

A

The Atlantic

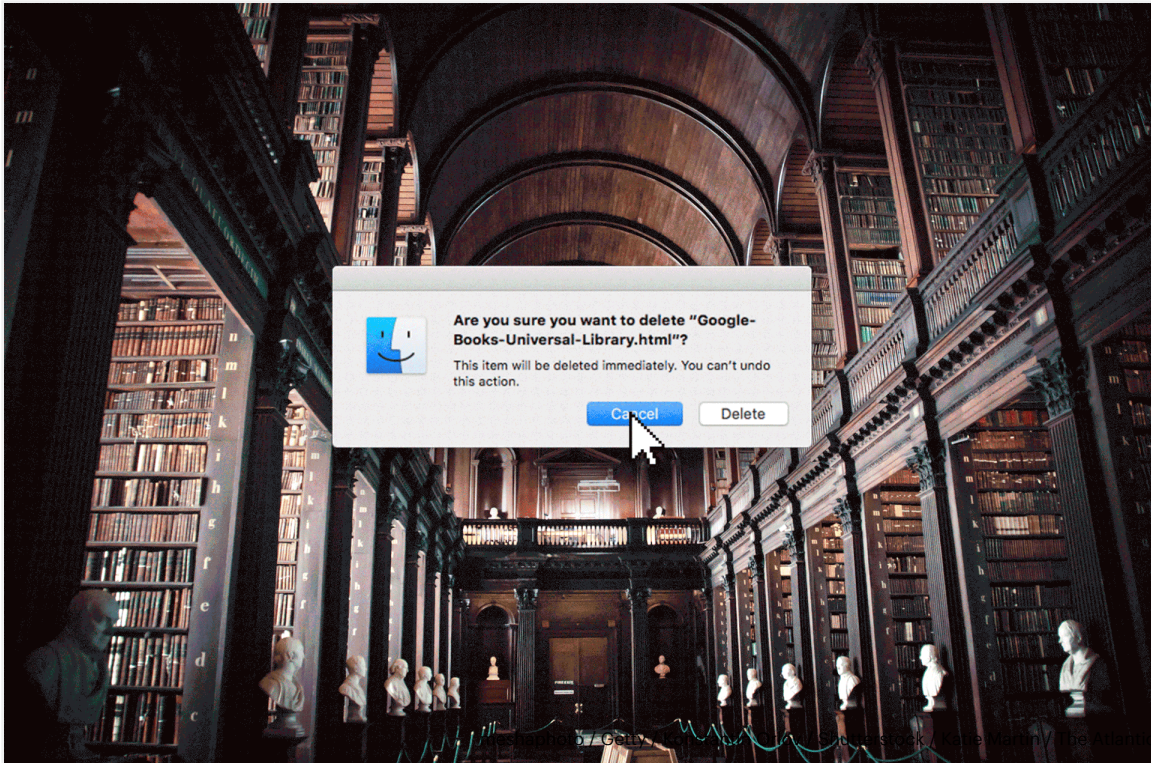
Sign In

Subscribe

The Atlantic

Torching the Modern-Day Library of Alexandria

“Somewhere at Google there is a database containing 25 million books and nobody is allowed to read them.”



Story by James Somers

| APRIL 20, 2017

| TECHNOLOGY



YOU WERE GOING to get one-click access to the full text of nearly every book that's ever been published. Books still in print you'd have to pay for, but everything else—a collection slated to grow larger than the holdings at the Library of Congress, Harvard, the University of Michigan, at any of the great national libraries of Europe—would have been available for free at terminals that were going to be placed in every local library that wanted one.

At the terminal you were going to be able to search tens of millions of books and read every page of any book you found. You'd be able to highlight passages and make annotations and share them; for the first time, you'd be able to pinpoint an idea somewhere inside the vastness of the printed record, and send somebody straight to it with a link. Books would become as instantly available, searchable, copy-pasteable—as alive in the digital world—as web pages.

It was to be the realization of a long-held dream. “The universal library has been talked about for millennia,” Richard Ovenden, the head of Oxford's Bodleian Libraries, has said. “It was possible to think in the Renaissance that you might be able to amass the whole of published knowledge in a single room or a single institution.” In the spring of 2011, it seemed we'd amassed it in a terminal small enough to fit on a desk.

