Cites & Insights

Crawford at Large

Libraries • Policy • Technology • Media

Volume 16, Number 3: April 2016 ISSN 1534-0937 Walt Crawford

The Front

Gold Open Access Journals 2011-2015: A SPARC Project

I'm delighted to announce that SPARC (the Scholarly Publishing and Academic Resources Coalition) is supporting the update of *Gold Open Access Journals 2011-2015* to provide an empirical basis for evaluating Open Access sustainability models. I am carrying out this project with SPARC's sponsorship, building from and expanding on *The Gold OA Landscape 2011-2014*.

The immediate effect of this project is that the dataset for the earlier project is publicly available for use on zenodo.org and on my personal website. The data is public domain, but attribution and feedback are both appreciated.

Here's what the rest of the project means:

- I am basing the study on the *Directory of Open Access Journals* as of December 31, 2015. With eleven duplicates (same URL, different journal names, typically in two languages) removed and reported back to *DOAJ*, that means a starting point of 10,948 journals. All journals will be accounted for, and as many as feasible will be fully analyzed.
- The grades and subgrades have been simplified and clarified, and two categories of journal excluded from the 2014 study will now be included (but tagged so that they can be counted separately if desired): journals consisting primarily of conference reports peer-reviewed at the conference level, and journals that require free registration to read articles.
- I'm visiting all journal sites (and using *DOAJ* as an additional source) to determine current article processing charges (if any), add 2015 article counts to data carried over from the 2014 project, clean up article counts as feasible, and add 2011-2015 article counts for journals not in the earlier report.
- Since some journals (typically smaller ones) take some time to post articles, and since some journals will not be analyzed for various reasons (malware, inability to access, difficulty in translating site or counting articles), I'll be doing a second pass for those requiring such a pass, starting in April 2016. My intent is to include as many journals as possible (although existence of malware is an automatic stopping point).
- The results will be written up in a form somewhat similar to *The Gold OA Landscape 2011-2014*, refined based on feedback and discussion.
- Once the analysis and preparation are complete, the dataset (in anonymized form) will be made freely available at appropriate sites and publicized as available.
- The PDF version of the final report will be *freely available* and carry an appropriate Creative Commons license.
- A paperback version of the final report will be available; details will be announced closer to publication.
- A shorter version of the final report will appear in *Cites & Insights*, and it's likely that notes along the way will also appear there.

My thanks to SPARC for making this possible.

Inside This Issue

Grade Changes from *The Gold OA Landscape 2011-2014*

After reviewing the numbers in *The Gold OA Landscape 2011-2014* and considering what I can and, more significantly, *cannot* reasonably ascertain and judge in non-English journals and in short visits to websites, and in

consultation with SPARC contacts, I made a number of changes in grades and, as a result, in exclusions.

I did *not* change the list of subjects and areas, although a few journals may have been assigned new subjects—and, as in the previous study, *PLOS One* is omitted from subject and area figures but included in overall discussions.

The fundamental meaning of Grade B has changed from "deserves attention" to "might be excluded from *DOAJ* or in some versions of Open Access."

Changes in Grade A Subgrades

All subgrades for Grade A have been eliminated. Subgrade C (ceased) is now a subgrade for Grade B. Subgrades D, E, H, O and S—all cases where *some* year other than the first had fewer than five articles—have been collapsed into Grade B, Subgrade F (few or no 2015 articles) if the article count for 2015 is less than 5 and simply Grade A otherwise.

Changes in Grade B Subgrades

Grade B consists of journals that may or may not belong, either in *DOAJ* or in a study of open access, depending on your definitions. The old subgrades all have to do with mild visual or editorial issues that now seem as though they're imposing my own values inappropriately.

There are four new subgrades—two from Grade A and two from Grade X, albeit with different letters.

- **C: Ceased**—journals that published at least one article later than 2010 but explicitly ceased during or before 2015, have merged with other journals, or show no articles more recent than 2012.
- **F: Few or no 2015 articles**—journals that published at least one article later than 2012 and published fewer than five articles in 2015. (By current *DOAJ* rules, these are subject to delisting.)
- **R: Conference and other reports**—journals consisting entirely or primarily of conference papers and other reports. These were previously excluded, in subgrade XN, as not OA.
- **S: Sign-in or registration required**—journals that require some form of free registration before reading articles. These were previously excluded, also in subgrade XN, as not OA.

Changes in Grade C Subgrades

Grade C, "avoid this journal," has been narrowed somewhat, specifically to eliminate subgrades that involve personal judgment or have so few journals that they're hardly worth noting. Specifically, subgrades E (very bad English), S (incoherent site) and T (absurd article titles—there were almost none of these) have been eliminated, leaving subgrades A (APC missing), F (clear falsehoods), O (mix of problems) and P (implausible peer review turnaround). Briefly, clear falsehoods are statements such as "the leading journal in this field" for a brand-new journal; implausible peer-review turnaround involves promises to complete all peer reviews in a couple of days.

Changes in Grade X Subgrades

Grade X, excluded journals, retains the same subgrades—but the two largest categories within subgrade N (not OA) have been moved to subgrades BR and BS.

A Partial Checkpoint

What are the consequences of these changes? In general, and combined with more exhaustive checking of some difficult situations, they should mean that more journals will be included in the full analysis. As for specific results, those won't be clear until the project is complete.

I thought it would be worth offering some glimpses into what *might* be happening at a natural breakpoint: essentially halfway through the first pass of data gathering (actually 5,500 of 10,948).

First pass? Yes indeed. There will be a second pass, beginning no earlier than April 1, 2016, for quite a few of the journals, for various reasons:

- Many smaller journals, especially in the humanities and social sciences, post online articles and issues with significant delays. In practice, even waiting a year won't get them all. I'm rechecking all journals that appear to be missing final issues for 2015; this gives them at least three months to get the articles posted.
- I'm rechecking all journals that couldn't be reached or that showed signs of malware, as well as those that showed as parking or ad pages or were unworkable.
- I'll take a second look at journals excluded for various reasons, trying harder to make sense of opaque cases and translation difficulties, looking more closely for apparently-missing APCs, rechecking whether certain journals are OA or not.

So far, it looks as though I'll need to recheck about one-fifth of the journals: 1,047 of the first 5,500. I'd be delighted if that percentage goes down in the second half—but I'd also be surprised.

All the rest of these numbers are truly tentative, since review of the journals may change their categorization.

Free and Pay

Some journals start imposing APCs that didn't have them previously (one large publisher dropped all of its free introductory periods); some (fewer) drop APCs; and some clarify the nature of their charges.

Overall, the percentage of no-APC journals (among journals where it's clear) among the first half dropped from 64.9% to 59.8%: there are more no-fee journals than in the previous study, but there are a *lot* more APC-charging journals. (There are also, to be sure, more journals in general: about 412 so far.) There are *fewer* journals (so far) where there is an APC but it's hidden.

The Newbies

Most journals that weren't in the 2014 study are simply A (that is, "nothing special here one way or the other"), but 30 have fewer than five articles in 2015, a few couldn't be contacted or were unworkable, a handful fall into various other categories—and, unfortunately, nine showed signs of malware.

Neutral Changes

Some changes in grade and subgrade are neutral: they're just redefinitions. That's true for the journals that changed from various A grades to BC (ceased explicitly or with no articles later than 2012): there are some 218 BC so far. It's also true for the various A subgrades that are now simply A (around 230 of them) and for a number of other changes including quite a few moving from B subgrades to A.

Some 300 journals had five or more articles in 2014 but not in 2015, moving them all to BF: some of those will add articles in a recheck.

Changes for the Good

Some 27 journals previously graded CA (APC missing or hidden) now have more clarity (and four changed to various X subgrades).

Quite a few journals with explicit falsehoods on their homepages have been cleaned up—at least 80 of them.

Half a dozen journals flagged for malware no longer seem to have that problem (but see later!).

Most "not OA" entries in the first half have moved elsewhere on re-examination or redefinition, including 35 journals oriented to conference programs (another seven that had been "A" appear to be *predominantly* conferences and have been moved here) and ten that require registration to read articles. Some two dozen moved elsewhere, including 17 that now appear to be proper OA journals.

Most journals that I previously found too difficult to count (XO) are now handled, and I hope to reduce the number (70 for this half in the previous study currently down to 28) even further.

Roughly half of the XT (couldn't understand the site well enough to measure it) cases have been cleared up: so far, there are only three such journals in the first half, and I'll try all of them again.

