Cites & Insights

Crawford at Large

Libraries • Policy • Technology • Media

Volume 16, Number 4: May 2016 ISSN 1534-0937 Walt Crawford

The Front

Short

A brief issue this time around, for two good reasons:

- I'm still mostly devoting time and energy to the SPARC-supported *Gold Open Access Journals 2011-2015*, moving from data gathering to analysis and writeup on or around May 2, 2016. There's more about this in the next essay—a somewhat surprising development, at least to me.
- I learned touch typing in junior high school (yes, they did teach typing when we weren't ducking the dinosaurs) and have been an 80wpm typist ever since I got my first electric typewriter. With the advent of PCs, I'm one of those who finds a fairly direct connection from the brain to the fingers, one reason I've been able to write so much. However…on March 29, 2016, I had a Schwannoma (benign nerve sheath tumor) removed from my right forearm, and apparently I didn't understand what the doctor was saying about recovery (this is the kind version). I still have limited functionality with the fingers of my right hand, including an inability to raise them to be level with the palm. The main effects at this point: I can't use chopsticks…and I can only type with six fingers. That latter one *substantially* slows my typing. (It didn't slow the data gathering much; a good right forefinger and left hand were enough.) I'm guessing 10 to 15 effective WPM, and I tire more easily. (Yes, I've ordered Nuance Dragon NaturallySpeaking; I'll give it a try—but I don't expect miracles.) If the physical therapists are right, and if the lateral nerve is just traumatized, not wholly severed, it may be another, oh. four to eight weeks or maybe five months… And with a book (or a book with two book-length supplements) to write, well, I only have so much half-hunt-and-pecking in me. (I suppose I could learn to use mostly words consisting primarily of qwertyasdfghzxcvbn, but that seems…counterproductive?)

I'm actually fairly excited about the book(s); I believe the 2011-2015 survey should and will be broadly meaningful in the OA world. So I'll carry on, but may be a little light on other writing for some time to come. (Yes, I know, given that I'm otherwise in excellent health at 70.5 years—no prescription medicines and just a touch of sciatica—I shouldn't complain. And I'm not, really: I'm explaining.)

Meanwhile, two brief essays—the second of which could be *very* long without much typing if I was so inclined.

Inside This Issue

Intersections

Two Worlds of Gold OA: APCLand and OAWorld

Treating all of gold open access as one monolithic entity can be seriouslu misleading. That's even true for serious gold OA, which I'll define as journals in the *Directory of Open Access Journals*. In my previous published attempts to survey most or all of gold OA, I found it useful to break things down into three broad subject segments: biomed, STEM (science, technology, engineering and mathematics) and HSS (humanities and social sciences). Each segment has very different characteristics in terms of APC prevalence and price levels and number of articles per year in journals.

While *Gold Open Access Journals 2011-2015* will retain those segments, there's another way to break down serious gold OA that I find compelling—so compelling that I'm writing about it in advance of the book and inviting your feedback and comments (to waltcrawford@gmail.com).

To wit, I now see the universe of gold OA as divided into two parts, which I'll call APCLand and OAWorld.

APCLand

APCLand consists of eleven publishers, each with more than 5,000 OA articles in 2015, each with a maximum potential 2015 APC revenue of more than $8 million (the actual revenue may be lower), and each with at least two-thirds of its 2015 articles in OA journals appearing in APC-charging journals,

APCLand accounts for 13% of the fully-analyzed *DOAJ* journals for 2015 and 29% of the 2015 articles in those journals. It also accounts for 74% of the maximum potential APC revenues.

In other words, although APCLand accounts for a bit more than one-eighth of the serious gold OA journals and somewhere between one-quarter and one-third of the articles, it takes in nearly **three-quarters** of the revenue.

APCLand includes the following publishers, listed alphabetically and using the publisher names used in *DOAJ* listings: BioMed Central, Dove Medical Press, Elsevier, Frontiers Media S.A., Hindawi Publishing Corporation, MDPI AG, Nature Publishing Group, Oxford University Press, Public Library of Science (PLoS), Springer and Wiley.

Noting that all numbers are subject to refinement as analysis continues, here are the actual numbers. For 2015, APCLand showed 1,391 gold OA journals (that could be fully analyzed) publishing 163,545 articles, with a total maximum potential APC revenue of $279,558,871.

Overall, 11% of the journals did not have APCs when checked in early 2016 (including journals funded through SCOAP[3]), but those journals published only 4% of the articles in APCLand. Average cost per article (assuming no waivers, discounts or less-expensive article categories) was $1,782; including the no-fee journals brings that down to $1,709.

OAWorld

OAWorld includes thousands of publishers (there are more than 5,400 names in a pivot table of *DOAJ* publisher fields, but it's clear that hundreds and possibly more than a thousand of them represent spelling or other minor variations). These publishers accounted for 87% of the journals and 71% of the articles, but only 26% of the revenues.

In terms of actual preliminary numbers, OAWorld accounts for 8.934 fully-analyzed journals in 2015 with 403,377 articles, with a maximum potential revenue of $97,173,704.

Here's perhaps the key point: in OAWorld, not only do 81% of the journals *not* charge APCs or equivalent fees, those journals account for 61% of the articles, In other words, in OAWorld *most articles did not involve author-side charges*.

Another key figure: for those articles that *did* involve fees, the average cost per article was $611 (six hundred and eleven dollars: I didn't leave out "1," after the dollar sign), just over one-third the average fee in APCLand. Averaged across *all* articles, the cost per article was $241—*one-seventh* the going rate for APCLand.

Most of the gold OA articles are published in OAWorld—but most of the gold goes to APCLand.

Totals and Feedback

If you're pulling out a calculator to see totals, I'll save you the trouble, again with the "preliminary figures" caveat: 10,325 journals, 71% of them free; 566,922 articles, 44% in free journals; $376,732,575 maximum potential 2015 revenue; $1,192 average cost per article in fee-charging journals; $665 average cost per article overall.

Consider that last figure: when you combine APCLand and OAWorld, the average cost for *all* articles is higher than it is for *paid* articles in OAWorld. That's the skewing effect of APCLand.

Before I offer more details, I'll ask again: If you don't think this is a useful breakdown or think it needs refining, let me know fairly soon. As things stand, it will play a significant role in the book.