“This is a watershed event and can serve as a catalyst for the reinvention of education, research, and intellectual life,” one eager observer wrote at

A*The Atlantic*[Sign In](#)[Subscribe](#)

unlocked a century's worth of books and peppered the country with access terminals to a universal library was rejected under Rule 23(e)(2) of the Federal Rules of Civil Procedure by the U.S. District Court for the Southern District of New York.

When the library at Alexandria burned it was said to be an “international catastrophe.” When the most significant humanities project of our time was dismantled in court, the scholars, archivists, and librarians who’d had a hand in its undoing breathed a sigh of relief, for they believed, at the time, that they had narrowly averted disaster.

* * *

GOOGLE’S SECRET EFFORT to scan every book in the world, codenamed “Project Ocean,” began in earnest in 2002 when Larry Page and Marissa Mayer sat down in the office together with a 300-page book and a metronome. Page wanted to know how long it would take to scan more than a hundred-million books, so he started with one that was lying around. Using the metronome to keep a steady pace, he and Mayer paged through the book cover-to-cover. It took them 40 minutes.

Page had always wanted to digitize books. Way back in 1996, the student project that eventually became Google—a “crawler” that would ingest documents and rank them for relevance against a user’s query—was actually conceived as part of an effort “to develop the enabling technologies for a single, integrated and universal digital library.” The idea was that in the future, once all books were digitized, you’d be able to map the citations among them, see which books got cited the most, and use that data to give better search results to library patrons. But

A*The Atlantic*[Sign In](#)[Subscribe](#)

from the World Wide Web.

By 2002, it seemed to Page like the time might be ripe to come back to books. With that 40-minute number in mind, he approached the University of Michigan, his alma mater and a world leader in book scanning, to find out what the state of the art in mass digitization looked like. Michigan told Page that at the current pace, digitizing their entire collection—7 million volumes—was going to take about a thousand years. Page, who'd by now given the problem some thought, replied that he thought Google could do it in six.

Every weekday, semi trucks full of books would pull up at designated Google scanning centers.

He offered the library a deal: You let us borrow all your books, he said, and we'll scan them for you. You'll end up with a digital copy of every volume in your collection, and Google will end up with access to one of the great untapped troves of data left in the world. Brin put Google's lust for library books this way: "You have thousands of years of human knowledge, and probably the highest-quality knowledge is captured in books." What if you could feed all the knowledge that's locked up on paper to a search engine?

By 2004, Google had started scanning. In just over a decade, after making deals with Michigan, Harvard, Stanford, Oxford, the New York

A*The Atlantic*[Sign In](#)[Subscribe](#)

them an estimated \$400 million. It was a feat not just of technology but of logistics.

EVERY WEEKDAY, SEMI TRUCKS full of books would pull up at designated Google scanning centers. The one ingesting Stanford's library was on Google's Mountain View campus, in a converted office building. The books were unloaded from the trucks onto the kind of carts you find in libraries and wheeled up to human operators sitting at one of a few dozen brightly lit scanning stations, arranged in rows about six to eight feet apart.

The stations—which didn't so much scan as photograph books—had been custom-built by Google from the sheet metal up. Each one could digitize books at a rate of 1,000 pages per hour. The book would lie in a specially designed motorized cradle that would adjust to the spine, locking it in place. Above, there was an array of lights and at least \$1,000 worth of optics, including four cameras, two pointed at each half of the book, and a range-finding LIDAR that overlaid a three-dimensional laser grid on the book's surface to capture the curvature of the paper. The human operator would turn pages by hand—no machine could be as quick and gentle—and fire the cameras by pressing a foot pedal, as though playing at a strange piano.

What made the system so efficient is that it left so much of the work to software. Rather than make sure that each page was aligned perfectly, and flattened, before taking a photo, which was a major source of delays in traditional book-scanning systems, cruder images of curved pages were fed to de-warping algorithms, which used the LIDAR data along with some clever mathematics to artificially bend the text back into straight lines.

A*The Atlantic*[Sign In](#)[Subscribe](#)

images into text; they wrote de-warping and color-correction and contrast-adjustment routines to make the images easier to process; they developed algorithms to detect illustrations and diagrams in books, to extract page numbers, to turn footnotes into real citations, and, per Brin and Page's early research, to rank books by relevance. "Books are not part of a network," Dan Clancy, who was the engineering director on the project during its heyday, has said. "There is a huge research challenge, to understand the relationship between books."

