University of Chicago Law School
Chicago Unbound

Coase-Sandor Working Paper Series in Law and Economics

Coase-Sandor Institute for Law and Economics

2013

Access and the Public Domain

Randal C. Picker

Follow this and additional works at: http://chicagounbound.uchicago.edu/law_and_economics

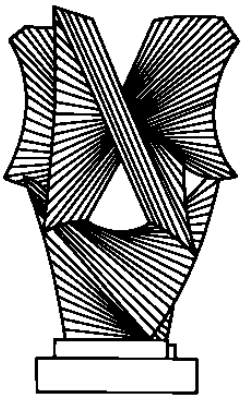
Part of the Law Commons

Recommended Citation
Randal C. Picker, "Access and the Public Domain" (Coase-Sandor Institute for Law and Economics Working Paper No. 631, 2013).

This Working Paper is brought to you for free and open access by the Coase-Sandor Institute for Law and Economics at Chicago Unbound. It has been accepted for inclusion in Coase-Sandor Working Paper Series in Law and Economics by an authorized administrator of Chicago Unbound. For more information, please contact unbound@law.uchicago.edu.

CHICAGO

COASE-SANDOR INSTITUTE FOR LAW AND ECONOMICS WORKING PAPER NO. 631
(2D SERIES)



Access and the Public Domain

Randal C. Picker

THE LAW SCHOOL
THE UNIVERSITY OF CHICAGO

February 2013

This paper can be downloaded without charge at:
The University of Chicago, Institute for Law and Economics Working Paper Series Index:
http://www.law.uchicago.edu/Lawecon/index.html
and at the Social Science Research Network Electronic Paper Collection.

Access and the Public Domain

Randal C. Picker[*]

One of the constitutive notions of intellectual property is the public domain. These are the ideas, notions and works that are usable by all where no person can invoke a right to block use by another. I will be focusing on copyright here, but of course the idea of the public domain is also at work in other areas of intellectual property. Defining the precise boundaries of the copyright public domain in the U.S. takes real work and there are obviously hotly-contested, high-level legal issues about those boundaries, such as those at stake in *Eldred* and *Golan*.[1]

My interests here are more mundane. I am interested in the mechanics of accessing the public domain and the consequences of the choices that we make about the operational effectiveness of various tools available for controlling access to the public domain. That matters both for the extent of access but also importantly for the scale and scope of competition in the provision of the public domain. This takes us to contract and terms of use; to the core of copyright in what it means to be original and also what it means to copy; to copyright's periphery in the form of the deposit requirement; to the DMCA and its tools of distant control; and to the Computer Fraud and Abuse Act and its restrictions on accessing materials and objects under local control. With the emergence of major digital scanning projects for works in the public domain—call these online public domain repositories (OPDRs)—we are at a point of possibly unparalleled practical access to the public domain. But the decisions we make about the tools just described will have a strong effect on the scope of the actual access to the public domain that emerges and in the extent of competition on the provision and use of the public domain.

We face standard conflicts about the limits of appropriability, efforts to expand appropriation and the possible consequences of those efforts for effective access to the public domain and overall welfare. Many institutions undertaking major scanning projects will have their interests first and foremost in mind. Some of those interests will relate to the substantial costs associated with digitizing works and hosting them. Those costs have to be financed and fully unfettered access to the public domain may be inconsistent with that financing.

---

[*] Copyright © 2013, Randal C. Picker. All Rights Reserved. James Parker Hall Distinguished Service Professor of Law, The University of Chicago Law School and Senior Fellow, The Computation Institute of the University of Chicago and Argonne National Laboratory. The substance of this paper was finalized as of August 3, 2012 and doesn't reflect events after that date, including the tragic death of Aaron Swartz. I thank the Burton and Adrienne Glazov Faculty Fund and the Microsoft Fund for their generous research support; Aimee Brown for able research assistance; David McGowan for his especially generous comments at the San Diego law and philosophy conference; and Peter Brantley, Justin Hughes, Richard McAdams and participants at the San Diego conference and the University of Chicago Law School Work-in-Progress workshop for comments.

[1] *Eldred v. Ashcroft*, 537 U.S. 186 (2003); *Golan v. Holder*, 537 U.S. 186 (2012).

That said, there is a broader set of questions that I am interested in, namely, exactly how unpropertied is the public domain? I confess that I am not sure that "property" as such is necessarily even the right framing for that question, but it does capture some of what is at stake. We think of the public domain as open to all and in that sense owned by no one (or owned by all). But that ignores the question of how the public domain is actually accessed. We have many tools that control that access and in that sense, through those tools, we re-establish a shadow control regime for the public domain. And we are at the early stages of architecting competition in digital libraries and one piece of that process will be assessing the mechanisms for controlling access to the public domain.

Section I of the paper sketches out the emerging public domain. Section II considers three conceptual questions for structuring use of the public domain focusing on the extent to which the public domain should be viral; on whether we should insist that the public domain be accessed only through the original artifacts embodying it; and on whether private appropriability incentives for distribution of public domain scans match overall social interests. Section III turns to the tools for restricting use of the public domain, to copyright, contract, the DMCA and the CFAA. Each of these matters for access to the public domain and for competition over it. Section IV considers one narrow question regarding the relationship between copyright's deposit requirement and a truly public public domain, while the last section briefly concludes the paper.

I. The Emerging Public Domain

We should probably start with a quick synopsis of the core features of the copyright public domain in the U.S. Works of the U.S. government enter the public domain immediately on creation.[2] Under long-standing case law, it is generally understood that other governmental works, including those of state actors, are in the public domain as well, though the statutory hook for that in the current copyright law is less clear than one might like.[3] Works published in the U.S. before 1923 are in the public domain as may be a variety of other works depending on how well the ins and outs of copyright formalities and renewal schemes were navigated over time.[4]

At a more conceptual level, ideas and facts are in the public domain as well.[5] You can't copy the entire expression in a new non-fiction text work, unless you are willing to make a pretty aggressive fair use argument, but separate from fair use, the noncopyrightability of fact and ideas means that most of the meat in a non-fiction text work is immediately available for use by all on publication. There will be interesting

[2] 17 U.S.C. § 105 (2011).

[3] The *Veeck* case is the most interesting serious recent look at this issue and it offers a tour of the relevant cases. *See Veeck v. Southern Building Congress Code Int'l*, 293 F.3d 791 (5th Cir. 2002) (en banc).

[4] The Cornell University Copyright Information Center maintains a very helpful guide on public domain status. *See Copyright Term and the Public Domain in the United States*, CORNELL (Jan. 1, 2012), online at http://copyright.cornell.edu/resources/publicdomain.cfm.

[5] 17 U.S.C. § 102(b); *Feist Publ'ns, Inc. v. Rural Tel. Serv, Co.*, 499 U.S. 340 (1991).

boundary cases—*Harper & Row*[6] is a good example—but nonfiction texts release an enormous amount of content into the public domain on publication. Use of textual fiction is much more limited and matters again get more complicated as we move from text to photographs, audio or video. To be sure, we could imagine a much larger public domain and the duration of a copyright has a direct and important effect on the size of the public domain, but the size of the public domain grows day by day.

But having work in the public domain and actually being able to use it are two quite different matters. Use requires access and copyright isn't a true access regime. I will address the deposit requirement in Section IV below but I think that is fair to say that the deposit rules haven't ensured widespread availability of copyrighted works, be those works that remain in copyright or works that have moved from copyright to the public domain. Instead, the public domain is distributed and is often scattered throughout libraries across the country. Library collection accession and deaccession practices over time will determine the practical availability of a public domain tied to physical formats.

You can get a sense of this casually by picking an obscure public domain work and searching WorldCat.org to see how the work is distributed physically. Of course, interlibrary loan practices will matter importantly in assessing the practical availability of the public domain. And that inquiry is framed as one in which you have a particular title in mind and just need to track down a copy of that work. Many inquiries—the kinds of searches we do everyday on Google and other search engines—are just unavailable on physical works. Of course, there have always been indexing services like the Reader's Guide to Periodical Literature or the Social Sciences Citation Index but even once those moved online, searching an indexed work is quite different from full-text searching on the underlying works.