For those of you who assume I'm just out to get Big Traditional Publishers: Note that fewer than half of the APCLand are (or at least started out as) traditional publishers. Furthermore, the APCLand publisher with the highest percentage of 2015 OA-journal articles in *free* journals is Elsevier (30%)—and Elsevier's average cost per article including free articles is a little below the overall APCLand average, for now. (Wiley and Oxford University Press have the highest figures here; Springer the lowest.)

Discovering APCLand

I'd love to tell you that I discovered APCLand through deeply expert analysis and synthesis of the big spreadsheet. In fact, it's sort of an accident that grew out of the desire to look at OA activity on a regional basis—that is, the idea that OA pricing and other characteristics differ regionally on more than a global-south-versus-north basis.

As I was gathering a list of countries represented in the fully-analyzed data (124 of them in all), since I'd made the list using a pivot table that also showed journal count and 2015 article count for each country, I was reminded that big publishers, primarily producing international journals, might bias the data for some countries and even regions. Discussing this with a colleague, we agreed that it might make sense to filter out the biggest publishers. In attempting to do that, I found that there was a sharp correlation between size in terms of article volume and "APCness" (tendency to charge APCs and to set them fairly high).

Filtering out APCLand journals, the two countries with the most OA journals continue to be Brazil (which passed the U.S. when I did a more comprehensive survey) and the United States…but within the top 15 there are some significant changes. The United Kingdom drops from third place to ninth; Egypt drops from fifth to sixtieth; Germany drops from seventh to thirteenth; and Switzerland drops from fifteenth to forty-first. Regions will be affected similarly—more so in Western Europe and the Middle East, less in other regions.

I *believe* the filtered regions are more indicative of general OA in regions (for example, roughly 80% of remaining Middle East OA articles in 2015 didn't involve APCs), and plan to discuss regions and countries within OAWorld.

Segment by Segment

As dramatic as the overall differences between APCLand and OAWorld are, the differences within broad subject segments are even more dramatic.

Biomed

APCLand is, as you'd expect, a big player here, with 29% of the journals and 39% of the articles. Only 5% of the APCLand biomed journals are free, and those journals account for only 2% of the 2015 articles. Average cost per article among APC-charging journals in 2015 was $2,035, coming down to $1,997 overall. APCLand published 80,706 biomed articles in 2015.

In OAWorld, where 126,356 biomed articles appeared in 2015, 68% of the biomed journals were free and those journals published 54% of the articles: even in the most APC-hungry subject segment, a majority of articles did *not* involve payment. Average cost per article among APC-charging journals was $854; the overall average was $396.

STEM

STEM is the largest segment overall, but not for APCLand, and APCLand only accounts for 16% of the journals, although those journals published 33% of the STEM OA articles in 2015. There's still not a lot of free activity in APCLand: 15% of the journals, publishing 6% of the 2015 articles. Average cost per article among APC-charging journals was $1,518; including free journals brings that down to $1,432, APCLand published 78,900 STEM articles in 2015—just slightly fewer than for biomed (but in a *lot* fewer journals).

STEM *is* the largest segment for OAWorld, with 161,562 articles in 2015; 72% of the journals didn't charge APCs, and those journals account for 52% of the articles. Average cost per article among APC-charging journals was $540; for all journals it was $259. (While APCLand has only 57% as many STEM journals as biomed journals, OAWorld has 25% more STEM journals than biomed.)

HSS

APCLand is almost wholly uninterested in the humanities and social sciences: it accounts for only 2% of the journals and 3% of the articles. Although just over half of those journals (51%) don't charge APCs, only 20% of the 4,009 articles in 2015 appeared in those journals. Average cost per article among APC-charging journals was $1,661; including non-APC journals, the cost per article comes down to $1,334—not a lot less than for STEM.

OAWorld published 115,389 HSS articles in 2015—the smallest segment but not by much. Very little of that involved APCs: 92% of the journals, publishing 80% of the articles, didn't charge them. Among the journals that did charge, average cost per article was $229—but the overall average was $45.

There are considerably more HSS journals than either biomed or STEM: 4,411 in all compared to 2,877 and 3,077 respectively. OAWorld accounts for 4,343 of those 4,411 journals.

The Brembs Dystopia

Note: while SPARC supported the 2011-2015 project, it has nothing to do with this section. These are entirely my own thoughts.

Björn Brembs set forth a dystopian scenario in an April 7, 2016 post: "How gold open access may make things worse." It's a cautionary tale that suggests, *correctly I believe*, that if all scholarly article publishing "flipped" to a gold OA model, but one in which existing commercial publishers (and especially ones with "aggressive" pricing models like Emerald) dominated the market and were free to raise APC charges as they saw fit, the result could be spending *even more* money than is now spent on subscriptions and APCs.

He's right. If we assume (for the sake of the discussion) 2.6 million articles per year and that publishers migrate to $5,000 APCs, that totals out to $13 billion, more than the $10 to $12 billion currently being spent. (How many articles actually are published each year? I've seen estimates from 1.5 million to 2.5 million to "who knows?")

But that assumes that funding agencies say "Charge whatever you think is appropriate and we'll pay it" with no controls or counterbalancing efforts. If that happens, we could indeed be worse off.

Let's see what happens if there *are* some limits, pressure points and countervailing forces. (What if funding sources asked for a clear explanation as to why publishing costs should be higher than, say, $396—or, for that matter, why biomed articles are more than 50% more expensive to publish than STEM articles, which in turn are at least *four times* as expensive to publish as HSS articles?)

- In the worst case of limiting fees to the average of APCLand charges, 2.6 million times $1,300 would be $3.38 billon: a lot, but still a considerable savings.
- If we assume that publishers should be as efficient as those of OAWorld (and remember that *nearly all* gold OA journals in the pricey United States—952 out of 990—are in OAWorld)—we get either $611 (APCs for everything!) or $241, or at worst worst $854 (paid biomed). That's $626 million to $2.22 billion, with $1.59 billion as a middle ground.

Yes, it could be worse—but only if there are no limits or pressures. Heck, even paying the average of the *most expensive* APCLand publisher, $2,294, would "only" cost $5.96 billion.

I'm neither a Proper Scientist nor a policy-maker, but I see better paths here than Brembs' ruinous possibility, as long as there's no nonsense about assuring that all "stakeholders" are fully protected, a view of stakeholders that seems to omit the public.

