Ethics and Information Technology **1:** 127–136, 1999.

On the Web, plagiarism matters more than copyright piracy

John W. Snapper
Dept. of Humanities SH218, Illinois Institute of Technology, Chicago Il 60616, USA. E-mail: snapper@iit.edu

Abstract. Although commonly confused, the values inherent in copyright policy are different from those inherent in scholarly standards for proper accreditation of ideas. Piracy is the infringement of a copyright, and plagiarism is the failure to give credit. The increasing use of Web-based electron publication has created new contexts for both piracy and plagiarism. In so far as piracy and plagiarism are confused, we cannot appreciate how the Web has changed the importance of these very different types of wrongs. The present paper argues that Web-based publication lessens the importance of piracy, while it heightens the need for protections against plagiarism. Copyright policy protects the opportunity for publishers to make a profit from their investments. As the cost of publication decreases in the electronic media, we need fewer copyright protections. Plagiarism is the failure to abide by scholarly standards for citation of sources. These standards assure us that information can be verified and traced to its source. Since Web sources are often volatile and changing, it becomes increasingly difficult and important to have clear standards for verifying the source of all information.

Plagiarism and piracy

Although commonly confused, the values inherent in copyright policy are different from those inherent in scholarly standards for the proper accreditation of ideas. Piracy is the infringement of a copyright, and plagiarism is the failure to give credit.[1] They are confused because the most common examples of these wrongs involve both sorts of wrongs. But it is not hard to give examples that separate them. It would be *plagiarism but not piracy* for me to take the works of an obscure 19th c. poet and try to pass them off as my own. Since the copyright will have expired on such works, this is not piracy. But it remains plagiarism of the sort that could be grounds for dismissal from a journalism post. It would be *piracy but not plagiarism* if I were to edit a volume of modern poetry and forget to get copyright permission for one item in the volume. Assuming that the credits were properly given to the author and source publication, this is not plagiarism. All the same, it would certainly be grounds for action under copyright law. We may base a more sophisticated distinction between plagiarism and piracy on the commonplace that copyrights grant control over the means of expression, but never grants control over information or ideas.[2] It may be plagiarism to take information without giving credit, even if there is no piracy of a form of expression.[3]

The increasing use of Web-based electronic publication has created new contexts for both piracy and plagiarism. Situations emerge daily for which we have no clear standard either for copyright or for scholarly accreditation. Consider for instance, the common practice of downloading html tags that define the lay-

[1] These contrasting definitions of plagiarism and piracy are consistent with most legal language, although there are some legal discussions where the notions are confused or the words are defined differently. See K.R. St. Onge *The Melancholy Anatomy of Plagiarism*, University Press, 1988, for a survey of various definitions.

[2] "In no case does copyright protection for an original work of authorship extend to any idea . . . " 1976 Copyright Act of 1976 #102b. The "idea/expression dichotomy" is central to all discussion of copyright law, but almost all commentators begin their discussion by noting that the dichotomy is at best fuzzy, and at worst simply confused. An US Supreme Court application of this dichotomy to a denial of protection for the content of an electronically published data base (in *Feist Publications, Inc. v Rural Telephone Service Co.*, 499 U.S. 340 (1991)) has led to considerable discussion and some legislative attempts (e.g. HR 2652, introduced 1997) to provide protection for data bases.

This is presently a very "hot topic" in intellectual property law.

[3] In Narell v. Freeman 872 F.2nd 907 nicely illustrates the point. Freeman wrote a book that freely borrowed from a prior study by Narell. This is a clear example of plagiarism. Since there was only minor direct quotation, however, the copyright claim was dismissed. Laurie Sterns calls this a "paradigm" of the distinction between plagiarism and copyright infringement, "Copy Wrong: Plagiarism, Process, Property, and the Law," *California Law Review* 80, 1992, p. 542.

out of a Web page. I see a page lay-out with moving figures that I like. I download that page and use it as a template for my personal publication. I do not include any words, pictures, or information from the original site, only the html tags and java script that give the site its special look. Is this action a copyright infringement? As a scholar, am I obligated to give credit to the source of the lay-out code? The situation is not at all clear. And we should not expect that the establishment of a copyright standard will settle the issue for scholarly accreditation, since these are separate sorts of wrongs. An analysis of this sort of example requires that we identify the values inherent in the condemnations of piracy and plagiarism, and then see if those values suggest a way to deal with the new situations.

Our investigation into values is much easier for the copyright issues than for accreditation issues. Copyright is now defined in statutory law. The US statutes have an explicit basis in the US Constitution.[4] International copyright law has an explicit basis in international agreements.[5] If we dislike some feature of the law or believe that the law has failed to promote the traditional copyright values, then we know how to go about revising the law. The investigation into plagiarism is, however, much more difficult. We are faced with a mishmash of differing academic statements of principles, an ill-defined tradition in the writing community, and an inconsistent history of recognition in common law.[6] Although in many cases, we 'know it when we see it,'[7] plagiarism remains a notion with no generally recognized body of classical examples. And if we believe that we have discovered a need for revision of its standards, then we have nothing better to do than to put forth our ideas in articles such as this. Perhaps due to the difficulty with investigating plagiarism, there is so little literature on the subject, especially in contrast to the huge literature on copyrights. The issue of copyright on the Web has received considerable attention and plagiarism per se is largely ignored. For these reasons, the present paper concentrates on a study of the values inherent in the condemnation of plagiarism, rather than the presently heated debate over piracy on the Web.

