

June 2010

Hurting Toward the Finish Line: Should the Google Book Settlement Be Approved?

Ivy Anderson

University of California, Oakland

Follow this and additional works at: <https://docs.lib.purdue.edu/atg>



Part of the [Library and Information Science Commons](#)

Recommended Citation

Anderson, Ivy (2010) "Hurting Toward the Finish Line: Should the Google Book Settlement Be Approved?," *Against the Grain*: Vol. 22: Iss. 3, Article 6.

DOI: <https://doi.org/10.7771/2380-176X.5545>

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact epubs@purdue.edu for additional information.

Hurling Toward the Finish Line: Should the Google Book Settlement Be Approved?

by Ivy Anderson (Director of Collections, California Digital Library, University of California, Oakland, CA 94607)



(The image presented here is a mashup: for the actual book in HathiTrust, see <http://hdl.handle.net/2027/uc1.b543888>)

Recently, **Google** and the plaintiffs filed their final briefs in defense of the **Google Book Amended Settlement Agreement (ASA)** that is before the New York Southern Federal District Court. As the rhetoric around the **Settlement** heats up to white-hot intensity, I'd like to offer a few personal thoughts from my vantage point at the **California Digital Library**.

The University of California Context

The **University of California Libraries** are **Google's** second-largest library digitization partner; we are also the second-largest book digitization partner of the **Internet Archive**, thanks to generous funding in the past from **Microsoft**, **Yahoo**, the **Alfred P. Sloan**, and **Kahle/Austin Foundations**, and other sponsors. In all, **UC Libraries** have now digitized 2.5M books from their collections through these projects, both in- and out-of-copyright. Within the **University of California**, some of our closest faculty colleagues are also among the **Settlement's** most prominent critics.

In our view, the proposed settlement is hardly perfect; as **Google** acknowledges in its brief, it's a compromise among parties with differing agendas and motivations. **CDL** is a staunch supporter of the underlying aims of the **Google Book** project to make the knowledge enshrined in the world's great libraries discoverable and accessible across the globe, and we support the public benefits that will ensue, including the benefits to libraries, if the **Settlement** is approved. At the same time, public criticism has been good for the **Settlement**, producing very real improvements in the amended version that is now before the court — improvements that would not have been made without that criticism. Long live democracy!

Digitization Partnerships: The Promise and the Peril

Like many of the objectors, participating libraries went through their own period of outrage and indignation when details of the **Settlement** first came to light. What! We would have to buy back access to our own books? Why did **Google** let us down in abandoning its fair use defense? Why should the parties be allowed to create an artificial revenue model for works that are long out of print, books that would no longer exist at all outside of used bookstores, if the libraries themselves hadn't purchased and maintained them at great expense over decades and indeed generations? How can they do this without our agreement as to terms, since it is we who have made these books available to them in the first place? Hasn't our stewardship paid for these books many times over? Isn't this why copyright law contains unique exceptions for libraries, in recognition of our mission to further the public good? Wasn't the appropriate use of our own copies in light of fair use principles our decision to make?

The problem with this view, of course, is that libraries did not initiate this enterprise, and we are not its only beneficiaries. The **Google** project placed two sets of commercial interests as loggerheads, with copyright law in the middle. Admittedly, libraries took a risk in engaging in a partnership so legally entangled.

But let's be honest: though few seem willing to admit it, revitalizing the world's heritage of books for a digital age — a task that many considered impossible only a few short years ago — appears within reach today almost entirely due to **Google's** enterprising vision. Even the **Open Content Alliance**,

which **CDL** joined a year before becoming a **Google** partner, was in some sense a response to **GBS** (although it had other important antecedents, as well, thanks to **Brewster Kahle's** equally inspired vision). When **Google's** competitors withdrew their support for that project, no other funders stepped in to fill the breach. The plain fact is that despite the idealistic adurations of some, the resources required to digitize our cultural book heritage on a grand scale are not likely to be marshaled in the U.S. by libraries and the public sector alone.

At **CDL**, we've done some estimating of what it would take to convert the roughly 15 million unique books in **University of California** library collections to digital form absent the **Google** enterprise, using the best alternative technology available today. The answer? Half a billion dollars, and one and a half centuries. And that is just the **University of California's** books.