I'm not going to go into an extended discussion of open access economics here. With luck, such a discussion will come in a future issue (I currently have 60 items tagged "oa-econ" in Diigo), but not until I've finished with this project. I will, however, point you to one interesting discussion that appeared *while* I was drafting this: Jeffrey Mackie-Mason's "Economic thoughts about 'gold' open access," posted April 23, 2016 at *madLibbing*. Well worth reading for an economist-librarian's take on realities and potentials.

Graphically Speaking

Here are two graphs to show the pay-vs-free picture over the five years covered in *Gold Open Access Journals 2011-2015*—followed by some notes about numbers that may surprise those of you familiar with *The Gold OA Landscape 2011-2014*.

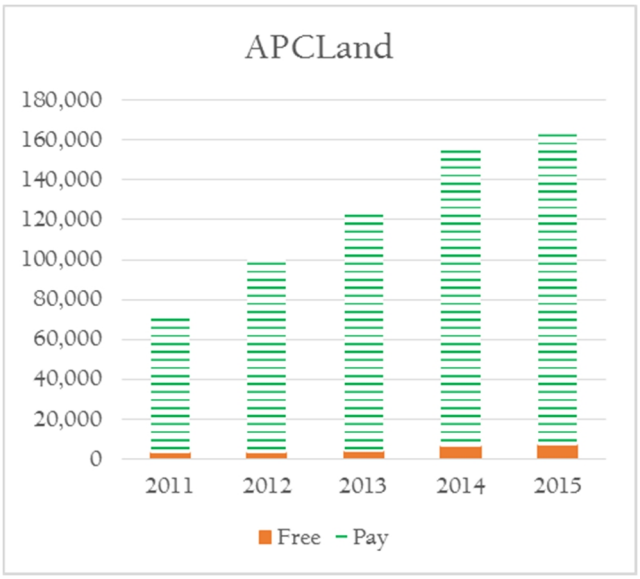

Figure 1. APCLand articles, 2011-2015

Figure 1 shows articles in no-fee journals (solid gold) and those in "pay" (fee-charging) journals (striped green) for each year 2011-2015—with the big caveat that pay or free is necessarily based on the journal's status in early 2016. While the free count has increased each year, actually increasing more percentagewise from 2014 to 2015 than did pay articles, it's never been a significant part of APCLand publishing.

Figure 2 uses the same colors and patterns (but not the same vertical scale!) for the much larger OAWorld. Here, you see that no-APC publishing has always been dominant—but that fee-based publishing is growing much more rapidly.

The apparent slight drop in free OAWorld article count from 2014 to 2015 may or may not be "real," for reasons discussed in the book and briefly below.

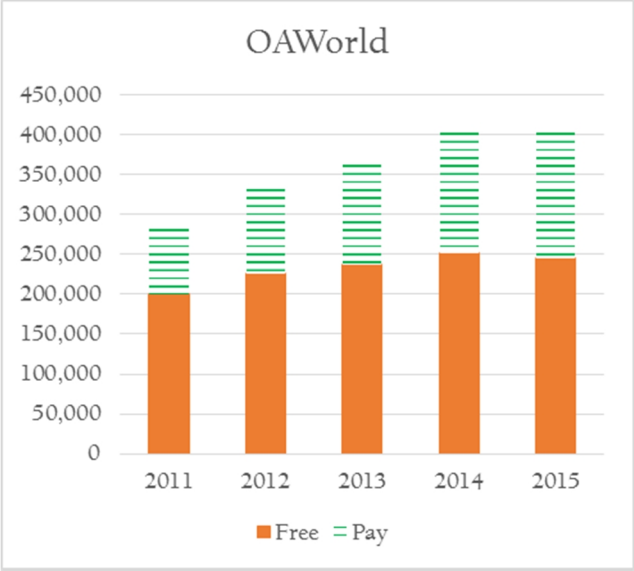

Figure 2. OAWorld articles, 2011-2015

The Big Increase, The Flattening

Let's close this little preview by considering some apparent anomalies as compared to the previous study.

First there's the matter of 2014 counts: 560,036 total here as compared to 482,361 last time around. How can that be?

- This study is a *lot* more complete, fully covering 10,325 "A" and "B" journals compared to 9,512 last time around.
- The newly-added journals (882 of them, most *not* starting in 2015 but newly added to *DOAJ*) published considerably more articles in 2014 than did those that disappeared (of which only 482 were fully analyzed)—about 8,000 more.
- This time around, I included journals publishing refereed conference papers and a few that require free registration to read articles (but not to see tables of contents: those are still excluded). I also counted issues of other journals that were devoted to conference papers (but not abstracts). The second group of journals added just over 1,800 articles to the 2014 count—but the first added some 17,000.
- I was more inclusive in counting, including reviewed/edited book reviews and shorter communications—which I always had done for publishers with article-count shortcuts such as MDPI, Dove, SciELI and many Iranian journals.
- There's the "late posting" factor, which also relates to the apparent slight drop in free OAWorld article counts: quite a few smaller journals, especially HSS journals, are issue-oriented and can take many months after the cover date to post issues.
- Finally—and probably not least—I used a lot fewer approximations (and I'd always estimated low when using approximations), with more fairly large journals being counted more precisely. In hundreds of cases I went back at least one year to provide better counts.

It's not just 2014. The revised good-journal totals for 2013 went from 440,843 to 493,475; for 2012, from 394,374 to 438,644; and for 2011, from 321,312 to 360,349. In all cases, I believe the new numbers—while still slightly incomplete—are more meaningful.

The Drop and the Flattening

There were *apparently* fewer no-fee articles in OAWorld in 2015 than in 2014: 244,219 compared to 250,966. Is that a real drop?

I don't know, but I will note this. At the completion of the first pass of journal visits, which took place from January 2, 2016 to around March 22, 2016, I showed 546,272 articles from 2014. At the end of the second pass—revisiting some 2,600 journals, including more than 1,000 where it looked as though there might be posting delays, between April 1 and April 21, 2016—I counted 560,036 articles from 2014. Some of that increase came from salvaging difficult-to-count journals, but some came from *very* delayed posting,

For 2015, the count went from 545,363 in the first pass to 566,922 in the second pass. If I was to revisit those journals in, say, October 2016, I would bet the count would go even higher, probably by anywhere from 5,000 to 15,000 articles but possibly by even more: quite possibly enough to show a (small) uptick in free OAWorld publishing, although I wouldn't bet on it.