At a time when the rest of Google was obsessed with making apps more "social"—Google Plus was released in 2011—Books was seen by those who worked on it as one of those projects from the old era, like Search itself, that made good on the company's mission "to organize the world's information and make it universally accessible and useful."

It was the first project that Google ever called a "moonshot." Before the self-driving car and Project Loon—their effort to deliver Internet to Africa via high-altitude balloons—it was the idea of digitizing books that struck the outside world as a wide-eyed dream. Even some Googlers themselves thought of the project as a boondoggle. "There were certainly lots of folks at Google that while we were doing Google Book Search were like, Why are we spending all this money on this project?," Clancy said to me. "Once Google started being a little more conscious about how it was spending money, it was like, wait, you have \$40 million a year, \$50 million a year on the cost of scanning? It's gonna cost us \$300 to \$400 million before we're done? What are you thinking? But Larry and Sergey were big supporters."

In August 2010, Google put out a blog post announcing that there were 129,864,880 books in the world. The company said they were going to

A

The Atlantic[Sign In](#)[Subscribe](#)

fell about a hundred-million books short of the moon. What happened was complicated but how it started was simple: Google did that thing where you ask for forgiveness rather than permission, and forgiveness was not forthcoming. Upon hearing that Google was taking millions of books out of libraries, scanning them, and returning them as if nothing had happened, authors and publishers filed suit against the company, alleging, as the authors put it simply in their initial complaint, “massive copyright infringement.”

* * *

WHEN GOOGLE STARTED scanning, they weren’t actually setting out to build a digital library where you could read books in their entirety; that idea would come later. Their original goal was just to let you *search* books. For books in copyright, all they would show you were “snippets,” just a few sentences of context around your search terms. They likened their service to a card catalog.

Google thought that creating a card catalog was protected by “fair use,” the same doctrine of copyright law that lets a scholar excerpt someone’s else’s work in order to talk about it. “A key part of the line between what’s fair use and what’s not is transformation,” Google’s lawyer, David Drummond, has said. “Yes, we’re making a copy when we digitize. But surely the ability to find something because a term appears in a book is not the same thing as reading the book. That’s why Google Books is a different product from the book itself.”

It was important for Drummond to be right. Statutory damages for “willful infringement” of a copyright can run as high as \$150,000 for each work infringed. Google’s potential liability for copying tens of



defense,” Pamela Samuelson, a law professor at UC Berkeley, wrote in 2011. Copyright owners pounced.

They had good reason to. Instead of asking for anyone’s permission, Google had plundered libraries. This seemed obviously wrong: If you wanted to copy a book, you had to have the *right to copy* it—you had to have the damn copyright. Letting Google get away with the wholesale copying of every book in America struck them as setting a dangerous precedent, one that might well render their copyrights worthless. An advocacy group called the Authors Guild, and several book authors, filed a class action lawsuit against Google on behalf of everyone with a U.S. copyright interest in a book. (A group of publishers filed their own lawsuit but joined the Authors Guild class action shortly thereafter.)

There’s actually a long tradition of technology companies disregarding intellectual-property rights as they invent new ways to distribute content. In the early 1900s, makers of the “piano rolls” that control player pianos ignored copyrights in sheet music and were sued by music publishers. The same thing happened with makers of vinyl records and early purveyors of commercial radio. In the 60s, cable operators re-aired broadcast TV signals without first getting permission and found themselves in costly litigation. Movie studios sued VCR makers. Music labels sued KazaA and Napster.

As Tim Wu pointed out in a [2003 law review article](#), what usually becomes of these battles—what happened with piano rolls, with records, with radio, and with cable—isn’t that copyright holders squash the new technology. Instead, they cut a deal and start making money from it. Often this takes the form of a “compulsory license” in which, for example, musicians are *required* to license their work to the piano-roll

A*The Atlantic*[Sign In](#)[Subscribe](#)

stream of income, and the public gets to hear their favorite songs on the player piano. “History has shown that time and market forces often provide equilibrium in balancing interests,” Wu writes.