I like to compare this to the building of the great Temple of the Sagrada Família in Barcelona, a city with which my family has an ancestral connection. When my husband's grandmother left Barcelona as a young girl in the late 19th century, the Sagrada Família had barely erected its first stone. In 2006, more than 125 years later, her great-granddaughter traveled to Barcelona for the first time, where she was able to observe **Gaudi's** monumental edifice, still under construction. At this writing, completion is projected for 2026.

The speed at which **Google** is converting this content is not without costs of its own. **Google's** iterative approach to building large-scale services has drawn criticism from some scholars accustomed to work that is honed and polished before it is released. This is, in part, an argument about means, not ends. Like those progressive JPEG images that start out blurry on the screen and become sharper as the details fill in, **Google's** services are improving over time as it continually upgrades and enhances its images and metadata. Over time we will be able to replace those missing or still-blurry pages with better versions. Where the value of the content warrants it, we can selectively invest in more meticulous rendering, textual markup, and other enhancements.

Two cases are illustrative here. **CDL** has digitized a large number of public domain books with the **Internet Archive**, some of which have also been digitized in our **Google** partnership. Although **CDL** had to suspend its **Internet Archive** book scanning project earlier this year after **Microsoft** withdrew its support and additional grant funding proved elusive, we have every expectation that we will take up comparable projects with **Internet Archive** in future, because its technology is better suited to certain types of uses (better artifactual rendering, for example). The **Early English Books Online (EEBO)** database

continued on page 20

marketed by **ProQuest** is another example in which, through an innovative partnership with libraries and scholars, basic scans are enhanced with detailed markup for a subset of carefully-selected works.

In the meantime, a great deal of value is already being derived from the **Google** work as it stands today. Students and scholars report finding much formerly-hidden material, journalists and etymologists are mining its content for historical information, and even some of **Google's** severest critics have said that they can no longer imagine life without **GBS**. This is neither an either-or proposition nor a zero-sum game. All of these services are fulfilling a niche, along with libraries, in a new information ecology that we are only now beginning to understand even as we participate in its unfolding.

Settlement Pro and Con

So in the long run, is the **Google Settlement** a good thing, or a bad thing? Before answering that question, let's look at just a few of the major criticisms that have been levied against the **Settlement**.

The Institutional Subscription will be come too expensive because it has no meaningful competition. Well, it's hard to know that, of course. In fact, we don't even know today how many books it will contain, nor what the scan or OCR quality of the content will be, given the variability of the overall corpus. But we do know that there are at least three checks on the institutional subscription price that should mitigate price-gouging. First, the broad distribution requirement in the **Settlement's** dual objectives means that prices cannot become so high that few choose to subscribe. Second, libraries themselves are savvy evaluators and negotiators of online content who can be expected to evaluate this offering rigorously and skeptically, and to eschew a subscription unless the price is acceptable for the benefit derived. Since none of us knows how our users will engage with this material, these assessments ought to be conservative. Third, the provisions for pricing arbitration built into the agreements between **Google** and the participating libraries will allow them to challenge price increases that they deem unwarranted; a provision that is intended to be exercised not on the basis of narrow self-interest among a small set of contributing libraries but on behalf of all libraries.

Academic authors want to release their books, not see them locked up. Indeed, no disagreement here; and the amended **Settlement** now explicitly provides for this (according to **Google** and the plaintiffs, this was always possible, but in the **ASA** it is now called out). We intend to work proactively with rights holders who would like to enable broader access to their books and to develop mechanisms that can help to make this straightforward.