What's the harm in plagiarism?

We may assess a standard through an appreciation of the harm done by an infringement of the standard. It would seem (at first glance, anyway) that a copyright infringement harms the copyright owner. The copyright owner suffers from loss of the revenue that is customarily paid for permission to copy. In contrast, it is not all clear at first glance that anyone is harmed by plagiarism?

The obvious candidate for a plagiarism harm is the author who receives no credit. But it is hard to see what harm that author may have suffered. Unless there is also copyright infringement, an author has few legal grounds for claiming economic loss for a plagiarized use of his work. There is no direct financial harm. And, given the strong tradition of refusing to grant property protection over ideas and information (as opposed to copyright protection over the means of expression), it is unlikely that we would want to grant an author any financial interest in the uncopyrighted content of a paper.[8] Perhaps there is an indirect financial harm to the author who fails to gain a reputation as ideas are taken without giving 'due credit.' But this harm is notoriously hard to assess.[9] Undoubtably, those scholars who provide citation counts as evidence of their scholarly reputations are "wronged"

[4] *U.S. Constitution*, Article I.8

[5] The Universal Copyright Convention of 1974, The Berne Convention of 1986, and the Copyright Law of 1976 amended by the Berne Convention Implementation Act of 1988, are all conveniently collected in *Selected Statutes and International Agreements on Unfair Competition, Trademark, Copyright and Patent*, ed P. Goldstein et al., Foundation Press, 1995.

[6] For an overview with many legal citations, see Laurie Stern, supra. This is an excellent article that should be studied with care by any reader interested in the issues raised in the present paper. Ms. Sterns provides an insightful critique of the plagiarism/piracy dichotomy that differs from the approach of the present article in several interesting ways.

[7] The most famous use of this criterion is Judge Stewart's claim that he knew pornography when he saw it, in *Jacobelis v. Ohio*, 378 US 197 (1964). That fact that it is extremely doubtful that Judge Stewart did indeed know pornography when he saw it should throw doubt on the criterion.

[8] This is a slight exaggeration, ignoring such areas of intellectual property as trade secret protection. In particular, consider the case of *International News Service v Associated Press*, 39 SC 38, in which a news service was forbidden to base its publication upon the reports of a rival news service, even though there was no copying of the form of expression and hence no copying. This is an example of what is commonly called "psuedo-property," based on common law and outside the limits of copyright protection. It would take us too far afield to consider these alternatives in this paper. For an introductory discussion see Charles McManis, *Unfair Trade Practices 3rd Ed.*, West Publishing, 1992, Ch. 6. Note that the *INS* case is not strictly a plagiarism case as defined in this paper, since proper citation would in no way diminish the pseudo-property claim of the prior news service.

[9] See Arthur Austin "The reliability of Citation Counts in Judgments on Promotion, Tenure, and Status" *Arizona Law Review*, 35, 1993. for an interesting discussion of how academics establish their reputation through a count of the number of times they are cited.

when their work is uncited. But this purported wrong certainly does not rise even near to the level of harm that demands legal protections through the criminalization of plagiarism. A possible loss of potential reputation is hardly sufficient grounds for the ethical indignation that academics express over incidents of plagiarism. And in the case of plagiarism from, for instance, the work of a defunct 19th century corporation, there seems to be no grounds whatsoever for worry about loss of potential reputation. We may also consider cases of 'self-plagiarism' in which an author uses material which he or she has previously published in another source. Although such action may be 'justified' by an author's sense of modesty, it does create a break in the citation trail. It is plagiarism in the sense of the present paper.

I suggest that the actual harm done by plagiarism is not harm to the author so much as harm to the reading public. When citations are left out of documents, the reader is deprived of one of the most fruitful ways of seeking additional resources related to the paper topic. It is also often of great importance for scholarship that sources can be traced backwards and verified. Consider, for instance, the study of the Bermuda Triangle[10] that traces the stories back to their unreliable sources. This trace is important for the assessment of the Triangle. Consider the importance of identifying the source of 'Protocols of the Elders of Zion,' and continuing to trace stories back to this source when certain views appear in modern works. Plagiarism destroys the scholarly trails and causes significant harm to the scholarly effort itself. Like the creation of false data and false histories, plagiarism cheats the public by presenting claims with a misleading or hidden provenance. Whereas piracy is a property violation, plagiarism is akin to 'fraud'[11] carried out against the reader.[12] Whereas a victim of piracy has the legal standing to institute a case of copyright infringement, it is the academic and journalistic communities themselves that protest plagiarism even in the absence of a 'victimized author.'

If this analysis of the harm of plagiarism is correct, then it would appear that the Web heightens the need for protections against plagiarism. The underlying problem is that the Web makes provenance difficult to establish, and consequently makes it more important that we work harder to preserve provenance. One of the problems is 'invisible revisability'.[13] To take a personal example, a biographer seeking information on the author of the present paper might have discovered that I received a PhD in 1967, according to the vita linked to my Web home page. That date was a typographic error that I have now corrected. The problem is not that some false information was published. False information is as common in the hard-copy print media as it is in the electronic media. The problem is that I corrected the date and left no trace of the correction. In contrast, a print correction would be visible by a comparison of earlier and later editions of the source. A biographer who used the false date taken from the Web might be accused of bad scholarship, as the Web source was invisibly altered. And a later scholar, tracing back the information might be puzzled about an odd discrepancy over dates with no apparent source.