But even if everyone typically ends up ahead, each new cycle starts with rightsholders fearful they’re being displaced by the new technology. When the VCR came out, film executives lashed out. “I say to you that the VCR is to the American film producer and the American public as the Boston strangler is to the woman home alone,” Jack Valenti, then the president of the MPAA, testified before Congress. The major studios sued Sony, arguing that with the VCR, the company was trying to build an entire business on intellectual property theft. But *Sony Corp. of America v. Universal City Studios, Inc.* became famous for its holding that as long as a copying device was capable of “substantial noninfringing uses”—like someone watching home movies—its makers couldn’t be held liable for copyright infringement.

“There was an opportunity to do something extraordinary for readers and academics in this country.”

The Sony case forced the movie industry to accept the existence of VCRs. Not long after, they began to see the device as an opportunity. “The VCR turned out to be one of the most lucrative inventions—for movie producers as well as hardware manufacturers—since movie

A*The Atlantic*[Sign In](#)[Subscribe](#)

Google to realize that there was enough middle ground to make everyone happy. This was especially true when you focused on the back catalog, on out-of-print works, instead of books still on store shelves. Once you made that distinction, it was possible to see the whole project in a different light. Maybe Google wasn't plundering anyone's work. Maybe they were giving it a new life. Google Books could turn out to be for out-of-print books what the VCR had been for movies out of the theater.

If that was true, you wouldn't actually want to stop Google from scanning out-of-print books—you'd want to encourage it. In fact, you'd want them to go beyond just showing snippets to actually selling those books as digital downloads. Out-of-print books, almost by definition, were commercial dead weight. If Google, through mass digitization, could make a new market for them, that would be a real victory for authors and publishers. "We realized there was an opportunity to do something extraordinary for readers and academics in this country," Richard Sarnoff, who was then Chairman of the American Association of Publishers, said at the time. "We realized that we could light up the out-of-print backlist of this industry for two things: discovery and consumption."

But once you had that goal in mind, the lawsuit itself—which was about whether Google could keep scanning and displaying snippets—began to seem small time. Suppose the Authors Guild won: they were unlikely to recoup anything more than the statutory minimum in damages; and what good would it do to stop Google from providing snippets of old books? If anything those snippets might drive demand. And suppose Google won: Authors and publishers would get nothing, and all readers



The plaintiffs, in other words, had gotten themselves into a pretty unusual situation. They didn't want to lose their own lawsuit—but they didn't want to win it either.

* * *

THE BASIC PROBLEM with out-of-print books is that it's unclear who owns most of them. An author might have signed a book deal with their publisher 40 years ago; that contract stipulated that the rights revert to the author after the book goes out of print, but required the author to send a notice to that effect, and probably didn't say anything about digital rights; and all this was recorded on some pieces of paper that nobody has.

It's been estimated that about half the books published between 1923 and 1963 are actually in the public domain—it's just that no one knows which half. Copyrights back then had to be renewed, and often the rightsholder wouldn't bother filing the paperwork; if they did, the paperwork could be lost. The cost of figuring out who owns the rights to a given book can end up being greater than the market value of the book itself. "To have people go and research each one of these titles," Sarnoff said to me, "It's not just Sisyphean—it's an impossible task economically." Most out-of-print books are therefore locked up, if not by copyright then by inconvenience.

The tipping point toward a settlement of *Authors Guild v. Google* was the realization that it offered a way to skirt this problem entirely. *Authors Guild* was a class action lawsuit, and the class included everyone who held an American copyright in one or more books. In a class action, the named plaintiffs litigate *on behalf of* the whole class (though

A*The Atlantic*[Sign In](#)[Subscribe](#)

about every author and publisher with a book in an American library. In particular, you could craft a deal in which copyright owners, as a class, agreed to release any claims against Google for scanning and displaying their books, in exchange for a cut of the revenue on sales of those books.

“If you have a kind of an institutional problem,” said Jeff Cunard, a partner at Debevoise & Plimpton who represented the publishers in the case, “you can address the issue through a class-action settlement mechanism, which releases all past claims and develops a solution on a going-forward basis. And I think the genius here was of those who saw this as a way of addressing the problem of out-of-print books and liberating them from the dusty corners to which they’d been consigned.”

It was a kind of hack. If you could get the class on board with your settlement, and if you could convince a judge to approve it—a step required by law, because you want to make sure the class representatives are acting in the class’s best interests—then you could in one stroke cut the Gordian knot of ambiguous rights to old books. With the class action settlement, authors and publishers who stayed in the class would in effect be saying to Google, “go ahead.”