Overall, there was growth from 2014 to 2015—but only about 6,900 articles or around 1.2%, as compared to 66,561 (or 13%) from 2013 to 2014; 54,831 (or 12.5%) from 2012 to 2013; and 78,295 (21.7%) from 2011 to 2012 (noting that 2011-2013 figures are likely to be somewhat less reliable than 2014-2015 numbers).

Has real growth dropped to somewhere between 1.2% and 4%? Quite possibly, and it's possible that biomed OA publishing has almost completely flattened out. That could be temporary, or it could be a serious issue for future changes to scholarly publishing. I'm mostly just trying to describe what's actually happening as thoroughly as possible.

Policy

Google Books: The Final Chapter?

On Monday, April 18, 2016, the U.S. Supreme Court declined to hear the Authors Guild appeal of a district court decision finding, once again, that Google Books Search is fair use. The brief denial, 15-849, is on page 12 of the PDF linked to above. Here it is, in its entirety:

> 15-849 AUTHORS GUILD, ET AL. V. GOOGLE, INC.
>
> The petition for a writ of certiorari is denied. Justice Kagan took no part in the consideration or decision of this petition.

That *should* be the final chapter in this decade-long epic case, and maybe I should stop right here.

But let's look at a couple of the early commentaries after the denial (two of many), then go back for the usual chronological citations and notes on items since the last *Cites & Insights* coverage of this legal marathon. The question mark in the essay's title? Well, the Authors Litigation Guild (the middle word isn't part of the name, but maybe it should be) seems as incapable of admitting defeat as it apparently is of recognizing that it only represents the interests of a few hundred or few thousand writers. And, of course, there's the enticing if unlikely counter possibility: what if Google asked to recover its legal costs, which must surely be in the millions of dollars?

The Denial

David Kravets provides a good same-day writeup at *ars technica* in "Fair use prevails as Supreme Court rejects Google Books copyright case." Portions, including key links(I'm not excerpting the linked briefs):

> The Supreme Court on Monday declined (PDF) to hear a challenge from the Authors Guild and other writers claiming Google's scanning of their books amounts to wanton copyright infringement and not fair use.
>
> The guild urged the high court to review a lower court decision in favor of Google that the writers said amounted to an "unprecedented judicial expansion of the fair-use doctrine." (PDF)
>
> At issue is a June decision (PDF) by the 2nd US Circuit Court of Appeals that essentially said it's legal to scan books if you don't own the copyright…
>
> Fair use is a concept baked into US copyright law, and it's a defense to copyright infringement if certain elements are met.

OK, so that last part isn't *quite* right: if it's fair use it is explicitly *not* an infringement. But it's closer than most commentaries get, and maybe close enough. (One comment gets this right.)

The last paragraph is interesting, spelling out that Google's not using either source of potential revenue in Google books searching: it doesn't display ads and it doesn't get a cut if you buy a book linking from the search. Those decisions may have helped Google consistently win in court.

More than 150 comments, most of them useful.

Authors Guild Petulantly Whines About How Wrong It Is That The Public Will Benefit From Google Books

For a slightly different take on the decision (or non-decision) and early reaction, here's Mike Masnick's April 19, 2016 piece at *techdirt*. He liked the rejection and was a little surprised by the Authors Guild reaction:

> It's no surprise that the Authors Guild—which has been tilting at this particular windmill for over a decade—was upset about the refusal to hear the case, but I wasn't quite expecting the level of ridiculous sour grapes that were put on display:
>
> > "Blinded by the public benefit arguments, the Second Circuit's ruling tells us that Google, not authors, deserves to profit from the digitization of their books," said Mary Rasenberger, executive director of the Authors Guild.
>
> Did you get that? The Authors Guild is so completely out of touch that it actually thinks that "public benefit arguments" have no place in copyright disputes, despite the very fact that the Constitutional underpinnings of copyright law is to maximize the public's benefit. And, of course, this all ignores the fact that the vast, vast majority of authors greatly benefit from such a searchable index in that it drives more sales of books.

There's more as Masnick partially fisks a remarkable tantrum, but you can read his commentary (and the AG statement) yourself. Still, I can't resist this one:

> "The price of this short-term public benefit may well be the future vitality of American culture," continued Rasenberger. "Authors are already among the most poorly paid workers in America; if tomorrow's authors cannot make a living from their work, only the independently wealthy or the subsidized will be able to pursue a career in writing, and America's intellectual and artistic soul will be impoverished."

> This is ridiculous on so many levels. First, most authors cannot make a living today because most books don't sell. That's not the fault of Google Books. In fact, as noted time and time again, Google Books acts as a *discovery mechanism* for many books and increases sales (I've bought dozens of books thanks to finding them via Google Book Search). Second, the gloom and doom predictions of legacy industries over new technologies is time-worn and **has never been even remotely correct**.

Bingo. And, as he notes, there are a lot more published authors now than there were in 2005. But the killer here, and the reason for my question mark, is that AG *will not give up:*

> Following the Supreme Court's order, the Guild vowed to remain vigilant to ensure that the Second Circuit's ruling is not taken as carte blanche for unfettered digitization of books. "The Second Circuit decision took pains to highlight that fair use was found based on the strict display restrictions and security measures currently employed by Google," said Authors Guild general counsel Jan Constantine. "We'll continue to monitor Google and its library partners to ensure these standards are met, as we will take appropriate action to ensure that fair use isn't abused."

Several dozen comments, a mix of useful commentary and confused trolling. An AG person even rings in, and apparently doesn't understand paragraphs (also suggests that "hobby writers" don't write great books, which would surprise literary scholars).

Two Weeks in October

The last time I wrote about the Google Books case was GOOGLE BOOKS: THE NEVERENDING STORY? in November 2015, part of an all-fair-use issue (actually written in late September 2015: Windows Explorer says I created the PDFs for that issue on October 5, 2015). That essay offers a probably-incomplete set of links to earlier *C&I* coverage dating back to December 2005—a set of links that may show why I'm not yielding to the temptation to extend this essay by copying in all the earlier ones. To wit, even omitting some smaller items not mentioned in November 2015, the background would be 170 traditional *Cites & Insights* pages or a roughly 300-page 6" x 9" book. Now, to be vainly honest, I think such a book—another C&I Reader—would be entertaining and mildly useful, but only if it was indexed. I can't justify the effort at this point, especially given the likelihood of absurdly low sales of a priced version—and I won't burden you with an insanely long almost-all-reprint issue.