And it is an online, full-text search public domain that is being created in the form of online public domain repositories (OPDRs). Consider briefly nine prominent projects:

- *The British Newspaper Archive*: The British Library is bringing 300 years of newspapers online. Newspapers are the day-to-day stories of our lives and provide an unparalleled way of accessing the past. With the November 2011 launch of the British Newspaper Archive, more than four million pages of scans came online, searchable for free and downloadable for a fee, as PDFs.[7]
- *ProQuest Historical Newspapers*: ProQuest offers an extensive—it boasts of nearly 30 million digitized pages—searchable collection of historical newspapers, including nineteen general interest U.S papers. For many newspapers, including *The Arizona Republican*, the *Cincinnati Enquirer* and the *Detroit Free Press*, the only content offered is pre-1923 public domain content. And if content nearing a century old isn't

[6] *Harper & Row, Publishers, Inc. v. Nation Enterprises*, 471 U.S. 539 (1985).

[7] See Amar Toor, *British Library Digitizes 300 Years Worth of Newspaper Archives, Brings 65 Million Articles Online*, ENGADGET (Nov. 30, 2011), http://www.engadget.com/2011/11/30/british-library-digitizes-300-years-worth-of-newspaper-archives/.

enough to excite you, read the *Nashville Tennessean*, which starts in 1812. For other newspapers, such as *The New York Times* and the *Wall Street Journal*, ProQuest offers public domain content and, pursuant to licenses, in-copyright content.[8] ProQuest is offered on a subscription basis to libraries.

- *The Internet Archive*: The Internet Archive is amassing a digital library of the web, movies, live music, audio recordings and texts. Its web archive offers the Wayback Machine: travel back in time on the Internet to see what your favorite website looked like on a particular date. But the archive also offers more than 3 million texts for reading and downloading.[9] Old works, such as *Little Goody Two Shoes*, published in 1766, are very much alive at the Internet Archive.
- *JSTOR*: Every academic knows and loves JSTOR, online at jstor.org. Most of the content digitized by JSTOR is in copyright and is digitized through a license with the copyright holder, but roughly 6% of the JSTOR content, nearly 500,000 articles, is in the public domain.[10] JSTOR terms this content Early Journal Content and, on September 7, 2011, announced a plan to unbundle the public domain content from the in-copyright content and make that public domain content available to all for free.[11]
- *Google Book Search*: When Google announced its new Google Print Service at the Frankfurt Book Fair in October 2004, it was clear that Google was taking a big step forward to fulfill its mission statement: "organize the world's information and make it universally accessible and useful."[12] That has been a bumpy process—and the litigation over it is ongoing—but Google hosts nearly 3 million public domain works in GBS.[13]
- *HathiTrust*: On October 13, 2008, two leading academic library consortia joined forces to announce the creation of the HathiTrust.[14] Libraries have been doing scanning on their own, but libraries who are partners in the Google Book Search project also get back copies of scans from Google, subject to restrictions detailed

[8] See *ProQuest Historical Newspapers*, PROQUEST, http://www.proquest.com/en-US/catalogs/databases/detail/pq-hist-news.shtml (last visited July 3, 2012).

[9] *See* INTERNET ARCHIVE, http://archive.org/details/texts (last visited July 3, 2012).

[10] *See Early Journal Content on JSTOR, Free to Anyone in World*, JSTOR, http://about.jstor.org/participate-jstor/individuals/early-journal-content (last visited July 3, 2012).

[11] *JSTOR—Free Access to Early Journal Content and Serving "Unaffiliated" Users*, JSTOR (Sep. 7, 2011, 3:02 AM), http://about.jstor.org/news-events/news/jstor%E2%80%93free-access-early-journal-content.

[12] *Google Mission Statement*, GOOGLE, http://www.google.com/about/company/ (last visited July 3, 2012).

[13] *Features of a Book on Google Play*, GOOGLE, http://support.google.com/googleplay/bin/answer.py?hl=en&p=books_features&answer=1062949 (last visited July 3, 2012).

[14] *Major Library Partners Launch HathiTrust Shared Digital Repository*, HATHITRUST (Oct. 13, 2008) http://www.hathitrust.org/press_10-13-2008.

below. As of mid-March 2012, the HathiTrust held more than 10 million volumes with almost 2.8 million public domain volumes.[15]

- *THOMAS:* Works of the federal government enter the public domain on creation.[16] That covers quite a bit, but note immediately that it covers all works of Congress and the federal court system. In 1995, the Library of Congress launched THOMAS—named in honor of Thomas Jefferson—as its window into congressional materials.[17]
- *Google Scholar*: THOMAS doesn't cover the work of federal courts. Although the U.S. Supreme Court now issues its opinions online as released, it has only backfilled to Volume 502 of the U.S. Reports, the official reporter for Supreme Court opinions. If you are interested in case law in the first 500 volumes, you are out of luck. Law libraries are full of shelf after shelf of published opinions and Westlaw and Lexis brought those opinions online, though often at very high prices. Google has now brought many of these opinions online and made them available for free at scholar.google.com.
- *Bodleian Library and the Bibloteca Apostolica Vaticana*: On April 12, 2012, these two libraries, based, respectively, at the University of Oxford and the Vatican, announced a plan to digitize 1.5 million pages of ancient texts held in their collections.[18]

I could add more, but you get the point. The public domain is being assembled, digitized, and made available to the public. These online repositories—online libraries?—represent the possibility of a huge step forward in the day-to-day practical availability of the public domain.[19]

II. Structuring Use of the Public Domain

I want to start by considering some preliminary conceptual questions about the use of the public domain. I want to address three issues. First, does the fact that a subsequent work incorporates the public domain have any consequence for the status of the first work? The strong version of this argument would embrace a notion of a viral public domain along the lines of the licensing approaches taken in open-source software. Second, to head to the other extreme, I want to address to what extent use of the public domain is subject to an original source requirement. Can you access the public domain through my republication of it or must you track down an original to access the public domain? Third, I want to consider the extent to which social and private incentives line up in the

[15] *See Statistics Information* HATHITRUST http://www.hathitrust.org/statistics_info (last visited July 3, 2012).

[16] 17 U.S.C. § 105.

[17] *See THOMAS*, THE LIBRARY OF CONGRESS, http://thomas.loc.gov/home/thomas.php (last visited July 3, 2012).

[18] *Bodleian and Vatican Libraries to Digitise Ancient Texts*, BODLEIAN LIBRARIES (Apr. 12, 2012), http://www.bodleian.ox.ac.uk/news/bodleian-and-the-vatican-libraries/.

[19] Do note that while both Google and HathiTrust have faced litigation for their efforts, that litigation relates to the digitization of in-copyright works and not for public domain works. Works in-copyright pose very different issues.

distribution of content, including public domain content. Limits on appropriability may mean that content possessors will take steps through contract, technical protection measures and the like to limit subsequent use inefficiently just because it boosts their ability to appropriate value from the content somewhat.

A. A Viral Public Domain?

We have seen a number of efforts over the last two decades to create viral ecosystems of property. The best known examples of these are the original GNU General Public License and its successors and the licenses of the Creative Commons. Somewhat ironically, these "copyleft" approaches rely on copyright itself to impose conditions on those who use prior work. The heart of the GPL is the fact that it is viral or infectious. To simplify considerably, if you use code subject to a GPL, the work you in turn create has to be just as useable by others as was the code that you started with. The idea is to plant a GPL code seed and watch the code base expand around it always increasing the amount of code available for use on GPL terms. It is the way in which use of the first software attaches conditions to the new software enabling further use of that software which makes the GPL viral.[20]

It is very much a counterfactual to imagine a fully viral public domain. Start with nonfiction. Would my inclusion of public domain facts somehow mean that my otherwise copyrightable expression would have to enter the public domain? That would seem to exclude completely copyright in nonfiction text. We could imagine freestanding new works of fiction—a new poem, perhaps—but we can imagine the line-drawing problems that would arise.

And the idea that use of any public domain material put the resulting work in the public domain would have dramatic consequences. Disney released its animated classic *Snow White and the Seven Dwarfs* in 1937.[21] The first full-length Disney animated film, cel by hand-drawn cel, Snow White's story was told again. Again, because, as everyone knows, her story was an old one going back in print to at least 1812 in the Grimm Brother's fairy tale *Little Snow-White*.[22] Her tale isn't a long one—6½ pages in an 1876 English edition—but Disney made the most of it in making an 83-minute film and unadjusted domestic grosses of almost $185 million.[23] And Hollywood is going back to

[20] For a detailed discussion, *see* Brett Smith, *A Quick Guide to GPLv3*, GNU OPERATING SYSTEM, http://www.gnu.org/licenses/quick-guide-gplv3.html (last updated June 10, 2012).