Naturally, they’d have to get something in return. And that was the clever part. At the heart of the settlement was a collective licensing regime for out-of-print books. Authors and publishers could opt out their books at any time. For those who didn’t, Google would be given wide latitude to display and sell their books, but in return, 63 percent of the revenues would go into escrow with a new entity called the Book Rights Registry. The Registry’s job would be to distribute funds to rightsholders as they came forward to claim their works; in ambiguous cases, part of the money would be used to figure out who actually

A*The Atlantic*[Sign In](#)[Subscribe](#)

individual authors don't make any money out of out-of-print books," Cunard said to me. "Not that they would have made gazillions of dollars" with Google Books and the Registry, "but they would at least have been paid something for it. *And* most authors actually want their books to be read."

What became known as the Google Books Search Amended Settlement Agreement came to 165 pages and more than a dozen appendices. It took two and a half years to hammer out the details. Sarnoff described the negotiations as "four-dimensional chess" between the authors, publishers, libraries, and Google. "Everyone involved," he said to me, "and I mean everyone—on all sides of this issue—thought that if we were going to get this through, this would be the single most important thing they did in their careers." Ultimately the deal put Google on the hook for about \$125 million, including a one-time \$45 million payout to the copyright holders of books it had scanned—something like \$60 per book—along with \$15.5 million in legal fees to the publishers, \$30 million to the authors, and \$34.5 million toward creating the Registry.

But it also set the terms for how out-of-print books, newly freed, would be displayed and sold. Under the agreement, Google would be able to preview up to 20 percent of a given book to entice individual users to buy, and it would be able to offer downloadable copies for sale, with the prices determined by an algorithm or by the individual rightsholder, in price bins initially ranging from \$1.99 to \$29.99. All the out-of-print books would be packaged into an "institutional subscription database" that would be sold to universities, where students and faculty could search and read the full collection for free. And in §4.8(a), the agreement describes in bland legalese the creation of an incomparable



Sorting out the details had taken years of litigation and then years of negotiation, but now, in 2011, there was a plan—a plan that seemed to work equally well for everyone at the table. As Samuelson, the Berkeley law professor, put it in a paper at the time, “The proposed settlement thus looked like a win-win-win: the libraries would get access to millions of books, Google would be able to recoup its investment in GBS, and authors and publishers would get a new revenue stream from books that had been yielding zero returns. And legislation would be unnecessary to bring about this result.”

In this, she wrote, it was “perhaps the most adventuresome class action settlement ever attempted.” But to her way of thinking, that was the very reason it should fail.

* * *

THE PUBLICATION OF the Amended Settlement Agreement to the *Authors Guild* case was headline news. It was quite literally a big deal—a deal that would involve the shakeup of an entire industry. Authors, publishers, Google’s rivals, legal scholars, librarians, the U.S. government, and the interested public paid attention to the case’s every move. When the presiding judge, Denny Chin, put out a call for responses to the proposed settlement, responses came in droves.

Those who had been at the table crafting the agreement had expected some resistance, but not the “parade of horrors,” as Sarnoff described it, that they eventually saw. The objections came in many flavors, but they all started with the sense that the settlement was handing to Google, and Google alone, an awesome power. “Did we want the greatest library that would ever exist to be in the hands of one giant

A*The Atlantic*[Sign In](#)[Subscribe](#)

said.

Darnton had initially been supportive of Google's scanning project, but the settlement made him wary. The scenario he and many others feared was that the same thing that had happened to the academic journal market would happen to the Google Books database. The price would be fair at first, but once libraries and universities became dependent on the subscription, the price would rise and rise until it began to rival the usurious rates that journals were charging, where for instance by 2011 a yearly subscription to the *Journal of Comparative Neurology* could cost as much as \$25,910.

Although academics and library enthusiasts like Darnton were thrilled by the prospect of opening up out-of-print books, they saw the settlement as a kind of deal with the devil. Yes, it would create the greatest library there's ever been—but at the expense of creating perhaps the largest *bookstore*, too, run by what they saw as a powerful monopolist. In their view, there had to be a better way to unlock all those books. "Indeed, most elements of the GBS settlement would seem to be in the public interest, except for the fact that the settlement restricts the benefits of the deal to Google," the Berkeley law professor Pamela Samuelson wrote.

