hopefully occur not only at this conference but continue after we have parted on our separate ways and lead to concrete results.

One remedy would be a reform of copyright law based on a proactive attitude towards new technology. As suggested by the examples of bottlenecks, it seems that the practice of copyright regulation is no longer consonant with the primary goal of promoting intellectual progress and that the secondary goals of regulating the publishing trade and enforcing author's rights have overshadowed the primary goal. The stance of copyright law towards this emerging technology may be described as reactive, the primary concern seeming to be prevention and circumvention of new forms of infringement. A most distressing aspect of this development has been the tendency to make copyright a matter of criminal rather than civil law (Harvard Law Review; Department of Justice, 1998, 2000, 2001, 2003, 2005; McCullagh 2003). As noted long ago by Mo Tzu and Montesquieu, increases in the severity of legal sanctions are rarely effective and may be taken as a sign that the government is not able to effectively deal with a particular menace. In such circumstances, it may be wiser to pull back and reconsider strategies rather than stubbornly and mindlessly pursue a hopeless course of action which in the end effects more woe in the form of collateral damage to innocent, law abiding bystanders than it averts by thwarting the intentions of the wicked.

Given that computer and network technology in the twentieth century has had at least as profound an effect on the dissemination, storage, and use of information as did the invention of printing with movable type in the fifteenth century, it seems that any legislation-amending approach which is based upon the conditions which surrounded the use of the older technology is woefully incommensurate with the end of harmonizing legal practice with the current state of technological affairs and is hence foreordained to lead to inefficient and ineffective results—no matter how skillfully and cleverly such means may be employed. Thus, a rational approach to the problem would be to (1) begin with the basic principles upon which intellectual property law rests, making

a clear distinction between primary and secondary goals, (2) reconsider the implementation of these principles in the circumstances afforded by the new technology, then (3) rebuild the edifice of copyright law according to a plan which is consonant with these principles.

It needs to be borne in mind that application of these laws will increasingly be made by the authors themselves rather than by specialists. It is not reasonable to demand that all or most authors invest the time needed to study and master the intricacies of copyright law or hire experts to sort out these intricacies on a regular basis; as documented earlier, this attitude has the effect that complex, often counterintuitive regulations serve to envelope the average author in a “fog of copyright” which can disorient and confuse them, often leading to disastrous outcomes. A more helpful approach would be to simplify and clarify at least those provisions of copyright law which arise routinely in everyday life and create some sort of “easy-filing form” which serves the everyday needs of author-publishers. Naturally, they might seek the advice of professional experts from time to time when more complicated or unusual circumstances arise, but on everyday matters they could proceed confidently on their own. Such an approach would respect the valuable time of authors, copyright professionals, and courts.

Just as copyright evolved from ad hoc contracts between publishers and authors to public law of license, so too perhaps it is time for free distribution to graduate from a notion privately defined in various free licenses to one which is acknowledged and legitimized in public copyright law. This would offer several advantages. In addition to solving the problem of “fog of copyleft,” such an approach would also have the benefit of quelling lingering doubts about the validity of free licenses and recognizing free distribution as a legitimate means of promoting intellectual progress.

As we have seen earlier, there is, within the scientific community, a broad consensus that the progress of mathematical sciences is best served by a relatively unhampered flow of information. This attitude has led to a *de facto* treatment of scientific literature as if it had been released under free license which clashes with the *de jure* interpretation to which most scientific publishers cling. Such a state of affairs is unacceptable and demonstrates that copyright regulation is out of touch with its constitutionally sanctioned goal. To be sure, the government has a responsibility to ameliorate this

situation, but perhaps the scientific community should take a more active role by forming a “free math” movement which seeks to improve the legal climate within which science operates by such means as supporting resources which make mathematical knowledge available under free license terms, putting pressure on publishers to permit alternative modes of distribution, developing means of peer review and reputation establishment which are not linked to paper journals, and advising courts and lobbying legislators to take into consideration the interests of the scientific community. To be sure, the formation of such a movement may require mathematicians to leave their comfort zones to assume a more active stance and require scientific associations to rethink their priorities, but these sacrifices would be warranted by the gravity of the situation and the threat posed to the mathematical progress.

It would help if the legislature were to recognize commons-based peer production and filesharing as institutions and, just as in the case of such establishments as restaurants, factories, and apartment houses, regulate these institutions by imposing guidelines designed to protect the general public and limiting the liability of proprietors who follow these guidelines in good faith. At present, the only form of regulation—if it may even be called regulation—is through the judiciary, which is ill-suited to the task because courts can only rule on the legality of an act after it has been done. As a result, those involved in innovative projects have felt unsure of what is legally permissible and of the limits of their liability, leading them to err on the side of prudence. Clarifying obligations and liability would lift this cloud of uncertainty and promote progress by letting pioneers of network-based production proceed with confidence under an arrangement which protects the interests of all parties involved.

Before concluding this essay, the author would like to draw the reader’s attention to the observation that, far from being an esoteric matter which is only of direct relevance to a handful of mathematicians involved in projects like PlanetMath and the Hyperreal Dictionary of Mathematics, the cause of free access to mathematical knowledge should be of vital importance to all who cherish the ideal of a free and democratic society. The very technology which led to this situation has its foundations in mathematical logic and, indeed, all off the technology which has become a part of everyday life—telephones, automobiles, televisions, refrigerators, electric lighting, washing machines—

owes its existence to the application the mathematical principles which govern the workings of the forces of nature. Since mathematical knowledge is the power which enables one to create and control this technology, it follows that, in a highly technological society such as ours, any group which is denied free access to mathematical knowledge becomes a group of second class citizens.

## CONCLUSION

The reforms suggested above cannot be implemented without a considerable expenditure of effort. Even if one disagrees with some or all of these suggestions, it is doubtful that anyone could come up with a solution which would not require a significant effort to implement—the problem is rather serious, with deep roots, and there are powerful entities with vested interests in the status quo who would oppose any efforts at reform. Any credible attempt at change will require a movement of a magnitude comparable in size to any of the great freedom movements.

To build such a movement requires commitment and solidarity. Those who are concerned about this issue cannot afford the luxury of assuming that someone else will take care of the problem or succumb to the illusion that they lack the skills necessary to contribute but need to contribute in whatever capacity they can—whether by documenting the situation and offering scholarly analyses or by preparing legal challenges and defending authors or by making knowledge available under free terms or by writing free computer programs or by publicizing the cause or by speaking out against abuses.

Success in such a campaign will entail uniting members of different intellectual communities—scientists, lawyers, educators, artists, philosophers, historians, librarians, writers—under the banner of a common cause in a manner which respects the differing, sometimes conflicting opinions of its members. But, though necessary, this alone will not be sufficient—the free culture movement must reach out to society at large and, if not gain mainstream attention, at least enter the mainstream of the counterculture. Free access to information has much in common with other freedoms; free culture activists have much to gain by dialogue and collaboration with other activists. In addition, it will be necessary to form alliances with those businesses and officials who are sympathetic to the cause. Only in this manner will it be

possible to obtain the leverage needed to affect the reforms required.

In the final analysis, the only real bottlenecks are apathy and division. If we allow them to constrict the flow of knowledge, we will have only ourselves to blame for the resulting intellectual drought and will rightly deserve the reproach of posterity.

## **ACKNOWLEDGEMENTS**

The author thanks Joe Corneli and Aaron Krowne for comments and discussion.

## **ENDNOTES**

1. The legal basis for these principles in U.S. law is to be found in sections 102 and 106, respectively of Title 17 U.S.C.
2. For a discussion of the technical and social aspects of this phenomenon please see Corneli and Krowne, in this volume
3. For more information on the history of copyright, please see (Birrell 1899; Patry 2000; U.K. Patent Office 2000; Tallmo 2005).
4. The Corpus Juris Civilis and other sources of Roman law make no mention of copyright. To be sure, there is the famous case of *Finnian vs. Columba*.
5. See Scweidler 2005 for a definition and history of the term “piracy” as applied to publishing and a discussion of the connotations of this term.
6. The primary genre of the research scientist is the journal article. Contributors to scientific journals do not receive royalties; to the contrary some journals require that authors pay page charges. To be sure, the author of a scientific treatise will receive royalties, but, since the audience for these works consists primarily of the same people who write such works, so there is no net economic gain. It is only when one comes to introductory text books and popular works that royalties assume a meaningful role, but only a small portion of the scientific community is involved in the production of such works. For more information on the economic aspects of scientific publishing please consult the following references: Guédon, 2001; Willinsky, 2002a, Berry and Stephen, 2001; Okerson 1986
7. For a description of this practice and for arguments why it should be considered as fair use and why it is beneficial to the progress of science please see Thomas et. al. 1993. (These arguments were ultimately rejected by the court.) From personal experience, the author would say that the practice of scientists photocopying articles

and even whole books is widespread; he estimates that, were the law enforced, more than 80 percent of mathematical scientists would be found guilty of infringement

8. For examples of such license agreements, please see Cornell, 2004a; Tufts 2005; Eisenhower Library 2005. Note that Cornell explicitly states that material in its digital math collection is in the public domain (Cornell, 2004b) whilst the material in the other two collections consists primarily of ancient and mediaeval texts, respectively. Nevertheless, all these licenses restrict access to the texts contained on these websites.
9. Private communication from Sarah Thomas, Cornell University Librarian.
10. Credit for this phrase “fog of copyright” goes to Aaron Krowne. For a first person account of an author’s disastrous encounter with this fog of copyright, please see Weisstein, 2001. For a third person account and discussion, please see Lessig, 2004.
11. To the best of the author’s understanding, the crux of the problem is as follows: To create a free license, one wants to impose a condition on anyone who might prepare a work derived directly or indirectly from a given work. The law grants the author no such right; it only grants the author the right to authorize the preparation of derivative works of that particular work (U. S. C. 17 106(2)). This is analogous to the distinction between ownership and copyright — whilst ownership of a book allows one to authorize copying of that particular copy of the book it does not allow one to control copying of the same text from any other copy of the same book. In the absence of copyright law (e.g. in classical antiquity), an author desirous of imposing restrictions on anyone making a copy of his text would have to do so via a legal instrument analogous to a free license which would leverage the right to permit copying of books in one’s possession so as to impose conditions on anyone who would make a copy of said work in the future.
12. To be sure, this problem is somewhat theoretical since, in practice, it is doubtful that an author releasing a work under one free license would sue another author who releases a derived work on a different free license which is compatible with the spirit if not the letter of the original license. Nevertheless, the fact that authors are not complying with the letter of the licenses they use undermines confidence in these licenses. As awareness of this issue is spreading, free licenses are being written so as to avoid such incompatibility issues.
13. To appreciate the full extent of the problem, one needs to bear in mind that there are a large number of free and open-source licenses in existence. For instance, the Open Source Initiative lists more than 50 licenses in its list (Open Source 2005).