There has been considerable effort in recent years by a number of organizations to address the problem of invisible revisability by the establishment of standards for scholarly references. For instance, the Modern Language Association,[14] the American Psychological Association,[15] the International Standardization Organization[16] are all involved in the creation of standards for Web citations, including guards against invis-

[10] Larry Kusche's study is a fine example of good search for sources that have gotten out of hand: *The Bermuda Triangle Mystery Solved*, Prometheus Books, 1995.

[11] The term "fraud" seems particularly apt, since like the term "plagiarism" it is only vaguely defined in the legal tradition. Broadly speaking, it is an intentional, material misrepresentation that actually misleads and causes damages, see Prosser On Torts. It would, indeed, be hard to treat plagiarism as fraud, since it would be hard to assess damages. On a related topic, see Stern, supra at note 49 on the relation between plagiarism and commercial "palming off."

[12] Peter Shaw, "Plagiarism," *American Scholar* 51, 1982, gives an example of a scientist who received a slap on the wrist for plagiarism, and then was fired when it was discovered that he had also faked data. It is clear which form of fraud is taken more seriously in science.

[13] The term is, so far as I can trace it, introduced by Andrew Harnack and Gene Kleppinger in "Beyond the MLA Handbook: Documenting Electronic Sources on the Internet," presented at the 23rd Conference, Kentucky Philological Association, March 1996, and posted at ⟨falcon.eku.edu/honors/beyond-mla⟩. Revised November 1996, accessed July 1998. This is an important source document for readers interested in citation standards.

[14] See Harnack and Kleppinger, *ibid*.

[15] See Mary Elleng Guffey "APA style Electronic Formats" Business Communication Quarterly March 1997, revised and posted at ⟨www.westwords.com/guffey/apa_z.html⟩, accessed July 1998. See Xia Li and nancy Crane *Electronic Style: A guide to Citing Electronic Information* (1993). Also see note #1 in Harnack and Kleppinger, supra, bemoaning the lack of discussion in the Publication Manual of the American Psychological Association.

[16] See "Excerpts from ISO Standard 690-2 (1997) posted at ⟨www.nlc-bnc.ca/iso/tc46sc9/standard/696-2c.ht ml⟩, accessed July 1998.

ible revisability.[17] Most of these guards are intuitive. In citing a Web publication, you should, for instance, refer to both the URL of site where you found the publication and also to a location where the publication is archived. This works, of course, if the archiving is secure, as it would be for a major newspaper published in both electronic and hard-copy versions. In response to the perceived need to archive, we now see the creation of a new industry that provides archiving services to Web users.[18] Some academic organizations strengthen the call for archived sources by refusing to accept citations to Web sources that are not securely archived. This is a fine standard, but a standard that may seem overly restrictive to someone seeking information that is most easily obtained through a search of my Web-based vita.

It is obvious that these standards for citation are definitionally tied to the notion of plagiarism: On one level, plagiarism is failure to follow the standards for citation. As we establish the standards for citation, we alter the notion of plagiarism. In order to guard against invisible revisability, the Web seems to require more stringent standards for citations. It would seem that more stringent standards create a higher scholarly obligation. And that higher obligation may entail a more serious sense of plagiarism with a heightened sense of the seriousness of the offense. This is a realistic scenario, which may in fact be played out as the Web, which is well recognized as a dangerous source for information, becomes a more important source of information. We shall see.

I would like to expand the discussion of invisible revisability to a more general concern for electronic provenance. Invisible revisability is only one example of the volatile nature of electronic information. Consider, as a further problem the new practice of publishing encrypted documents with a key made available to subscribers. If a publisher permits a "cryptolope" document to "go out of print" without ever publishing an unencrypted form of the document, there is a real possibility that the document will have disappeared as completely as the ancient Greek books lost at the destruction of the library at Alexandria. Electronic publications can be even more ephemeral than hard-copy publications.[19] And as a final concern, let us note that an author can easily alter archived primary sources. For instance, consider a claim based on an email correspondence. (This is an area of concern addressed by all the organizations that are presently looking into citations for Web based sources of information.) The fraudulent author can alter the archived version email message that stands as evidence for a claim. This fraud is simply the electronic version of the use of forged documentary evidence in traditional print media, except that the electronic version is much more difficult to expose. As the problems of electronic provenance make fraud easier and more dangerous to scholarship, it would seem reasonable for the scholarly community to heighten its sensitivity to plagiarism. This may take many forms, including a higher standard for citation against which to assess plagiarism, or a heightened reaction among academics to incidents of plagiarism.[20]

What's the use of copyright?