Certainly Google's competitors felt put out by the deal. Microsoft, predictably, argued that it would further cement Google's position as the world's dominant search engine, by making it the only one that could legally mine out-of-print books. By using those books in results for user's long-tail queries, Google would have an unfair advantage over competitors. Google's response to this objection was simply that anyone could scan books and show them in search results if they wanted—and

A*The Atlantic*[Sign In](#)[Subscribe](#)

was, in fact, fair use.)

“There was this hypothesis that there was this huge competitive advantage,” Clancy said to me, regarding Google’s access to the books corpus. But he said that the data never ended up being a core part of any project at Google, simply because the amount of information on the web itself dwarfed anything available in books. “You don’t need to go to a book to know when Woodrow Wilson was born,” he said. The books data was helpful, and interesting for researchers, but “the degree to which the naysayers characterized this as being the strategic motivation for the whole project—that was malarkey.”

Amazon, for its part, worried that the settlement allowed Google to set up a bookstore that no one else could. Anyone else who wanted to sell out-of-print books, they argued, would have to clear rights on a book-by-book basis, which was as good as impossible, whereas the class action agreement gave Google a license to all of the books at once.

This objection got the attention of the Justice Department, in particular the Antitrust division, who began investigating the settlement. In a statement filed with the court, the DOJ argued that the settlement would give Google a de facto monopoly on out-of-print books. That’s because for Google’s competitors to get the same rights to those books, they’d basically have to go through the exact same bizarre process: scan them en masse, get sued in a class action, and try to settle. “Even if there were reason to think history could repeat itself in this unlikely fashion,” the DOJ wrote, “it would scarcely be sound policy to encourage deliberate copyright violations and additional litigation.”

Google’s best defense was that the whole point of antitrust law was to protect consumers, and, as one of their lawyers put it, “From the

A*The Atlantic*[Sign In](#)[Subscribe](#)

inaccessible online; now there'd be a way to buy them. How did that hurt consumers? A person closely involved in the settlement said to me, "Each of the publishers would go into the Antitrust Division and say well but look, Amazon has 80 percent of the e-book market. Google has 0 percent or 1 percent. This is allowing someone else to compete in the digital books space against Amazon. And so you should be regarding this as pro-competitive, not anti-competitive. Which seemed also very sensible to me. But it was like they were talking to a brick wall. And that reaction was shameful."

The DOJ held fast. In some ways, the parties to the settlement didn't have a good way out: no matter how "non-exclusive" they tried to make the deal, it was in effect a deal that only Google could get—because Google was the only defendant in the case. For a settlement in a class action titled *Authors Guild v. Google* to include not just Google but, say, every company that wanted to become a digital bookseller, would be to stretch the class action mechanism past its breaking point.

This was a point that the DOJ kept coming back to. The settlement was *already* a stretch, they argued: the original case had been about whether Google could show snippets of books it had scanned, and here you had a settlement agreement that went way beyond that question to create an elaborate online marketplace, one that depended on the indefinite release of copyrights by authors and publishers who might be difficult to find, particularly for books long out of print. "It is an attempt," they wrote, "to use the class-action mechanism to implement forward-looking business arrangements that go far beyond the dispute before the Court in this litigation."

The DOJ objections left the settlement in a double bind: Focus the deal



The lawyers who had crafted the settlement tried to thread the needle. The DOJ acknowledged as much. “The United States recognizes that the parties to the ASA are seeking to use the class action mechanism to overcome legal and structural challenges to the emergence of a robust and diverse marketplace for digital books,” they wrote. “Despite this worthy goal, the United States has reluctantly concluded that use of the class-action mechanism in the manner proposed by the ASA is a bridge too far.”

Their argument was compelling, but the fact that the settlement was ambitious didn’t mean it was illegal—just unprecedented. Years later, another class-action settlement that involved opt-out, “forward-looking business arrangements” very similar to the kind set up by the Google settlement was approved by another district court. That case involved the prospective exploitation of publicity rights of retired NFL players; the settlement made those rights available to an entity that would license them and distribute the proceeds. “What was interesting about it,” says Cunard, who was also involved in that litigation, “was that not a single opponent of the settlement ever raised Judge Chin’s decision or any of the oppositions to it with respect to that settlement being ‘beyond the scope of the pleadings.’” Had that case been decided ten years ago, Cunard said, it would have been “a very important and substantial precedent,” significantly undercutting the “bridge too far” argument against the *Authors Guild* agreement. “It demonstrates that the law is a very fluid thing,” he said. “Somebody’s got to be first.”