14. For example, see the New User's Guide on Planet Math—the longest section of this document is devoted to copyright issues and only begins to scratch the surface of this topic.
15. For a discussion of the advantages of this mode of production, please (see Krowne 2003).
16. For a discussion of the difficulties of implementing filtering and of its effectiveness, please see the computer scientists' brief in *M.G.M vs. Grokster*.
17. Such as, for instance, a file in a proprietary format which contains samples of copyrighted texts against which submissions would be checked.
18. From *Sony vs. Universal studios* and *Britannica vs. Crooks* it is rather clear that videotaping television shows without permission is not permitted even when the copies are not distributed, but only used for viewing at a more convenient time. Generalizing these precedents, it seems clear that preparing a digital copy of a book even for personal use would require permission. Also, it is worth noting that license agreements for electronic versions of books often only permit the reader to read the book online but do not grant other rights.

## REFERENCES CITED

- Bachrach, Steven, R. et al. 1998. "Who should own scientific papers?" *Science* 281: 1459-1460. Available at <http://www.sciencemag.org/cgi/content/full/281/5382/1459>.
- Baez, John. 2003. What We Can Do About Science Journals? <http://math.ucr.edu/home/baez/journals.html>.
- Balint, Kathryn. 2005. "Snap judgments San Diego," *Union-Tribune*. May 30, 2005.
- Barlow, John Perry. 1994. "The Economy of Ideas: A Framework for Rethinking Patents and Copyrights in the Digital Age," *Wired*, March 1994.
- Benowitz, Steven. 1996. "Is Corporate Research Funding Leading To Secrecy In Science?" *The Scientist*. 10, 7:1.
- Berry, R. Stephen. 2001. "Is electronic publishing being used in the best interests of science? The scientist's view," *Electronic Publishing in Science II, UNESCO HQ, Paris*, (eds. Sir R. Elliot and D. Shaw). <http://www.unesco.org/science/publication/electronicpublishing2001/proceedingsrecommendations.shtml#haank>.
- Beschler, Edwin F. 1998. "Pricing of Scientific Publications: A Commercial Publisher's Point of View," *Notices AMS*, November 1998.
- Birrell, A. *Seven Lectures on the Law and History of Copyright in Books*, Rothman Reprints Inc., 1899 (1971 reprint).



- Biss, Daniel K. 2003. "The Elephant in the Internet," *Notices AMS*, November 2004. "The Digital Mathematics Library," Allyn Jackson, *Notices AMS*, August 2003.
- Borland, John. 2003. "Teamwork brings P2P spying app closer," September 10, 2003. *CNET* <http://news.com.com/Teamwork+brings+P2P+spying+app+closer/2100-10273-5074429.html?tag=st.rn>.
- Borland, John. 2004. "P2P companies say they can't filter," *CNET*, January 28, 2004. <http://news.com.com/2100-10383-5149720.html>.
- Boynton, Robert. S. "The Tyranny of Copyright," *New York Times Magazine*, January 25, 2004
- Branin, Joseph J. and Mary Case, "Reforming Scholarly Publishing in the Sciences: A Librarian Perspective," *Notices AMS*, April 1998. California State University, East Bay Library Copyright and Intellectual Property. <http://www.library.csuhayward.edu/staff/asoules/copyrightpublicdomain.htm>
- CARL/ABRC Task Force on Academic Libraries and Scholarly Communication 1995 TOWARDS A NEW PARADIGM FOR SCHOLARLY COMMUNICATION Association of Universities and Colleges of Canada [http://library.uwaterloo.ca/documents/scholarly\(aucc-carl\).html](http://library.uwaterloo.ca/documents/scholarly(aucc-carl).html).
- Carlson, Scott. 2004. "In the Copyright Wars, This Scholar Sides with the Anarchists," *chronicle.com*, *Information Technology* Volume 51, Issue13:A29.
- Chopra, S. and S. Dexter. A Comparative Ethical Assessment of Free Software Licensing Schemes. <http://www.sci.brooklyn.cuny.edu/sdexter/copyleftethics.html>
- Consumer Electronics Association. 2005. CEA Expresses Concerns Over SupremeCourt Grokster Ruling Consumer Electronics Association. Available at <http://www.eff.org/IP/P2P/MGMvGrokster/cea.php>
- Consumer Federation of America, Consumers Union, and Free Press. 2005 Statement of the Consumer Federation of America, Consumers Union and Free Press on the Supreme Court's Decision in MGM vs. Grokster Mark Cooper, Gene Kimmelman, Ben Scott. Available at <http://www.eff.org/IP/P2P/MGMvGrokster/statementconsumer.php>
- Corneli, J. 2005. What is the Hyperreal Dictionary of Mathematics? <http://www.ma.utexas.edu/jcorneli/d/>
- Cornell University Library. 2004. Guidelines for Using Text and Images from Cornell Digital Library Collections. <http://cdl.library.cornell.edu/guidelines.html>.
- CRC PRESS, LLC, vs. WOLFRAM RESEARCH, INC., STEPHEN WOLFRAM and ERIC WEISSTEIN, UNITED STATES DISTRICT COURT SOUTHERN DISTRICT OF FLORIDA CASE NO. 00-893-CIVGRAHAM/TURNOFF. <http://mathworld.wolfram.com/docs/CRCComplaint.html>.

- Creative Commons 2005 Homepage <http://creativecommons.org/>
- Darch, Colin 2001 The 'Alexandrian' Library, Digital Resources, and the Shrinking Public Domain: the Current Model for Delivering Academic Information in Africa. In *LIASA Conference on African Renaissance through Libraries Pretoria*, 24-28 September 2001.
- Dyson, Esther. Intellectual Value Wired 3.07 Available at [http://www.wired.com/wired/archive/3.07/dyson\\_pr.html](http://www.wired.com/wired/archive/3.07/dyson_pr.html)
- Electronic Frontier Foundation. 2005 MGM vs. Grokster [http://www.eff.org/IP/P2P/MGM v Grokster/](http://www.eff.org/IP/P2P/MGM_v_Grokster/) Fred von Lohmann 2004 Tech Dodge a Bullet LAW.COM 08-27-2004 <http://www.law.com/jsp/article.jsp?id=1090180416759>
- Electronic Frontier Foundation 2005 Supreme Court Ruling Will Chill Technology Innovation Electronic Frontier Foundation. Available at [http://www.eff.org/news/archives/2005\\_06.php#003748](http://www.eff.org/news/archives/2005_06.php#003748).
- Felten vs. RIAA US DC NJ Case #CV-01-2669(GEB) 2001
- Free Software Foundation. 2002. GNU Free Documentation License <http://www.gnu.org/copyleft/fdl.html>.
- From Preprints to E-prints: The Rise of Electronic Preprint Servers in Mathematics Allyn Jackson, Notices AMS, January 2002.
- Ginsburg, Jane C. 2002. "Essay - How Copyright Got a Bad Name For Itself," *Columbia Journal of Law and the Arts*, 26:1, 2002.
- Ginsparg, Paul 1996 Los Alamos XXX APS News November 1996 Available at <http://www.aps.org/apsnews/1196/11718.cfm>.
- Ginsparg, Paul 1996 Winners and Losers in the Global Research Village, 1996 Conference held at UNESCO HQ, Paris, 19-23 Feb 1996.
- Goldstone, David. 2001. Deciding Whether to Prosecute an Intellectual Property Case United States Attorneys. *USA Bulletin*. March 2001 Vol. 49, No.2 [http://www.cybercrime.gov/usamarch2001\\_1.htm](http://www.cybercrime.gov/usamarch2001_1.htm).
- Gudon Jean-Claude. 2001. "In Oldenburg's Long Shadow: Librarians, Research Scientists, Publishers, and the Control of Scientific Publishing," Proc. ARL Membership Meeting, May 2001. Available at <http://www.arl.org/arl/proceedings/138/guedon.html>.
- Harper, Georgia. 2001. Copyright and the University Community Austin: The University of Texas System Office of General Counsel <http://www.utsystem.edu/ogc/intellectualproperty/mono4.htm> Harper, Georgia 2001 Professional Fair Use after Texaco Austin: The University of Texas System Office of General Counsel <http://www.utsystem.edu/ogc/intellectualproperty/tex2.html>.
- Harper, Georgia 2004 WILL WE NEED FAIR USE IN THE TWENTY-FIRST CENTURY? Austin: The University of Texas System Office of General Counsel [http://www.utsystem.edu/ogc/intellectualproperty/FAIR USE.HTM](http://www.utsystem.edu/ogc/intellectualproperty/FAIR_USE.HTM).
- Hart, Michael S. 1992. History and Philosophy of Project Gutenberg <http://www.gutenberg.org/about/history> Harvard Law Review. "The

Criminalization of Copyright Infringement in the Digital Era,”  
*Harvard Law Review*. Vol. 112: 1705-1722.