[21] THE INTERNET MOVIE DATABASE, http://www.imdb.com/title/tt0029583/ (last visited July 3, 2012).

[22] See D.L. Ashliman, *Little Snow-White*, http://www.pitt.edu/~dash/type0709.html#snowwhite (last visited July 3, 2012).

[23] And adjusted for ticket price inflation, *Snow White and the Seven Dwarfs* grossed roughly $853,000,000, tenth on the all-time list. *See All Time Box Office, Domestic Grosses*, BOX OFFICE MOJO, http://boxofficemojo.com/alltime/adjusted.htm (last visited July 3, 2012).

the well again with the release in 2012 of *two* Snow White movies.[24] It is hard to imagine any of this work getting created with a viral public domain.

And of course we don't take a GPL-like approach to the public domain in the U.S. The copyright statute itself doesn't really address this, but we might find an analog in Section 103. That section addresses situations in which new copyrighted material and old copyrighted material are mixed together. Assuming that you have permission to use the old material, as a creator, you receive copyright in the new material and not the old material. This is what we should anticipate: we would create very odd incentives if incorporation of prior material stripped new material of its copyright protection. New works would need to be vacuum-packed and carefully sealed to ensure that no public domain content had somehow snuck in.

As to digital scans, we should anticipate that when scans are done well, nothing new is added to them. That may limit the extent to which copyright can protect the scans themselves—I address this below—but the fact that I include public domain content into a new work doesn't in and of itself somehow limit the scope of protection I can enjoy for that work. I get what I get and the fact that the work is somehow derived from work in the public domain in no way limits the protection that I enjoy for my work. This line of analysis should limit the idea that some consequence attaches to the use of public domain work in the new digital scans, something I think of as the "the works were in the public domain and therefore" argument. I don't think copyright itself fills in anything after the therefore.

B. Original Sources, Derivative Sources and Copying the Public Domain

As we scan public domain works, we face something of a conundrum as to what it means to copy a work. Suppose that we were to conclude that a scan itself was an independently copyrightable object. For concreteness, consider a poem, say *Casey at the Bat*, written by Ernest Thayer and first published in the *San Francisco Examiner* on June 3, 1888. A search on Google Books turns up many copies of the poem, some published in books still in copyright, while others, such as the 1911 *America's National Game* by Albert G. Spalding, have entered the public domain.[25]

Suppose that Google claimed a copyright in the new digitized object that it has created. Where would that put us? What would that copyright limit? The poem is presented in full on the screen so that I can read it. I might be able to take a screen capture of the poem and then set about extracting the text of the poem from the screen capture. Were I to succeed in doing that, I would end up with just the text of the poem

[24] *Mirror Mirror*, starring Julia Roberts, was released on March 30, 2012 (*see* BOX OFFICE MOJO, http://boxofficemojo.com/movies/?id=untitledsnowwhite.htm). *Snow White and the Huntsman*, with Charlize Theron as the evil queen, was released July 1, 2012 (*see* BOX OFFICE MOJO, http://boxofficemojo.com/search/?q=snow%20white%20and%20the%20huntsman).

[25] Spalding recounts the initial publication of the poem and the public's reaction to it. See Albert G. Spalding, *America's national game*, 449-54 (American Sports Publishing Co., 1911).

and the poem itself, of course, is in the public domain. Even if I wasn't able to bring to bear the technical wizardry of capture and extraction, I could just type the poem into my computer by hand. Again, in each case, I have copied the poem to be sure and that would be a copyright violation save for the fact that the poem is in the public domain.

But what have I done vis-à-vis the new digitized object? We could imagine a regime that barred access to the public domain through the scan. Obviously, there are always proof problems, though in these circumstances, creators often seed their content with unimportant, intentional mistakes in an effort to make proof of copying easier. A regime that required subsequent firms to duplicate the efforts of the first firm in tracking down an original artifact containing the public domain work would obviously maximally protect the efforts of the first firm in doing just that. If tracking down and copying the public domain is expensive, we might need an original source—or, put differently, no-access-through—regime to ensure that the public domain was rediscovered in the first place. Of course, requiring duplication of effort means that we are rebuilding the wheel each time we want a competing copy of the public domain scan. This is a key point in how we structure competing uses of the public domain.

Prior to *Feist*, there was a strand of analysis in the cases that came close to this regime. The line drawn barred a subsequent user from extracting the public domain content from a particular text but instead required the second user to gather her own version of the public domain facts. Once those facts were in hand, then the second user could use the first public domain copy as a tool for checking the new second accumulation of facts.[26] But in *Feist*, the "industrious collection" of facts was found to be insufficient for copyright and full-blown copying of public domain content was permitted even if that took place through a new copyrighted work.[27] It seems clear as a matter of copyright proper that the public domain original can be accessed through a scan without running afoul of copyright even if we were to assign a copyright to the scan itself.[28]

C. Appropriability Incentives and the Public Domain

We know that spillovers—positive externalities—are an important feature of intellectual property works. The fact that use of the works is nonrivalrous means that from the perspective of creators, without more, they will only capture a fraction of the value that

[26] *See, e.g.*, *Rockford Maps Publishers, Inc. v. Directory Serv. Co. of Colorado, Inc.*, 768 F.2d 145, 149 (7th Cir. 1985) ("The second compiler must assemble the material as if there had never been a first compilation; only then may the second compiler use the first as a check on error.").

[27] *Feist*, 499 U.S. at 348 ("Thus, if the compilation author clothes facts with an original collocation of words, he or she may be able to claim a copyright in this written expression. Others may copy the underlying facts from the publication, but not the precise words used to present them.").

[28] This idea works most naturally when we are talking about accessing text through a scan. Matters become more interesting if we switch from text to images, but sorting through that isn't really the point of this paper. I talk about some of these issues in a blog post, *Fairey v. Associated Press: Yes He Can*, THE UNIVERSITY OF CHICAGO LAW SCHOOL FACULTY BLOG (Feb. 10, 2009), http://uchicagolaw.typepad.com/faculty/2009/02/fairey-v-associated-press-yes-he-can.html.

they create and much of the value will spill over to third parties. To just be slightly more concrete about that in the context here, in a nonfiction text work, the author controls only her expression of the underlying facts and ideas but those facts and ideas themselves enter the public domain and thereby spill over. The spillover point, of course, means that there may be insufficient private incentives to create socially-valuable works. But the spillover point is general. Here I want to consider the extent to which a creator might be willing to destroy social value to boost appropriability. The point isn't to destroy value, of course, but more that given incomplete appropriability, the creator won't take into account destroyed social value in her efforts to grab a larger slice of the pie.

Take a simple situation to see these issues. A creator can take a step to create a work. Doing so, absent more, will result in value of 6 to the creator and an additional value of 10 to the rest of society. The 10 represents external benefits from the creation of the work. There is an additional tool available to the creator—in the section below I will discuss technical protection measures, contractual provisions and other means—that would that would boost her take from 6 to 7 but at the cost of reducing the external benefits to 5. How should we assess this additional restriction?

We can't say. If we knew that the creator would create the work even without the provision—receiving 6 was enough of a motivation—then we would do better to have a pie of 16 created, assuming we are neutral about how benefits are distributed. The creator would have a private incentive to adopt the restriction, but that would be a bad outcome because she would be destroying 4 units of value just to get 1 more. But if 6 is an insufficient motivation to get the work created and 7 will do the trick, then we are clearly better off to have the restriction imposed.

We can see the problem for law. On the one hand, when appropriability is incomplete or imperfect, creators will be willing to destroy social value sometimes at a high cost in an effort to boost appropriability. Creators may be willing to inflict high social losses even for modest gains to themselves. But, to go to the other hand, barring these restrictions will shrink overall welfare if the incremental appropriability is actually marginal for the creation of the work in the first place. The question is how to operationalize these notions. I will try to do more of that below as I talk through particular examples.

III. The Tools of Controlling Access to the Public Domain

Assembling libraries of public domain materials involves substantial amounts of time and effort. Given that, we can hardly be surprised when digitizers look for tools to protect their investments. I will consider four plausible tools available to digitizers: copyright; contracts and terms of use; the DMCA; and the CFAA.