In the end, the DOJ’s intervention likely spelled the end of the settlement agreement. No one is quite sure why the DOJ decided to take a stand instead of remaining neutral. Dan Clancy, the Google

A*The Atlantic*[Sign In](#)[Subscribe](#)

competitors but “sympathetic entities” you’d think would be in favor of it, like library enthusiasts, academic authors, and so on—that ultimately flipped the DOJ. “I don’t know how the settlement would have transpired if those naysayers hadn’t been so vocal,” he told me. “It’s not clear to me that if the libraries and the Bob Darntons and the Pam Samuelsons of the world hadn’t been so active that the Justice Department *ever* would have become involved, because it just would have been Amazon and Microsoft bitching about Google. Which is like yeah, tell me something new.”

Whatever the motivation, the DOJ said its piece and that seemed to carry the day. In his ruling concluding that the settlement was not “fair, adequate, and reasonable” under the rules governing class actions, Judge Denny Chin recited the DOJ’s objections and suggested that to fix them, you’d either have to change the settlement to be an opt-in arrangement—which would render it toothless—or try to accomplish the same thing in Congress.

“While the digitization of books and the creation of a universal digital library would benefit many,” Chin wrote in his decision, “the ASA would simply go too far.”

* * *

AT THE CLOSE of the “fairness hearing,” where people spoke for and against the settlement, Judge Chin asked, as if merely out of curiosity, just how many objections had there been? And how many people had opted out of the class? The answers were more than 500, and more than 6,800.

Reasonable people could disagree about the legality of the settlement;

A*The Atlantic*[Sign In](#)[Subscribe](#)

on. What seemed to turn the tide against the settlement was the reaction of the class itself. “In my more than twenty-five years of practice in class action litigation, I’ve never seen a settlement reacted to that way, with that many objectors,” said Michael Boni, who was the lead negotiator for the authors class in the case. That strong reaction was what likely led to the DOJ’s intervention; it turned public opinion against the agreement; and it may have led Chin to look for ways to kill it. After all, the question before him was whether the agreement was fair to class members. The more class members came out of the woodwork, and the more upset they seemed to be, the more reason he’d have to think that the settlement didn’t represent their interests.

The irony is that so many people opposed the settlement in ways that suggested they fundamentally believed in what Google was trying to do. One of Pamela Samuelson’s main objections was that Google was going to be able to *sell* books like hers, whereas she thought they should be made available for free. (The fact that she, like any author under the terms of the settlement, could set her own books’ price to zero was not consolation enough, because “orphan works” with un-findable authors would still be sold for a price.) In hindsight, it looks like the classic case of perfect being the enemy of the good: surely having the books made available at all would be better than keeping them locked up—even if the price for doing so was to offer orphan works for sale. In her paper concluding that the settlement went too far, Samuelson herself even wrote, “It would be a tragedy not to try to bring this vision to fruition, now that it is so evident that the vision is realizable.”

“This is not important enough for the

A*The Atlantic*[Sign In](#)[Subscribe](#)*law.*

Many of the objectors indeed thought that there would be some other way to get to the same outcome without any of the ickiness of a class action settlement. A refrain throughout the fairness hearing was that releasing the rights of out-of-print books for mass digitization was more properly “a matter for Congress.” When the settlement failed, they pointed to proposals by the U.S. Copyright Office recommending legislation that seemed in many ways inspired by it, and to similar efforts in the Nordic countries to open up out-of-print books, as evidence that Congress could succeed where the settlement had failed.

Of course, nearly a decade later, nothing of the sort has actually happened. “It has got no traction,” Cunard said to me about the Copyright Office’s proposal, “and is not going to get a lot of traction now I don’t think.” Many of the people I spoke to who were in favor of the settlement said that the objectors simply weren’t practical-minded—they didn’t seem to understand how things actually get done in the world. “They felt that if not for us and this lawsuit, there was some other future where they could unlock all these books, because Congress would pass a law or something. And that future... as soon as the settlement with *Guild*, nobody gave a shit about this anymore,” Clancy said to me.