The first group of items is from two weeks in October 2015.

Appeals court rules that Google book scanning is fair use

That's Joe Mullin on October 16, 2015 at *ars technica* on the 2nd District Court of Appeals decision that the Supreme Court just decided not to review.

> The Authors' Guild sued Google, saying that serving up search results from scanned books infringes on publishers' copyrights, even though the search giant shows only restricted snippets of the work. The authors' group said that Google's book search isn't transformative, that the snippets provide an illegal free substitute for their work, and that Google Books infringes their "derivative rights" in revenue they could gain from a "licensed search" market.
>
> In its opinion (PDF), a three-judge panel rejected all of the Authors' Guild claims in a decision that will broaden the scope of fair use in the digital age. The immediate effect means that Google Books won't have to close up shop or ask book publishers for permission to scan. In the long run, the ruling could inspire other large-scale digitization projects.

Mullin reviews the court's analysis of Fair Use's four factors. The court found scanning for an index (and the ngrams tool) highly transformative, along with the snippets view. "Nature of the copyrighted work" wasn't especially significant. The third factor—amount of the work used—is key:

> "While Google *makes* an unauthorized digital copy of the entire book, it does not reveal that digital copy to the public," the opinion states. "The copy is made to enable the search functions to reveal limited, important information about the books."
>
> The judges also carefully analyze Google's use of "snippet view." The snippets are small, normally an eighth of a page. No more than three snippets are shown for any searched term, and no more than one per page. Google also "blacklists" some parts of each page—and one full page out of each ten—excluding them from snippet view entirely. Finally, snippet view isn't available in cases where a snippet might entirely satisfy a reader's needs, such as in dictionaries or cookbooks.

Finally, the court didn't find potential loss of sales to be a convincing factor.

There's more, none of it helpful to the Authors Guild, which responded predictably:

> "The Authors Guild is disappointed that the Court has failed to reverse the District Court's flawed interpretation of the fair use doctrine," said Authors Guild Executive Director Mary Rasenberger. "Most full-time authors live on the edge of being able to

keep writing as a profession, as our recent income survey showed; a loss of licensing revenue can tip the balance... We are very disheartened that the court was unable to understand the grave impact that this decision, if left standing, could have on copyright incentives and, ultimately, our literary heritage.

A mixture of comments, some worth reading.

Libraries laud appeals court affirmation that mass book digitization by Google is 'fair use'

Kara Malenfant posted this press release on *ACRL insider* on October 16, 2015, on behalf of the Library Copyright Alliance (ALA, ACRL and ARL), and since it's clearly intended to be quoted as fully as possible, I'll quote the whole thing:

> The U.S. Court of Appeals for the Second Circuit today ruled in *Authors Guild v. Google* that Google Book's mass digital indexing of books for use in creating a searchable online library constituted a legal "fair use" of copyrighted material rather than an infringement. Statements by members of the Library Copyright Alliance may be attributed as follows:
>
> **Sari Feldman, president, American Library Association (ALA):**
>
> "The Court's decision today does much more than affirm the critical importance and clear legality of digitally indexing books on a large scale as fair use, hugely important as that is. The ruling's broader and potentially landmark legacy is that vague fears of speculative harm due to possible copyright infringement cannot and must not be permitted to deprive every sector of our society of the very real and identifiable benefits of fair use and other legal limits on copyright. In other words, the assumption that maximum restriction in copyright is the path to maximum benefit which for too long has animated US copyright policies, laws and treaties has been soundly and rationally rejected."
>
> **Ann Campion Riley, president, Association of College & Research Libraries (ACRL):**
>
> "I join my colleagues in applauding this ruling as it strongly supports fair use principles, allowing scholars and others to discover a wealth of resources. This is a tremendous opportunity for our communities, in particular for students and others with visual disabilities, as Google Book search makes millions of books searchable."
>
> **Larry Alford, president, Association of Research Libraries (ARL):**
>
> "The Association of Research Libraries applauds this victory for fair use regarding the Google Books project, which involved partnerships with many of our libraries. This important project supports the goal of the copyright system to expand knowledge by supporting access for those with print disabilities with the creation of machine-readable digital copies, and by providing a new research tool for users who can, among other uses engage in text-and-data mining to understand in a deeper way the record of human culture and achievement contained in these books that would not otherwise be possible without the large, searchable database created by Google Books."

Maybe I'm biased, but I think those are all fine statements that don't require further commentary.

Second Circuit issued opinion in Google Books case!

I'm linking to this October 16, 2015 post by Nancy Sims at *Copyright Librarian* not because I plan to excerpt and comment, but rather because it's an interesting Storify essay, a series of Sims' live tweets as she was reading the decision.

I recommend reading the piece: it's not long and it's quite interesting—including Sims' reminder that one reason this all took so long was because Google tried so hard to *avoid* fair-use based litigation, preferring to negotiate a deal instead. Also interesting for other reasons.

Google Books, Fair Use, and the Public Good

Kevin Smith offered this take on the decision on October 18, 2015 at *Scholarly Communications @ Duke*. As you might expect, Smith was not displeased: "The decision was a complete vindication of the District Court's dismissal of the case, affirming fair use and rejecting all of the counterarguments offered by the Authors Guild."

He offers links to some other useful analyses and adds points he finds "especially important":

> First, Judge Pierre Leval, who wrote the opinion, does a nice job of drawing a line from the idea of transformative uses to the public purpose of copyright law… Judge Leval reminds us quite forcibly that the primary beneficiary intended by copyright law is the public, through "access to knowledge" (p.13) and "expand[ed] public learning" (p. 15). Economic benefits for authors are instrumental, not the ultimate goal of the copyright monopoly… Then Judge Leval explains how this analysis of transformation serves those goals, clarifying why fair use is an essential part of copyright's fundamental purpose…
>
> Another important thing we can learn from Judge Leval's opinion is about the difference between a transformative use and a derivative work. The Author's Guild (really some individual authors set up as plaintiffs because the AG has been found to lack standing to sue in this case) argues that allowing the Google Books' search function usurps a right held by those authors to license indexing of their works. This is ridiculous on its face, of course — imagine

> the effect such a right would have on libraries — but the judge does a nice job of explaining why it is so wrong. The decisions rest heavily on the idea/expression dichotomy that is fundamental in copyright, and stresses that what is presented in the Google Books "snippet view" is more information about books (facts) rather than expressive content from those books…

There's more, to be sure, including his note that AG *already* intending to appeal to the Supreme Court shows "the AG simply does not know when to cut its losses and stop wasting the money provided by its members" and his assertion that the "Supreme Court is not likely to take the case anyway"—with reasons:

> [T]his is just a case about a greedy plaintiff who wants to be given an even bigger slice of the copyright pie, which the courts have determined repeatedly it does not deserve. This is not the sort of issue that attracts the very limited attention of the Supreme Court. In fact, reading the Court of Appeals' ruling leaves one with a sense that many of the AG's arguments were rather silly, and there is no reason to believe they would be less silly when presented to the Supreme Court in a petition for certiorari.