- Jackson, Allyn. 2000. The Slow Revolution of the Free Electronic Journal, *Notices AMS*, September 2000.
- Jefferson, Thomas 1813 Letter from Thomas Jefferson to Isaac McPherson August 13, 1813 <http://odur.let.rug.nl/>.
- Kasunic, Rob. Solving the P2P “Problem” - An Innovative Marketplace Solution Stanford University Libraries [http://fairuse.stanford.edu/commentary and analysis/2004 03 kasunic.html](http://fairuse.stanford.edu/commentary%20and%20analysis/2004_03_kasunic.html).
- Kelty, Christopher M. Free Software/Free Science *First Monday*, volume 6, number 12 (December 2001) Available at URL: [http://firstmonday.org/issues/issue6 12/kelty/index.html](http://firstmonday.org/issues/issue6_12/kelty/index.html).
- Kendall, Rob 2005. DRM Creep: Fair Use Dies Not with a Bang, but with a Whimper *Yale University Law Meme* Friday, June 17 2005. <http://research.yale.edu/lawmeme/modules.php?name=News&file=article&sid=1720>.
- Kuperberg, Greg. 2002. “Scholarly mathematical communication at a crossroads”, arXiv:math.HO/0210144, *Nieuw Arch. Wisk.* 5,3:262–264.
- Lessig, Lawrence. 2004. *Free Culture How Big Media Uses Technology and the Law to Lock Down Culture and Control Creativity* The Penguin Press. Available at [http://www.easylum.net/freecultureLibraryofCongress2005. About Copyright and the Collections](http://www.easylum.net/freecultureLibraryofCongress2005>AboutCopyrightandtheCollections) <http://www.loc.gov/homepage/legal.html>
- Long, Roderick T. 1995. The Libertarian Case Against Intellectual PropertyRights “Formulations” Autumn 1995 Free Nation Foundation Available at <http://libertariannation.org/a/f3111.html>
- Loren, Lydia Pallas. 1997. REDEFINING THE MARKET FAILURE APPROACH TO FAIR USE IN AN ERA OF COPYRIGHT PERMISSION SYSTEMS *Journal of Intellectual Property Law*, Volume 5 Fall 1997 1:1 Available at <http://www.lclark.edu/loren/articles/fairuse.htm>
- Loren, Lydia Pallas. 2000. The Purpose of Copyright. *Open Spaces Quarterly* 2:1 Available at <http://open-spaces.com/article-v2n1-loren.php>
- Majerus, Laura A. 2003. Court Evaluates Meaning of “Derivative Work in an Open Source License Findlaw [http:// library.findlaw.com/2003/Jun/ 16/ 132811.html](http://library.findlaw.com/2003/Jun/16/132811.html)
- McCullagh, Declan. 2003. Perspective: The new jailbird jingle *CNET* January 27, 2003 [http://news.com.com/Perspective+The+new+jailbird+jingle/2010-1071 3-982121.html?tag=nl](http://news.com.com/Perspective+The+new+jailbird+jingle/2010-1071_3-982121.html?tag=nl)

- McCullagh, Declan. 2003. "Share 'True Crime,' do the time," *CNET* November 12, 2003. <http://news.com.com/Share+True+Crime+%2C+do+the+time/2100-10263-5106684.html?tag=st.rn>.
- McLaren, Carrie and Siva Vaidhyanathan. Copyrights and Copywrongs Interview with Siva Vaidhyanathan. By Carrie McLaren. *Stay Free!* Issue #20 <http://www.stayfreemagazine.org/archives/20/sivavaidhyanathan.html>
- McSherry, Corynne, 2001. *Who Owns Academic Work? Battling for Control of Intellectual Property*. Cambridge, Mass.: Harvard University Press. Milton S. Eisenhower Library and Pierpont Morgan Library CONDITIONS FOR USE OF THIS SITE <http://rose.mse.jhu.edu/pages/introframeset.htm>
- Minow Mary. 2004. How I Learned to Love FAIR USE Stanford University Libraries <http://fairuse.stanford.edu/commentaryandanalysis/200307minow.html>.
- Moglen, Eben. 2001. Enforcing the GNU GPL <http://www.gnu.org/philosophy/enforcing-gpl.html>.
- Negativland. Changing Copyright <http://negativland.com/changingcopyright.html>.
- Netanel, Neil Weinstock. "Impose a Noncommercial Use Levy to Allow FreeP2P File-Swapping and Remixing" (Nov. 2002). Unpublished manuscript, available at: <http://www.utexas.edu/law/faculty/nnetanel/Levieschapter.pdf>
- Nimmer, DAVID. 2003. "FAIREST OF THEM ALL" AND OTHER FAIRYTALES OF FAIR USE *Harvard Law Review*, 112(7), 1705-1722. *Law & Contemp. Probs.* 66:263 (Winter/Spring 2003) Available at [http://www.law.duke.edu/shell/cite.pl?66+Law+%26+Contemp.+Probs.+263+\(WinterSpring+2003\)](http://www.law.duke.edu/shell/cite.pl?66+Law+%26+Contemp.+Probs.+263+(WinterSpring+2003))
- Okerson, Ann. 1986. PERIODICAL PRICES: A HISTORY AND DISCUSSION Originally appeared in: *Advances in Serials Management* JAI Press Volume 1 (101-134) <http://www.library.yale.edu/~okerson/pricing.html>
- Okerson, Ann. 1991. With Feathers: Effects of Ownership on Scholarly Publishing <http://www.library.yale.edu/>
- Open Source Initiative 2005 The Approved Licenses. <http://www.opensource.org/licenses/index.php>
- Patry, William F. 2000. Copyright Law and Practice The Bureau of National Affairs, Inc. <http://digital-law-online.info/patry/patry1.html>
- Raymond, Eric. 1999. *The Cathedral and the Bazaar: Musings on Linux and Open Source by an Accidental Revolutionary*. Cambridge, Mass.: O'Reilly. <http://www.catb.org/esr/writings/cathedral-bazaar/>
- Reichman, Jerome H. and Paul F. Uhlir, 2001. "Promoting Public Good Uses of Scientific Data: A Contractually Reconstructed Commons for Science and Innovation," Proceedings of the Conference on the

- Public Domain, Duke Law School, 9–11 November 2001 Available at <http://www.law.duke.edu/pd/papers/ReichmanandUhlir.pdf>
- Richard Stallman 2002 Misinterpreting Copyright <http://www.gnu.org/philosophy/misinterpreting-copyright.html>
- Samuelson, Pamela. 1996. The Copyright Grab *Wired News* Issue 4.01 - Jan1996 Available at <http://www.wired.com/wired/archive/4.01/white.paper.pr.html>
- Schweidler, Christine and Sasha Costanza-Chock. 2005. Piracy vecam <http://www.vecam.org/edm/article.php3?id=93>
- Scott, Brendan 2001 Copyright in a Frictionless World: Toward a Rhetoric of Responsibility. *First Monday*, 6:9 <http://firstmonday.org/issues/issue6.9/scott/index.html>
- Söderberg, Johan 2002 Copyleft vs. Copyright: A Marxist critique by First Monday, volume 7, number 3 (March 2002), Available at <http://firstmonday.org/issues/issue7.3/soderberg/index.html>
- Sony Corp. of America vs. Universal City Studios, Inc.* Wikipedia [http://en.wikipedia.org/wiki/Sony\\_Corp\\_vs\\_Universal\\_City\\_Studios](http://en.wikipedia.org/wiki/Sony_Corp_vs_Universal_City_Studios)
- Stallman, Richard. 1985. The GNU Manifesto <http://www.gnu.org/gnu/manifesto.html>
- Stallman, Richard. 2004. Did You Say “Intellectual Property”? It’s a Seductive Mirage <http://www.gnu.org/philosophy/not-ipr.xhtml>
- Stedman, Kimberly. So what do I do about it? *Stay Free!* Issue #20 <http://www.stayfreemagazine.org/archives/20/take-action.html>
- Stim Richard. 2004. Getting Permission: How to License & Clear Copyrighted Materials Online & Off Nolo, Berkeley, CA 2004 Also see [http://fairuse.stanford.edu/Copyright\\_and\\_Fair\\_Use\\_Overview/index.html](http://fairuse.stanford.edu/Copyright_and_Fair_Use_Overview/index.html)
- Tallmo, Karl-Erik. 2005. The History of Copyright: A Critical Overview With Source Texts in Five Languages work in progress <http://www.copyrighthistory.com/index.html>.
- The Oyez Project. 1984. Oyez: *Sony Corp vs. Universal City Studios*, 464 U.S.417 (1984), available at: <http://www.oyez.org/oyez/resource/case/768>.
- The UK Patent Office. 2000. A history of copyright The UK Patent Office <http://www.intellectual-property.gov.uk/std/resources/copyright/history.htm>.
- Thomas, Ritchie T. et al. 1993. \*AMICUS CURIAE\* BRIEF OF ASSOCIATION OF RESEARCH LIBRARIES, AMERICAN ASSOCIATION OF LAW LIBRARIES, SPECIAL LIBRARIES ASSOCIATION, MEDICAL LIBRARY ASSOCIATION, AMERICAN COUNCIL OF LEARNED SOCIETIES, NATIONAL HUMANITIES ALLIANCE, AND ASSOCIATION OF ACADEMIC HEALTH SCIENCES LIBRARY DIRECTORS Available at

<http://fairuse.stanford.edu/primarymaterials/cases/texaco/joint.amicus.html>.

Three Views of Peer Review, Steven G. Krantz, Greg Kuperberg, Alf van derPoorten, *Notices AMS*, June/July 2003.

Tufts University Perseus Copyrights <http://www.perseus.tufts.edu/copyright.html>.

UNITED STATES COURT OF APPEALS FOR THE NINTH CIRCUIT.  
2004. *MGM Studios, Inc. vs. Grokster, Ltd.*, 259 F. Supp. 2d 1029 (C.D. Cal. 2003), appeal argued, No. 03-55894 (9[th] Cir. Feb. 3, 2004). Available at <http://www.techlawadvisor.com/docs/mgmgrokster.html>.

United States Department of Justice 1998 The No Electronic Theft (“NET”) Act Relevant portions of 17 U.S.C. and 18 U.S.C. as amended (redlined) <http://www.usdoj.gov/criminal/cybercrime/17-18red.htm>

United States District Court for the Southern District of New York 1991 *BasicBooks, Inc. vs. Kinko’s Graphics Corporation* 758 F. Supp. 1522 Available at <http://fairuse.stanford.edu/primarymaterials/cases/c758FSupp1522.html>.

United States District Court for the Western District of New York 1983 *Encyclopaedia Britannica vs. Crooks* 58 F. Supp. 1247 Available at <http://fairuse.stanford.edu/primarymaterials/cases/c558FSupp1247.html>.

*United States Supreme Court 1984 Sony Corp vs. Universal City Studios* 464.

U.S. 417 Docket Number: 81-1687.