The Settlement will give Google a monopoly over orphan works and is anti-competitive. It's hard for me to see how **Google's** activities to date can be viewed as anti-competitive when **GBS** is almost single-handedly

responsible for the eBook explosion that is swirling all around us, with new entrants popping up every day. That may be a controversial assertion, but eBooks and eBook readers were a languishing backwater until **Google** stimulated the market by putting books online through its library and publisher partner programs. If anything, **Google's** entrance into the retail space is likely to engender even fiercer competition. It seems cynical at best for rival behemoths **Microsoft** and **Amazon** to decry **Google's** impending monopoly over a sliver of the eBook market — much of it of uncertain commercial value — under the noble-sounding rubric of the **Open Book Alliance**. But then, competition makes strange bedfellows. As to orphan works, the **Settlement** should, if anything, goad us all the more toward a legislative solution. It is as irksome to me, as it is to other critics, that **Google** should be uniquely empowered to collect royalties on behalf of absent rights holders who may have long ago relinquished any economic interest in their works. Still, the **ASA** addresses this in a far more satisfying manner than its predecessor. Finding a better long-term solution to the orphan works problem is something we all can get behind.

When the purposes that we first envisioned when embarking on these projects — all arguably fair uses of this content — are reviewed against the **Settlement** impacts, it's hard to view the **Settlement** as anything but a positive development. More books will be available in full view, both to libraries and to consumers. New services will be developed for print-disabled users and for large-scale computational analysis, further unlocking digitization's transformative potential. Disclosure of rights information through a central registry (at least for U.S. books) is likely to have far-reaching impacts, facilitating the eventual orderly release of books into the public domain. **Google's** competitors are likely to join the push for orphan works legislation, increasing its chances of success. And with the **Settlement** behind us, we can all proceed in an environment of greater certainty.

What if the Settlement is not approved?

For libraries, failure of the agreement would hardly be a crisis. The benefits that we initially envisioned — improved discovery and full-text search of our vast legacy collections, and broad public availability of works that are out-of-copyright or otherwise released by their copyright owners — will still be realized. The fears of some **Settlement** objectors — of monopolistic pricing and the forced commercialization of materials that are long out-of-print — will melt away like the elusive Vancouver snow. Participating libraries may still choose to undertake novel services, without the unwelcome restrictions imposed by the **Settlement**. As long as **Google** and others continue to partner with us, we will go forward in reinvigorating our collections for a new digital age.

The **Google Settlement** is fundamentally about whether **Google** and rights holders will be allowed to implement a particular set of business

models for a certain set of books. I believe the **Settlement** should be approved, because it will create new and valuable services for libraries as well as consumers. But many of **Google's** participating libraries have their own plans for these books, plans that do not ultimately depend on the outcome of the **Settlement**. The greatest risk for libraries if the **Settlement** is not approved is that further legal setbacks might lead **Google** to abandon its interest in library digitization altogether. If that were to happen, a unique opportunity would be lost that is not likely to be repeated in our lifetime.

Life Beyond Google Book Search

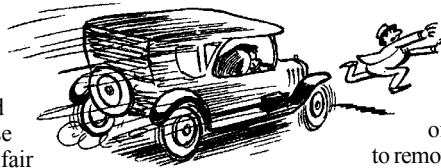
What of our relationship to the **Google Book** project itself? Some of the concerns we hear from faculty have nothing to do with the **Settlement per se**, but rather with the long-term implications of **GBS** for library collections and services. Let me close with a few words about some of those concerns.

To our scholars who worry that we are about to throw our physical collections overboard in favor of digital surrogates of sometimes uneven quality, I want to say: not to worry. True, libraries everywhere find themselves having to consign more and more of their physical collections to remote storage as campus space

grows increasingly scarce and user preferences migrate online. And some libraries — the **UC's** far less than others — are addressing the space crunch by de-accessioning low-use materials that are widely held with the knowledge that they can borrow these items from another library if need be. (Many cooperative initiatives are now underway to share such information and ensure that enough copies are retained throughout the nation's system of libraries to protect the integrity of the scholarly record.) That train has already left the station, and it's happening independently of large scale digitization. What digitization offers is a valuable complementary mitigation strategy: we can now make those remote collections eminently browsable, saving time and expense both for users and for libraries. As a library user, you can now determine whether that book is really what you're looking for before you request it, not afterward — and in some cases, the digital surrogate may indeed be all that you need. Libraries can promote these "hidden" volumes more effectively to their users, while limiting delivery costs to just those items that are truly wanted. This browsable and/or searchable digital surrogate — which is the quality level that most of the **Google** mass digitized scans are aimed at — is not a replacement for the original print book, and was never intended to be.