A. Copyright

The average digital scan is in some basic sense a photograph of a piece of paper, so we should review quickly the basics of copyright in photographs. We start, of course, with the Supreme Court's 1884 decision in *Burrow-Giles* and Sarony's photograph of Oscar Wilde.[29] Burrow-Giles helped itself to that photograph and then defended its use on the ground that the photograph was a "mere mechanical reproduction" and hence insufficiently original to qualify for copyright protection. The Court acknowledged that that might be the case for the "ordinary production of a photograph" though it declined to conclude that no copyright would attach even then. But Sarony had done much more in creating the photograph of Wilde: Sarony had composed the picture, arranged Wilde's pose and the setting for it, adjusted the lighting and shading, and all of that together represented an original creation by Sarony.

But the line that *Burrow-Giles* seemed to suggest—we should look for originality in the creation of the subject of the photograph—went by the by as courts embraced copyright in uncreated subjects (such as outdoor settings like the New York Public Library).[30] As soon as we took that path, we were going to have huge numbers of original photographs. Go to a party, take a bunch of pictures, create original copyrighted works. Indeed, the more interesting question quickly becomes: what does an unoriginal photograph look like? Burrow-Giles could of course have copied the Sarony photograph of Wilde by getting Wilde to repose and by making sure that every aspect of the new photograph matched the old one. That would be as much a copy of the Sarony photograph—and therefore unoriginal—as what Burrow-Giles did in the actual case.

Next, have Burrow-Giles take a photograph of Sarony's photograph of Wilde. That is just as much a copy of the original photograph as when Burrow-Giles recreated the picture per my hypo. That photograph of the photograph should be treated as a copy of the original photograph and unoriginal. You could, of course, make an original photograph that incorporated the Sarony photograph—perhaps with the photograph surrounded by other objects—but a pure photo of the photo should be treated as a copy. When you take such a photo of a copyrighted photo, you infringe. When the original photo is no longer copyrighted because it has entered the public domain, your photo doesn't infringe but it doesn't cease to be a copy and it still lacks originality.

Where does that put digital scans of public domain texts? We are starting to see skirmishes over photographs and scans of public domain works. The British National Portrait Gallery got into a spat with Wikipedia when Derrick Coetzee, a Wikipedia participant, uploaded into Wikipedia digital images created by the NPG of public domain

---

[29] *Burrow-Giles Lithographic Co. v. Sarony*, 111 U.S. 53 (1884).

[30] *Pagano v. Chas. Beseler Co*., 234 F. 963 (S.D.N.Y. 1916). Justin Hughes regards this expansive approach to the copyrightablity of photographs as a mistake. See Justin Hughes, *The Photographer's Copyright—photograph as art, photograph as database* (fortmcoming HARV. J. L. TECH.)

works in its collection.[31] Actual case law is scarce, with *Bridgeman Art Library v. Corel Corp.*, a 1999 federal district court decision, as a leading case.[32] The court characterized the dispute as one over "'slavish copies' of public domain works of art" and concluded that such copies lacked the spark of originality and therefore could not be copyrighted.[33]

As a freestanding proposition, it is hard to see how we could think of a digital scan as having sufficient originality to enjoy copyright protection. The essence of a high-quality digital scan of a text is perfect fidelity to the underlying physical page. A bad scan introduces imperfections and systematic deviations from the original physical page might be original, but those aren't the scans that we hope our digital curators are making.

But retrace our steps as to copyrights in photographs to see how we have done and focus not on the single digital scan but instead consider a continuum of original photographs. Put the fully-posed and staged Sarony photograph of Wilde perhaps at one of the continuum. Move to photographs of objects such as the Grand Canyon or the White House. Consider a rose in your back yard. At least as conventionally understood, copyright will subsist in each of those photographs. You might think of a photograph of a rose in your backyard as probably nothing more than a slavish copy of the rose and yet it is hard to imagine that a court would find insufficient originality in the photo. Yet a photo of a photo will probably be found to be nothing more than a copy of the original work, infringing or not depending on whether the work has entered the public domain, but almost certainly lacking sufficient originality for the second photo or scan to be independently copyrightable.

B. Terms of Use and Contract

As noted above, we should expect those digitizing public domain content to use contract to limit how the resulting scans are used. That is exactly what we see in practice. We need to map out the TOS-and-contract space a little before examining it in detail. Some OPDRs are privately-held and operate for profit but are generally available to the public (Google Book Search). That access may be conditional in that a TOS or contract may apply at various points of use. Other OPDRs are operated by non-profits—JSTOR, HathiTrust and the Internet Archive—but may offer different levels of access to members and non-members (HathiTrust and JSTOR prior to its recent unbundling of its public domain works) or may offer equal access to the public (the Internet Archive) again subject to possible TOS or contract limits at the point of actual use.

As all of that suggests, the breadth of access to the public is a complex pattern. Because Google uses a two-sided financing model—advertisers pay to reach

---

[31] Rory Cellan-Jones, *Wikipedia Painting Row Escalates*, BBC NEWS (July 17, 2009) http://news.bbc.co.uk/2/hi/8156268.stm. Some of the correspondence in the case can be found online at http://commons.wikimedia.org/wiki/User:Dcoetzee/NPG_legal_threat.

[32] 36 F.Supp. 2d 191 (S.D.N.Y. 1999).

[33] *Id.* at 197.

consumers—public access is full though it comes with standard Google issues regarding privacy and tracking, but, to be clear, you can use Google Book Search anonymously. ProQuest, our other private firm, charges users directly and that means much more limited access to the public. HathiTrust's limits on the public domain to non-members appear to be at least partially derived from the fact that many of their public domain scans are traceable to underlying contracts with Google and, as detailed below, those contracts imposed restrictions on Google's library partners. For most of its life, JSTOR had operated with a bundled access model: access to in-copyright works was bundled with access to public domain works. Even though JSTOR is a nonprofit, it needs to finance its operations and thus charges members fees for access to its databases. JSTOR has a rich approach to licensing—you can mix and match across multiple collections—but it had not separately unbundled a public domain collection and made that available to the public until 2011. All of that describes the general structure of access to the OPDRs. Actual use is subject to more detailed restrictions.

1. JSTOR

Consider that favorite of the academic researcher, www.jstor.org.[34] JSTOR provides scans of academic works in many fields. Of course, much of that work is in copyright, but the pre-1923 U.S. work is in the public domain.[35] And JSTOR makes some of that work available for free to the public. But when you try to download a public domain work, you are presented with terms. You are told that "Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions," with a link to those terms. If you don't click in acceptance, you can't download the public domain work. And if you want to read the terms, get a cup of coffee first: it runs 7,524 words.

Consider just the special codicil for what JSTOR terms "Early Journal Content." This is content published in journals prior to 1923 in the U.S. or prior to 1870 if initially published internationally.[36] JSTOR "encourages broad use" of this content, but while users are "free to copy, use, and redistribute" this content that use is limited to non-commercial purposes, and JSTOR asks that you attribute the content to JSTOR.[37] JSTOR also addresses bulk downloads: "Please be considerate of other users and do not use robots or other devices to systematically download these works as this may be disruptive to our systems."[38]

---

[34] For background on JSTOR, *see* Roger C. Schonfeld, *JSTOR: A History* (Princeton Univ. Press 2003).

[35] *Early Journal Content on JSTOR, Free to Anyone in World*, JSTOR (Sept. 6, 2011), http://about.jstor.org/participate-jstor/individuals/early-journal-content.

[36] *Terms and Conditions of Use*, JSTOR, http://www.jstor.org/page/info/about/policies/terms.jsp (last visited July 3, 2012).

[37] *Id.*

[38] *Id.*

2. GOOGLE BOOK SEARCH

Public domain downloads from the Google Book Search project come with similar limitations set forth in the first page of the file. Google offers something of a paean to the public domain. The public domain book that you have downloaded, you are told "is a digital copy of a book that was preserved for generations on library shelves before it was carefully scanned by Google as part of a project to make the world's books discoverable online."[39] Google goes on to explain that "… a public domain book is one that was never subject to copyright or whose legal copyright term has expired. … Public domain books are our gateways to the past, representing a wealth of history, culture and knowledge that's often difficult to discover." And "[p]ublic domain books belong to the public and we are merely their custodians."

And, with one word, the tone and content shift: "Nevertheless." Nevertheless meaning that the scanning of the public domain is something like a quarter of a billion-dollar undertaking and Google has an investment to protect. "Nevertheless, this work is expensive, so in order to keep providing this resource, we have taken steps to prevent abuse by commercial parties, including placing technical restrictions on automated querying." You are requested to use the files for only "personal, non-commercial purposes." You are told to "refrain from automated querying" and to maintain the Google watermark embedded in each file. The watermark "is essential for informing people about this project and helping them find additional materials through Google Book Search."