U.S. Constitution Article I, Section 8, Clause 8.

U.S. Copyright Office Report On Legal Protection For Databases U.S. COPYRIGHTOFFICE August 1997 <http://www.copyright.gov/reports/dbase.html>.

U.S. Department of Justice 1999 Defendant Sentenced for First Criminal Copyright Conviction Under the “No Electronic Theft” (NET) Act for Unlawful Distribution of Software on the Internet November 23, 1999 <http://www.cybercrime.gov/levy2rls.htm>.

U.S. Department of Justice 2001 Man Sentenced in Michigan for Offering Software Programs for Free Downloading on “Hacker Hurricane” Web site <http://www.cybercrime.gov/baltutatsent.htm>.

U.S. Department of Justice 2001 Nine Indicted in Chicago in \$1 Million “Fastlane” *Software Piracy Conspiracy* Feb. 16, 2001 <http://www.cybercrime.gov/fastlane.htm>.

U.S. Department Of Justice 2005 Federal Law Enforcement Announces Operation D-Elite, Crackdown on P2P Piracy Network: May 25, 2005. <http://www.cybercrime.gov/BitTorrent.htm>

- U.S. Department of Justice Virginia Man Sentenced for Violation of the “No Electronic Theft” (NET) Act for Unlawful Distribution of Software on the Internet March 3, 2000 <http://www.cybercrime.gov/thorntonSent.htm>
- U.S. Supreme Court. 1991. *FEIST PUBLICATIONS, INC. vs. RURAL TEL.SERVICE CO.*, 499 U.S. 340 (1991)
- Vaidhyathan, Siva. 2001. *Copyrights and Copywrongs: The Rise of Intellectual Property and How It Threatens Creativity*. New York: New York University Press
- Vaidhyathan, Siva, 2002. “Copyright as Cudgel,” *Chronicle of Higher Education* (2 August), B7–B9.
- Vaknin, Sam. 2002. *The Second Gutenberg United Press International* Available at <http://www.planetebook.com/mainpage.asp?webpageid=376>
- von Lohmann, Fred. 2005. 1984 COURT RULING LAUNCHED A TECHNOLOGICAL REVOLUTION *Mercury News*. Available at [http://www.eff.org/IP/P2P/MGM v Grokster/betamax 20th.php](http://www.eff.org/IP/P2P/MGM_v_Grokster/betamax_20th.php)
- Weisstein, Eric W. 2001 Author’s Detailed Commentary [http://mathworld.wolfram.com/about/erics\\_commentary.html](http://mathworld.wolfram.com/about/erics_commentary.html)
- Willinsky, John. 2002. “Copyright Contradictions in Scholarly Publishing,” *First Monday*, 7:11 (November 2002), URL: <http://firstmonday.org/issues/issue711/willinsky/index.html>.
- Willinsky, John, 2002. “Scholarly Associations and the Economic Viability of Open Access Publishing,” unpublished paper, Vancouver: University of British Columbia, at <http://research2.csci.educ.ubc.ca/workingpapers/association.pdf>

# How Free Culture Will Save Digital Libraries

Aaron Krowne (Emory University)

**Abstract:** Today, we are watching as web search engines—especially Google—push libraries further down the service-provision hierarchy, towards roles as anonymous information-silos, and therefore diminished importance. The production of free culture in a digital library context will reverse this trend, making digital libraries themselves compelling, useful, intentionally-visited virtual places, and ensuring the continuing importance of libraries in the future. We will see that this outcome largely follows from digital libraries doing what we already expect libraries to do—but translated into a digital context, where forces of commons-based peer production (CBPP) operate in a milieu of free culture.

## INTRODUCTION


A few years ago, as a graduate student mining my advisor’s lengthy CV for interesting publications, I came across one called “How Digital Libraries Will Save Civilization” (Fox *et al.* 1998). While I had access to nothing more than the citation string of this article until recently, the title struck me as embodying a wonderful sort of sunny optimism, which—being someone working in the field of digital libraries—I found very motivating. And still do.

Nothing in the Fox article attempts to rigorously substantiate the claim implicit in the title. However, the ideas in the paper, when projected to their greatest-possible impact, could realistically attain this end. The possibilities have tantalized technologists and dreamers as far back as Vannevar Bush (Bush 1945).

The current paper is written in the spirit of that article—and as sort of a prelude to it. While I won’t attempt to seriously substantiate the implicit claim of the title here, I hope that my optimism for free culture and commons-based peer production (CBPP) will similarly inspire others. And perhaps the outcome will be that the digital library will persevere as a socially-significant institution, and

---

M. Halbert (Ed.): *Free Culture and the Digital Library Symposium Proceedings*. Atlanta, Georgia: MetaScholar Initiative at Emory University, 2005. pp. 300-324.

 This work is covered by the Creative Commons Attribution-NonCommercial-NoDerivs 2.5 License.



maybe—through bringing people together to build our common culture of knowledge—actually save civilization.

### **A Note on “Digital Libraries”**

It will be useful for this paper to establish what, exactly, I mean when I say “digital libraries” (DLs). In the present context, I am not speaking about systems which have basically all the functions and functionality of what are formally defined to be digital libraries (*e.g.*, with the 5S framework, introduced in Gonçalves et al., 2004). Indeed, under such a definition, the current incarnation of Google even qualifies as a digital library—as well perhaps it should. Instead, however, I mean systems that self-consciously call themselves “digital libraries,” and which are likely to be supported in a top-down fashion, from large, official, often national funding organizations, or as outgrowths of physical libraries.

These systems, while inarguably important, tend to be governed by different forces than commercial or grassroots public projects. And it is largely to them I direct this lecture, out of a concern for seeing them maximize their potential, and deliver the most value to the general public—from which their support and lifeblood ultimately derives.

I should also point out here that this article will to a great extent conflate classical libraries and digital libraries. This is largely unavoidable for two reasons: the first being that many of the trends and principles discussed here affect both kinds of libraries, and the second being that classical libraries are increasingly *becoming* digital libraries. The latter is evidenced by a number of trends, including web-based OPACs that basically behave like digital library search interfaces, and an increasing number of all-electronic holdings.

The reader should exercise common sense as to which kind of library I mean in particular passages, whether or not cited unambiguously.

### **THE “THREAT” TO DIGITAL LIBRARIES**

In recent years, the chagrin and frustration of librarians has been raised by web search engine Google (*e.g.*, Gorman 2004). The problem is that library patrons are using web search engines, especially Google, as the primary means of meeting their research needs (Lippincott and Kyriallidou 2004). The reality is that almost every research undertaking starts at Google—whether or not it

ends up at a digital library record or a physical library. Librarians want patrons to use their electronic library catalog search systems, their specialized domain database search interfaces, and their digital library search and browse services. But patrons are staying away in droves.

To add insult to injury, if a search starts on the internet, there's a good chance the research task will end there, completely bypassing the library or digital library. But even when the research task continues on to some electronic library interface, it is usually because it is *mostly done*. The user knows what "record" or records they want; the role of the library is simply to "serve it up." This is a significantly diminished role for libraries, compared to their historical purpose and present aspirations.

A major part of the explanation for this is that web search engines have essentially become universal metasearch engines. These search engines are not limited by mere "kinds" of records; if it's on the web and the typical web surfer can see it, the search engine can get them to it. Google has provided the template for how this is to be done "right:" a single search text box, in which the user can enter keywords (even natural language is safe to use), with a minimalistic interface design, free for all to use, supported by low-key text advertisements, and providing instantaneous response. By contrast, library search engines are still all-too-often fragmented over disparate databases, utilize legacy, Boolean-based OPAC query syntax, have cluttered and unclear interfaces, and are slow to boot.

## **RESPONSES FROM THE DIGITAL LIBRARY WORLD**

### **The Technical Approach**

Libraries and DLs are beginning to respond with their own metasearch solutions. But these may be a day late and a dollar short. From the user's perspective, it is difficult to determine why they should go first to their local library's metasearch interface (possibly one of many local libraries) when they can simply "google it"—and probably find all the information they could possibly need.

Seeming to realize this reality, in the past few years, the digital library community has been working to find ways to make DLs important and useful to end users. Digital library researchers have even embraced Google as a way to get users in "through the back door," as services such as DP9<sup>1</sup> illustrate (Liu *et al.* 2002). But

most effort has been in digital library architecture— in essence coming up with reasons for users to show up (or at least, return) through “the front door.” This effort has been focused on two broad technical categories: (1) exploiting domain specificity, (2) and metadata-based services.<sup>2</sup>

Domain specificity refers to building digital libraries *for X*, where X is some subject. Metadata-based services refers to providing ways of retrieving and organizing information that web search engines can't, because the digital library has access to (and “understands”) richer metadata. These two categories are not unrelated, *e.g.*, often most of the advantage in building domain-specific digital libraries is in the fact that such a library can support metadata elements which are unique to the domain. But this need not be the case.

Much progress has been made along both of these technical fronts.<sup>3</sup> Surely they are a part of the solution. However, it is unclear that these efforts alone are paying off in significant “front door” use of digital libraries.

One problem is that digital library researchers and developers have an incredibly tough task: they must make these value-added technical services so compelling, so convenient, and so easy to use, that they will be as attractive and high-priority to users as Google. But there is an inherent conflict here: there are many digital libraries, yet as a universal meta-search system, Google precisely is a machine for avoiding the repetition of searches through many interfaces. The problem is not just convenience, but the “economics” of attention and efficiency.

A second problem is that the technical route to shoring up digital libraries implicitly expects to “out-innovate” Google in the areas of information retrieval and information integration in general. One could call this a technical “arms race” with Google—but Google has such a war-chest of resources and talent that it doesn't actually seem to be facing any serious challenge (from the most established of tech companies, let alone digital libraries—*e.g.*, Vogelstein 2005).

Indeed, Google gets mentioned by name in this paper, by librarians, and even in vernacular verbiage, because it cannot be treated the same as “other” internet search services. Google says, ever ambitious, that its mission is “to organize the world's information,” and it has proven it is very serious about this goal. The problem is that this is basically the mission of libraries too, so

almost every step of progress Google makes toward this goal seems to shrink the universe of importance of libraries (*especially* DLs).