A focus on the rights of the producer rather than the expectations of the reader is what distinguishes piracy from plagiarism in the sense of the present paper. As noted above, copyright piracy harms the copyright owner. The harm is almost always viewed as a crime against property. The most common harm is the victim's loss of revenue from unauthorized copying, which is commonly called 'theft.' We may also consider unauthorized alteration of a copyrighted text. Although this is also a crime against property, it more closely resemble vandalism than theft. But in any event, it is clear that the copyright owner is the victim in these property crimes and torts.

The observation that copyright shows concern for the owner rather than the user, however, is only a starting point for a study of the issues regarding copyrights in Web publication. Our copyright policies are legal conventions that establish the relevant notion

[17] The "Electronic References and Scholarly Citations of Internet Sources" Web site at ⟨www.gu.edu.au/gint.WWWVL/OnlineRefs.html⟩ is an invaluable source for research into changing standards for Web citations.

[18] For a lengthy discussion of the needs and problems in archiving and of the industry response, see Cross Industry Working Team white paper (May 1997) on "Managing Access to Digital Information: An Approach Based on Digital Objects and Stated Operations," posted ant ⟨www.xiwt.org/documents/MagagAccess/ManagAcess.htm l⟩, accessed July 1998. This white paper draws attention to a wide range of interesting intellectual property issues for Web publications.

[19] It is a mark of the ephemeral nature of Web information that I have not been able to access a primary reference on cryptolopes: the IBM InfoMarket 1995 site at ⟨www.infomarket. ibm.com⟩. To all appearances, the site has been eliminated, without links to a new site. The problem of cryptolopes is also discussed in the XIWT white paper, supra.

[20] Academic responses to scholarly improprieties are notoriously mild. Although universities tend to scold errant professor and sweep wrongs under the carpet, newspapers and magazines take plagiarism and piracy much more seriously. For an interesting discussion of academic practices see Marlilyn V. Yarbrough "Do As I Say, Not As I Do: Mixed Messages for Law Students," Dickenson Law Review 100.3 (1996).

of property. The real issue for copyright is the social utility of these conventions.[21] In particular, the problem before us is whether those conventions are as useful in the context of electronic Web publication as in the context of hard-copy publication. It is simply impossible in a short paper to do justice to the huge volume of recent studies of this copyright issue. The study of copyrights on the Web is almost an industry, with a steady stream of excellent law review articles, books, organization white papers, and suggestive legal decisions. I will make no attempt here to address even a reasonable portion of the range of complex copyright issues raised by the Web publication. Rather I will draw attention to one particular aspect of electronic publication that, I believe, suggests that Web publication can remain a viable industry with fewer copyright protections. The argument is that a shift from hard-copy publication to Web-based publication creates a new economic environment in which slightly weakened copyright protections can still provide an adequate economic incentives for the publication industry. At the start, however, let me note three important qualifications to this argument.

First, it is my personal (unscientific) impression that the majority of legal scholars propose strengthened, rather than weakened, intellectual property protections for electronic publication. All reasonable commentators recognize that copyright policy must find a balance between (a) overly stringent protections that interfere with an author's right to base new studies on previously copyrighted works and (b) overly weak protections that fail to provide economic encouragement for the creation of new copyrightable works. My review of the literature leads me to the impression that the mood among most (but not all) legal scholars is that the Web creates an environment in which we need a new balance with additional protections against the ease with which works are electronically copied and distributed. We may take the attitude of the Working Group on Intellectual Property Rights as typical in this regard. Responding to a number of arguments (although not the argument of this paper) the WGIPR concludes that "weakening copyright owners's rights in the NII [national information infrastructure] is not in the public interest; nor would a dramatic increase in their rights be justified."[22] In contrast to the WGIPR call for a undramatic increase in rights, the present paper goes against the majority view by suggesting an undramatic reduction in rights.

Secondly, the present paper seems to echo a generally unconvincing argument presented by S. Breyer in 1970.[23] Breyer argued that changes in technology lowered the cost of publication to the point that most copyright protections were no longer needed to encourage publication. Although the argument received some attention, it is generally believed that Breyer failed to appreciate the economics of the publication industry. The present argument differs from Breyer's argument in several ways. Obviously Web publication is a far cry from the emerging technology contemplated by Breyer in 1970. But more to the point, whereas Breyer advocated an broad overhaul of the copyright system, the present argument only advocates some slight weakening of copyright protection in some contexts.

Thirdly, as most complex policies, copyright policy has a myriad of social utilities. Its intended utility, as stated in the US Constitution for US copyright policy, is that copyright policy is 'to promote the progress of science and the useful arts,' which suggests that copyright policy is to be judged on whether that policy is an incentive to the publication of new scientific and technical work. Although we may count the increased publication of scientific information among the social utilities of a successful copyright policy, this is obviously not the sole utility of copyright policy in today's world. Among the other utilities, for instance, are the encouragement of work in the fine arts and (in the software industry) the encouragement of the development of industrial processes. Without in the least bit suggesting that this is the only social utility to copyright policy, however, I focus here how copyright policies that provide financial incentives for the publication of original documents. This supposed utility is indeed the first social utility of copyright that is

[21] This remark assumes a utilitarian assessment of copyright policy, and ignores any suggestion that an author might have a natural right to the fruit of his or her labor. This paper is no place to argue this fundamental issue, and we simply adopts the standard utilitarian approach without argument. A natural rights attitude may be common among novelists and journalists who take pride in their work, but the mainstream of legal and philosophical literature is united in agreeing with Thomas Jefferson's observation (in the context of a patent dispute) that intellectual property is "given not of natural right, but for the benefit of society," the "Letter to Isaac M'Pherson, Aug 13, 1813," *The Writings of Thomas Jefferson*, Taylor & Maury, 1854, vol 6, p. 181. As an incidental matter, it reinforces the argument of this paper to note that I first took this passage from a secondary source that misquoted Jefferson. I have corrected the quotation.