A brief item on October 19, 2015 at *Inside Higher Ed* quotes the AG's Mary Rasenberger, who flatly calls Google Books "Google's seizure of property" and, of course, calls it a threat to "our intellectual culture."

Another Big Win for Google Books (and for Researchers)

Rick Anderson offered his own analysis of the decision and its significance in this October 21, 2015 piece at *the scholarly kitchen*.

It's a good discussion, making good use of "But wait." at several points to introduce important points and undermine some seeming myths. In one case, Anderson sees the flipside of a problem I've been aware of (through my wife, from when she was a college library director): the notion that one of the four Fair Use tests trumps all the others,

> But wait. Google is a highly profitable commercial entity. How can its wholesale copying of copyrighted texts, undertaken for an explicitly commercial purpose, possibly be considered fair use? Here the court points out something that many of us in libraries have tried to help students, faculty members, and sometimes even publishers understand for many years: the fourfold test for fair use doesn't contain any trump tests—fair use is determined by considering the proposed use holistically. Professors and students who take the "Purpose and Character of Use" test for a trump card and therefore believe they can do anything they want with copyrighted works as long as the use is educational are falling victim to this misunderstanding, but the same is true of copyright holders who believe that fair use is always fatally undermined by commercial reuse. As the court puts it in this decision, "Our court has… repeatedly rejected the contention that commercial motivation should outweigh" an otherwise convincing fair use argument. In this case specifically, "we see no reason… why Google's overall profit motivation should prevail as a reason for denying fair use over its highly convincing transformative purpose, together with the absence of significant substitutive competition, as reasons for granting fair use."

Nicely put, as is the rest of the discussion. Anderson closes by wondering whether some librarians may be unhappy Because Google, although he notes that early librarian responses are cases where "my colleagues mostly agree with me: this finding is a win for Google, for libraries, and for researchers."

The close:

> The Authors Guild may not believe it, but on balance it's almost certainly a win for authors and publishers as well.

Right on both counts, as far as I can tell: it's a win…and AG won't (can't?) believe that.

Skitch comments are almost always more thoughtful and varied than most sites' "Don't read the" comments; in this case, a commenter immediately conflates library support of OA (an entirely different topic) with lack of respect for authors' intellectual property rights. Later, a consultant blasting away at Google essentially says that librarians would like to see copyright eliminated entirely. It is, in fact, an *interesting* comment stream…but perhaps not very enlightening.

Authors, Nay, EVERYONE Will Get Screwed If The Authors Guild Wins the Google Books Lawsuit

So says Nate Hoffelder in this October 20, 2015 item at *The Digital Reader*—and he's looking at potential outcomes from a very different perspective.

He starts with some comments on AG's statement after the Court of Appeals verdict:

> The statement is chock full of ridiculous statements like "America owes its thriving literary culture to copyright protection."
>
> That, my dear, is utter nonsense. Until the 1976 Copyright Act was passed, the US copyright law limited the copyright term to 56 years and required that a copyright be registered and renewed after 28 years, otherwise the work was in the public domain.

> By modern standards, that would best be described as a lack of copyright protection, because it let publishers ignore foreign copyrights unless that copyright was also registered in the US, and it made every work more than 56 years old fair game.
>
> The Authors Guild is wrong in almost everything they say in that statement, and they are also completely wrong to pursue this doomed effort to force Google to pay for the use of excerpts.

After noting some other reasons AG is wrong, he offers a number of examples to support his major argument: "**Google will not pay the fees.**"

As evidence, he cites a German attempt to require micropayments for news snippets (Google basically said "if we can't show snippets we won't cover your newspapers," which strikes me as an appropriate response); a Belgian case where publishers *won* in court—and Google *did* remove the publishers, which (naturally) convinced them to change their tune; and Spain, where legislation required fees for news snippets…resulting in Google *shutting down* Google News in Spain. Hoffelder reiterates:

> **Google will not pay a fee for giving away free advertising - not to web publishers, nor to book publishers and authors.**

He says Google would shut down Google Books rather than pay authors and publishers. Is he right? Fortunately, we shouldn't have to find out. (Of course, in the early years of this marathon Google *was* willing to pay authors—but that involved a new semi-monopolistic revenue stream for Google as well,)

Hoffelder links to an October 19, 2015 essay by J.A. Konrath that includes a fisking of the AG statement—and I tagged that essay for discussion, but I think I'll just link to it. You have to appreciate Konrath's style. I suspect his prescriptions for an Authors Guild that actually served authors are pretty good, but that's a different discussion.

After 10 Years, Google Books Is Legal

That's an unfortunate title for Robinson Meyer's October 20, 2015 article at *The Atlantic:* Google Books was legal all along, at least based on all the court rulings—and the October 2015 decision wasn't the final possible step. But never mind. The essay is better than the title.

> Fair warning: this is a Traditional Media Website: you'll be ~~plagued~~ interrupted by advertising in an intrusive manner that doesn't happen in actual traditional print media like *The Atlantic*.

Meyer takes us back 25 years to Pierre Leval's landmark *Harvard Law Review* article on Fair Use, and specifically Leval's argument that transformation should be a major factor in determining fair use:

> A quotation of copyrighted material that merely repackages or republishes the original is unlikely to pass the test; in Justice Story's words, it would merely "supersede the objects" of the original. If, on the other hand, the quoted matter is used as raw material, transformed in the creation of new information, new aesthetics, new insights and understandings, this is *the very type of activity* that the fair use doctrine intends to protect for the enrichment of society.
>
> A quotation of copyrighted material that merely repackages or republishes the original is unlikely to pass the test; in Justice Story's words, it would merely "supersede the objects" of the original. If, on the other hand, the quoted matter is used as raw material, transformed in the creation of new information, new aesthetics, new insights and understandings, this is *the very type of activity* that the fair use doctrine intends to protect for the enrichment of society.