Changes for the Bad

A few journals have changed home pages such that I can no longer find an APC (but am sure they have one), but it's a tiny number.

Some 70 journals that were reachable the last time around are either unreachable or unworkable when I checked this time; they'll all be rechecked, but it's unfortunate that there are so many.

Finally there's the most unfortunate group, in my opinion: journals that now show signs of malware—frequently, I suspect, because they include ad networks that don't have proper standards. A journal gets flagged for malware if Malwarebytes or McAfee Site Advisor or Windows Defender flags it or some of its components as malware; cases include phishing attempts and deliberate malware downloads. There are now *twice* as many of these as there were (for this subset of journals) in the previous study, and that's about 72 too many.

Summing Up

Hundreds of new journals; a much shorter and simpler set of grades; adding literally thousands of peer-reviewed articles that were given as conference papers.

Far fewer journals falling by the wayside because I only read English (thanks, Google!) or because I can't or am unwilling to count them (with true broadband, I'm willing to open up a dozen PDFs a year to see how many articles there are).

There will still be some approximate counts, but fewer (and better approximations) than last time around.

And, of course, the results will be freely available to everybody. In a few months.

Changes for the Better?

Do you have suggestions that will help make *Gold Open Access Journals 2011-2015* even better than *The Gold OA Landscape 2011-2014?*

If so, now's the time to suggest them—any time between now and May 1, 2016 (the earliest date I'm likely to start working on data analysis and the book manuscript). Suggestions should go to me at waltcrawford@gmail.com.

You say you haven't purchased the book yet, either in paperback or PDF ebook form? You still can,

and it will still be worthwhile when the new book comes out.

Alternatively, you can get a good idea of the general approach and tables used in the excerpt published as the October 2015 *Cites & Insights*, although that version lacks any graphs.

I've appended pages 39 through 73 of *The Gold OA Landscape 2011-2014* to the end of this issue. That segment includes almost all varieties of tables and graphs used in the book. The online version is an *exact* replica of the print book; the print (two-column) version is just slightly smaller, so that four pages of the 6x9" book fit on each 8.5x11" sheet rather than having loads of waste space.

The Basics

Basically, the data used for analysis includes for each journal the year reported to *DOAJ* (which is not always the start of publication), the country of publication (again as reported to *DOAJ*), one of 28 subjects and three broad areas that I've derived from the subjects, keywords and journal/article titles for the journals, and the data I went looking for: whether there's an author-side fee (usually called an APC or Article Processing Charge but they're not all that straightforward) and how much it is, and the number of published articles (and similar items) for each year 2011 through 2015. There's also a two-letter code (or "grade and subgrade") for special cases, but most journals don't have special codes. I also derive some measures: the peak article number during the five years and, if there are APCs, the maximum revenue for 2014 (2015 this time around).

Last year, after an overall discussion of maximum revenues, overall article counts, and special cases, I looked at journals by annual article volume for each of the three major areas (which have very different characteristics), fee and revenue levels, starting dates for free and APC-charging journals, and a number of measures by country of publication. I also provided one set of pie charts breaking down free and pay journals by major area.

For each of the three major areas (biomed, STEM, and humanities and social sciences) I looked at cost per article by year, journal and article volume by year (and free percentage of each), revenue brackets for journals, article volume brackets, and APC level brackets. A bar graph showed free and pay articles for each year.

For each subject within an area—using the revenue and article volume brackets appropriate for that area—I showed journals and articles for each year (and free percentage), the free/pay article bar graph, journals by article volume (and percent free), journals and articles by APC range, a line graph showing free and pay journals by starting date, and a table showing the countries with the most published 2014 articles for that subject.

At the end of the book, I provided a few subject summaries—percentage of free journals, percentage of articles in no-fee journals, change in article volume, change in free article volume, journals changing article volume by 10% or more from 2013 to 2014, average APC per paid article and for all articles, *median* APC per paid article and all articles, and the median, first quartile, and third quartile articles per journal for 2014.

Data Changes for 2015

There's another year of data—more journals and more data for existing journals. I'm taking some pains to include more journals (and defining "articles" somewhat more inclusively and, I believe, consistently).

Beyond that, there may be one new category of derived data: publisher category—breaking journals down into what seem to be five reasonable groups based on what's in the *DOAJ* publisher field:

- **Academic**, published by universities and colleges, including university presses.
- **Society**, published by societies and associations.
- **Traditional***, published by publishers that also publish subscription journals.
- **OA publisher***, published by groups that don't appear to publish subscription journals (and that publish at least a handful of journals—see notes on the "*" below)
- **Miscellany**, everybody else.

About the asterisk on **Traditional** and **OA publisher**: there are 5,983 different "publisher names" (that is, distinct character strings in the *DOAJ* publisher field). That's more than one "publisher" for every two journals. The vast majority of those, all but 919, publish a single *DOAJ*-listed journal.

I think it's reasonable to limit the two "publisher" categories (Traditional and OA) to firms that publish at least a handful of journals, and lump the others in as Miscellany. (If nothing else, it makes this added data feasible.)

What's a handful? If the cutoff is "five or more," it involves only 221 publishers in all, accounting for 4,128 journals. If the cutoff is "four or more," it involves 316 publishers—and, naturally, adds 380 journals for a total of 4,508. Dropping it to "three or more journals"

brings us up to 486 publishers and 5,018 journals. I suspect the final cutoff will be either four or five.

Incidentally, if I add that column, it *will* be in the anonymized spreadsheet made publicly available at the end of this project. Other than the list of journal titles apparently containing malware, it will be possible for anybody else to replicate any or all of the graphs and numbers in the book.

Probable Changes

I believe it will make sense to devote a chapter to publisher categories—whether there are major differences in article volume, APC charges (existence and amount) and, possibly, domination in some countries.

I'm fairly certain the pie charts will go away: I don't believe they add enough information to justify the space. I could be convinced otherwise. (Note that the print paperback will, of necessity, be black and white to keep production costs down, so really attractive pie charts aren't feasible.)

Possible Changes

What else should I consider? Which existing tables and graphs don't seem especially valuable—and what would work better? (Assume that this year's book can be larger than last, but not enormously larger.)

I'm open to suggestions, which I'll discuss with my contacts at SPARC (and I anticipate suggestions from SPARC as well).

I would offer a free PDF version of this year's book as a reward for good suggestions—but since this year's PDF version will be free in any case, that's not much of an offer. I will acknowledge especially useful suggestions (unless you prefer anonymity).

Why Anonymize?

The project plan for this project calls for me to make an anonymized version of the master spreadsheet freely available—and as soon as the project was approved, I made an anonymized version of the 2014 spreadsheet available.

Two people raised the question "Why anonymized?"—why don't I just post the spreadsheet including all data, instead of removing journal names, publishers and URLs and adding a simple numeric key to make rows unique?

The short answer is that doing so would shift the focus of the project from patterns and the overall state of gold OA to specifics, and lead to arguments as to whether the data was any good.

Maybe that's all the answer that's needed. Although I counted very little use of the 2014 spreadsheet in January and February 2016, it's been used more than 900 times in the first half of March 2016—but I have received no more queries as to why it's anonymized. For any analysis of patterns, of course, journal names don't matter. But maybe a slightly longer answer is useful.

That longer answer begins with the likelihood that some folks would try to undermine the report's findings by claiming that the data is full of errors—and the *certainty* that such folks could find "errors" in the data.

Am I being paranoid in suggesting this would happen? Thanks to Kent Anderson, I can safely say I'm not, since within a day or two of my posting the spreadsheet, he tweeted this:

Anderson didn't say "Am I misunderstanding?" or "Clarification needed" or any alternative suggesting that more information was needed. No: he went directly on the attack with "Errors exist" (by completely misreading the dataset, as it happens: around 500 gold OA *journals* began publication, usually not as OA, between 1853 and 1994).

It's not wrong, it's just different

To paraphrase Ed and Patsy Bruce (they wrote the song, even though Willie Nelson and Waylon Jennings had the big hit with it)…

If somebody else—especially someone looking to "invalidate" this research—goes back to do new counts on some number of journal, they will probably get different numbers in a fair number of cases.