[22] From the last page of the Introduction to The Report. The Report was prepared by the Working Group on Intellectual Property Rights (chaired by Assistant Secretary of Commerce, Bruce Lehman) for the White House National Information Infrastructure Task Force (chaired by Secretary of Commerce, Ronald Brown), and submitted to Congress in 1995. It is available at ⟨www.uspto.gov/web/offices/com/doc/ipnii⟩, assessed Aug 1998.

[23] S. Breyer, "The Uneasy case for Copyright," *Harvard Law Review 84*, 1970.

considered by most commentators. In particular, I ask what level of protection against 'free riders' is needed to provide publishers an incentive to publish new documents in the Web environment.

A focus on publishers may seem odd to some creative writers. These writers are more interested in providing incentives to authors to write than in providing incentives to publishers to publish. From that writer's perspective, it has been argued that the Web opportunity for more self-publication by writers on their works shifts copyright concern away from protections for publishers to the need to protect writers.[24] In contrast, the following analysis shows little concern for authors who post their own works on their own Web sites. (I must admit a certain lack of sympathy for authors who loudly demand economic rights to works self-published on the Web. This sort of activity hardly seems to require the economic incentives of copyright protection.[25]) My emphasis on economic protections for the publication industry may betray my academic background. Whereas the writers of cookbooks and mystery novels are in the business for the collection of royalties on their works, academics in fields such as my own rarely write in the expectation of direct profit from their efforts. With some outstanding exceptions, the fact is that successful academic writers rarely depend on royalties to do much more than finance a nice summer vacation or a new car now and then. Although academics write a great deal and profit from their writings with academic recognition, they rarely expect to make much money from copyright licenses. What I need, as an academic, is a healthy academic publishing industry that wants to publish my writings. So the question raised by Web technology is what sort of copyright policy is needed to encourage electronic publishers?[26] Are these the same copyright policies as are needed to encourage hard-copy publishers?

Once the issue is presented in terms of the economic incentives for the publication industry, we see an immediate difference between the needs of the industry in the hard-copy and electronic realms. Electronic publication lowers both the cost and the financial risk of publication, and therefore suggests that the electronic publication industry needs fewer copyright protections than the hard-copy publication industry. In preparing this paper, I spoke to several publishers who have moved or are moving into electronic publication. A typical story concerns the publication of a small journal in Chicago that focuses on African-American fine arts. The publisher informs me that the annual budget in 1997 for the publication of his journal in hard-copy was about $300k, of which about half goes to support the editorial staff, marketing, solicitation of articles, etc. The other half goes to the production of the journal, including type setting, purchase of paper, printing, etc. This latter cost is reduced from $150k to $10k by the shift to electronic publication. Basically, his costs are cut in half by the shift to electronic publication. This, however, is only one part of the saving. In addition, the publication risks are immensely reduced. In hard-copy publication, the editor must estimate the size of a printing run. If he overestimates, he suffers a considerable financial loss as the edition sits unsold in the warehouse. If he underestimates, he suffers the financial loss of lost opportunity, including the opportunity to raise his advertising rates for a more widely distributed journal. Run-size is not at issue for electronic publication. Whether he distributes his product over the Web or through the sale of CD-ROMS, the publisher needs only a small inventory and can produce the product quickly to meet demand. And finally, the publisher sees an new form of profit opportunity through on-line sale of subsidiary products. The publisher sees the possibility of both direct marketing from the journal site and of some form of compensation from advertisers who sell items over the Web to customers who link from the journal site. For the publisher of a small art journal, electronic publication provides new economic opportunities at considerably lower economic risk.

If we view the copyright as providing such protections against piracy as are needed to ensure a profit for the publication of documents, then the above example seems to suggest that we can lower the level of copyright protection. The example does not, however, suggest elimination of copyright protection. Even if publishing costs are cut by something over half, they do not go away. The small journal publisher still needs some protections against piracy. We cannot use this example to suggest the elimination of copy-

[24] Joan Latchaw and Jeffrey Galin argue in "Shifting Boundaries of Intellectual Property: Authors and Publishers Negotiating the WWW," *Computers and Composition*, 15.12, 1998 that the Web creates an context in which the "power" of authors and publishers is shifting to give greater emphasis to the concerns of authors.

[25] Given that damages for copyright infringement are based on a calculation of lost income or potential income, it is doubtful that there would be significant reward for a copyright infringement for non-commercial Web publication of a self-published free-access manuscript. It would be another matter, of course, if there was some potential commercial value to the original site. See note 34 below on the need to recognize loss commercial value when the original site is usually accessed by way of a link from a site that includes commercial advertisements.