The article was cited in the Supreme Court's ruling in one landmark Fair Use case (the "2 Live Crew case") and the doctrine has been used in a number of other cases, to the extent that the Ninth Circuit (which includes Silicon Valley) appeared to be a hotbed of expansionary fair use.

And guess who is on the Second Circuit Court of Appeals—and who wrote the decision for Authors Guild's appeal of its lower-court loss? Judge Pierre Leval. Who believes Google got it exactly right.

There's more to this article (which quotes James Grimmelmann among others), and it's worth reading. I would note that this article also (correctly) argues that the Supreme Court wouldn't hear an appeal:

> And not only that, but the case is likely resolved for good. In 2012, a district court ruled that Hathitrust, a university consortium that used Google Books's scans to make books accessible to blind students, was not only a legal form of fair use but also required by the Americans with Disabilities Act. Experts say that the Supreme Court is unlikely to hear an appeal, because so many district court judges, and two different federal circuits, have found themselves so broadly in agreement about the nature of transformative use online.
>
> "The Authors Guild is deluding itself to think that this is an area that is open and controversial in the view of the lower courts," Grimmelmann said.

What I especially like about Leval's arguments is that he views fair use in the light of Constitutional copyright law. Quoting the article's closing paragraph:

> A muscular fair use is precisely what Leval had in mind in his original 1990 article. "The copyright is

> not an inevitable, divine, or natural right that confers on authors the absolute ownership of their creations," he wrote. "It is designed rather to stimulate activity and progress in the arts for the intellectual enrichment of the public."

You will not be surprised that comments include long and factually inaccurate attacks on Google and the decision by at least one author, who among other things makes the simply outrageous statement that "Those who defend Google Books are those who want 'something for nothing." He apparently also believes that ideas are or should be copyrightable.

What the Google Books Victory Means for Readers

To close out the October 2015 segment, here's a piece by Dan Cohen (of the Digital Public Library of America, which might usefully be the subject of a complex essay, but not now and probably not by me) on October 22, 2015 at *The Atlantic*.

The tease for the article: "Is a universal library finally within reach?" You can probably guess my answer: No.

My somewhat snarky feelings about Cohen and DPLA, and to some extent this article, aren't helped a lot by Cohen's seeming digital triumphalism—he touts not only claimed big increases in percentages of Americans reading ebooks but also, and seemingly as approvingly, a supposed drop in print-book reading. (Both findings are from Pew Research polls—and the linked discussion doesn't have the digital-triumphalist bent that Cohen offers. For one thing, the rise in ebook reading ends in 2014, with a tiny *drop* in 2015.)

OK, so I have issues with how DPLA is sometimes presented and with Cohen's "Yay digital! And look, print books are disappearing!" attitude. That said, it's an interesting essay.

Once again, odd sets of comments—including one author who not only bewails increased fair use but also fairly explicitly says that the rights of creators are in opposition to the rights of "the public"—such as readers. In another heated comment, this author says about Google: "They are not a library. You are giving control of our literary heritage, and possibly our current literature, to a single entity, for FREE." One of those two sentences is absolutely true. The other is nonsensical unless all libraries, bookstores (e and otherwise) and other sources of "our literary heritage" have somehow been destroyed or placed under Google's control.

January-February 2016

Then came the ~~threatened~~ promised appeal. Technically, that appeal (and the first coverage) happened on December 31, 2015, so "January" above is a rounding option.

The Authors Guild Files Supreme Court Appeal in Google Books Case

Begin with this December 31, 2015 piece by Nate Hoffelder at *The Digital Reader*.

> The Authors Guild has suffered defeat at every stage of their decade-old legal battle over Google's book-scanning project, including losing an appeal in October, but that hasn't stayed their resolve.
>
> Earlier today The Author's Guild announced that they had filed an appeal with the US Supreme Court.

There's some history (in which Hoffelder says the book-scanning portion of Google Books is "now defunct") and links to AG's own Q&A, a remarkable document that seems to praise Google Book Search while stating flatly:

> Google copied 20 million books to create a massive and uniquely valuable database, all without asking for copyright permission or paying their authors a cent. It mines this vast natural language storehouse for various purposes, not least among them to improve the performance of its search and translation services. The problem is that before Google created Book Search, it digitized and made many digital copies of millions of copyrighted books, which the company never paid for. It never even bought a single book. That, in itself, was an act of theft. If you did it with a single book, you'd be infringing.

I suggest reading the Authors Guild Q&A; maybe you'll find it more persuasive than I do. Do note that the document uses scare quotes around Fair Use, and that it makes the remarkable assertion that Google's provision of visibility for and snippet views of books means that "the market to bring back out of print books is completely devalued."

Hoffelder concludes:

> If The Authors Guild is successful, we're all going to lose. Google has shown that they will not pay for the use of news articles in its search results, and there is no reason to think they will change their mind when it comes to books.
>
> If the fair use ruling is overturned, then Google Books will be stripped down to public domain titles and the works that publishers and authors uploaded on their own accord. It will be an anemic shadow of itself, and not half as useful as it is now.

Sounds about right to me.

End of the Line for Google Books Lawsuit?

Carl Straumheim's January 5, 2016 story at *Inside Higher Ed* has a different web title, as you'll see in the hyperlink, but it carries the same message: copyright experts didn't think the Supreme Court would hear this case.

This piece does link to the appeal itself (embedded within the webpage linked to), a 185-page document that I did not read in its entirety and will not comment on. Nor was the brief itself the basis for some of the opinions:

> "The Supreme Court typically takes cases when there are important unsettled issues of law that need to be decided or in cases of overwhelming importance," said James Grimmelmann, professor of law at the University of Maryland at Baltimore. "This case might have seemed like a case of overwhelming importance a decade ago, but it has dragged on for so long and the ground has moved so much in copyright that it doesn't have that urgency."

(Later in this story Grimmelmann does comment on the brief.)

Not many comments; Sanford Gray Thatcher takes a dim view of what he calls a transformation of fair use.

Leading authors press for Supreme Court review of Google's digitised library

So says Alison Flood in this February 8, 2016 story at *The Guardian*, discussing an amicus brief in the name of seventeen authors and dramatists.