Why? Several reasons:

- **Inclusiveness:** Which items in journals—and which journals—do you include? The 2014 count tended to be more exclusive when I had to count each article individually; the 2015 count tends to include all items subject to some form of review, including book reviews and case reports. Similarly, the 2015 report includes journals that consist of (reviewed) conference reports (although I'll note the subset of such journals).
- **Shortcuts:** I did not in fact look at each and every item in each and every issue of each and every journal, compare it to that journal's own criteria for reviewed or peer-reviewed, and determine whether to include it. To do that, I'd estimate that a single year's count would require

at least 2,000 hours *exclusive of* determining APC existence and levels and all other overhead—and, of course, a five-year study would require four times that amount (fewer journals and articles in earlier years). That's not plausible under any circumstances. Instead, I used every shortcut I could: publication-date indexes or equivalent for SciELO, J-Stage, MDPI, Dove and several others; DOI numbers when it's clear they're assigned sequentially; numbered tables of contents; Find (Ctrl-F) counts for distinctive strings (e.g., "doi:" or "HTML") after quick scans of the contents tables. For the latter, I did make rough adjustments for clear editorials and other overhead.

- **Estimates:** In some cases—fewer in 2015 than in 2014, but still some—I had to estimate, as for instance when a journal with no other way of counting publishes hundreds of articles each year and maintains page numbering throughout a dozen issues. I might count the articles in one or two issues, determine an average article length, and estimate the year's total count based on that length. I also used counts from *DOAJ* in many cases, when those counts were plausible based on manual sampling.
- **Errors**: I'm certain that my counts are off by one or two in some cases; that happens.
- **Late additions**: Some journals, especially those that are issue-oriented and still include print versions, post online articles *very* late. Even though I'm retesting all cases where the "final issue" of 2015 seemed to be missing when checked in January-March 2016, it's nearly certain that somebody looking at some journals in, say, August 2016 will find more 2015 articles than I did.

In practice, I doubt that any two counts of a thousand or more OA journals will yield precisely the same totals. I'd guess that I'm very slightly overcounting articles in some journals that provide convenient annual totals—and undercounting articles in some journals that don't.

For the analysis I'm doing, and for any analysis others are likely to do, these "errors" shouldn't matter. If somebody claimed that overall numbers were 5% lower or 5% higher, my response would be that this is quite possible. I doubt the differences in counts would be greater than that, at least for any aggregated data.

Making the case

If you believe I'm wrong—that there are real, serious, worthwhile research cases where only the unanonymized version will do—let me know (waltcrawford@gmail.com).

Obviously, anonymized datasets aren't unusual; I don't know of any open science advocate who would seriously argue that medical data should be posted with patient names or that libraries should keep enough data to be able to do analysis such as "people who borrowed X also borrowed Y." In practice, there may be special use cases for an open copy of the master spreadsheet. On the other hand, except for the list of journals flagged as having malware on their sites, I'll be doing my analysis with the anonymized spreadsheet—it's what's needed for this work, and won't distract me with individual journal titles and how I might feel about their publishers.

Second Checkpoint

If this essay has an unusually stream-of-consciousness feel to it, there's a reason: as many of you will have recognized, all of it up to here has appeared as a series of posts on *Walt at Random* during the course of the first pass—including the checkpoint of sorts when I was halfway through.

The first pass is now complete, and that seems like an appropriate time for a second checkpoint, with the caveat that all of these figures are preliminary: there will *certainly* be significant changes during the second pass. Otherwise, there wouldn't be a second pass!

Pay, Free and Unknown

The trend toward fee-charging journals in the first half did *not* show up to such an extent in the second half. Right now, and including some journals that may be excluded from final analysis (C and X codes), there are around 7,450 apparently-free journals, around 3,050 APC-charging journals, and around 140 where I *believe* there's a fee but either can't verify it or the amount is not clearly stated.

Thus, the free-to-pay percentage is around 71% free: lower than in 2014, but not massively so. (Journals that stay in the third group will be excluded.)

As far as I can tell, about 400 journals now impose APCs that formerly didn't, around 60 have eliminated APCs, and 34 have moved to the unknown category; further research may eliminate most or all of these. A few journals that couldn't be analyzed fully in 2014 now show up as either free or pay, mostly free. Of journals that are new to *DOAJ*, 183

have APCs, 679 do not—and four are currently in the unknown category.

Journals by Code

Code? What's that? It's a simpler way of referring to the combination of grade and subgrade. At this point, around 8,540 journals have the only one-letter code, "A," which is really a non-code: a journal with nothing to raise cautions or red flags, that published at least five articles in 2015 and meets all expected standards.

B codes—included in the full analysis but requiring special notice—currently include around 270 BC (cancelled or with no articles later than 2012); around 1,170 BF (fewer than five articles in 2015—the group most likely to shrink during the second pass); around 60 BR (primarily conference papers); and around 20 BS (requiring free registration to read papers).

C codes—journals that I believe should be avoided—include around 130 CA (APC not stated). The other codes have fewer than six journals each and only ten total; those codes may disappear in the final analysis.

X codes—journals excluded from analysis for specific reasons—are mostly cases that I hope will shrink during the second pass, although that's unlikely for the first group. XE (empty, no articles later than 2010) accounts for some 40 journals. XM (journals that are flagged as malware by Malwarebytes, McAfee Site Advisor or MS Office) accounts for some 130, a disturbing number and one I hope will go down on recheck; it's twice as high as last year, an unhealthy trend. XN (journals that don't appear to be properly OA) is down *enormously* from 177 to fewer than 60. XO (journals where I was either unable or unwilling to do article counts by year) is already down from 149 to fewer than 50, even though more journals are involved, and I anticipate reducing that number further. XT (journals where I couldn't get a good enough translation to do the analysis) is also down, from 20 to seven, and that number might go down a bit further. (In any case, having only seven inadequate cases out of nearly 11,000 journals is a testament to Google/Chrome's translation software.) XU (unworkable sites) is down slightly, from 95 to fewer than 70—but, unfortunately, XX (unreachable sites) is up a *lot*, from 195 to more than 320. (The related XP, parking or ad pages, is up from 30 to around 45.) I'm hopeful that XX will come down in the second pass; all C and X codes are rechecked in the second pass, as are all BF journals.

Sometimes journal situations do change for the better, such as a dozen or more cases of malware that seem to have been cleared up and nearly 170 sites that were either unreachable or unworkable and are now OK.

Second Pass and Article Estimates

At this point, the second pass involves around 2,650 journals—all journals coded BF, X (other than XE and XN) or CA, and other journals where I used approximate counting methods that I may be able to improve. Additionally, if time permits, I plan to recheck around 440 of the larger biomed journals (those with at least 100 articles in 2015) in PubMed Central, which may refine some counts; it's possible that recheck will extend a little further down.

As for article estimates, I can say with some confidence that the final count will be well above 500,000 articles for 2015 and probably less than 600,000 articles.

Beyond that, it gets very fuzzy; that's partly what the second pass is for. The journals I do *not* plan to recheck published around 388,000 articles in 2014, growing to around 398,000 articles in 2015 (noting that the 2015 total includes more journals). The journals I *do* plan to recheck—excluding possible biomed rechecks—published around 65,000 articles in 2014, but only about 33,000 in 2015. But that group includes journals that will be excluded or that were reachable before but aren't now; some 15,000 articles in 2014 are down to around 3,000, but that's because the articles can't be counted (unless they show up in *DOAJ*) since the journals were either unreachable or flagged as malware this time around. The other huge drop is in code BF, and that's where I'd expect to see the biggest change, as many smaller journals or journals that still issue print editions add 2015 articles in 2016. Right now, the totals went from nearly 14,000 in 2014 to not much more than 1,000 in 2015; that should change substantially. Finally, the 438 largish biomed journals show reasonable increases, going from around 119,000 articles in 2014 to around 127,000 in 2015: either or both of those numbers may change in either direction.

One thing that's clear, as I found myself rechecking 2014 numbers for quite a few journals: a fair number of journals have substantial delays in posting articles. I'm allowing for a three-month delay—but I have no doubt that rechecking all the journals in late 2016 would yield even more 2015 articles than I'll find in April 2016. Probably not tens of thousands more, but almost certainly thousands more.

That's where it stands. If somebody asked "What's your best guesstimate for total serious gold

OA publishing in 2015?" I'd say "somewhere between 550,000 and 590,000 articles" and resist efforts to narrow that guesstimate further.

Side note: during the first pass and some cleanup work, I've found four more duplicates—same URL, different titles—bringing the total down to 10,944.

The Front (also) Readership Notes

A little delayed, but here it is: some notes on *Cites & Insights* readership for 2015. Overall readership is down quite a bit from 2014, but still good enough: just under 126,000 PDF downloads (a figure that's probably about a 3% undercount).

