

Wikipedia readers get shortchanged by copyrighted material

February 13 2017



When Google Books digitized 40 years worth of copyrighted and out-of-copyright issues of Baseball Digest magazine, Wikipedia editors realized they had scored. Suddenly they had access to pages and pages of player information from a new source. Yet not all information could be used equally: citations to out-of-copyright issues increased 135 percent more than issues still subject to copyright restrictions.

Those are the results of a new study, "[Does Copyright Affect Reuse? Evidence from Google Books and Wikipedia](#)," conditionally accepted in *Management Science*. By studying how [copyright](#) laws restrict the free exchange of [information](#), author Abhishek Nagaraj also found pages that could benefit from copyrighted information received 20 percent less traffic than pages that could benefit from out-of-copyright information. That presents a significant disadvantage to Wikipedia readers.

Copyrighted images suffered even more lack of distribution or reuse because they cannot be paraphrased and repurposed like written information.

Perhaps more importantly, the study's findings suggest how an Internet without copyrighted material may be better used to create new content, and not just allow people to consume what's already out there.

"There is a big debate about what copyright restrictions do to the diffusion of knowledge. Some people say copyright laws have not caught up with the digital age," says Nagaraj, an assistant professor of management at UC Berkeley's Haas School of Business.

With just about everything available online now, Nagaraj chose to study *Baseball Digest* for several reasons. First, it is one of only a small number of publications that Google Books digitized in its entirety in 2008. Second, *Baseball Digest*'s copyright status changed over time; the copyright of issues published before 1964 was never renewed and therefore, all pre-1964 issues entered the public domain 28 years after their respective publication dates. At the same time, issues published in 1964 and after are not subject to renewal and remain under copyright, at least until 2020. These conditions gave Nagaraj the ability to study citation variation—under copyright and not under copyright—of the same publication. Third, Nagaraj contends that baseball's popularity would make his experiment "economically meaningful."

Nagaraj created two samples based on the digest's publication years and on 541 players' Wikipedia pages. The players were all nominated for the Baseball Hall of Fame and made their professional debuts between 1944 and 1984. By creating a "quality metric" for each player based on the number of times they played in an all-star game, Nagaraj ensured that each player in the sample had a significant baseball career. The result was a dataset that counts the number of citations to *Baseball Digest* on

each player's Wikipedia page as well as the number of images and word citations.

The data revealed three primary results: 1) There was no variation in using information from copyrighted and out-of-copyright sources before the Google Books digitization process; 2) After Baseball Digest was digitized, Wikipedia editors started using both non-copyrighted and copyrighted information but more so of the former; and 3) The effects varied by the type of content. Text material was reused regardless of its copyright status. For example, factual information that Babe Ruth hit a homerun moved from the Digest to Wikipedia smoothly because it could be rewritten. However photos of players and teams were reused more rarely because they could not be reproduced with any variation unrestricted by copyright protection.

"Well-known players like Yogi Berra were less affected by this variation because there are enough alternative sources of information besides Baseball Digest," explains Nagaraj. "But there are many players for whom we have limited information. People seeking information about these players are most hurt by [copyright law](#)."

This deficiency in the transfer of knowledge impacts not only Internet users who are looking for information but also users seeking to create new content. Nagaraj hopes his work will provide evidence for re-evaluating the value of copyright laws.

"The loss from future copyright extensions is likely to be high. If we want to incentivize new creative work using historical information, we need to fix the system," says Nagaraj.

Provided by University of California - Berkeley

Citation: Wikipedia readers get shortchanged by copyrighted material (2017, February 13)
retrieved 20 April 2024 from

<https://phys.org/news/2017-02-wikipedia-readers-shortchanged-copyrighted-material.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.