It certainly seems unlikely that someone is going to spend political capital—especially today—trying to change the licensing regime for *books*, let alone old ones. “This is not important enough for the Congress to somehow adjust copyright law,” Clancy said. “It’s not going

A*The Atlantic*[Sign In](#) [Subscribe](#)

perhaps the only plausible venue for this kind of reform: Google was the only one with the initiative, and the money, to make it happen. “If you want to look at this in a raw way,” Allan Adler, in-house counsel for the publishers, said to me, “a deep pocketed, private corporate actor was going to foot the bill for something that everyone wanted to see.”

Google poured resources into the project, not just to scan the books but to dig up and digitize old copyright records, to negotiate with authors and publishers, to foot the bill for a Books Rights Registry. Years later, the Copyright Office has gotten nowhere with a proposal that re-treads much the same ground, but whose every component would have to be funded with Congressional appropriations.

I asked Bob Darnton, who ran Harvard’s library during the Google Books litigation and who spoke out against the settlement, whether he had any regrets about what ended up happening. “Insofar as I have a regret, it is that the attempts to out-Google Google are so limited by copyright law,” he said. He’s been working on another project to scan library books; the scanning has been limited to books in the public domain. “I’m in favor of copyright, don’t get me wrong, but really to leave books out of the public domain for more than a century—to keep most American literature behind copyright barrier,” he said, “I think is crazy.”

The first copyright statute in the United States, passed in 1790, was called An Act for the Encouragement of Learning. Copyright terms were to last fourteen years, with the option to renew for another fourteen, but only if the author was alive at the end of the first term. The idea was to strike a “pragmatic bargain” between authors and the reading public. Authors would get a limited monopoly on their work so they could make a living from it; but their work would retire quickly into the



Somewhere at Google there is a database containing 25 million books and nobody is allowed to read them.

Copyright terms have been radically extended in this country largely to keep pace with Europe, where the standard has long been that copyrights last for the life of the author plus 50 years. But the European idea, “It’s based on natural law as opposed to positive law,” Lateef Mtima, a copyright scholar at Howard University Law School, said. “Their whole thought process is coming out of France and Hugo and those guys that like, you know, ‘My work is my *enfant*,’” he said, “and the *state* has absolutely no right to do anything with it—kind of a Lockean point of view.” As the world has flattened, copyright laws have converged, lest one country be at a disadvantage by freeing its intellectual products for exploitation by the others. And so the American idea of using copyright primarily as a vehicle, per the constitution, “to promote the Progress of Science and useful Arts,” not to protect authors, has eroded to the point where today we’ve locked up nearly every book published after 1923.

“The greatest tragedy is we are still exactly where we were on the orphan works question. That stuff is just sitting out there gathering dust and decaying in physical libraries, and with very limited exceptions,” Mtima said, “*nobody* can use them. So everybody has lost and no one has won.”

AFTER THE SETTLEMENT failed, Clancy told me that at Google “there was

A*The Atlantic*[Sign In](#)[Subscribe](#)

U displaying snippets of copyrighted books was fair use, the company all but shut down its scanning operation.

It was strange to me, the idea that somewhere at Google there is a database containing 25-million books and nobody is allowed to read them. It's like that scene at the end of the first *Indiana Jones* movie where they put the Ark of the Covenant back on a shelf somewhere, lost in the chaos of a vast warehouse. It's *there*. The books are *there*. People have been trying to build a library like this for ages—to do so, they've said, would be to erect one of the great humanitarian artifacts of all time—and here we've done the work to make it real and we were about to give it to the world and now, instead, it's 50 or 60 petabytes on disk, and the only people who can see it are half a dozen engineers on the project who happen to have access because they're the ones responsible for locking it up.

I asked someone who used to have that job, what would it take to make the books viewable in full to everybody? I wanted to know how hard it would have been to unlock them. What's standing between us and a digital public library of 25 million volumes?

You'd get in a lot of trouble, they said, but all you'd have to do, more or less, is write a single database query. You'd flip some access control bits from off to on. It might take a few minutes for the command to propagate.



The Atlantic

Sign In [Subscribe](#)