3. INTERNET ARCHIVE

The Archive is an important, free-standing nonprofit effort to create a digital repository for the artifacts of the Internet Age. The Archive offers a number of key tools. Want to see how the Google homepage has evolved over time? Go visit IA's Wayback Machine. But, as noted above, the Archive also has an extensive collection of public domain works, and all of that comes with a terms of use.[40] You are to use the archive for "scholarship and research purposes only."[41] Presumably, commercial use is forbidden. The TOU doesn't seem to address automated querying or bulk downloading directly but it does bar you from doing anything "to interfere with the work of other users or Archive personnel, servers, or resources."[42] And the Archive wants attribution for its contribution to your research.[43]

---

[39] These quotes are taken from the front page that Google attached to LEWIS CARROLL, ALICE'S ADVENTURES IN WONDERLAND 1 (The MacMillan Co. eds.,1898) *available at* http://books.google.com/books?id=CLoNAAAAYAAJ&printsec=frontcover#v=onepage&q&f=false

[40] *Internet Archive's Terms of Use, Privacy Policy, and Copyright Policy*, INTERNET ARCHIVE (Mar. 10, 2001), http://archive.org/about/terms.php.

[41] *Id.*

[42] *Id.*

[43] "In addition, we request that, according to standard academic practice, if you use the Archive's Collections for any

4. HATHITRUST

The HathiTrust imposes a number of key limits on the public domain works that it hosts. These limits reflect the origins of many of the works which arose through the deals cut by Google with its academic library partners. Under those deals, libraries would be entitled under some circumstances to receive back their own digital copies but often subject to substantial contractual restrictions. Hathi has emerged as a key way for those libraries to navigate their obligations to Google. In addition to that, Hathi has its own cost model and imposes restrictions on use to support that model.[44] For the public domain works, this means that these "volumes are freely accessible to the public and can be downloaded in their entirety with authentication by persons affiliated with partner institutions."[45]

***

We see a number of limits that run across these situations. One bars copying at scale: you aren't allowed to simply copy all of the public domain documents and replicate the original service. From one perspective, you might think that these services would welcome wholesale copying. Some of the downloads would be diverted to other sites and the original sites would save some of the download and hosting costs. Yet our OPDRs routinely bar bulk downloading.

It seems clear that the OPDRs value control over the public domain works they have amassed. Some of this goes to the revenue opportunities associated with these works. Google could serve ads and collect information through the hosting of works. That seems less true of the Internet Archive, but the Archive seems to value attribution and may believe that that interest is more likely to be vindicated if users have to come to the Archive to see works. Hathi's limits may be derivative of the original relationship with of Hathi's members with Google. And JSTOR's approach on this has evolved from the original bundled model—the public domain works bundled with in-copyright works—to now allowing broad access to its public domain materials.

We see a second limit frequently as well: a restriction to noncommercial use of the materials or research use. We should think of this as an option on commercial use of the materials. By that I mean that a person desiring to make commercial use of the materials would be required to negotiate for rights to use the works in a commercial fashion. The repository would hold an option on those commercial uses in that it could negotiate for a chunk of the upside of the new project in exchange for its consent to allow use of its public domain materials. This really does take us back to a version of the independent sourcing notion discussed above in Section II.B. We are talking here of commercial use of public domain materials. If you have independent access to these materials, you need

research that results in an article, a book, or other publication, you list the Archive as a resource in your bibliography." *Id.*

[44] *See Cost*, HATHITRUST DIGITAL LIBRARY, http://www.hathitrust.org/cost (last visited July 3, 2012).

[45] *See Features and Benefits*, HATHITRUST DIGITAL LIBRARY, http://www.hathitrust.org/features_benefits (last visited July 3, 2012).

not pay a fee to use them, but absent that access, a new commercial user might be stuck paying the repository for access to the public domain.

C. DMCA

The Digital Millennium Copyright Act is often referred to as a paracopyright statute in that its approach to protecting copyrighted work differs substantially from more traditional approaches to copyright. The DMCA implements a lock-and-key system for copyrighted works. To simplify considerably, individuals are barred from trying to unlock locked works and, to step a level up, the statute also limits trafficking in tools that facilitate unlocking locked works.

Digital rights management (DRM) of the sort protected by the DMCA is controversial. DRM is extensively used by Amazon as part of its Kindle platform, by Apple in running iTunes through its FairPlay software, and was baked into the DVD platform in the form of the Content Scramble System (CSS).[46] DRM defenders focus on, among other things, the way in which lock-and-key systems can support useful cross-product subsidies between hardware and content, while DRM detractors focus on the use limits that arise and the way in which DRM can limit after-the-fact competition, both limits that they regard as objectionable given what they see as the ease by which DRM schemes can be evaded by the hardcore.

What role should a digital lock-and-key system play for digital scans of public domain works? We face our usual analysis. For-profit scanners will want to control the works and, as noted above, may be willing to sacrifice overall social value to boost appropriability. At the same time, absent a digital lock, for-profit scanners may be unwilling to see their scans distributed in the clear. Doing so will may make it easier for competitors to scoop up those scans without having to incur the original costs of scanning. Many users might not find digital locks problematic and would prefer to have a downloaded copy rather than one permanently tied to an online digital repository. That can be about convenience—no need to have Internet access—but also about privacy or perceived privacy. And a downloaded copy might be more stable than a copy served through the OPDR.

Lest all of this be thought a bit hypothetical, consider how Google has approached the scans in Google Book Search and how its competitors would like access to those scans. The DMCA comes with many limits, including that the statute calls for a rule-making every three years undertaken by the Librarian of Congress and the Register of

[46] For info on these, *see* Jeremy Greenfield, *Bookseller Backed by Big Publishers Advocates Abandoning Digital Rights Management*, DIGITAL BOOK WORLD (Jan. 25, 2012), http://www.digitalbookworld.com/2012/bookseller-backed-by-big-publishers-advocates-abandoning-digital-rights-management/; Bryan Bishop, *Apple's FairPlay DRM for iBooks Cracked by Requiem App*, THE VERGE (Feb. 25, 2012), http://www.theverge.com/2012/2/25/2823218/apples-fairplay-drm-ibooks-cracked-requiem-app; *Universal City Studios, Inc. v. Corley*, 273 F.3d 429 (2d Cir. 2001).

Copyrights.[47] The target of the rulemaking is "any class of copyrighted works" such that the digital locks validated by the DMCA are, or are likely, to adversely affect noninfringing uses of those classes of work.[48] Note that the rulemaking powers call off unlocking liability but not liability for making or trafficking in tools of unlocking.

On December 1, 2011, in the current rulemaking proceeding being conducted by the U.S. Copyright Office, the Open Book Alliance has asked the Copyright Office to address digital scans of public domain works. The Open Book Alliance's mission is, in its words, "to assert that any mass book digitization and publishing effort be open and competitive."[49] The OBA is a comprised of a mix of tech firms such as Amazon, Microsoft and Yahoo!; author groups, such as the American Society of Journalists and Authors and National Writers Union; and reader facing organizations such as the Internet Archive and the New York Library Association.

To be faithful to its statutory mandate, the Copyright Office classifies the requests it receives and it characterizes the OBA request as involving "literary works in the public domain that are made available in digital copies."[50] The OBA in turn sees its request as directed at "prying open Google's closed books.[51] As detailed in its December 1, 2011 filing with the Copyright Office, the OBA describes the measures that Google has undertaken to lock up the scanned books.[52]

The OBA focuses on the contracts that Google signed with libraries to set up the rules of the game for digitizing the books held by the libraries. Those contracts are quite useful because they give a good sense of the institutional undertaking involved in making the public domain available. Of course, there is a way in which much of the public domain is just *sitting* there somewhere. These are the physical copies of the works that have entered the public domain. But there is a long, long road from sitting somewhere to being meaningfully available to all and Google's contracts with the libraries make that crystal clear. That is especially true given that Google and the libraries have been creating contracts amidst substantial uncertainty over the copyright foundations of the GBS project.

Our focus here is on the DMCA and the technical locks contemplated in the agreements with the libraries. Google announced its partnership with key libraries in mid-

---

[47] 17 U.S.C. § 1201.

[48] *Id.*

[49] *Mission*, OPEN BOOK ALLIANCE, http://www.openbookalliance.org/mission/ (last visited July 3, 2012).

