



ROBINSON MEYER - Robinson Meyer is a writer and musician based near Chicago and a regular contributor to TheAtlantic.com.

MIT Economist: Here's How Copyright Laws Impoverish Wikipedia

By Robinson Meyer

JUL 18 2012, 4:45 PM ET 7

Using a little-known copyright rule and a trove of baseball-related trivia, an MIT economist figured out how current copyright laws specifically affect one online community.



Public Domain/Baseball Digest

Everyone knows that the flow of information is complex and tangled in society today -- so thank goodness for copyright law! Truly, no part of our national policy is as coherent, in the interest of the public or as updated for the Internet age as that gleaming tome in the US Code.

Not.

Unless you're reppin' the MPAA, you probably know that the modern copyright regime doesn't work. You don't have to believe in radical copyleftism -- or even progressivism -- to understand this. But it's hard to know how the current body of law governing copyright and intellectual property affects individual works, simply because of the way communication, and ideas in general, work. One thing connects to another, and pulling apart the causes from the effects requires an Aristotle-like familiarity with contemporary culture.

But one MIT economist, Abhishek Nagaraj*, who recently presented his work at Wikimania, has **found a way to test how the copyright law affects one online community -- Wikipedia -- and how digitized, public domain works dramatically affect the quality of knowledge.**

How? The story begins in 2008. That year, Google Books digitized a number of magazines, including *Ebony*, *Popular Mechanics* and *New York*. Google also digitized the oldest and longest-running journal of matters baseball-related: *Baseball Digest*, published since 1942 in Evanston, Illinois. A huge number of issues, July 1945 to 2008, had gone online. And the magazines were full of images of the players.

A small group of Wikipedians, dedicated to improving the project's baseball articles, discovered the trove. Their editing, plus the huge, new body of baseball knowledge, soon dramatically improved the encyclopedia. After the digitization, Nagaraj found articles on four decades of All-Stars between 1944 and 1984 **grew by about 5,200 words per article.**

But his research was able to go further. Because of a small clause in copyright law, all the issues of *Baseball Digest* from before 1964 had fallen in the public domain -- meaning, that though all of the *Baseball Digest* articles from 1944 to 1984 were online in full on the *Baseball Digest* site, Wikipedia editors could only use the images from the earlier years. So Nagaraj created, from his set of All-Stars, two historical sets: a "control" group of players who first played in a game between 1964 and 1984 (and thus likely have *Baseball Digest* material that remains privately-owned), and a "treatment" group of All-Stars who first played in the big game between 1944 and 1964.

By comparing the two groups, Nagaraj could see the direct effects of copyright on the articles in terms of length, number of images, and traffic. That first metric -- length -- proved resilient to the copyright divide. Words are easy to rescue from private-ownership, and the Wikipedia authors simply rewrote the information still owned by the *Digest*. *Every* article, post-digitization, became on average much longer.

But Nagaraj found was that the availability of public domain material dramatically improved the article's *images*. Before the digitization, players from between '44 and '64 had an average of .183 pictures on their articles. The '64 to '84 group had about .158 pictures. But after digitization, those numbers dramatically changed: there were 1.15 pictures on each of the older group's articles -- but only .667 in the new group. **More recent players, covered by privately-owned parts of *Baseball Digest*, had half as many images on their pages as did old-timers.**

And the effects of this -- of just having an image on the page -- cascaded to other metrics. "Out-of-copyright" players' pages saw **a significant boost in traffic**. Articles from the pre-'64 that

were already in the top 10 percent saw their hits increase more than 70 percent. Articles from that group in the least-popular ten percent saw traffic to their articles increase by 25 percent. Those pages were more frequently edited across the board, too. And this makes sense: Google rewards updated content, and it rewards images. The out-of-copyright players provided more of both.

Nagaraj controlled for much in his study: the talent of players, their left-handedness, the duration of their careers, and he even controlled for the general drop-off of editing on Wikipedia. His report is clear: **Copyright law affects to some degree what information makes its way onto Wikipedia, but what it more strongly affects is how we use that information once it's there. In other words, digitizing *any* knowledge increases an article's text, but only digitizing public domain images makes articles more frequently updated and visited.** This may be in part due to the particularities of Google's algorithm, which rewards updates and images. Nagaraj is studying this next, in fact, comparing an article's Page Rank to its *Digest* copyright status.

And those results are exciting, because Nagaraj's found a way to do something rare. His *Baseball Digest* is a probe we need: into how copyright law controls one community, into how it impoverishes one set of knowledge, and into how it makes all knowledge less usable. Ain't no *Nicomachean Ethics* required.

** In the original version of this article, Abhisek Nagaraj's first name was missing, and there were typos present. Both have been fixed.*