To our scholars who worry that we are outsourcing our library collections and services to **Google**, again I want to say: please don't worry on this score, either. Far from abrogating our mission as stewards of the cultural record, we who have opened up our collections to digitization are shouldering this role with vigor. While **Google** and others are making these books discoverable online to a general audience, the **University of California**, along with other

continued on page 22



peer institutions, is creating a robust shared access and preservation service for our mass digitized books, one that adheres to professional standards, through our partnership in a groundbreaking enterprise called the **HathiTrust**. If you haven't heard of **HathiTrust** yet, you soon will. No UC library user need go to **Google** to search the full text of our books, or to find accurate bibliographic information, or to view and download those that are in the public domain: s/he can go to <http://catalog.hathitrust.org/> and be reassured that those books will be there, in ever-improved versions, for the long-term. **HathiTrust** now numbers 5.4 million volumes from 26 libraries and is growing at a rapid rate, all searchable, all viewable if in the public domain (or otherwise rights-cleared), and all designed to inure to the long-term benefit of the nation's libraries and their users. The digital library of the future resides not with **Google**, but with us. And we are building it today.

At the same time, **Google**, **Internet Ar-**

Ivy Anderson was appointed to the **California Digital Library** effective January 2, 2006. In this role, **Anderson** coordinates and provides systemwide leadership in buying and licensing digital content. Her efforts also extend beyond licensed information, to the acquisition and management of select printed materials and reformatting. She also plays an important part in the university's strategic goal of addressing the profound problems in the marketplace for scholarly publications. Prior to joining the **CDL**, **Anderson** worked at **Harvard University Library**, where she developed a digital acquisitions program. She is widely known for her work in advancing the goals of libraries and library consortia. 🐼

chive, and others, are providing an invaluable service in bringing the vast holdings of the great research libraries to a worldwide audience and integrating that content with general-purpose internet search services and other content. As one colleague has written, "Who among us has not benefited from a **Google** search?" In participating in these efforts, we are fulfilling our long-standing public service mission. The **Google Settlement**, if approved, will further these aims by providing more content, in more ways, to an even wider audience.

But in the end, approval of the Settlement is not a make or break event for libraries. Despite the claim that the **Google Settlement** promises

to build "the greatest library in history," libraries are not leaving the future of information to **Google** and these other partners alone. Nor need we wait, Godot-like, for fugitive national legislation to begin the work of serving up our cultural heritage in digital form. Through a combination of efforts, including public-private partnerships such as that of libraries with **Google**, we can go forward in this transformative enterprise together.

This piece was initially published in **Anderson's** blog, at: <http://www.cdlib.org/cdlinfo/2010/02/16/hurling-toward-the-finish-line-should-the-google-books-settlement-be-approved/>. 🐼

Booklover — Sea

Column Editor: **Donna Jacobs** (Research Specialist, Transgenic Mouse Core Facility, MUSC, Charleston, SC 29425) <jacobsd@muscu.edu>

On the last Sunday of April, in Mt. Pleasant, SC, the shrimp trawlers nose out of Shem Creek festooned with brightly colored flags. They parade by the pier at the new Memorial Park to receive their blessing. Many feel this annual event is essential, prior to the beginning of the shrimping season. Each boat is called by name and given a special blessing for good harvest and safe return to harbor. Then the festivities begin. Music, food, shrimp eating contests, dancing, and arts and crafts entertain the crowd that has gathered to witness this fishing community tradition.

My husband and I used to be a part of this shrimping community. We owned a shrimp trawler that was tied up at the Geechee Seafood Dock on Shem Creek. Every spring brought an intense time of repair and maintenance culminating in our blessing. The season opened soon after the event, and the long hours of trawling for shrimp began. We have long ago traded our life in the mercurial shrimping community for a somewhat more regular life in the marine research field. But every year during spring there is the intense time of repair and maintenance before my husband returns to the sea, taking the marine biologists wherever along the South Atlantic coastline they want to go. His love for the sea extends from his profession into his recreation, as he loves to fish and dive.