[50] *Exemption to Prohibition on Circumvention of Copyright Protection Systems for Access Control Technologies*, U.S. COPYRIGHT OFFICE, http://www.copyright.gov/1201/2011/initial/ (last visited July 3, 2012).

[51] *Prying Open Google's Closed Books*, OPEN BOOK ALLIANCE (Dec. 2, 2011), http://www.openbookalliance.org/2011/12/prying-open-google%E2%80%99s-closed-books/.

[52] Open Book Alliance, *Comments of the Open Book Alliance (OBA); Docket No. RM 2011-7* (Dec. 01, 2011), http://www.copyright.gov/1201/2011/initial/open_book_alliance.pdf [Open Book Alliance Comment].

December 2004.[53] To take just one example, Google and the University of Michigan signed their original agreement on June 15, 2005[54] and that was subsequently amended on May 20, 2009 to adjust the agreement after the (attempted) settlement of the Google Book Search litigation.[55]

As Google scans, Google and Michigan each get a copy.[56] The contract envisions that Michigan's copy will be made available through Michigan's website and will also be available through cooperative arrangements with partner research libraries. In doing that, Michigan is required to "implement technological measures (e.g., through the use of the robots.txt protocol) to restrict automated access" to Michigan's copies of the scans.[57] And Google and Michigan together are to cooperate to develop methods to ensure that there are no bulk downloads of the digital copies at Michigan. Under the original contract, the restrictions on the use of Michigan's copy of the scans are independent of the copyright status of the books, and works in the public domain faced exactly the same limits as works in copyright. The 2009 amendment adjusted this somewhat given the proposed class-action settlement, though that settlement seems dead for now.

At GBS, public domain books are usually available for download as a pdf or in the epub format. It isn't clear to me whether those books, once downloaded, come subject to some sort of technological protection measure. But it is also clear that independent of that, what really troubles the OBA is technical protection measures that prevent large-scale downloads of the scans of public domain works. As the OBA puts it in its filing:

> While Google at the moment allows users to manually download individual PDFs of public domain works at the Google Books site, the stringent TPMs it imposes with respect to automated access prevent broader use of these files by competing search engines, digital libraries, and other online providers. It simply would not be feasible for an organization to make manual downloads of over three million books.[58]

To assess this, we should start with law and then turn to policy. We need to start with Section 1201 on circumvention of copyright protection systems. When you read Sec. 1201(a)(1) it is hard to avoid one conclusion: it simply doesn't apply to public domain works. Subparagraph (A) of that section bars any person from circumventing a

---

[53] *See* Press Release, *Google Checks Out Library Books*, GOOGLE INC. (Dec. 14, 2004), http://www.google.com/press/pressrel/print_library.html.

[54] *Cooperative Agreement*, UNIVERSITY OF MICHIGAN, http://www.lib.umich.edu/files/services/mdp/um-google-cooperative-agreement.pdf (last visited July 3, 2012).

[55] *See* the timeline set out at *Michigan Digitization Project*, UNIVERSITY OF MICHIGAN, http://www.lib.umich.edu/michigan-digitization-project (last visited July 3, 2012). The settlement of the class action brought by the Authors Guild against Google was rejected and that litigation is still pending. *See The Authors Guild v. Google Inc.*, 770 F.Supp 2d 666 (S.D.N.Y. 2011).

[56] *See Cooperative Agreement* Sec. 2.5

[57] *Id.* at Sec. 4.4.1

[58] *See* Open Book Alliance Comment, *supra* note 52, at 24.

technological measure that "effectively controls access to a work protected under this title." That of course is a reference to the copyright title and, presumably, to a copyrighted work. The next subparagraph confirms that certain classes of copyrighted works are exempt from the prior limits if the Librarian of Congress chooses to exempt them pursuant to the terms of subparagraph (C). That exemption is implemented by having the Librarian publish "any class of copyrighted works" that are to receive the benefit of the exemption scheme for a three-year window. None of the text applies naturally to works in the public domain and Section 1201(a)(1) is best read not to apply to public domain works.

Step back and assess what all of that means. Suppose that I download a public domain scan to my computer and I discover that it comes subject to DRM protection. Assuming that the scan itself is not a new copyrighted work—see Section III.A above—I don't violate Section 1201(a)(1)(A) if I circumvent that measure, as that section only applies to measures that protect works protected under title 17, meaning works in copyright. At the same time, the anticircumvention rulemaking process conducted by the Librarian of Congress only addresses copyrighted works, so nothing there seems to empower the Librarian to somehow exempt efforts to crack public domain scans.

All of that means that the DMCA is not likely to be the source of either protection for scans or for exemptions for people seeking to circumvent whatever DRM is embedded in the scans though the DMCA does limit the overall development of decryption tools and that may offer some protection even for public domain scans. Whether that is good or bad policy is a question that I am skeptical that we are well-situated to answer. This takes us back to the appropriability discussion in Section II.C. There are substantial costs to scan books. I am not sure what the right number is, but I have seen figures as high as $100 per book.[59] The OBA suggests that Google has scanned 3 million public domain books and that would get us quickly to $300 million dollars or exactly the estimated budget of *Pirates of the Caribbean: At World's End*.[60] Real money to be sure, but the sort of money that big firms spend with frequency.

The OBA filing argues that "[i]t simply would not be feasible for an organization to make manual downloads of over three million books." That is an interesting claim. Apparently, it was feasible for an organization to digitize 3 million public domain books—that would be Google of course—but downloading those copies one-by-one is the real stumper. This seems difficult to understand. I assume that the most expensive undertaking is organizing and scanning the public domain, the least expensive a bulk download of someone else's scans. One-by-one downloads sits in the middle. Doable, but not as cheap as bulk downloading.

---

[59] The Authors Guild mentions this figure in a filing in its lawsuit against the HathiTrust. *See The Authors Guild, Inc. v. HathiTrust*, No. 11 Civ. 6351 (HB), Memorandum of Law in Support of Plaintiffs' Motion for Partial Judgment on the Pleadings, Feb. 28, 2012, p5.

[60] *See* BOX OFFICE MOJO, http://boxofficemojo.com/movies/?id=piratesofthecaribbean3.htm (last visited July 3, 2012).

Of course, Google doesn't want to face a first-mover disadvantage where it bears the cost of scanning and then second-movers free ride on those scans. We might welcome competition in the scans given the complaints about the quality of the scans, but if the scans were perfect, we shouldn't want a second set of scans to be made. Of course in that circumstance, we might imagine that Google might license the scans to potential entrants where Google and the entrant would split the cost of doing the second set of scans. The cost of duplicating those scans would define the size of the pie available to be split in a licensing deal.

Of course, without legal protection for the locks, we can envision a number of unattractive outcomes. One is a standard cat-and-mouse arms race, where Google invests in locking the scans and the attackers seek to unlock them. Although we can learn things of interest during these kinds of wars, we should think that most of the resources in these tech wars are wasted. If we validate the locks—as the DMCA does for copyrighted works but not seemingly for public domain works—we cut off the unlocking tech war. If we don't validate the locks, then scanners will take other steps that we may not like. That may mean that they won't allow downloading generally and instead operate the access to the scans as a service. That possibility would mean that we would not only have DMCA issues but probably issues under the Computer Fraud and Abuse Act as well.

D. CFAA

As passed, the Counterfeit Access Device and Computer Fraud and Abuse Act of 1984 put in place a targeted regime to protect against unauthorized access to computers. That statute was quite detailed but it picked up certain information relating to national defense and foreign relations; certain financial records; and other information on computers operated by the government. That is actually pretty sizable, but amendments to the statute greatly expanded its reach and that is our interest here. The current version of the statute is complicated but to take it in its broadest formulation it treats as a criminal "[w]hoever … intentionally accesses a computer without authorization or exceeds authorized access, and thereby obtains … information from any protected computer."[61] "Protected computer" is a defined term as is "computer" but, to simplify, that now picks up any computer used in interstate commerce.[62] Meaning any computer on the Internet. The targeted regime of the 1984 act has become quite broad and the breadth of that is now being sorted out in the courts.

On July 14, 2011, the federal government issued a four-count indictment against Aaron Swartz alleging wire fraud, computer fraud, unlawfully obtaining information from a protected computer and recklessly damaging a protected computer. Swartz was

[61] 18 U.S.C. § 1030(a)(2)(C).