Volume 15

Here's the actual table, in descending order, followed by some notes. Note that I combine one-column and two-column versions for these counts; in general, the print-oriented version predominates (two or three to one) *unless* I specifically point people to the single-column version.

Issue	Count
civ15i9.pdf	3,135
civ15i1.pdf	2,889
civ15i4.pdf	2,701
civ15i3.pdf	2,438
civ15i6.pdf	1,612
civ15i5.pdf	1,410
civ15i2.pdf	1,159
civ15i10.pdf	1,069
civ15i11.pdf	978
civ15i7.pdf	968
civ15i8.pdf	511

Figure 1. 2015 issues

I consider any first-year total of 700 or more to be fully successful, and I won't argue with 300 or more at this point. Given that, there's only one issue that wasn't a full success: the fairly bizarre August/September issue—pretty clearly a holding pattern while I was working on *The Gold OA Landscape 2011-2014*—with its different filler for the two versions. Frankly, I'm pleasantly surprised that the issue(s) was (were) downloaded more than 500 times so far.

At the other extreme, with more than three thousand downloads, was the large excerpt from *The Gold OA Landscape 2011-2014*. Also no surprise, and I'd be delighted if one out of a hundred readers actually purchased a copy. (There have been more copies sold, but it's still in the low double digits.)

The other issues with more than two thousand downloads so far have the following primary essays: The "Third Half" (more of *DOAJ* journals); the economics of open access; and ebooks and pbooks (and one last chunk of *DOAJ*).

The four with more than 1,000 and fewer than 2,000 downloads feature: "who needs open access, anyway?"; essays on Twitter and the LSW diaspora; Deathwatch 2014! and copyright extremism; and a fair use trilogy—Google Books, HathiTrust and more.

So: not quite "all OA all the time" but the most-read issues do follow a trend.

Overall 2015

Which other issues did especially well in 2015?

Issue	2015
civ14i4.pdf	4,828
civ10i8.pdf	2,832
civ14i7.pdf	2,393
civ8i7.pdf	1,890
civ12i2.pdf	1,585
civ13i6.pdf	1,518
civ13i8.pdf	1,435
civ10i9.pdf	1,407
civ12i8.pdf	1,259
civ9i5.pdf	1,251
civ6i10.pdf	1,246
civ6i2.pdf	1,212

Figure 2. Most-read non-2015 issues in 2015

Figure 2 shows the dozen earlier editions that had at least 1,200 downloads in 2015. The most astonishing, of course, is v14i4: "forecasts and futurism." Oh, right, it also had "The Sad Case of Jeffrey Beall," which may have had something to do with it. Possibly.

Two others in excess of 2,000 feature: an odd mix of essays, the longest on Facebook; and journals, "journals" and wannabes (there's Beall again).

The trio between 1,500 and 1,999 include the July 2008 issue, with no especially distinctive essay (but maybe it's "One, Two, Some, Many: Search Results & Meaning"); another one where the only distinctive essay is on social networks; and a long single-topic essay, "Hot Times for Open Access."

The next six: another social network essay (although I'm fond of "Differences" as well); a five-essay issue, the longest being on social media and social networks; an essay on blogging; one almost evenly split between blogging, writing about reading, and

open access; "looking at liblogs: the great middle; and the all-time champion (which hasn't actually been directly available for a while) on Library 2.0. I'm amazed that some ten-year-old stuff is still getting so many downloads. And pleased, to be sure.

Overall numbers? Since these statistics only go back to October 2013, I'm not sure how meaningful these are, but here are the top eleven—all issues downloaded more than 4,000 times between October 2013 and December 2015. If there are issues here you don't recognize, you should know where to go.

Issue	Total
civ14i4.pdf	11,266
civ6i10.pdf	7,409
civ10i8.pdf	7,196
civ8i7.pdf	6,206
civ9i11.pdf	5,757
civ12i2.pdf	5,672
civ14i7.pdf	5,486
civ12i8.pdf	4,870
civ11i1.pdf	4,314
civ14i5.pdf	4,282
civ9i5.pdf	4,232

Figure 3. Most-downloaded Issues since October 2013

No, *Cites & Insights* will not become an all-OA journal, even though OA is dominating my time for the first half of the year. Who knows? I might even write about blogging again one of these days.

Intersections

"Trust Me": The Other Problem with Beall's Lists

Here's the *real* tl;dr: I could only find any discussion at all in Beall's blog for 230 of the 1,834 journals and publishers in his 2016 lists—and those cases don't include even 2% of the journals in *DOAJ*.

Now for the shorter version…

As long-time readers will know, I don't much like blacklists. I admit to that ~~prejudice~~ belief: I don't think blacklists are good ways to solve problems.

And yet, when I first took a hard look at Jeffrey Beall's lists in 2014, I was mostly assessing whether the lists represented as massive a problem as Beall seemed to assert. As you may know, I concluded that they did not.

But there's a deeper problem—one I believe applies whether you dislike blacklists or mourn the passing of the ***Index Librorum Prohibitorum.*** To wit, Beall's lists don't meet what I would regard as minimal standards for a blacklist *even if you agree with all of his judgments*.

Why not? Because, in seven cases out of eight (on the 2016 lists), Beall provides no case whatsoever in his blog: the journal or publisher is in the lists Just Because. (Or, in some but not most cases, Beall provided a case on his earlier blog but failed to copy those posts.)

Seven cases out of eight: 87.5%. 1,604 journals and publishers of the 1,834 (excluding duplicates) on the 2016 versions have no more than an unstated "Trust me" as the reason for avoiding them.

I believe that's inexcusable, and makes the strongest possible case that nobody should treat Beall's lists as being significant. (It also, of course, means that research based on the assumption that the lists are meaningful is fatally flawed.)

The Short Version

Since these key numbers will appear first as a blog post on *Walt at Random* and much later in *Cites & Insights*, I'll lead with the short version.

I converted the two lists into an Excel spreadsheet (trivially easy to do), adding columns for "Type" (Pub or Jrn), Case (no, weak, maybe or strong), Beall (URL for Beall's commentary on this journal or publisher—the most recent or strongest when there's more than one), and—after completing the hard work—six additional columns. We'll get to those.

Then I went through Beall's blog, month by month, post by post. Whenever a post mentioned one or more publishers or independent journals, I pasted the post's URL into the "Beall" column for the appropriate row, read the post carefully, and filled in the "Case" column based on the most generous reading I could make of Beall's discussion. (More on this later in the full article, maybe.)

I did that for all four years, 2012 through 2015, and even January 2016.

The results? In 1,604 cases, I was unable to find any discussion whatsoever. (No, I didn't read all of the comments on the posts. Surely if you're going to condemn a publisher or journal, you would at least mention your reasons in the body of a post, right?)

If you discard those on the basis that it's grotesquely unfair to blacklist a journal or publisher without giving *any* reason why, you're left with a list of 53 journals and 177 publishers. Giving Beall the benefit of the doubt, I judged that he made no case at

all in five cases (the fact that you think a publisher has a "funny name" is *no case at all*, for example). I think he made a very weak case (e.g., one questionable article in one journal from a multijournal publisher) in 69 cases. I came down on the side of "maybe" 43 times and "strong" 113 times, although it's important to note that "strong" means that at some point for some journal there were significant issues raised, not that a publisher is forever doomed to be garbage.

Call it 156 reasonable cases—now we're down to less than 10% of the lists.

Then I looked at the spreadsheets I'm working on for the 2015 project (note here that SPARC has nothing at all to do with this little essay!)—"spreadsheets" because I did this when I was about 35% of the way through the first-pass data gathering. I could certainly identify which publishers had journals in *DOAJ*, but could only provide article counts for those in the first 35% or so. (In the end, I just looked up the 53 journals directly in *DOAJ*.)

Here's what I found.

- Ignoring the strength of case, Beall's lists include 209 *DOAJ* journals—or 1.9% of the total. But of those 209, 85 are from Bentham Open (which, in my opinion, has cleaned up its act considerably) and 49 are from Frontiers Media (which Beall never actually made a case to include in his list, but somehow it's there). If you eliminate those, you're down to 75 journals, or 0.7%: Less than one out of every hundred *DOAJ* journals.
- For that matter, if you limit the results to strong and maybe cases, the number drops to 37 journals: 0.33%, roughly one in every three hundred *DOAJ* journals.
- For journals I've already analyzed (since I'm working by publisher name, that includes most of these—at this writing, January 29, I just finished Hindawi), total articles were just over 16,000 (with more to come on a second pass) in 2015, just under 14,000 in 2014, just over 10,000 in 2013, around 8,500 in 2012, and around 4,500 in 2011.
- But most of those articles are from Frontiers Media. Eliminating them and Bentham brings article counts down to the 1,700-2,500 range. That's considerably less than one half of one percent of total serious OA articles.
- The most realistic counts—those where Beall made more than a weak case—show around 150 articles for 2015, around 200-250 for 2013 and 2014, around 1,000 for 2012 and around 780 for 2011 (Those numbers will go up, but probably not by much. There was one active journal that's mostly fallen by the wayside since 2012.)