[26] "The motive force behind legal protection for published works did not emanate from authors, scholars, and scientists; it came from booksellers, printers, and publishers." Stern, supra at 535.

right protection. What the example might suggest, however, is a lower level of protection. In particular, I suggest that we might expand the notion of "fair use" to permit more unauthorized electronic copying of articles, for instance for distribution as an electronic text to students in a university class. We might also loosen the standards for 'similarity' so as to permit unauthorized publication of an article that is substantially similar although not identical to a copyrighted electronic manuscript. We might generally recognize that an author and publisher both independently have copyright authority to authorize republication, and no longer recognize the sale by an author of full copyright authority. In any event, the example suggests that piracy is less important in the context of Web publication than it is in the context of hard-copy publication.[27] And this contrasts directly with the conclusions drawn above on the importance of plagiarism in the Web environment.

The example is suggestive, but by itself, it hardly provides a conclusive argument. From my personal discussions with a number of publishers, I think the example is fairly typical for journals shifting from hard-copy to Web publication. The situation radically changes if we consider, for instance, the somewhat futuristic possibility of distributing movies over the Web. The major cost to the film studios are costs of making the master copy. Although the Web opens up the possibility that studios may do their own distribution over the Web (which would be a significant change from the present financial structure of the industry), we could not make the same claims as we did for the journal that costs are cut by over half. The Web-based distribution of computer games,[28] the Web-based distribution of computer software, the Web-based distribution of music, and the Web-based presentation of college classes[29] each have special needs that distinguish them from the electronic journal industry, and we must be careful not to overgeneralize from the experience of electronic journals. The computer pornography industry is an interesting case in point for our study: it functions well and earns a profit for its publishers, even though the industry makes few attempts to claim copyrights on its Web sites. Projections on the emerging industry of electronic book publication also provide evidence that Web-based publication does not need the same level of copyright protections against piracy as does hard-copy publication.[30]

The analysis above treats piracy is a form of 'free-riding' in which the pirate takes advantage of the efforts of the original author without having the investment of resources by the original author. Although this is the most common focus of discussion of copyright issues, we must recognize that are certainly other ways to analyze the harm of piracy. In particular, we should take note of the popular use of copyright to protect against unauthorized alteration of a manuscript. The most recent revisions of the copyright law add recognition of this 'moral right' of the author to the traditional economic rights of the copyright holder. Exercising this right, one of the Web-based papers that I read in preparation for this paper, included the notice that the author held the copyright, but would permit replication on condition that 'the paper is copied fully, without alteration.' Given the usual academic standard for 'fair use' quotation of passages from published papers, it is unclear what sort of copying this author intended to preclude. All the same, this sort of concern is clearly in the mind of many authors,[31] and it is not primarily

[27] Fred Cate et al. in "Copyright Issues in Colleges and Universities" argue the opposite point: "... copying a page from an article or a frame from a motion picture might previously have been thought fair because it involved copying only a small portion of a larger work where no market existed for that portion alone. Today however a use might be found.... As computers create markets for smaller and smaller fragments of works, uses that were fair in print may cease to be so in the context of digital information." *Academe*, May/June 1998, p. 44.

[28] The protection needs of each industry in this list is discussed in the report of the Working Group on Intellectual Property Rights. The Report is rapidly acquiring the status of a benchmark for discussions of the varying needs for copyright protection.

[29] There has been considerable theoretical discussion and some legislative action addressing the need to accommodate electronic classrooms through redefinition of fair use. Ellen Schrecker in her editorial introduction to the "Technology and Intellectual Property" issue of *Academe*, May/June 1998, argues that Web-based courses and "distance learning" in general are quite expensive, largely due to equipment costs. She rejects the commonplace claim that distance learning is a cheap alternative to traditional education. Henrietta Shirk and Howard Smith provide a nice overview of changing standards for fair use in electronic classes in "Emerging Fair Use Guidelines for multimedia: Impoicationns for the Writing Classroom," in *Computers and Composition* 15.2 1998.

[30] The *Wall Street Journal* tells us that a publisher of electronic books predicts that electronic distribution will permit him to cut retail price by 30% and still double profits. This is entirely in keeping with the projections for the shift to electronic journal publishing that are reported above.

Joshua Kwan, "Nascent Electronic Book Field Already Seems Crowded," *Wall Street Journal*, July 30, 1998.

[31] The US Copyright Act of 1976, as originally enacted, shows little concern for an author's "moral right" (in contrast to "economic rights") to prevent alterations to a work. This concern, however, received considerable attention in other countries. The Berne Convention Article 6b is is a very powerful statement of moral rights: "Independent of the author's economic rights, and even after the transfer of said

a financial concern. In fact, this is a concern that views piracy as something very close to plagiarism, except that it views alteration of a text as a harm against the author rather than against the reader. In both cases, the issue is the purity of the textual information source.