For example:

- The launch of Google Scholar<sup>4</sup> in the past few months shows that Google is acting more like a universal metasearch service than any other entity. The function of this service is to provide searching of scholarly materials—from library and digital library catalogs—through Google’s interface. If the scholarly materials don’t happen to be available for free online, then the user, properly authenticated, can access it through a library interface. Hence, the library’s role here: filling in the blanks at the final step of the research process.
- The library world is also watching in slow motion as Google undertakes its audacious book digitization effort. This venture demonstrates the power of Google’s clout—not only can they fund such a massive effort, but they are completely unafraid of the inevitable copyright headaches. Solutions will probably have to be found purely to accommodate Google.

There are other reasons to have pause about pursuing technical “competition” with Google. Google Local and some binding with the Open Directory web categorization hierarchy shows that the company can indeed provide domain-specific information services when it wants—even over a generalized web search. Binding with category hierarchies could be even tighter, with minor additional engineering. Link analysis could be used to automatically discover and guide searchers to subject domains (Flake *et al.* 2000, 2002; Gibson *et al.* 1998; Reddy and Kitsuregawa 2001). And the nascent treatment of metadata, at which Google Scholar hints, could be extended.

Libraries and digital libraries can of course always rest assured that they will be required elements of the research process, inasmuch as patrons require access to copyrighted items which are distribution-restricted. However, this may have to be in the most pedestrian of roles: doing nothing but delivering records, once found. In fact, libraries may face an “identity-stripping,” whereby tools meant (innocently) to interface web search with digital libraries have the unintended result of almost completely abstracting away and blackboxing the provider of the end artifact (*i.e.*, the library or

digital library). This is an inevitable consequence of tools which make research easier for the user, which we can already see prototyped by the WAG localizer (Singer 2005).<sup>5</sup>

### **The Closed-Access Approach**

Even if the above is not considered an issue, relying on proprietary digital content as the *raison d'être* of digital libraries still seems to be a risky bet. As the Sabo “Public Access to Science Act,” H.R. 2613 indicates,<sup>6</sup> the tide of public sentiment is turning against the present practice of letting remain generally-unavailable scholarly content which is produced using public monies. And research libraries rely heavily on this content—mostly expensive paper and electronic journals—for patronage. While H.R. 2613 seems stalled, what happens when a bill in the same spirit finally succeeds, and publicly-funded research is unshackled from restricted distribution? When one considers that there is very little private scholarly research left, and on top of this the fact that this kind of law would necessarily force blanket open access policies on most journals, it is difficult to see which scholarly content could legally remain closed.

There are other problems as well. As pointed out in (Regazzi 2004),

The early 1970s was a time when, for the most part, research libraries could buy all new research material, thus keeping up with virtually all R&D developments. But for the 20-year period from 1975 to 1995, university library expenditures increased only at the rate of 2.2%, while research and development spending increased by 4.6%, nearly double that of the library. The result is a huge gap in the university library's ability to keep up with the production of research and development.

Exacerbating this, as also reported in (Regazzi 2004), library expenditures as a measure of total university spending have decreased from 3.7 percent to 2.8 percent annually since 1982. Thus, there is more to buy, and less money being allocated to buy it. This has resulted in a situation where it is becoming less likely an individual will be able to access a given journal article, because their member institution probably doesn't have a subscription to it (or perhaps not the right subscription).<sup>7</sup>

Thus, it seems closed-access is unwanted, inefficient, and a poor way for libraries and DLs to make a meaningful identity for themselves. Table 1 speaks to these points: collaborative (CBPP) digital libraries are compared with open access but strictly-controlled digital libraries, as well as one completely closed one. According to this data, it appears that closed libraries such as ACM

DL are being seriously challenged by open, collaborative ones, such as CiteSeer.

Perhaps more important than the above reasons to avoid putting all of the DL “eggs” in the proprietary-content “basket,” I want to suggest that relying on these materials to give DLs a purpose runs completely counter to the philosophy of libraries. This philosophy is one of disseminating knowledge as widely as possible and furthering scholarly activities. I would like to put forth here that the way out of this narrow and possibly terminal future for digital libraries is to embrace free culture, and in doing so, embrace their true calling.

### **IN FREE CULTURE, A BETTER SOLUTION**

That a fuzzy, social concept like “culture” is part of the solution I am proposing signals a radical departure from the extant, technical attempts at solutions outlined in part above. The cue that culture is a part of the solution comes from perhaps the least likely of places one would expect to find inspiration for surviving in the digital age: classical, brick-and-mortar libraries.

The key fact about classical libraries is that they are not seen as, or used as, information retrieval *machines*. They are seen as social and cultural places. People go there not only to retrieve information in the form of books, but to study it, to conduct work derived from the knowledge these books contain, to discuss with others the ramifications of what they are reading and researching, or to interact with the library staff to help give direction to their research activities. In short, they go to act in a scholastic way, in a social context, with peers and experts.

This notion has been all but lost in digital libraries—or at least, systems that self-consciously call themselves digital libraries (as distinguished above). I will argue in this paper that replacing this “lost” social notion is the key to adding compelling value to digital libraries, and sustaining them as a useful, meaningful institution.

As someone who works in a library, I routinely observe that people will come to physical libraries to act studiously and scholastically, despite alternatives that let them stay at home—because of the special and social nature of the space. This continuing fact is reported in (Lippincott and Kyriallidou 2004). In fact, for all intents and purposes, Starbucks and Borders benefit from the same phenomenon. The practice of providing a social atmosphere for

intellectual activities seems to be alive and well, and if anything, growing.

The key, then, to “saving” digital libraries is to similarly re-establish a notion of a social place—within the context of the digital library. In essence, this allows the patron to undertake intellectual, cultural activities, resulting in the actual creation of culture.

### **Free Culture, CBPP, and Digital Libraries**

For this article, I define free culture as the social milieu of information artifacts which may be disseminated and modified without permission (*libre* free), for which there is also zero structural monetary cost to do so (*gratis* free).<sup>8</sup> Note that open access is necessary but not sufficient for free culture; it provides the *gratis*, but may lack the *libre* component.

The tie-in of free culture to digital libraries as social, culture-producing places is for two main reasons. The first is that, in the digital context, all interaction is potentially subject to copyright restrictions. That is, every communication is an artifact created, and the dissemination of such cultural communications by the digital library is generally reduced or ruled-out entirely under the current, permission-default copyright regime (Lessig 2004). I will talk more about this in a later section.

The second reason, which is the focus of this paper, is that free culture both enables and is produced by commons-based peer production (CBPP). CBPP is the name given to the distinct mode of production of intellectual artifacts which has emerged in the past few years, enabled by the internet and the appropriate software layer on top of it. The GNU/Linux operating system and Wikipedia are prominent examples of this phenomenon. CBPP is considered a mode of production as distinct, and perhaps important, as markets or firms. For more on CBPP see (Benkler 2002) and (Iannacci and Mitleton-Kelly 2005).<sup>9</sup>

Making intellectual artifacts free (in both the *libre* and *gratis* senses) enables CBPP, which then produces more free intellectual artifacts. The upshot of this feedback loop is a powerful economics of intellectual production, which we will explore in detail later. A simple illustration of this feedback loop is shown in Figure 1.

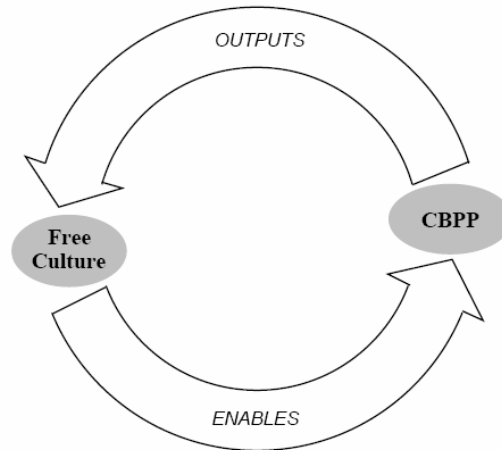


Figure 1: The relationship between free culture and commons-based peer production (CBPP).

In Figure 2, the “pillars” which support free culture are shown. These pillars are not identical with CBPP—they are “made out of” not only the technical elements of by CBPP systems, but also social protocols within and above these systems, and copyright law and licensing which permit the necessary productive activities to take place. This figure also shows how production of free culture progresses “through” the pillars, then feeds back on itself (as in Figure 1).

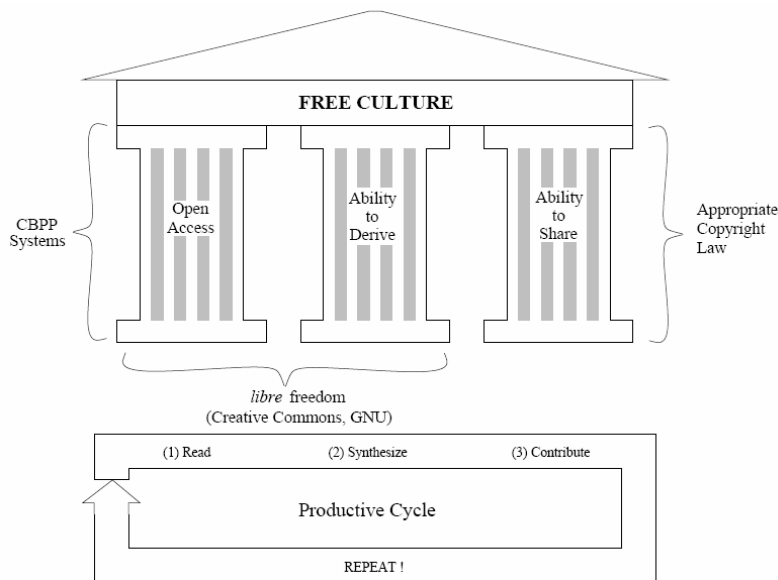


Figure 2: The “pillars” which support free culture. Each pillar is enabled both by CBPP systems and social conventions, as well as the proper copyright and license law.



I contend that free culture is so compelling that users will be drawn to virtual places that allow them to work with it, use it, manipulate it, and bring it into the context of their lives and their interests. By integrating CBPP services, digital libraries can become “engines” of free culture, crossing a threshold where their *services* and their *communities* become more important than their collections. Such a transformation undermines the deleterious effects (to DLs) of the recent Google-fueled trend toward commoditization of collections I described earlier.