The conclusion to this too-long short version: Beall's lists are mostly the worst possible kind of blacklist: one where there's no stated reason for things to be included. If you're comfortable using "trust me" as the basis for a tool, that's your business. My comment might echo those of Joseph Welch, but that would be mean.

Oh, by the way: you can download the trimmed version of Beall's lists (with partial article counts for journals in *DOAJ*, admittedly lacking some of them). It's available in .csv form for minimum size and maximum flexibility. Don't use it as a blacklist, though: it's still far too inclusive, as far as I'm considered.

Added Notes

When I wrote the preceding as a blog post, I anticipated fleshing out these quick notes to a considerably longer article. I no longer see much point in doing that. There are more useful things to write about, and of course there are still thousands more *DOAJ*-listed journals to examine.

I do wonder about some of the things Jeffrey Beall sees as signs of ppppredatory journals and publishers, however:

- In the 21st century, where an online-only journal that only accepts electronic submissions is going to do *everything* over the internet, what's wrong with having a house, an apartment or even a mail drop as the editorial headquarters for a journal or publisher? The editorial employees are probably working from home (and quite possibly part-time). What purpose is served by having a Proper Building? I conclude that looking at Google Maps for a postal address has no evidentiary value for establishing the questionablity of a journal or publisher.
- I'm not sure about the fascination with "probably a made-up name," unless one assumes a certain amount of ethnocentrism, since the "probably made-up names" are typically very much Western.
- In at least a couple of instances where a journal is found at fault for publishing "junk science" in the form of articles linking glyphosate to cancer—well, maybe Jeffrey Beall knows more about these matters than the World Health Organization, but I'm not convinced.
- Then there are so-called "junk" fields—and here, I'll use Elsevier as the bar. I suspect

there's at least as much scientific evidence for the efficacy of ayurvedic treatments as there is for homeopathy, even if either or both might primarily be sophisticated ways of making placebos more effective. (Actually, I'm surprised there's *not* an *International Placebo Journal*, since there should be a lot of room to make highly-effective non-medicines even more effective. Of course, it may be tough to get grants to find better ways to non-medicate…)

- Lately, Beall loves to toss out admonitions about not needing new journals in field X or field Y—and, you know, as soon as Elsevier stops introducing new journals at a crisp pace (75 in 2015!), maybe I could take that seriously. Personally, I'm not aware of the magic Correct Number of Journals for any given field—but I am aware that, for non-profit Gold OA to ever chip away at the strongholds of the major publishers, there need to be enough good Gold OA journals to serve as quality outlets.
- Let's not get into the whole ppppredatory thing. Who's the predator? A billion-dollar corporation driving up subscription prices and buying up smaller competitors, or some small-time operators serving the needs of scholars whose needs tend to be overlooked by "high-impact" Western journals?
- I must admit to find reading through four years of Beall's opinions and findings wearisome; fortunately, I didn't do it all at once. I used spare time after I'd done the maximum number of *DOAJ* journal checks I allow myself for any given day—and finished it when I hit a major milestone (4,000 journals) somewhat ahead of time.

Finally, lest there be any doubt, I am *not* saying there are no questionable publishers and journals out there. It depends partly on who's asking the question, of course, but I can think of one big publisher (pardon me while I meditate for a bit and perhaps drop the first letter of a traditional alternative name for graphic novels) that seems pretty well established as sketchy—one that has no journals in *DOAJ* at all, as far as I can tell.

But I'm also seeing publishers that have cleaned up their acts; publishers that have had one sketchy journal and taken care of the problem; and publishers and platforms providing yeoman service to making research outlets available to the global South and to making research work available to all. I find that much more interesting and worth investigating.

Speaking of Which…

If you haven't already done so, do read THE FRONT in this issue, with its announcement of (and some details about) the SPARC project I'm spending most of my time on these days.

The Back

I'll admit this issue is a little light on new material for those who read *Walt at Random* (although the last section of the first essay regularly, "Second Checkpoint," only appears here). You can probably guess where most of my time and attention are going. The truly cynical among you might even suggest that THE BACK only appears here to fill out an even number of pages, before appending the 35 pages of *The Gold OA Landscape 2011-2015*.

The truly cynical among you might even be right. With that, here are a few random chunks of snark. The first is a true gem…

Homeopathy successfully turns water into a placebo

That's the title of this February 20, 2016 piece by Beth Mole at *ars technica*. The tease: "Shocking analysis finds water is not medicine—and doesn't have a memory."

The story recounts "a thorough evaluation of 57 scientific reviews that encompassed 176 studies on 68 illnesses," and the result's a little more nuanced: homeopathy is *at best* a placebo, when it's not potentially harmful.

This shouldn't be a big surprise: after all, homeopathic "remedies" typically dilute the active agent so much that there are unlikely to be *any* molecules of it left in the final product.

There's more to the piece (reporting on an Australian study), and it's worth reading.

Let me check: Yep, Elsevier still publishes *Homeopathy Journal*; it even has an Impact Factor. (And two other impact/rank numbers, which apparently are OK to publicize as long as your're a Big Traditional Publisher, although one sign of being pppredatory if you're OA scum.)

The Clown Show at VW

Fortune is one magazine that retains the fine old tradition of putting lighthearted stuff on the last editorial page—that is, on The Back. In *Fortune*'s case, it's

Stanley Bing's column; this one's in the November 1, 2015 edition—and yes, you can read it online.

> So you ask yourself, What in the world were those guys at Volkswagen allegedly thinking? I'm using the word "allegedly" here not about what they allegedly did but about their thinking, which was most certainly alleged.

He offers several possibilities, beginning with "They're a bunch of clowns" and moving on from there. Some of Bing's more pointed suggestions (each with a following paragraph): They thought everybody else was stupid; Maybe they thought they could get away with it forever?; and…to be sure…And they figured they would be able to handle it if the thing blew up. His final suggestion: "So they're crooks, right?"

Of course, if there still is a VW at the end of all this that bears any relationship to the firm that spent so much time and money helping to pollute the air, then I'd have to say the clowns at VW will have verified that they *can* handle the thing blowing up: maybe a few billion dollars poorer, but hey, that's business, right?

Anyway: Bing's always funnier than I am…but this is perhaps the least humorous column he's written. For good reason.

My Favorite Whatever?

Let's finish this odd little section with a really nice piece by John Scalzi at *Whatever* on August 19, 2013: "These Are a Few Of My Favorite Things, On Tiers." I'm not making fun of it; I'm recommending it, but it seems to fit here better than anywhere else. It's a timeless piece, fortunately.

> **I get asked a lot** what about what my favorite book/movie/album/creatively generated object might be (or my favorite author/filmmaker/musician/creative type), and I find as I go along in life I get progressively more annoyed with the question. This is usually not the fault of the person asking the question, who is generally trying to make innocuous conversation and is doing so by opening up a socially-approved line of trivial conversation.
>
> It is, however, the fault of the question itself, which is unsophisticated, naive and annoying. Like most people over the age of twelve, who both had time to expand their creative palates and who recognize that life is not always a zero-sum *Highlander*-like experience, in which There Can Be Only One, I don't have a single favorite book, or movie, or album, etc, or a favorite author or filmmaker or musician, or so on. I like a lot of different things (and artists) almost equally for reasons that are often *not* equivalent or comparable.

That's the setup. He then describes his six "tiers" for creative works and for artists. It's an interesting approach, and his annoyance with the question itself is one I share. I'll quote the first and sixth tier from each category; you should read the intervening ones and the rest of the piece yourself.

For creative works:

> **First tier:** The works of art that, for lack of a better term, regenerate me: I take them in and they make me feel like a better person for having gone through them.
>
> **Sixth tier:** Deeply annoyed some of the precious few moments in which I exist as a conscious being in this universe have been wasted on this crap.

And for the artists:

> **First tier:** I consider these folks as my personal artistic pantheon.
>
> **Sixth tier:** Abstractly okay with the concept that these people are allowed to express themselves in a manner that looks like creativity if you don't think about it too hard, but honestly, what the hell.