Bringing it all together

The above observations provide evidence for a claim that the Web creates an environment that heightens the dangers of plagiarism and lessens the dangers of piracy. Let us not overstate the situation. The evidence is suggestive, but does not justify a radical revision of our present system. Indeed, at some level our observations justify a continuation of the present legal system that formally recognizes the crime of piracy while largely ignoring plagiarism. It is much easier to quantify and calculate the harm caused by piracy than by plagiarism, as these wrongs are defined above. In calculating damages for piracy, the courts calculate the loss in value to the original work, as well as the unfair profit made by the free-riding pirate. It is not clear how we would calculate the harm caused by plagiarism. The courts certainly do not want to enter into sterile arguments over who gets credited for what. And therefore, we should expect that the law should deal with the former more aggressively than the latter. In contrast, we should not be surprised by the fact that academics seem willing to infringe copyrights as an everyday matter at their photocopy machines, while those same academics complain loudly when they see signs of plagiarized data in academic papers. Although plagiarism is aggravated and piracy is lessened by the use of electronic publication, it does not follow that our present legal and ethical responses to these wrongs need an immediate overhaul. It is more likely that, as the Web becomes a more and more important tool for research and distribution of information, we will see a slight shift in the importance that is attached to piracy and plagiarism. To return, for instance, to the question of citation counts, we can take note of the fact that electronic data bases are making it easier to establish citation counts. Whereas it used to be the case that only the legal profession had the sort of reference records that made citations easy to establish, it is now the case that electronic data bases are making citation counts possible in fields such as theoretical physics. Although it remains that case that this concern for a loss of potential reputation cannot justify the seriousness with which academics treat plagiarism, we must note how Web-based data bases are changing the game to heighten concern for the reference trail. And we may need to respond to the change with new standards.

The point is nicely illustrated by a look at the practice of digital archiving of files taken from the Web. One obvious guard against the volatility of electronic provenance is the archiving of Web-based information. If it is plagiarism to present information without citing the source, then scholars should be able to keep a copy of the source as a guard against its disappearance in change on the volatile Web. But it may be piracy to make the copy needed to avoid the charge of plagiarism, particularly if we make a complete copy of a lengthy document. It would seem reasonable, therefore, that copying for the sake of archiving electronic information be seen as 'fair use.' Indeed this is very close to the definition of fair use as traditionally defined in sections 106 and 107 of the Copyright Act that permits a copy for preservation which does not diminish 'the potential market for or value of the copyrighted work.' The doctrine concerning fair use archiving was not originally designed to meet the present problem, but rather recognized the practice by 'libraries and archives' (as opposed to individuals) that make (photographic, but not electronic) copies for 'the purposes of preservation and security.' It would appear that the need to strengthen guidelines for Web-based citation in the Web environment entails a broader recognition in copyright policy for electronic archival copying of files in personal archives that are not strictly speaking libraries. Indeed the present trend in copyright law is to recognize more fair use copying for archival purposes.[32] In order to protect the scholarly community against plagiarism in cyberspace, we may need to lessen the property protections of the copyright owner in cyberspace.

I suggest that a heightened concern for Web plagiarism combined with a lessened need for economic protection of Web publishers might entail a number of Web-based extensions of the fair use policy. For example, we might consider the apparent volatility of a Web site as relevant to whether it is fair use to digitally download a file from that site. On this grounds, it is fair use to download a Web home-page which is not obviously archived in any manner. In contrast, books published by MIT press are not frequently available both on-line and in bookstores. The existence of the hard-copy version would speak against the fair

rights, the author shall have the right to claim authorship of the work and to object to any distortion, mutilation, or other modification of, or other derogatory action in relation to, the said work, which would be prejudicial to his honor or reputation." When the US Copyright Act was amended to bring it into agreement with the Berne Convention, a strong statement of moral rights was added as section 106A.

[32] S 1146 introduced in the Senate in 1997 amends the fair use section of the copyright law to explicitly recognize digital archiving.

use right to digitally copy the book. That a work is firmly archived (on standards that are yet to be defined) becomes part of the copyright owner's expectation for economic protection. This is, of course, only one example of the way in which the Web environment may lead to a broadened notion of free use and of weakened copyright protection. These sorts of examples suggest that as a general matter, many of the new issues for cyberspace use of files should be decided in favor of broader rather than more restricted fair use policy.[33]

We must be very leery of any extension of the present argument that would lead to copyright policies that seriously harm the economic incentive for commercial Web publication. I do not suggest a policy under which a scholar who cited a Web source could make a downloaded version of the source available as an attachment to a scholarly study.[34] This is explicitly not fair use as presently defined. It would practically eliminate all copyright protections, greatly diminishing the revenue opportunity of a pay-to-access Web site that holds a copyrighted source file. (We can imagine a system where secondary users could distribute originally sources under a system of mandatory licensing of copyrighted electronic documents, with automatic distribution of royalties such as are presently used by the music industry. That we can imagine these new systems of protections for Web sites, of course, does not mean that we should implement them. For now, it is enough to take note of them.) The present paper suggests the we can loosen the notion of fair use for some cases of digital copying, not that we embark on a rampant elimination of copyright protections.

Let us now return to the issue raised in the first section of this paper: Let us consider the uncited and unauthorized copying of the lay-out code that establishes the look of a Web site, without the copying of the words or pictures that appear on the site. I suggest that this should not be seen as a serious incident of plagiarism. The point is that, for most purposes, there is little or no scholarly value in knowing the original source of the html tags. We may be able to construct examples where a historian of the fine arts has an interest in the creator of the code, but these examples are a bit recherche. On the whole, this sort of copying seems unimportant on the values inherent in guards against plagiarism. In contrast, however, we might very well see this as piracy. On the theory that a copyright protects the programmer's opportunity to make a profit from the labor that goes into the writing of a manuscript, we may recognize that the code to define a graphics template is under certain circumstances 'copyrightable subject matter.' Our discussion, however, suggests that as the Web makes distribution easier of documents easier and cheaper, then, we can afford to set fairly high criteria for a text is copyrightable subject matter. On a 'sweat-of-the-brow' test for copyright claim on lay-out code,[35] we should only extend copyright protection to lay-out code if the code is of considerable length and represents considerable development cost.