### **Exemplary Information Systems**

In this section I give a brief tour of digital library and information systems which integrate CBPP, either completely or to a large extent. The intent here is to demonstrate that CBPP can create or enrich this type of system, making them widely-used, engaging, useful, and sustainable.

- *Wikipedia* - <http://en.wikipedia.org/> - Wikipedia is a collaborative, general-information encyclopedia. It is made up of articles which are created and improved by volunteers (who can be anyone on the internet—a login is not even required to make edits). The English edition of Wikipedia, which is the largest, has over 600,000 articles. In essence, Wikipedia “scales up” the encyclopedia concept, distributing the production work, adding hyperlinking, removing limitations on topical coverage, and adding the typical benefits of digitization and internet-accessibility.
- *ArXiv* - <http://www.arxiv.org/> - The ArXiv was started at the beginning of the 90s by Paul Ginsparg, while he was at Los Alamos National Lab, as a pre-print server for physics research. This service allowed physicists to share each other’s work before journal publication, allowing more extensive feedback to be received earlier, bypassing the lengthy delays in the publication process (sometimes measured in years), and avoiding the suppression of legitimate research.<sup>10</sup> Today, ArXiv (and Ginsparg) are at Cornell University, and the subject coverage of the service has spread into mathematics, computer science, and areas of chemistry and biology. In physics, the service is almost ubiquitous, with nearly every journal

article (and then some) having an arxiv.org incarnation. Articles are often simply cited by their ArXiv URL.

- *PlanetMath* - <http://planetmath.org/> - PlanetMath is a mathematics community featuring a collaborative mathematics “encyclopedia” contributed by volunteers. It was started by myself and Nathan Egge, shortly before I began grad school (in 2001).<sup>11</sup> The concept is similar to Wikipedia, except that mathematics is taken as the focus, the authoring language is LaTeX (the de facto standard among the mathematics community), linking is automated, and a more academic authorship model is utilized (accounts are needed to edit and articles have owners who exert a high level of creative control over them).
- *Slashdot* - <http://slashdot.org/> - Slashdot is a news web log (or “blog”)—in fact one of the first (dating back to 1996). It chiefly serves the “geek” community—those interested in science and technology. News stories are submitted by the general public, vetted by editorial staff, and posted. However, the real “magic” begins after this point, as each article is also a discussion area. Hundreds or thousands of people comment on the typical article, and comments are scored and filtered collaboratively with Slashdot’s “karma” system. The result is that funny and insightful comments “float” to the top. Slashdot in essence filters news stories through the “hive mind,” often leading to surprisingly original and penetrating analyses the mainstream media tends to miss.
- *CiteSeer* - <http://citeseer.ist.psu.edu/> - CiteSeer is an autonomous scientific literature digital library. It is built through web crawlers (or “spiders”) which traverse the web looking for documents likely to be research papers (*i.e.* PDFs, PostScript files, etc.). Those that pass some basic machine learning filters are given metadata (also through machine learning—see Han *et al.* 2003), posted, and interlinked with the rest of the collection. Although this digital library is automated, it is collaborative in the sense that contributors “post” to it simply by posting their research on the web (for instance, on their home pages, or at university technical reports archives). Further, users can suggest new locations for the system to crawl (and in essence, manually add papers), rate articles, and correct

metadata. However, the most popular and effective services by far are the central acquisition, listing, and integration services.

How well are these largely-collaborative information systems doing? Some examples follow. As pointed out in Table 1, CiteSeer is more popular than its closed counterpart, ACM DL, at least in the lens of PageRank. ArXiv.org is just as popular as APS's site, which hosts the massive complex of journals run by the American Physical Society. Slashdot is such an 800-pound gorilla in the "geek-o-sphere" that its news posts routinely bring the servers hosting the linked stories to a halt. There is even a verb for this—it is called getting "slashdotted." Additionally, Wikipedia is clearly incredibly popular by a number of metrics, besides PageRank, as it has an Alexa ranking of 65, and has the distinction of being highlighted separately in Google searches and within Amazon's A9. There have also been convincing (or at least thoughtprovoking) benchmark studies and qualitative arguments that Wikipedia is the "best" encyclopedia out there—and it took only four years to build.<sup>12</sup>

	CBPP DL	Non-CBPP DL	
CiteSeer	8	7	ACM DL <sup>a</sup>
Wikipedia <sup>b</sup>	9	9	Britannica
PlanetMath	6	8	MathWorld <sup>c</sup>
ArXiv.org	9	9	APS

<sup>a</sup>The PageRank of the IEEE Computer Society's DL was unavailable, possibly because of how their web site redirects and mangles URLs.

<sup>b</sup>The english edition of Wikipedia, at en.wikipedia.org, was used.

<sup>c</sup>For perspective, note that MathWorld is about three times the size of PlanetMath (in terms of content), and predates it by at least a decade.

Table 1: Interesting PageRanks: CBPP DLs overall score impressively in "PageRank" (Google's derived metric of "importance" on the internet, see Brin and Page (1998); Page et al. (1998)), generally at least matching their Non-CBPP counterparts. Perhaps tellingly, the only fully closed-access DL on the list, ACM DL, was the only non-CBPP DL to achieve a PageRank strictly lower than its CBPP counterpart. Underscoring this, CiteSeer's Alexa (<http://www.alexa.com/>) traffic rank can be found to be 1,378, versus ACM DL's 8,591 (higher is more-visited).

## Research Results

Some encouraging formal research results are beginning to be reported, which suggest that making digital libraries more

participatory fosters pedagogy, engagement, uptake, utility, and so forth. This goes a significant ways towards studying what happens in a collaborative, free culture information environment. A few of such reports follow:

- In (Zhang and Quintana 2005), a system called IdeaKeeper is evaluated. This system gives students a structured way to analyze digital library learning objects. For example, they can give feedback about whether a viewed item was related to their question, enter in the main idea of the presentation, list the supporting evidence, give feedback about bias and expertise, highlight the specific information that answers the initial question, and so forth. The research shows that uptake was very good and that there were significant positive effects on learning among students who used this tool over students who just used the learning objects without IdeaKeeper.
- In (Brusilovsky *et al.* 2005), the Knowledge Sea system is evaluated. This system focuses on social navigation, whereby users of the digital library give feedback (both implicit and explicit) which is used to facilitate the discovery process of others. Utilizing novel visualization techniques, the system provides cues for which resources are most popular, active, and of the highest quality. Annotations, discussions, and ratings are supported and help provide the basis for the corresponding visualized indicators. The system's efficacy was tested with students, and it was found that there was a significant positive correlation between resources indicated as of high quality or of interest and the resources which were most utilized.
- In (Milson and Krowne 2005) in this volume, we show that CBPP systems can be compatible with the formal education setting, while still yielding the extra benefit of creating useful learning objects. This was done with a small pilot study, whereby the Noösphere system was deployed for classroom support for a small mathematics graduate class. Students were loosely given assignments to produce mathematical articles, and were given credit for activity within the system. From the articles which were written by the students, a set of collaborative course notes were compiled. By conventional evaluation metrics, the performance of the students was found not to

have been diminished, and they were additionally exposed to an aspect of scholarly work they otherwise would not have encountered.

- In (Efron and Sizemore 2003), a pilot study of iBiblio is done. iBiblio is a large, public-access, collaborative archival digital library.<sup>13</sup> In the study, the authors ask the question “do increased contributor efforts lead to increased collection popularity?” They find a strong answer in the affirmative: the more collaborative maintenance activity there is in a collection, the more popular it becomes.

Results like these should be no surprise, and I believe we will be hearing more like them. The emerging pattern seems to be that free culture through CBPP is beneficial in two ways: (1) for certain types of individuals (more in some settings than others), it turns them from passive consumers of information to engaged constructors of the shared knowledge environment, and (2) everyone benefits from the distribution of teaching and sharing away from small, central, often intellectually-homogeneous knowledge “oligarchies” to anyone who has the ability, motivation, and expertise to teach, help, and share.<sup>14</sup>

Much of the work towards these ends comes under the rubric of “personalization” research in the digital library research world. I would encourage the continuation of this thread of work, but would also add that researchers should consider enabling the sharing of the effects of personalization services whenever possible (if not making sharing the default). This is because each individual’s “sense-making” of the information in the library produces valuable secondary information which will likely be of use to many other users of the library. Sequences and groups, annotations, ratings, categorization, or even views and activation all provide generally useful information about the collection, even though the individual user may think of them primarily as means for customizing it.

### **How DLs Can Support Free Culture**

The descriptions of the exemplary systems and published research above suggest a little bit about how CBPP can be employed by digital libraries. Some types of collaborative services that can be employed are:

- annotations/discussions;

- list-making and categorizing (forming associations);
- ratings (for collaborative filtering/recommenders);
- reputations services;
- reviews and moderation (i.e. include/exclude judgments);
- content authoring/creation;
- correction/enhancement.

DLs need not be *entirely* based on CBPP in order to provide compelling free culture value—they can pick and choose from the above menu of services (and surely beyond) to determine their overall constitution of collaborative, automated, and controlled labor.

It is worth briefly mentioning a few more CBPP projects which illustrate creative combinations of the above collaborative services. For example, the Distributed Proofreaders<sup>15</sup> project implements collaborative correction. Systems like Furl,<sup>16</sup> CiteULike,<sup>17</sup> and Delicio.us<sup>18</sup> implement list-making, categorizing, and annotations.<sup>19</sup> Amazon.com implements list-making, reviews, and ratings. eBay implements ratings and reputations services. Observe that all of these sites are very influential and successful, or at least are “upstarts” making a large splash. In fact, it was by integrating the latent social information of hyperlinks that Google leapfrogged the competition—in effect becoming a collaborative filtering service by exploiting millions of “endorsement” judgments on the web.

There is almost certainly a niche for DLs which have a “static” content base (either “frozen” due to being historical or slow-changing due to being centrally-vetted) yet which include a high-impact CBPP component. Figure 3 shows how CBPP services can still be layered on top of such a content base, being bound to the underlying artifacts via identifiers and standard internet interface conventions such as links and frames.