As he notes, even that oversimplifies. Read the essay. And, of course, read the comments.

Masthead

Cites & Insights: Crawford at Large, Volume 16, Number 3, Whole # 193, 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/civ16i3.pdf

5. Starting Dates

In earlier reports, I suggested that there was a gold rush of APC-charging journals between 2006 and 2010, based on the rapid increase in such journals during that period. This report includes many more journals and quite a few journals that began publishing in 2014 but weren't included in *DOAJ* by May 2014. That may change the situation.

The tables and figures in this chapter are *not* directly comparable to Tables 4.1-4.5 in *Open-Access Journals: Idealism and Opportunism* because, in order to accommodate a significant number of 2014 journals, I used two-year periods beginning in odd-numbered years rather than even-numbered years. I've also collapsed early journals somewhat, into two buckets rather than four: up to and including 1980, and 1981-1990, rather than the earlier pre-1960, 1960-69, 1970-79, and 1980-89.

There's one more caveat, which may also apply to the earlier report and help explain the sheer rapidity of recent growth in OA journals: starting dates in *DOAJ* are as reported by journals, and it's clear that in some cases the dates represent either the start of OA availability (in journals that previously didn't have OA) or the start of OA through that publisher. In a few cases, I modified the date based on actual article appearances, but never back earlier than 2011.

As you'll see in the overall tables and graph, the apparent gold rush is much less evident when all journals are included: newer non-English journals appear much less likely to charge APCs. While there was certainly a big jump in APC-charging journals in 2009 through 2012 (dropping sharply in 2013-14), with 1,402 APC-charging journals in those four years (compared to 373 in the previous four years), there was also a fairly big (but not quite as marked) jump in free OA journals: 2,427 in those four years, compared to 1,680 in the previous four years.

Years	Journals	Free%
Up to 1980	175	82.3%
1981-90	166	86.1%
1991-92	65	86.2%
1993-94	103	90.3%
1995-96	210	90.5%
1997-98	303	90.4%
1999-2000	481	91.5%
2001-02	601	90.3%
2003-04	677	87.3%
2005-06	901	85.6%
2007-08	1,152	78.9%
2009-10	1,883	64.3%
2011-12	1,846	60.5%
2013-14	949	58.8%

Table 5.1. Starting dates for OA journals

Ignoring the first two rows in Table 5.1 for now, note the steady increase in new journal introductions in each two-year period, which speeds up in 2005-06 but jumps much more rapidly in 2009-10, then declining almost imperceptibly in 2011-12 and dropping by nearly half in 2013-2014. Note also the percentages: the percentage of new journals that do *not* charge APCs dropped below 78% for the first time in 2009-2010, and has continued to drop since then, although the drop in 2013-2014 is minimal.

Figure 5.1 shows free and pay journals by starting date. Unlike the similar Figure 4.1 in the earlier report, where the two lines almost touch in 2008-09 (only 51% of new OA journals were free in that period), there's still a healthy gap throughout the last decade.

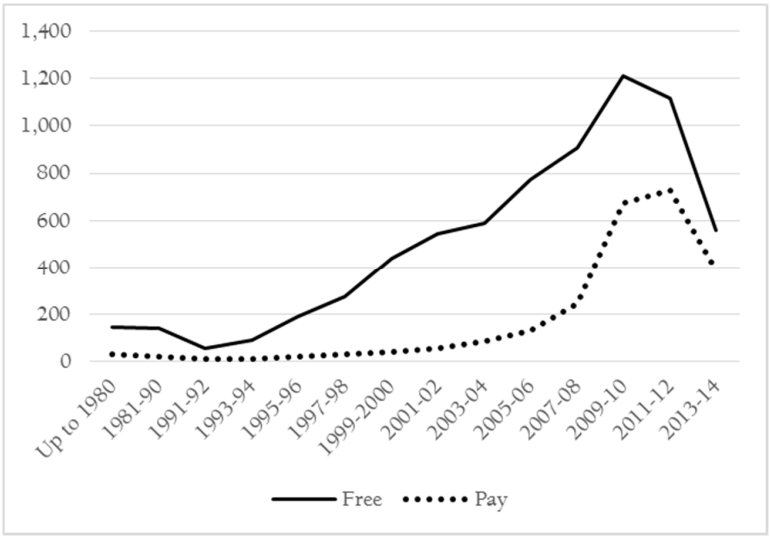

Figure 5.1. OA journals by starting date, free and pay

Years	Journals	Articles	Art/J
Up to 1980	167	10,478	63
1981-90	153	7,235	47
1991-92	64	2,266	35
1993-94	94	4,943	53
1995-96	198	12,721	64
1997-98	278	18,996	68
1999-2000	449	20,469	46
2001-02	537	29,403	55
2003-04	614	60,445	98
2005-06	816	39,539	48
2007-08	1,045	47,935	46
2009-10	1,737	111,035	64
2011-12	1,696	86,115	51
2013-14	912	30,781	34

Table 5.2. OA journals that published articles in 2014, by starting date

Table 5.2 shows those OA journals that actually published articles in 2014, the number of 2014 articles in those journals, and the average articles per journal. Other than a couple of low spots (and, of course, 2014 journals may have begun in *late* 2014 with very few articles), the only standout here has an easy explanation: 2003-04 was the start of *PLOS One*.

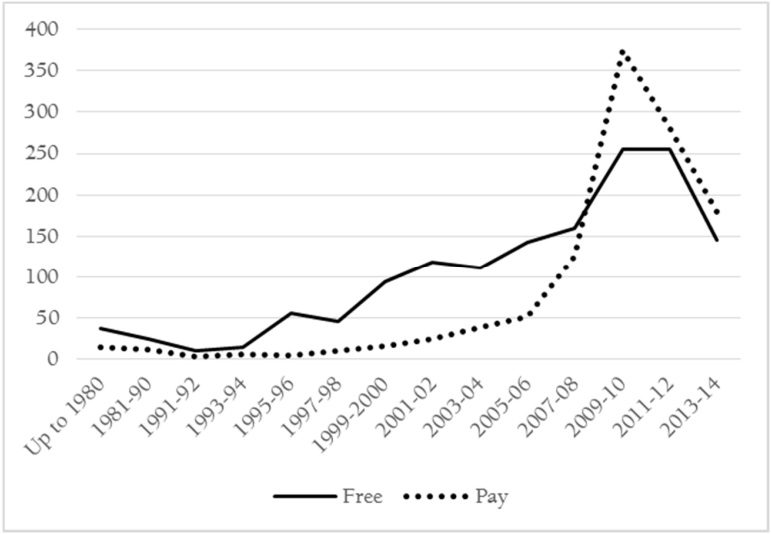

Figure 5.2. Biomed OA journals, free and pay, by starting date

If there has been a gold rush, it's in biology and medicine, as shown in Figure 5.2, showing free and pay journals by starting dates—and, after all, that's clearly where the gold is. Note that, after running well below the free line until 2003-04, APC-charging journals charge ahead to the point where they're well ahead of free journals in 2009-10 (375 new APC-charging journals, 255 free ones—and if the free line looks flat for 2011-12, that's because it is: another 255).

Figure 5.3 shows comparable information for science, technology, engineering and mathematics journals. In this case, APC-charging journals started to surge in 2009-10 and outpaced free journals in 2011-12, but only slightly (327 to 290—and, yes, the number of new free journals in 2009-10 is identical to the number of APC-charging journals in 2011-12).

