

## Discovery, Cannibalization, or Both: How Digitization Affects Demand for Physical Copies

**Author :** Christopher J. Buccafusco

**Date :** April 16, 2019

Abhishek Nagaraj & Imke Reimers, *Digitization and the Demand for Physical Works: Evidence from the Google Books Project* (2019), available at [SSRN](#).

From 2004 until 2009, the Google Books Project (GBP) digitized thousands of books from the collection of Harvard University's library and made them available online. According to Google and proponents of the GBP, digitization would introduce readers to books that they otherwise couldn't find or obtain, increasing access to and interest in the digitized works. But according to some authors and publishers, the creation of free digital copies would usurp the demand for print copies, undermining an important industry. This dispute was at the heart of a decade of litigation over GBP's legality. After all of that, who was right?

According to a recent empirical study by economists Abhishek Nagaraj and Imke Reimers, the answer is: both of them. The paper, [Digitization and the Demand for Physical Works: Evidence from the Google Books Project](#), combines data from several sources to reveal some key features about the effects of digitization on dead-tree versions of books. The story they tell suggests that neither of the simple narratives is entirely correct.

Google worked with Harvard to scan books from 2004 to 2009, proceeding in a largely random fashion. The only limitation was that Google only scanned books that had been published prior to 1923, because these works were in the public domain and, thus, could be freely copied. Works published in 1923 or later might still be covered by copyright, so Google chose not to scan those initially. Nagaraj and Reimers obtained from Harvard the approximate dates on which the pre-1923 books were scanned.

Harvard also provided them with the number of times between 2003 and 2011 that a book was checked out of the library. Book loans serve as one of the ways in which consumer demand for books is supplied, so these data enabled the researchers to test whether digitization affected demand for printed versions of works. The researchers also obtained sales data for a sample of approximately 9,000 books that Google digitized, as well as data on the number of new editions of each of these books. With these data, Nagaraj and Reimers engage a difference-in-differences method to compare loans and sales of digitized books to those of non-digitized books, before and after the year in which books were digitized.

If the GBP's opponents are correct, then digitization should lead to a decrease in loans and sales, as cheaper and more easily accessed digital versions substitute for physical copies of books, especially for consumers who prefer digital to physical copies. According to the substitution theory, consumers basically know which books they want, and if they can get them for free, they will. If GBP's proponents are correct, by contrast, consumers do not always know which books they want or need, and finding those books can entail substantial search costs. Digitization reduces the costs of discovering books and will lead some consumers to demand physical copies of those books.

Nagaraj and Reimers find that digitization *reduces* the probability that a book will be borrowed from the library by 6.3%, reducing total library loans for digitized books by about 36%. Thus, some consumers who can get free and easy digital access choose it over physical access. The figures for marketwide book sales are, however, reversed. Digitization *increases* market-wide sales by about 35% and the probability of a book make at least one sale by 7.8%. Accordingly, some consumers are finding books they otherwise wouldn't have and are purchasing physical copies of them.

To further explore these effects, Nagaraj and Reimers disaggregate the data into popular and less popular books, and here the effects are starker. For little known works, digitization drastically decreases the costs of discovering new titles, and consumers purchase them at a 40% higher rate than non-digitized books. Discovery benefits trump substitution costs. But for popular works, where digitization does little to increase discovery of new works, sales drop by about 10%, suggesting substantial cannibalization.<sup>1</sup>

What do these findings mean for copyright law and policy? One implication is that substitution effects may not be that great for many works *even when the whole work is available*. Thus, the substitutionary effect of Google's "snippet view," which shows only about 20% of a work should be much smaller still. Also, it's important to realize that these data help prove that otherwise forgotten or "orphan" works still have substantial value, if only people can find them. Consumers were willing to pay for less popular works, once they discovered their existence.

Ultimately, however, because the data does not tell a simple story, they may not be able to move the legal debate much. The study confirms both publishers' fears about the works they care the most about (popular works) and GBP's proponents' hopes about the works they care the most about (orphan works). One possibility, however, is that we may see a more sophisticated approach to licensing works for digitization. Publishers may be more willing to allow Google or others to digitize unpopular works cheaply or for free, while choosing to release popular titles only in full price editions. This could provide the access that many people want to see while enabling publishers to stay in business.

1. I find one of the authors' robustness checks unconvincing. They consider the effect of digitization using the 1923 public domain date as a discontinuity to look for differences in loans and sales. The periods they consider for loans and sales are 2003-04 and 2010-11. They find that relative to 2003-04, post-1923 were loaned more often in 2010-11, and that relative to the earlier period, post-1923 were sold less often in the later period. I doubt the explanation, because the recession of 2008 occurred between the two periods, substantially decreasing consumers' willingness to pay for works, and thus, increasing their willingness to borrow them. Because post-1923 books are more expensive than pre-1923 books, a change in consumer willingness-to-pay would produce exactly the results that the authors demonstrate.

Cite as: Christopher J. Buccafusco, *Discovery, Cannibalization, or Both: How Digitization Affects Demand for Physical Copies*, JOTWELL (April 16, 2019) (reviewing Abhishek Nagaraj & Imke Reimers, *Digitization and the Demand for Physical Works: Evidence from the Google Books Project* (2019), available at SSRN), <https://ip.jotwell.com/discovery-cannibalization-or-both-how-digitization-affects-demand-for-physical-copies/>.