This time, I did read the 28-page document, which uses the first several pages to wow us with the awards and bona fides of the "authors" of the brief. I scare-quote authors because it's pretty clear that Ursula LeGuin and Stephen Sondheim aren't legal scholars; they signed on to a brief prepared by lawyers, which is perfectly appropriate. I'm not entirely sure why Nobel, Pulitzer and Booker awards mean somebody's legal opinions should carry extra weight, but I suspect this kind of thing happens in many celebrity-based briefs. In any case, you can read the brief for yourself, including the (unsurprising) swipes at libraries and HathiTrust.

The article quotes some of the brief, and I find the first quoted passage in this paragraph especially telling:

> The authors, who also include Thomas Keneally, Ursula K Le Guin, Tracy Chevalier, Yann Martel and Richard Flanagan, write in their filing that "copyright protection was included in the constitution to reward authors and provide incentives for them to continue writing", and that the fair use doctrine was not intended "to permit a wealthy for-profit entity to digitise millions of works and to cut off authors' licensing of their reproduction, distribution, and public display rights.

You know, I've looked at the Constitution carefully for "to reward authors" and damned if I can find that clause. For that matter,

> To promote the Progress of Science and useful Arts, by securing for limited Times to Authors and Inventors the exclusive Right to their respective Writings and Discoveries;

(copied from the official site for the Constitution's text and images) doesn't say "for very long terms so the *same* authors and inventors will keep writing."

Steal this book?

Read this February 11, 2016 post by Kevin Smith at *Scholarly Communications @ Duke.*

Maybe that's all I should say, but what fun would that be?

Smith apparently had a chance encounter with a woman who turned out to be the president of the Authors Guild.

> She began to explain to me why Google was wrong, but that the author for whom I was doing the research should be allowed to rely on fair use. When I introduced myself as a lawyer and copyright specialist for the Libraries, the conversation came to a polite but stilted conclusion.

But then he got to read the woman's (Roxana Robinson) *Wall Street Journal* column with its oh-so-nuanced title "How Google Stole the Work of Millions of Authors." I won't link to it since all but the first paragraph is behind a paywall and quickly covered with an ad.

> Ms. Robinson, a novelist and biographer, unfortunately proves what I suspected at the time of our encounter, that her perspective on fair use is based on a preconceived idea about who are good users entitled to rely on fair use (authors) and who are bad, unworthy users (Google), rather than on an understanding of the careful legal analysis of specific uses that actually underlies these decisions.
>
> The WSJ column employs some interesting rhetoric, starting with its title, which is clearly intended to provoke a visceral response. Many people have noted that the language of theft and stealing is inappropriate when the issue is copyright infringement. This point is made in great detail in William Patry's book "Moral Panics and the Copyright Wars." As is true for most crimes, the definition of theft includes an intention, a mental state or "mens rea" that is a required element of that crime. For theft this intention

> is "to deprive the true owner of [the personal property]" (definition from Black's Law Dictionary, Seventh edition). Because of the nature of intellectual property, copyright infringement never meets this definition; that is why the law has a different word — infringement — for the unauthorized taking of someone else's IP.

That's just a bit of a wonderfully informative (and, of course, well-written) discussion. Smith pushes something that frequently gets overlooked in discussions of copyright and "crime":

> It is worthwhile, nevertheless, to think a minute about the logic structure of the argument that what Google has done is infringement. Ms. Robinson makes the point that there are many books that were scanned by Google, that Google is a profitable company, and that no authorization for the scanning was asked for or given by the authors of the works that were scanned. All of this is true, of course, but it does not amount to an argument that Google has infringed any copyrights. What is missing, at least as I see it, is any notice that the authors have been harmed. The rhetoric of the column clearly tells us that the Authors Guild, and at least some individual authors who are involved in the lawsuit, are angry. But it does not explain a fundamental element of any tort action — harm.

Have the seventeen authors in the brief or the 800 (or 8,000) members of the Authors Guild been harmed in any way by Google Books? The courts have said no, and the loss of a possible "licensing market" is, well, Smith uses the word "silly."

And there's this, which I'd missed along the way: despite continued claims in various comment streams that you can view whole books by clever use of the Google Book views and snippets, "**the AG's own experts** were unable to retrieve as much as 16% of any work using word searches and snippet results, and even that amount of text was randomized in a way that made reading a coherent piece of the work impossible." [Emphasis added.]

Why the Authors Guild Is Still Wrong About Google's Book Scanning

Now that we've seen excellent discussion based on a commentary in one business-oriented publication (but one that won't show me the content), let's finish with commentary in another business-oriented publication: Mathew Ingram's February 8, 2016 commentary in *Fortune*. You'll have to put up with ads (including video ads) but you'll be able to read the article.

> Scanning books has a clear public benefit, which is what copyright law was designed for.

That's the lead paragraph, and I'm guessing some authors immediately thought "Nonsense! Copyright law is there to protect our income!" Ingram's right, of course: see my earlier quotation of the underlying Constitutional text.

OK, so Ingram gets it wrong later on when he says this:

> Copyright infringement can be allowed under U.S. law if a court decides that it constitutes "fair use" of the original material.

He shouldn't be scare-quoting fair use—but in any case, the law clearly states that if it's fair use it is *not an infringement*: it's a fully legal exception. I think that's an important distinction, which is why keep mentioning it.

Otherwise, this is a solid brief commentary. And as good a place as any to wrap this up.

Pay What You Wish

Cites & Insights carries no advertising and has no sponsorship. It does have costs, both direct and indirect. If you find it valuable or interesting, you are invited to contribute toward its ongoing operation. The Paypal donation button (for which you can use Paypal or a credit card) is on the *Cites & Insights* home page. Thanks.

For readers of the traditional two-column version: I'm leaving the final page blank so you can take notes…well, actually because I don't feel like padding it out with a one-page THE BACK, although I have some topics in mind (such as a high-minded geography association running jet trips that use a 757's carbon footprint to transport 75 masters of the universe to seven cities around the world over 22 days for a mere $72K, global warming be damned…).

Masthead

Cites & Insights: Crawford at Large, Volume 16, Number 4, Whole # 194, ISSN 1534-0937, a periodical of libraries, policy, technology and media, is written and produced, usually monthly, by Walt Crawford.

Comments should be sent to waltcrawford@gmail.com.

URL: citesandinsights.info/civ16i4.pdf