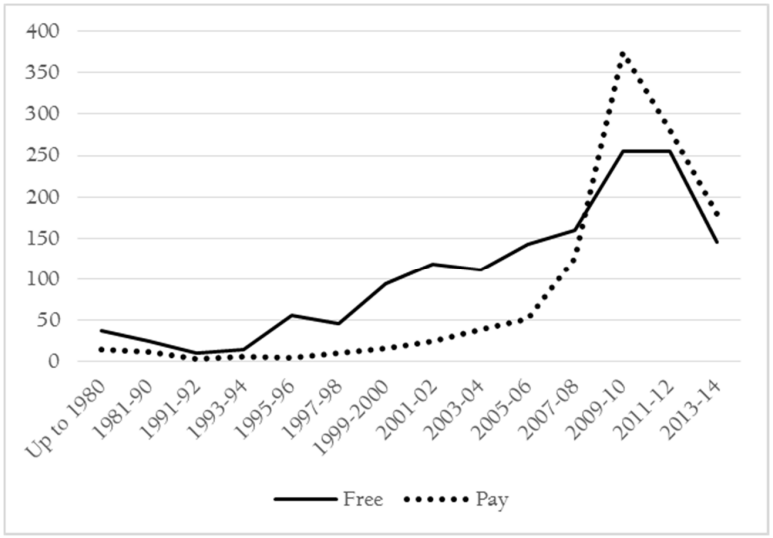

Figure 5.3. STEM OA journals, free and pay by starting date

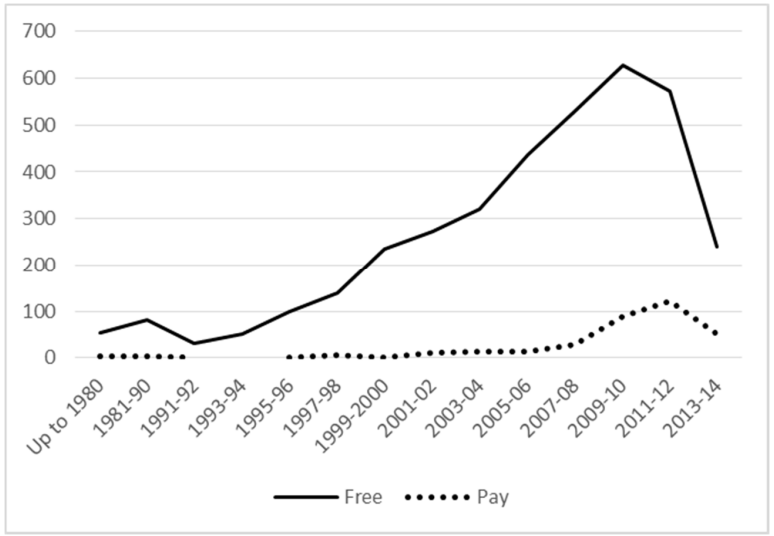

Figure 5.4. HSS OA journals, free and pay by starting date

Finally, Figure 5.4 shows new OA journals in the humanities and social sciences by starting date—and if you have access to the earlier

report, the comparison with Figure 4.5 in that report may be interesting. Although the shape of the lines is vaguely similar (except that free journals have a single spike in 2010-11 in the earlier report, where it's spread out from 2007 through 2012 in Figure 5.4), the actual numbers are much higher: the top line of the graph was 400 for the earlier report, while it's 700 this time around, with 529, 628, and 572 new free journals respectively in 2007-08, 2009-10, and 2011-12. It's fairly clear that there's very strong growth in non-English free humanities and social science journals over the past few years.

6. Country of Publication

Where do open access journals come from? In most cases, *DOAJ* includes the country of publication as stated by the publisher. In all, 121 nations and territories have at least one Gold OA journal that appears to be a good publication (that is, is graded A or B). Here's the full list, in descending order by total number of A and B OA journals, with the percentage of journals that are free.

Country	**Journals**	**Free %**
United States	996	62.3%
Brazil	929	94.4%
United Kingdom	649	28.8%
Spain	517	98.1%
Egypt	493	46.2%
India	438	46.3%
Germany	315	67.9%
Romania	285	88.8%
Italy	277	90.6%
Iran, Islamic Republic of	269	90.0%
Turkey	260	88.8%
Poland	258	91.1%
Canada	254	63.8%
Colombia	242	98.8%
Switzerland	216	43.1%
France	166	96.4%

Argentina	146	91.8%
Mexico	146	95.2%
Chile	138	92.8%
Indonesia	136	61.8%
New Zealand	115	21.7%
Australia	111	87.4%
Russian Federation	100	93.0%
Netherlands	99	53.5%
Croatia	97	94.8%
Serbia	95	96.8%
Japan	94	68.1%
Czech Republic	84	81.0%
Portugal	76	89.5%
Pakistan	75	56.0%
South Africa	70	64.3%
Malaysia	67	76.1%
Sweden	63	58.7%
Ukraine	62	91.9%
Cuba	61	100.0%
Venezuela, Bolivarian Republic of	47	100.0%
Slovenia	45	97.8%
Peru	44	95.5%
South Korea	42	33.3%
Denmark	41	100.0%
China	40	67.5%
Greece	40	87.5%
Norway	40	97.5%
Austria	38	94.7%
Costa Rica	37	100.0%
Finland	36	69.4%

Hong Kong	36	55.6%
Slovakia	33	93.9%
Lithuania	32	93.8%
Hungary	31	93.5%
Singapore	31	41.9%
Bangladesh	30	66.7%
Nigeria	30	16.7%
Bulgaria	29	65.5%
Taiwan, Province of China	25	72.0%
Belgium	24	95.8%
Estonia	23	100.0%
Nepal	17	82.4%
Bosnia and Herzegovina	15	86.7%
Macedonia, the Former Yugoslav Republic of	15	86.7%
Ireland	13	92.3%
Israel	13	84.6%
Thailand	13	84.6%
United Arab Emirates	13	23.1%
Philippines	12	100.0%
Sri Lanka	11	100.0%
Ecuador	10	100.0%
Uruguay	9	100.0%
Bolivia, Plurinational State of	7	100.0%
Iraq	7	71.4%
Jordan	7	71.4%
Qatar	7	71.4%
Kenya	6	66.7%
Latvia	6	100.0%
Moldova, Republic of	6	100.0%
Morocco	6	83.3%

Algeria	5	100.0%
Cyprus	5	80.0%
Ethiopia	5	100.0%
Malta	5	100.0%
Montenegro	5	100.0%
Saudi Arabia	5	60.0%
Georgia	4	100.0%
Iceland	4	100.0%
Armenia	3	100.0%
Guatemala	3	100.0%
Kosova	3	0.0%
Kuwait	3	100.0%
Nicaragua	3	100.0%
Uganda	3	66.7%
Albania	2	50.0%
Azerbaijan	2	100.0%
Belarus	2	100.0%
Bhutan	2	100.0%
Kyrgyzstan	2	100.0%
Libya	2	50.0%
Mauritius	2	50.0%
Oman	2	100.0%
Paraguay	2	100.0%
Puerto Rico	2	100.0%
Tanzania, United Republic of	2	100.0%
Yemen	2	50.0%
Zambia	2	50.0%
Bahamas	1	100.0%
Barbados	1	100.0%
British Virgin Islands	1	100.0%

Brunei Darussalam	1	100.0%
Burundi	1	100.0%
Democratic Republic of the Congo	1	100.0%
Dominican Republic	1	100.0%
Ghana	1	0.0%
Guam	1	100.0%
Jamaica	1	100.0%
Kazakhstan	1	0.0%
Korea, Republic of	1	100.0%
Luxembourg	1	100.0%
Madagascar	1	100.0%
Martinique	1	100.0%
Palestine, State of	1	0.0%
Rwanda	1	100.0%
Tunisia	1	100.0%

Table 6.1. OA journals by country of publication

It's hardly surprising that the U.S. has more OA journals than any other nation; some of the other high-ranking nations may (or may not) be more surprising.

That's a very long list; additional tables will still be long, but limited to countries publishing at least ten OA journals graded A or B.

First, consider the countries most involved in truly free OA. Table 6.2 shows countries that publish ten or more gold OA journals, ranked in descending order by the percentage of free journals (and by the number of free journals in case of ties). It's a *very* different list.