References

Arthur Austin. The Reliability of Citation Counts in Judgments on Promotion, Tenure, and Status. *Arizona Law Review* 35, 1993.

S. Breyer. The Uneasy Case for Copyright. *Harvard Law Review* 84, 1970.

Fred Cate et al. Copyright Issues in Colleges and Universities. *Academe*, May/June 1998.

Cross Industry Working Team. Managing Access to Digital Information: An Approach Based on Digital Objects and Stated Operations. <www.xiwt.org/documents/MagagAccess/ManagAcess.html>, May 1997, accessed July 1998.

Electronic References and Scholarly Citations of Internet Sources. <www.gu.edu.au/gint.WWWVL/OnlineRefs.html.

Mary Ellen Guffey. APA style Electronic Formats. *Business Communication Quarterly*, <www.westwords.com/guffey/apa_z.html>, March 1997, accessed July 1998.

Andrew Harnack and Gene Kleppinger. Beyond the MLA Handbook: Documenting Electronic Sources on the Internet. 23rd Conference, Kentucky Philological Association, <falcon.eku.edu/honors/beyond-mla>, November 1996, accessed July 1998.

[33] A broad range of relevant issues are discussed in the Interim Report of the Conference on Fair Use (1996/97) available at ⟨www.usptol.gov/web/offices/dcom/olia/confu⟩, assessed August 1998. CONFU does note the scholarly need for archives. It goes much further to consider issues for the definition of fair use on the Web that go beyond the scope of the present paper, including quire specific proposals for fair use guidelines. It suggests, for instance, a 1000 word limit on passages copied from lengthy documents.

[34] The issue here is files that are downloaded and attached to a published document. A closely related issue is the practice of linking a reference to the site which is the original source of the reference. On this matter, see Walter A. Effross, "Withdrawal of the Reference: Rights, Rules, and Remedies for Unwelcome Web-Linking", *South Carolina Law Review* 49, Summer 1998. This volume is available over the Web on a link from ⟨www.law.sc.edu/sclr/vol49_4.htm⟩. It is noteworthy that I was made aware of this article through a reference that went directly to the location of the html version of the paper, rather than to the above site of the *South Carolina Law Review*. The article discusses the journal's remedies for loss of commercial value in such direct linkages, which ignore the mother journal.

[35] A sweat-of-brow test has a long an controversial history in copyright law. Generally the test has not stood up well in recent decisions, see Feist supra. The present suggestion is quite radical in this respect.

International Standardization Organization. Excerpts from ISO Standard 690-2, <www.nlc-bnc.ca/iso/tc46sc9/standard/696-2c.html>, 1997, accessed July 1998.

Thomas Jefferson. Letter to Isaac M'Pherson, Aug 13, 1813. *The Writings of Thomas Jefferson*. Taylor & Maury, vol 6. 1854.

Larr Kusche. *The Bermuda Triangle Mystery Solved*. Prometheus Books, 1995.

Joshua Kwan. Nascent Electronic Book Field Already Seems Crowded. *Wall Street Journal*, July 30, 1998.

Joan Latchaw and Jeffrey Galin. Shifting Boundaries of Intellectual Property: Authors and Publishers Negotiating the WWW. *Computers and Composition* 15(12), 1998.

Lia Li and Nancy Crane. *Electronic Style: A guide to Citing Electronic Information*, 1993.

Charles McManis. *Unfair Trade Practices 3rd Ed.* West Publishing, 1992.

l.P. Prosser et al. *Casebook on Torts*. Foundation Press, 1982.

Ellen Schrecker. Technology and Intellectual Property. *Academe*, May/June 1998.

Peter Shaw. Plagiarism. *American Scholar* 51, 1982.

Henrietta Shirk and Howard Smith. Emerging Fair Use Guidelines for Multimedia: Implications for the Writing Classroom. *Computers and Composition* 15(2), 1998.

K.R. St. Onge. *The Melancholy Anatomy of Plagiarism*. University Press, 1988.

Laurie Sterns. Copy Wrong: Plagiarism, Process, Property, and the Law. *California Law Review* 80, 1992.

Working Group on Intellectual Property Rights for the White House National Information Infrastructure Task Force. <www.uspto.gov/web/offices/com/doc/ipnii>, 1995, accessed Aug 1998.

Marilyn V. Yarbrough. Do As I Say, Not As I Do: Mixed Messages for Law Students. *Dickenson Law Review* 100(3), 1996.

Law Cases

Feist Publications, Inc. v Rural Telephone Service Co., 499 U.S. 340, 1991.

International News Service v Associated Press, 39 SC 38, 1918.

Jacobelis v. Ohio, 378 US 197, 1964.

Narell v. Freeman 872 F.2nd 907.