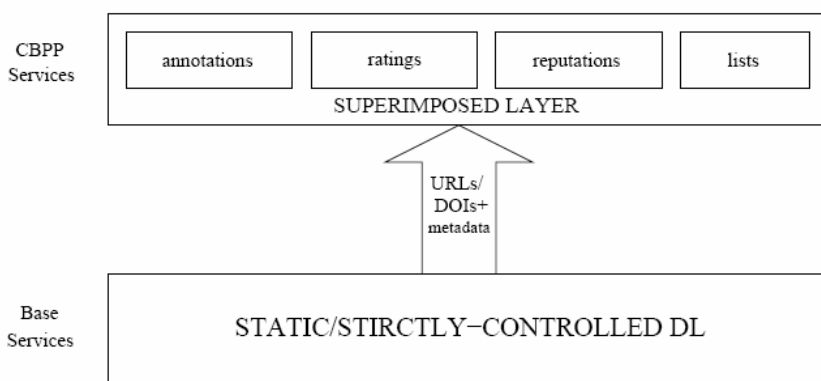


Figure 3: Superimposed CBPP: How digital libraries can employ CBPP without becoming completely collaborative. An underlying, central content base is “linked” to services and collaboratively-produced data at a superimposed layer. The example superimposed CBPP services given are annotations, ratings, reputations (systems), and user-created lists (which group resources in the DL).

## ROADBLOCKS AND CHALLENGES

### Perceptions of CBPP

There is an immense amount of consternation in certain quarters about the prospect of CBPP becoming a major force in the production of our information landscape (Biss 2004; Gorman 2005; McHenry 2004). This, I think, is ultimately rooted in fear.

Popular culture in the west for the past century has not traditionally been free—nearly all aspects of it have been strictly controlled by corporations. Yet we now have a digital, networked landscape, where regular individuals suddenly have the authoring and publishing power once reserved only to powerful entities with extraordinary resources. This infrastructure for free culture—the internet—is now setting the stage for a struggle for realization of free culture, between the people and the powerful (Vaidhyathan 2004).

The prospect of the “little guy” having an impact on culture is considered by some to be a scary thing. There are worries of confusion, information overload, inaccuracy, fraud, degeneracy, and vandalism. Yet, the systems toured in this article strongly suggest that these concerns are not terminal, if even significant.

In sum, opposition to CBPP and free culture is largely illegitimate: it is the result of fear of the new and unknown and/or an intuitive understanding from vested interests that they are being threatened.

But the reality is that markets and firms are simply not enough to produce a vibrant, diverse, healthy culture, in a world where firms dominate the individual and the commons. This kind of world is an alien thing to the unchanging, social nature of humankind. A free culture which includes widespread CBPP systems offers a more natural alternative.

### **The Copyright Situation**

As discussed earlier, for free culture to work through CBPP, people must be free to access, free to extend, and free to contribute. At any one of these steps, there commonly exist in the present day numerous, often insurmountable legal pitfalls.<sup>20</sup> A rough (and probably incomplete) taxonomy of scenarios for these copyright pitfalls is:

- *Orphaned works* - Archives hold reams and volumes of digitized works, to which the copyright holder is unknown or unavailable. Due to how copyright terms have been extended, and how copyright defaults to “all rights reserved,” these works are therefore unavailable for sharing or making derived works.
- *Lock-up of significant cultural and knowledge works* - The well-known examples here are “free the mouse” (Economist, 2002) (i.e., how Disney has for three-quarters of a century retained iron-clad control over how their creations are used, despite their status as popular culture) and scholarly journals (which are typically closed to the general public).
- *No protection for CBPP efforts* - The “blessing” of distributed authorship of CBPP projects turns into a curse when current copyright law is applied. As the number of contributors increases (something which is in fact a reasonable and desirable goal), the probability of one of them intentionally or unintentionally causing the entire project to run afoul of copyright law approaches near-certainty. While I believe this is a small problem logistically, it is a huge problem legally.
- *Elimination of analogous fair use in the digital world* - Since every operation in the digital world is a “copy,” copyright technically forbids almost all operations. Thus, it is generally not permitted to “lend” an e-book to a friend, or to send to them some of the music on your iPod.



Those who control information have exploited this unforeseen technicality to apply legal restrictions to what was once fair use, in order to ensure their own profit.

- *Re-capture of public domain works* - Related to the previous item, many information providers (even digital libraries) assert copyright which they have dubious claim to, over public domain materials which they carry in some form. This has been called “thin copyright,” and it is unclear whether it is really allowed (Puzio 2005).
- *Criminalization of “circumvention” (technicalization of copyright)* - As Vaidhyanathan and Lessig point out (Lessig, 2004; Vaidhyanathan, 2004), the DMCA has radically altered the nature of copyright by criminalizing any activity which circumvents copy protection technology (no matter how flimsy this technology is). Since this is now a criminal offense separate from normal copyright violation, and as aforementioned, every operation is a copy, the government has in effect outsourced thinking with respect to copyright to private companies, while giving them unlimited access to government power of reprisal.
- *Stillborn works* - As I realized when listening to Apple’s Bud Tribble deliver a speech (Tribble 2005), some works are locked out of usage upon creation, due to the innate attributes of the author or the circumstances surrounding creation. As Tribble pointed out in his anecdote, kids who created valuable multimedia learning resources as a part of Apple’s school outreach programs were unable to share these works— because children are not “authorized” to assign usage license to their own creations. Consequently, one free educational digital library that could have existed did not.
- *Buried works* - Works that were once widely available for free (*i.e.*, through libraries) often become more scarce after they are digitized (you read that correctly). Once paper copies of old works are tossed or sent to long-term storage, researchers must rely more heavily on their digitized representations. However, as discussed earlier, the odds of having a subscription to access these locked-down works can be quite poor (and worsening). For example, during the writing of this paper, a friend of mine

doing his PhD work approached me for help finding a classic mathematics article from the 70s. The work was digitized and from a prominent journal, but he had been unable to access it through any of the six institutional libraries he tried!<sup>21</sup>

Most of these scenarios (except perhaps the last two) are well-covered in the rest of this volume. However, within this symposium and elsewhere, they have generally been seen as isolated copyright problems. This is understandable, given how we are all most familiar with the situations we've encountered, but it is not the case—they are all simply a result of the same underlying, outdated, flawed model of intellectual property, and a copyright regime built upon it. Thus, they should be properly taken as scenarios. Efforts to address them can then be focused on the common (and dysfunctional) underlying rules, which establish under what conditions permission (to disseminate or derive) is needed.<sup>22</sup>

It is my hope that the confluence of minds and visibility of issues afforded by this symposium will accelerate progress in “solving copyright,” thus making all of the above “nightmare scenarios” go away.

### **Technical Challenges**

Impediments to CBPP are not all perceptions and abstract legal conditions. There actually are technical challenges to making CBPP work—and critics latch onto these challenges as if they disprove the utility of the entire mode of production.

Previously, “information oligarchies” were naturally induced because the technical challenges of authoring (especially authoring *en masse*) and publishing were intractable otherwise (Vaidhyathan 2004). Now, we have powerful computer processing network architectures which enable solutions of most of these fundamental problems. So far, we've witnessed the birth of CBPP as a consequence of these advances. But more could be done to improve CBPP systems and increase their applicability, especially in the area of quality control.

Quality control, in fact, seems to be the last safe harbor for the CBPP naysayers. Even when CBPP resources (such as Wikipedia) are quite clearly high-quality, the uncertainty of how this was achieved seems to make the fact invisible to some. The “sleight of hand” is that quality in such systems is emergent, either through

collaborative filtering, voting, or the sheer domination of good contributions over bad (Krowne 2005). Making these emergent systems work is much more difficult than simply giving some authority the power to include or exclude portions and to give explicit indicators of quality. Further, explaining how this process happens in CBPP is more complicated than “because the editors (or moderators) say it is so.”

While real, I propose that quality control in CBPP is now more a problem of *degree* than of *kind*. The computer and network revolution is not over. New methods have yet to be developed and applied. Free toolkits for performing generalized CBPP quality-control functions will be developed and replicated many-fold.<sup>23</sup> CBPP systems can and will be taken farther, and will displace more and more information resources that could formerly only be produced centrally. Those who want the book to be closed now are setting themselves up for disappointment.

## CONCLUSION

Perhaps the greatest success of the digital library community to date has in fact been Google. Yet, this paper has argued that even with this major (but essentially accidental) creation, the promise and potential of digital libraries has not been fully met. To remain as relevant and useful as Google and additionally fulfill this latent potential, digital libraries will need to support free culture by integrating commons-based peer production services.

While digital libraries that do not do this may not up and “vanish” overnight, in the near future, they may find their social impact shrinking relative to alternative resources. Technically superior resources, as Google has demonstrated, will not necessarily be libraries. And free culture resources will provide a compelling alternative to static, oligarchic, top-down-controlled information silos—a model which the digital library world has subscribed almost exclusively to thus far.

As there is much benefit to be had from the sustainable infrastructure of officially-funded efforts, it is my hope that digital libraries within such environments will embrace the benefits of CBPP and an ideal of free culture. By using pooled resources to provide a kind of knowledge to everyone which also considers worthwhile input from everyone, overhead efficiency as well as social impact would be maximized. In such a world, digital libraries could better-foster widespread equity in knowledge, empowerment to create and use it, and social harmony resulting

from cooperatively doing so. Perhaps then the digital library truly could be the savior of civilization.

## ACKNOWLEDGEMENTS

I would like to extend special thanks to those who reviewed and made suggestions to earlier versions of this article, in particular, Ross Singer, Joe Corneli, Raymond Puzio, and Robert Milson.