[62] 18 U.S.C. §§ 1030(e)(1)-(e)(2).

then a twenty-four year old Harvard researcher with a strong record of digital activism.[63] Swartz had authored the Guerilla Open Access Manifesto in which he set forth a call to liberate the world's knowledge from restrictions.[64]

Swartz argued that those with access to materials had a moral obligation to share those materials with others by trading passwords and by downloading files for friends. But Swartz envisioned a more systematic response to what he termed the "private theft of public culture:"

> We need to take information, wherever it is stored, make our copies and share them with the world. We need to take stuff that's out of copyright and add it to the archive. We need to buy secret databases and put them on the Web. We need to download scientific journals and upload them to file sharing networks. We need to fight for Guerilla Open Access.

According to the indictment, over a four-month period in 2010 and early 2011, Swartz set out to do exactly that by downloading millions of scholarly works from JSTOR.

Swartz had the ability to access JSTOR legitimately through his Harvard position, but, as detailed above, that access came with many restrictions. Swartz sidestepped those by going to MIT and accessing their network through a guest network access set up. Somewhat surprisingly, the guest access rules at MIT seemed to allow access to JSTOR and use that access he did to "download an extraordinary volume of articles from JSTOR."[65]

Both JSTOR and MIT responded to this, and this set the usual pattern of response/counter-response that we see in these situations. As Swartz was temporarily denied access to JSTOR and MIT resources, Swartz pursued other paths. He spoofed the MAC address on his laptop—used to provide a unique identifier for someone accessing a network—switched laptops and eventually hard-wired a link into the network in an MIT communications closet. Given the volume of downloads, it was hard for Swartz's activities to go on undetected: In two months, Swartz used one laptop to make two million downloads at MIT, a volume that was more than one hundred times as great as all of the legitimate downloads at MIT combined.[66]

We can now return to our public domain scans. It is clear that the CFAA applies to a much broader set of information than the DMCA. The core concern of the CFAA is the invasion of a computer by an outsider to get at information. Very little of that information may be in copyright, as it may consist of raw facts and, under *Feist*, copyright doesn't attach to facts. A number of CFAA cases have concerned scraping of content from public

---

[63] John Schwartz, *Open-Access Advocate is Arrested for Huge Download*, THE NEW YORK TIMES (July 19, 2011), http://www.nytimes.com/2011/07/20/us/20compute.html.

[64] Aaron Swartz, *The Guerilla Open Access Manifesto*, P2P FOUNDATION BLOG (Aug. 7, 2011), http://blog.p2pfoundation.net/the-guerilla-open-access-manifesto/2011/08/07.

[65] Swartz Indictment ¶ 16.

[66] *Id.* ¶25.

websites allegedly in violation of the terms of service of those websites. Exactly how the authorization provisions of the CFAA work is a matter of controversy and the Ninth Circuit's recent en banc decision in *Nosal* created a conflict in concluding that the CFAA did not apply to violations of terms of use.[67]

***

Taken together, we can see the issues posed by these tools. As described in Section II.C, amassers of the public domain will have strong incentives to restrict use to prevent competition. And beyond direct competition, they will seek to restrict use to preserve an option on future noncompeting uses. That option represents a tax on future innovative activities, though one cabined to some extent by the possibility of tracking down an original artifact containing the public domain work. Of course, that self-help limit works best when we are talking about a single work, especially if the work isn't embodied in a unique physical artifact. The rarer the artifact or the more artifacts that must be accessed, the harder it will be to duplicate.

At the same time, it is very hard for us to assess how much propertization is necessary to ensure that our OPDRs have successful financing models. The common restriction on bulk downloading directly bars the emergence of easy competition at the same scale but also is fully understandable as a way of supporting the financing model of amassing the public domain. Absent that, we would have a classic second-mover advantage collective action problem—"no you scan, no, no why don't you go first"—and we can't all be free riders.

The underlying contract regime then ties into the technical protection regimes of the DMCA and the CFAA. As written, the DMCA seems to offer little direct protection for public domain works subject to DRM-type schemes, though the way in which the DMCA limits overall development of decryption tools may provide some indirect protection for DRM-wrapped public domain works. With the *Nosal* decision, the CFAA seems to be in a state of uncertainty. The CFAA seems to entwine together technical protections and authorization and that takes us exactly to the boundary where our OPDRs will operate.

IV. A Public Public Domain

In the U.S., we are seeing the emergence of a meaningful private public domain through the efforts of Google, JSTOR, ProQuest, the Internet Archive, the HathiTrust and others. This is an interesting and nice mix of for-profit and non-profit organizations. And if we focus on the scope of actual access by the public to the public domain, a great deal has been accomplished. Google has made nearly 3 million volumes available to all for free, subject, to be sure, to advertising and privacy concerns, though, again, GBS can be used fully anonymously. Google could close GBS tomorrow and it is reported that its scanning efforts have slowed down, though that would be expected at some point in the project and

[67] *United States v. Nosal* 676 F.3d 854 (9th Cir. 2012) (en banc).

the slowdown doesn't necessarily reflect the overhang of the ongoing litigation over GBS.

JSTOR has made its slice of the public domain available to the public, though again nothing would prevent JSTOR from taking back and limiting access to the public domain. The ProQuest treasure trove—and I have mentioned only the historical newspapers but ProQuest has much more, including a historical database of company annual reports—is a big piece of the public domain not available to the public generally, though ProQuest is eager to license its databases to public libraries. Of course, libraries make acquisition trade-offs all of the time and a library not licensing ProQuest's public domain materials is making exactly that sort of choice.

But none of these are true public organizations like the great public libraries and each runs an access model that reflects the need to pay the bills. The natural alternative is a government-funded public library, the digital equivalent of the public libraries in communities across the country. In the United States, this might be housed at the Library of Congress—go to www.digitalpreservation.gov to see their efforts—or might be the nascent Digital Public Library of America. And we are likely to see efforts across the globe, such as europeana.eu.

We should step back briefly to consider the broader idea of public property and then situate the public domain in that notion. There is a long-standing idea that certain property is held in a public trust by the government. As the U.S. Supreme Court noted recently: "The public trust doctrine is of ancient origin. Its roots trace to Roman civil law and its principles can be found in the English common law on public navigation and fishing rights over tidal lands and in the state laws of this country."[68] The public trust doctrine operates today to police the boundary between private property and the rights of the public most frequently on waterways and beaches.[69] At its core, this is the idea of a public commons open to all.

We speak about copyright's public domain in exactly those terms but that is a shorthand and one that ignores both the private origins of the public domain and the differences that arise between intellectual property and physical property given the nonrivalnous of the former and the limited nature of the latter. As the public trust fights over beaches demonstrate, beach space is scarce and if a private party is able to fence off part of the beach, the balance of the public loses real access to the beach.

Copyright's public domain doesn't operate that way. First, the content that enters the public domain started its life as private content. The underlying objects in which the work was embodied—paper for years but other media today—were typically privately owned. The public had no right to the blank paper and our hypothetical author's ownership of the paper no more removed it from the public than occurs with any other private property.

---

[68] *PPL Montana, LLC v. Montana*, xxx U.S. xxx, xxx (2012).

[69] See, e.g., *Raleigh Avenue Beach Ass'n v. Atlantis Beach Club, Inc.*, 879 A.2d 112 (N.J. 2005).

The public's right to that paper didn't change when words were first put on paper. Under U.S. copyright law before 1976, an author could create a work and never publish it and thereby retain a perpetual common-law copyright under state law.[70] These private works would never enter the public domain. So unlike the waterways that make up the core of the public trust doctrine, the works that make it into copyright's public domain first started in private hands and were in no sense somehow removed from public hands.[71]

We then turn to the nonrivalnous of intellectual property. As noted before, moving beach-front property out of public hands reduces the public's ability to access that land. In contrast, allowing digitizers to enforce the restrictions that emerge from the tools described in Section xxx above does not remove access to the public domain from the public. The key word there is "remove" as the restrictions clearly limit access relative to the full, unfettered access that would occur from a truly public public domain. After digitization, the public's access to the public domain is almost always expanded: the public continues to have whatever prior access it had to the physical objects in which the public domain is embedded plus the public picks up the access created by the digitizer. There are, to be sure, wrinkles here. We do need to attend to the extent to which holders of public domain objects exit from those objects given the digitization. It is expensive for libraries to manage these physical objects and they may deaccess them—that is libraryspeak for tossing them—given the existence of digital alternatives.