The love affair that men have with the sea is like no other, and two short novellas, *The Old Man and the Sea* by **Ernest Hemingway** and *The Story of the Shipwrecked Sailor* by **Gabriel García Márquez**, completely immerse the reader in the power and seduction of the sea. **Hemingway's** novella is probably one of the best known works of a **Nobel Literature Laureate**, as well as of the author. The lesser known story by **García Márquez** is remarkable in that it is true.

The Old Man and the Sea accompanied me on a recent trip to Boston. Periodically it is nice to reacquaint yourself with an old friend. It is not long before I am balancing myself in a fishing boat, smelling the salt air, and hoping for a good catch.

Everyone who has ever made a living from the sea can identify with **Santiago** as the old fisherman feels the tug of the fishing line after over 80 days without a fish. In simple declarative sentences, **Hemingway** gives the reader the intensity of every ache as **Santiago** first prays for the fish to come, then patiently triumphs over the fish, and ultimately loses his prize to the sharks. **Hemingway's** remarkable use of metaphors and imagery creates a story that commands a large presence in the literary world, not unlike the fish that is lashed to the side of **Santiago's** boat. Everyone who reads this gem of fiction has been touched by the novella and left with a sense of the sea's fickle nature.

While in Boston I stumbled upon the **Brattle Book Store** on West Street. Established in 1825 it claims to be the largest bookstore in the New England area and one of the oldest and largest antiquarian bookstores in America. Boston is a place where you could spend your life exploring the birth of independence, America's revolutionary history, and yet I gravitate to the bookstore. There I found *The Story of the Shipwrecked Sailor* by **Gabriel García Márquez**. **García Márquez** is one of my all time favorite authors so I decided to add this little book to my collection. On the return flight to Charleston I had the good fortune to be upgraded to Business Class, where I settled into the comfort of a large seat with ample legroom and began this tale of the shipwrecked sailor. Every seafarer's family members' nightmare is to learn of a loss at sea. Life rafts, flotation devices, GPSs, EPRBs, and radio communication are among the safety devices aboard every vessel. But the sea is temperamental, and what is beautiful can become deadly in an instant.

Luis Alejandro Velasco is stationed aboard the Colombian military destroyer *Caldas*. The ship is being repaired in Mobile, Alabama and will be returning to its home port of Cartagena de Indias, Colombia. As the date for departure nears, **Luis** is burdened with a strong sense of foreboding that he cannot reconcile. He goes through the motions of

good-byes in his last days in Mobile and boards the ship. His "radar" continues to tell him that something is off as the ship feels funny under his feet. All of the American "gifts" that crew members have purchased for their families back home in Colombia are

stowed on the deck and, thus, the seaworthiness of the *Caldas* is compromised. **Luis** along with eight other crew members find themselves struggling on the deck during a violently rough sea. A wave washes over the deck, and all are instantly swimming. They fear the boat has sunk, but as **Luis** reaches a life raft he sees the *Caldas* crest a wave and continue along its course. One by one his shipmates are lost as the ones he can hear and see are unable to make it to the raft in the boiling sea. For the next 10 days the reader shares his tiny life raft with no water, food, or protection from the grueling sun, all told in simple declarative sentences. **García Márquez** was the ghost writer for **Luis's** recounting, and the story of this phenomenal event was published in 1955 as a series of newspaper articles in the *El Espectador*. The revelation that contraband was being transported on the military vessel was an embarrassment to the dictatorial government of **General Gustavo Rojas Pinilla**, which resulted in the closure of the newspaper and led to **García Márquez's** "nomadic and somewhat nostalgic exile that in certain ways also resembles a drifting raft." In 1970 when **García Márquez** wrote the forward to the book which compiled the series he mused that the publisher was more interested in **García Márquez's** notoriety as a **Nobel Laureate** than the incredible story which he had the opportunity to recount for the world.

So now it is April, and the sea once again sings a siren's song to the fishermen, sailors, and seafarers. My husband packs for his trip, and I have stashed my copy of *The Story of the Shipwrecked Sailor* in his bag. I will pray for his safe return and that he will never be lost at sea, only lost to the lure of the sea. 🐼