## ENDNOTES

1. DP9 makes Open Archives repository records automatically accessible as web pages. An immediate consequences of this is that web search engines can crawl and index them. See <http://dlib.cs.odu.edu/dp9/>.
2. One needs only to peruse the DLI-1 and DLI-2 initiative projects or the proceedings of JCDL in the past few years to get a feel for this trend.
3. For example, see our MetaCombine (<http://www.metacombine.org/>) and OCKHAM (<http://www.ockham.org/>) efforts. These projects seek to more meaningfully combine digital library resources and services, and to methodically propagate these services throughout the library world.
4. See <http://scholar.google.com/>.
5. The WAG-The-Dog web localizer, from Georgia Tech, smoothly integrates library and digital library holdings with Google Scholar and other web sites. It is available at <http://rsinger.library.gatech.edu/localizer/localizer.html>.
6. See <http://thomas.loc.gov/cgi-bin/query/z?c108:H.R.2613:>.
7. The “solution” of inter-library loan for this problem still incurs a significant convenience hit. It also seems a bit convoluted, given the already-digitized state of most of the inaccessible articles.
8. By *structural* I mean costs which are built into *access*. Carrying costs of dissemination (such as paying for media and handling) do not count. The litmus test would be: if the distributor or seller of the work has a legitimate complaint in preventing do-it-yourself copying or dissemination of the work, then it is not truly free of structural dissemination costs, and is hence not *gratis* free.
9. CBPP can actually be used in a closed fashion, *e.g.*, for internal corporate collaboration solutions. Applied in such situations, it is still an innovative approach, due to its flattening of productive hierarchies. Thus, the reader should not get the impression that CBPP cannot be used for more narrow communities and with more

limitations on permitted activities. For a more technical and generic treatment of CBPP frameworks, see (Corneli and Krowne 2005) in this volume. For this article, however, I deal with the more widely open sense of CBPP.

10. Due to space constraints or “political” feuding.
11. Significantly, PlanetMath was started when a similar, earlier resource went offline due to being *gratis* free but not *libre* free. Therefore, the distinction is important, and this illustrates how free culture will fail to flourish without both senses of “free.”
12. Strangely, it seems only the Germans have been interested in actually performing scientific studies on the quality of Wikipedia (perhaps McHenry (2004) was never translated to German). Two such studies have been done. See [http://meta.wikimedia.org/wiki/Content\\_reviews](http://meta.wikimedia.org/wiki/Content_reviews) for details.
13. <http://www.ibiblio.org/>.
14. This can be considered to foster the kind of “sense-making” services that are called for in (Regazzi 2004).
15. <http://www.pgdp.net/>.
16. <http://www.furl.net/>.
17. <http://www.citeulike.org/>.
18. <http://www.deliciou.us/>.
19. The upshot of these kinds of sites is a collaborative filtering effect, whereby each user’s efforts to organize and categorize the web leads to an emergent aggregate organizational effect called a *folksonomy* (Mathes, 2004).
20. Copyright roadblocks are less a problem for purely-superimposed, “feedback” style CBPP services (such as a ratings) than for the contribution of extensive content objects (as in article-writing or review contribution) or any modification of primary content objects (as in metadata enhancement).
21. Even if he eventually gets a copy of the work (in any form), he still will never get back the time wasted on searching and waiting. And if he was successful, he’d technically be violating terms-of-use of network access at any but his home institution. And further, if I found the article, I’d be running afoul of copyright law by sharing it with him. It is difficult to see how this model furthers the public good.
22. See, especially, (Lessig 2004) for more.
23. For example, see CoFE (<http://eecs.oregonstate.edu/iis/CoFE/>), an open source collaborative filtering system.

## REFERENCES CITED

- Yochai Benkler. “Coase’s penguin, or, linux and the nature of the firm.” *Yale Law Journal*, 112:369–446, 2002.

- Daniel K. Biss. "The elephant in the internet." *Notices of the AMS*, November 2004.
- Sergey Brin and Lawrence Page. "The anatomy of a large-scale hypertextual Web search engine." *Computer Networks and ISDN Systems*, 30(1-7):107-117, 1998. URL <http://citeseer.ist.psu.edu/brin98anatomy.html>.
- Peter Brusilovsky, Rosta Farzan, and Jae wook Ahn. "Comprehensive personalized information access in an educational digital library." In *Proceedings of JCDL 2005*, June 2005.
- Vannevar Bush. "As we may think." *The Atlantic Monthly*, 176(1):101-108, July 1945. URL <http://www.theatlantic.com/doc/prem/194507/bush>.
- Joe Corneli and Aaron Krowne. A scholia-based document model for commons-based peer production. In *Proceedings of the Symposium on Free Culture and the Digital Library*, October 2005.
- Economist. "Free Mickey Mouse." *The Economist*, Oct. 10 2002. URL [http://www.economist.com/printedition/displayStory.cfm?Story\\_ID=1378700](http://www.economist.com/printedition/displayStory.cfm?Story_ID=1378700).
- Miles Efron and Donald Sizemore. "Link attachment (preferential and otherwise) in contributor-run digital libraries." In *JCDL 2003*, pages 369-371.
- Gary Flake, Steve Lawrence, and C. Lee Giles. "Efficient identification of web communities." In *Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 150-160, Boston, MA, August 20-23 2000. URL <http://citeseer.ist.psu.edu/flake00efficient.html>.
- Gary William Flake, Steve Lawrence, C. Lee Giles, and Frans Coetzee. "Self-organization of the web and identification of communities." *IEEE Computer*, 35(3):66-71, 2002. URL <http://citeseer.ist.psu.edu/flake02selforganization.html>.
- Edward A. Fox, Neill Kipp, and Paul Mather. "How digital libraries will save civilization." *Database Programming & Design*, 11(8):60-64, August 1998.
- David Gibson, Jon M. Kleinberg, and Prabhakar Raghavan. "Inferring web communities from link topology." In *UK Conference on Hypertext*, pages 225-234, 1998. URL <http://citeseer.ist.psu.edu/gibson98inferring.html>.
- Marcos Gonçalves, Edward A. Fox, Layne T. Watson, and Neill A. Kipp. "Streams, structures, spaces, scenarios, societies (5s): A formal model for digital libraries." *ACM Trans. Inf. Syst.*, 22(2):270-312, 2004.
- Michael Gorman. "Google and God's mind." *LA Times*, December 2004.

- Michael Gorman. "Revenge of the blog people." *Library Journal*, February 2005. URL <http://www.libraryjournal.com/article/CA502009>.
- Hui Han, C. Lee Giles, Eren Manavoglu, Hongyuan Zha, Zhenyue Zhang, and Edward A. Fox. "Automatic document metadata extraction using support vector machines." In *JCDL 2003*, pages 37–48.
- F. Iannacci and E. Mitleton-kelly. "Beyond markets and firms." *First Monday*, 10(5), May 2005. URL [http://www.firstmonday.org/issues/issue10\\_5/iannacci/index.html](http://www.firstmonday.org/issues/issue10_5/iannacci/index.html).
- Aaron Krowne. "The FUD-based encyclopedia." *Free Software Magazine*, (2), March 2005. URL [http://www.freesoftwaremagazine.com/free\\_issues/issue\\_02/fud\\_based\\_encyclopedia/](http://www.freesoftwaremagazine.com/free_issues/issue_02/fud_based_encyclopedia/).
- Lawrence Lessig. *Free Culture: How Big Media Uses Technology and the Law to Lock Down Culture and Control Creativity*. Penguin Press, 2004.
- Sarah Lippincott and Martha Kyrillidou. "How ARL university communities access information: Highlights from LibQUAL+TM." *ARL Bimonthly Report*, 236, October 2004. URL <http://www.arl.org/newsltr/236/lqaccess.html>.
- Xiaoming Liu, Kurt Maly, Mohammad Zubair, and Michael L. Nelson. "DP9: an OAI gateway service for web crawlers." In *JCDL 2002*, pages 283–284. URL [http://www.cs.odu.edu/~liu\\_x/dp9/dp9.pdf](http://www.cs.odu.edu/~liu_x/dp9/dp9.pdf).
- Adam Mathes. Folksonomies - cooperative classification and communication through shared metadata, December 2004. URL <http://www.adammathes.com/academic/computer-mediated-communication/folksonomies.html>.
- Robert D. McHenry. "The faith-based encyclopedia." *Tech Central Station*, November 2004. URL <http://www.techcentralstation.com/111504A.html>.
- Robert Milson and Aaron Krowne. "Adapting CBPP platforms for instructional use." In *Proceedings of the Symposium on Free Culture and the Digital Library*, October 2005.
- Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. The pagerank citation ranking: Bringing order to the web. Technical report, Stanford Digital Library Technologies Project, 1998. URL <http://citeseer.ist.psu.edu/page98pagerank.html>.
- Raymond Puzio. "On free math and copyright bottlenecks." In *Proceedings of the Symposium on Free Culture and the Digital Library*, October 2005.
- P. Krishna Reddy and Masaru Kitsuregawa. "An approach to relate the web communities through bipartite graphs." In *WISE*, pages 301–310, 2001. URL <http://citeseer.ist.psu.edu/reddy01approach.html>.

- John J. Regazzi. "The battle for mindshare: A battle beyond access and retrieval (miles conrad memorial lecture)." In *46th NFAIS Annual Conference*, February 2004. URL [http://www.nfaeis.org/publications/mc\\_lecture\\_2004.htm](http://www.nfaeis.org/publications/mc_lecture_2004.htm).
- Ross Singer. Google scholar and the dawn of web localization, 2005. URL <http://rsinger.library.gatech.edu/papers/WebLocalizing.html>.
- Bud Tribble. (keynote speech). JCDL, June 2005.
- Siva Vaidhyanathan. *The Anarchist in the Library*. Basic Books, 2004.
- Fred Vogelstein. "Gates vs. Google: Search and destroy." *Fortune*, April 2005. URL <http://www.fortune.com/fortune/print/0,15935,1050065,00.html>.
- Meilan Zhang and Chris Quintana. "Facilitating middle school students' sense making process in digital libraries." In *Proceedings of JCDL 2005*, June 2005.





## **ABOUT THE METASCHOLAR INITIATIVE**

The MetaScholar Initiative is a center for collaborative digital scholarship projects involving research libraries and faculty from institutions around the United States. Based at the Robert W. Woodruff Library at Emory University, MetaScholar is creating new models for sharing and organizing metadata, tools for the preservation of at-risk digital objects, and services for scholars in focused research areas. This program encompasses more than ten digital library projects undertaken in the past six years, and has received funding from sources that include the Andrew W. Mellon Foundation, the Institute for Museum and Library Services (IMLS), the National Science Foundation (NSF), and the Library of Congress.

<http://MetaScholar.Org>