All of that suggest why there is some push for a true public public domain in the form of a national—international?—digital public library. Private digitization will come with restrictions and it isn't clear that anyone is doing a good job of internalizing the issue of how broad access to the public domain should be achieved. Individual libraries destroying old newspapers won't take into account the way in which that action deprives the public of public domain access. There is no obvious steward for the public domain.

A public digital library could take many forms. A full-blown undertaking would mean that the library would create its own search engine for content and would undertake all of the work required to stock the library with digital content. That would mean contracts with current, active copyright holders, some approach to orphan works and scanning of the public domain. The government would also have to build search technology or contract for it. A much more modest effort would be a scanned corpus created by the government and made available to all users. The government could host those works but could also allow private parties to incorporate those scans into their own offerings.

---

[70] For discussion, see *Bobbs-Merrill Co. v. Straus*, 210 U.S. 339, 347 (1908).

[71] Given that, as might be expected, the effort to extend the public trust doctrine to copyright's public domain has so far been unsuccessful. See *Eldred v. Reno*, 74 F.Supp. 2d 1, 3-4 (D.D.C. 1999), *aff'd*, 239 F.3d 372 (D.C. Cir. 2001) *aff'd sub nom*. *Eldred v. Ashcroft*, 537 U.S. 186 (2003). For additional discussion, *see* Maureen Ryan, *Cyberspace as Public Space: A Public Trust Paradigm for Copyright in a Digital World*, 79 Or. L. Rev. 647 (2000).

Describing a public digital library is obviously a large topic and I want to avoid it here, save for one point, namely the relationship between access to the public domain and copyright's deposit requirement. Our first federal copyright statue, the 1790 Copyright Act, brought with it a dual deposit requirement. Authors were required to deposit a copy of the work prior to publication with their local district court and within six months after publication with the Secretary of State.[72] Section 407 sets out the current version of the deposit requirement. Copyright owners are required to deposit with the Copyright Office two copies of the work within three months of publication, unless the Register of Copyrights has exempted the work from the deposit requirement.[73]

The critical question is whether we should think that the deposit requirement is by design or effect the way in which we stock our digital public library. We could imagine this as central quid pro quo for copyright protection: authors would receive copyright protection but do so subject to the requirement that they turn over multiple copies of their work to the government. You can easily imagine how we might adjust this right. At one extreme, the deposit copies would go instantly into our digital public library and would be available for check out. If we treated these as we would physical books, the digital book could only be accessed once it was returned by the prior user. Of course, the transaction costs of check out and return would be much lower; not quite a friction free environment but close.

An alternative approach would be to treat the deposit copies as just stocking the public domain in advance. Think of the deposit copies as a type of public domain escrow. In-copyright works in the escrow wouldn't be available for use by the public until they entered the public domain. At that point, the access problem would be solved and we would, presumably, allow downloading and use of these public domain works without restriction.

The library idea and the public domain escrow notion are obviously quite different. The library could operate as a powerful tax on copyrighted works and a tax borne exclusively by authors. The books in the digital library would substitute for books that would otherwise be purchased by libraries and might substitute for private purchases as well. We don't typically have special in-kind taxes. Law professors need not go teach free law classes as a condition for having full rights to the revenues generated by their regular teaching of classes. The deposit requirement as in-kind tax is a vision of copyright as a special kind of government protection—privilege and not right—and given that status as a privilege one that the government is entitled to attach conditions to. The public domain escrow notion recognizes that assembling the public domain is real work with substantial public benefits.

---

[72] Secs. 3 and 4 of the 1790 Act.

[73] *See Mandatory Deposit*, U.S. COPYRIGHT OFFICE, http://www.copyright.gov/help/faq/mandatory_deposit.html (last updated Jan. 24, 2011).

Conclusion

Like the future, as the saying goes, the public domain is already here, it is just unevenly distributed. The public domain is bound to physical artifacts and even though the public domain is free to use to all, its physical instantiation is tied to particular locations. In some cases, for especially rare works, access may be particularly limited.

But with the rise of mass digitization, actual access to the public domain is growing dramatically. But amassing, digitizing and hosting the public domain takes real resources and it is hardly surprising that the entities doing this work simultaneously impose restrictions on the use of the public domain. In turn, how law responds to those restrictions will determine the structure of competition over the use and delivery of the public domain.

In structuring that competition, we are likely to encounter a wide range of legal tools that implicate that competition. These include the laws of contract, copyright, the DMCA and the Computer Fraud and Abuse Act. Freestanding decisions will be made in each of these areas, often without careful consideration of the consequences for the potential public domain competition that animates this paper.

Readers with comments should address them to:

Professor Randal C. Picker
University of Chicago Law School
1111 East 60th Street
Chicago, IL 60637
rpicker@uchicago.edu

Chicago Working Papers in Law and Economics
(Second Series)

For a listing of papers 1–599 please go to Working Papers at http://www.law.uchicago.edu/Lawecon/index.html

600. Joshua Elliott, Ian Foster, Sam Kortum, Gita Khun Jush, Todd Munson, and David Weisbach, Unilateral Carbon Taxes, Border Tax Adjustments, and Carbon Leakage, June 2012
601. David A. Weisbach, Should Environmental Taxes Be Precautionary? June 2012
602. Saul Levmore, Harmonization, Preferences, and the Calculus of Consent in Commercial and Other Law, June 2012
603. David S. Evans, Excessive Litigation by Business Users of Free Platform Services, June 2012
604. Ariel Porat, Mistake under the Common European Sales Law, June 2012
605. Stephen J. Choi, Mitu Gulati, and Eric A. Posner, The Dynamics of Contrat Evolution, June 2012
606. Eric A. Posner and David Weisbach, International Paretianism: A Defense, July 2012
607 Eric A. Posner, The Institutional Structure of Immigration Law, July 2012
608. Lior Jacob Strahilevitz, Absolute Preferences *and* Relative Preferences in Property Law, July 2012
609. Eric A. Posner and Alan O. Sykes, International Law and the Limits of Macroeconomic Cooperation, July 2012
610. M. Todd Henderson and Frederick Tung, Reverse Regulatory Arbitrage: An Auction Approach to Regulatory Assignments, August 2012
611. Joseph Isenbergh, Cliff Schmiff, August 2012
612. Tom Ginsburg and James Melton, Does De Jure Judicial Independence Really Matter? A Reevaluastion of Explanations for Judicial Independence, August 2012
613. M. Todd Henderson, Voice versus Exit in Health Care Policy, October 2012
614. Gary Becker, François Ewald, and Bernard Harcourt, "Becker on Ewald on Foucault on Becker" American Neoliberalism and Michel Foucault's 1979 *Birth of Biopolitics* Lectures, October 2012
615. William H. J. Hubbard, Another Look at the Eurobarometer Surveys, October 2012
616. Lee Anne Fennell, Resource Access Costs, October 2012
617. Ariel Porat, Negligence Liability for Non-Negligent Behavior, November 2012
618. William A. Birdthistle and M. Todd Henderson, Becoming the Fifth Branch, November 2012
619. David S. Evans and Elisa V. Mariscal, The Role of Keyword Advertisign in Competition among Rival Brands, November 2012
620. Rosa M. Abrantes-Metz and David S. Evans, Replacing the LIBOR with a Transparent and Reliable Index of interbank Borrowing: Comments on the Wheatley Review of LIBOR Initial Discussion Paper, November 2012
621. Reid Thompson and David Weisbach, Attributes of Ownership, November 2012
622. Eric A. Posner, Balance-of-Powers Arguments and the Structural Constitution, November 2012
623. David S. Evans and Richard Schmalensee, The Antitrust Analysis of Multi-Sided Platform Businesses, December 2012
624. James Melton, Zachary Elkins, Tom Ginsburg, and Kalev Leetaru, On the Interpretability of Law: Lessons from the Decoding of National Constitutions, December 2012
625. Jonathan S. Masur and Eric A. Posner, Unemployment and Regulatory Policy, December 2012
626. David S. Evans, Economics of Vertical Restraints for Multi-Sided Platforms, January 2013
627. David S. Evans, Attention to Rivalry among Online Platforms and Its Implications for Antitrust Analysis, January 2013
628. Omri Ben-Shahar, Arbitration and Access to Justice: Economic Analysis, January 2013
629. M. Todd Henderson, Can Lawyers Stay in the Driver's Seat?, January 2013
630. Stephen J. Choi, Mitu Gulati, and Eric A. Posner, Altruism Exchanges and the Kidney Shortage, January 2013
631. Randal C. Picker, Access and the Public Domain, February 2013