Country	Free	Pay	Free %
Cuba	61		100.0%
Venezuela, Bolivarian Republic of	47		100.0%
Denmark	41		100.0%
Costa Rica	37		100.0%
Estonia	23		100.0%

Philippines	12		100.0%
Sri Lanka	11		100.0%
Ecuador	10		100.0%
Colombia	239	3	98.8%
Spain	507	10	98.1%
Slovenia	44	1	97.8%
Norway	39	1	97.5%
Serbia	92	3	96.8%
France	160	6	96.4%
Belgium	23	1	95.8%
Peru	42	2	95.5%
Mexico	139	7	95.2%
Croatia	92	5	94.8%
Austria	36	2	94.7%
Brazil	877	52	94.4%
Slovakia	31	2	93.9%
Lithuania	30	2	93.8%
Hungary	29	2	93.5%
Russian Federation	93	7	93.0%
Chile	128	10	92.8%
Ireland	12	1	92.3%
Ukraine	57	5	91.9%
Argentina	134	12	91.8%
Poland	235	23	91.1%
Italy	251	26	90.6%
Iran, Islamic Republic of	242	27	90.0%
Portugal	68	8	89.5%
Turkey	231	29	88.8%
Romania	253	32	88.8%
Greece	35	5	87.5%

Australia	97	14	87.4%
Bosnia and Herzegovina	13	2	86.7%
Macedonia, the Former Yugoslav Republic of	13	2	86.7%
Israel	11	2	84.6%
Thailand	11	2	84.6%
Nepal	14	3	82.4%
Czech Republic	68	16	81.0%
Malaysia	51	16	76.1%
Taiwan, Province of China	18	7	72.0%
Finland	25	11	69.4%
Japan	64	30	68.1%
Germany	214	101	67.9%
China	27	13	67.5%
Bangladesh	20	10	66.7%
Bulgaria	19	10	65.5%
South Africa	45	25	64.3%
Canada	162	92	63.8%
United States	621	375	62.3%
Indonesia	84	52	61.8%
Sweden	37	26	58.7%
Pakistan	42	33	56.0%
Hong Kong	20	16	55.6%
Netherlands	53	46	53.5%
India	203	235	46.3%
Egypt	228	265	46.2%
Switzerland	93	123	43.1%
Singapore	13	18	41.9%
South Korea	14	28	33.3%
United Kingdom	187	462	28.8%
United Arab Emirates	3	10	23.1%

New Zealand	25	90	21.7%
Nigeria	5	25	16.7%

Table 6.2. Countries with highest percentage of free OA journals

Cuba and Venezuela may not be surprising; Denmark and Costa Rica, somewhat more so. Do note Brazil and Spain—while not the highest *percentage* of free OA journals, these two nations have the largest *number* of free OA journals, considerably more than the U.S.

Articles by Country

Country	**Articles**	**Free%**
United States	89,485	17.3%
United Kingdom	60,838	7.3%
India	42,227	22.3%
Brazil	38,069	88.1%
Switzerland	25,039	9.3%
Egypt	20,000	20.9%
Germany	14,755	34.9%
Iran, Islamic Republic of	12,181	81.6%
Turkey	10,950	87.6%
Spain	10,602	95.8%
Canada	10,431	30.1%
Romania	10,185	74.0%
Poland	8,996	80.7%
Russian Federation	7,328	88.3%
Italy	6,030	82.2%
Colombia	5,847	99.0%
Netherlands	5,532	48.4%
Japan	5,323	54.3%
Indonesia	4,814	37.3%

Ukraine	4,794	75.1%
Chile	4,597	86.9%
South Korea	4,537	8.4%
China	4,417	57.9%
Pakistan	4,218	32.9%
Mexico	4,158	96.2%
Serbia	3,294	93.0%
France	3,122	94.2%
Australia	2,806	54.6%
South Africa	2,757	59.9%
Argentina	2,747	90.0%
Czech Republic	2,733	62.3%
Croatia	2,552	93.9%
Cuba	2,542	100.0%
Hong Kong	2,414	47.4%
Singapore	2,364	19.0%
Finland	2,097	19.0%
Malaysia	1,918	84.4%
Nigeria	1,778	5.2%
Bangladesh	1,723	28.6%
Sweden	1,531	54.1%
Austria	1,466	83.3%
Bulgaria	1,412	47.7%
United Arab Emirates	1,400	10.6%
Portugal	1,389	86.5%
Venezuela, Bolivarian Republic of	1,345	100.0%
New Zealand	1,322	30.0%
Macedonia, the Former Yugoslav Republic of	1,228	21.3%
Uganda	1,227	12.1%
Slovenia	1,164	95.2%

Greece	1,134	81.3%
Morocco	1,110	39.8%
Lithuania	1,012	88.7%
Peru	1,001	94.2%

Table 6.3. OA articles by country of publication

Table 6.3 shows countries with more than 1,000 OA articles in 2014. In some ways, the actual numbers are less interesting than the free percentages, including the very low free percentage in the UK and surprisingly low percentage in India, compared to the very high free percentages in Brazil, Iran, Turkey and Spain.

Biomed Articles by Country

Country	**Articles**	**Free%**
United Kingdom	44,816	4.4%
United States	22,006	22.5%
India	14,850	38.3%
Switzerland	12,392	7.5%
Brazil	11,987	88.5%
Egypt	10,903	15.9%
Iran, Islamic Republic of	8,750	78.7%
Turkey	4,247	98.8%
Netherlands	3,493	53.4%
Japan	3,171	53.2%
Poland	2,665	57.3%
China	2,360	63.5%
Canada	2,289	13.5%
Italy	2,056	59.8%
Spain	1,912	83.6%
Cuba	1,763	100.0%

Germany	1,549	65.2%
Colombia	1,505	100.0%
South Korea	1,407	8.8%
Chile	1,276	63.1%
Uganda	1,227	12.1%
Bangladesh	1,199	33.3%
Pakistan	1,141	83.3%

Table 6.4. Biomed articles by country

Table 6.4 shows the number of 2014 articles in biology and medicine from each country with at least 1,000 such articles—22 of them, as compared to 52 overall. It also shows the percentage of articles that appeared in no-fee journals, astonishingly low for the UK and Switzerland (and, oddly enough, higher for the U.S. than its overall free-article percentage).

STEM Articles by Country

Country	Articles	Free%
United States	25,293	14.5%
India	23,076	12.3%
United Kingdom	12,770	4.4%
Brazil	12,186	76.2%
Germany	11,389	21.3%
Switzerland	10,280	9.4%
Egypt	8,513	25.4%
Poland	4,024	87.5%
Russian Federation	3,675	86.2%
Romania	3,532	78.8%
Indonesia	3,416	28.1%
South Korea	2,613	8.2%
Canada	2,563	40.9%

Pakistan	2,487	10.3%
Turkey	2,464	73.5%
Iran, Islamic Republic of	2,438	84.5%
Ukraine	2,087	96.1%
Japan	1,961	52.4%
China	1,955	49.0%
Czech Republic	1,926	56.9%
Spain	1,875	96.5%
Colombia	1,695	97.8%
Serbia	1,549	86.8%
Chile	1,522	91.4%
Mexico	1,503	94.5%
Finland	1,449	15.0%
Singapore	1,316	26.0%
Netherlands	1,304	14.2%
Hong Kong	1,231	59.3%
Italy	1,184	82.3%
Malaysia	1,160	93.3%
Bulgaria	1,133	45.4%
Morocco	1,082	38.3%
Croatia	1,015	90.6%

Table 6.5. STEM Articles by Country

Table 6.5 shows the number of science, technology, engineering and mathematics articles for each of the 33 countries with more than 1,000 such articles in gold OA journals in 2014, and the usual free percentage.

There may not be a lot more to say here; I do note the UK's astonishing low free-article percentage.

HSS Articles by Country

Country	Articles	Free%
Brazil	13,896	98.3%
United States	10,304	66.6%
Spain	6,815	99.0%
Romania	5,909	70.1%
Canada	5,579	31.9%
India	4,301	21.1%
Turkey	4,239	84.6%
United Kingdom	3,252	58.1%
Italy	2,790	98.7%
Russian Federation	2,693	87.0%
Colombia	2,647	99.1%
Switzerland	2,367	18.4%
Poland	2,307	95.7%
Ukraine	2,306	51.8%
Mexico	2,248	96.6%
France	2,032	99.4%
Germany	1,817	94.1%
Chile	1,799	100.0%
Argentina	1,376	99.6%
Australia	1,376	81.0%
South Africa	1,358	37.8%
Croatia	1,113	98.1%
Singapore	1,048	10.2%

Table 6.6. HSS Articles by Country

Finally, Table 6.6 shows 2014 OA articles in the humanities and social sciences for the 22 nations publishing more than 1,000 such articles—which, oddly enough, is the same number of countries as for medicine, although the numbers and percentages are very different.

Brazil publishes more OA humanities and social sciences articles than any other country—and all but a few appear in no-fee journals. Two-thirds of the U.S. articles also appear in no-fee journals. Note that the UK, first or third in volume of biomed and STEM articles, drops down to eighth for HSS. Meanwhile, Spain—14th in biomed and 21st in STEM—jumps to third place, with considerably more HSS articles than in the other two segments combined.

7. Segments and Subjects

Most of the rest of this report focuses on the three segments and 28 subject, how each of them looks in terms of gold OA and how they differ from the norm.

The three segments are sufficiently distinctive that it makes sense to alter what looks like the norm for individual subjects, specifically the bracketing of journal volume and APC levels.

The first chapter for each segment offers a little information on the segment itself, in addition to segment-specific tables and discussion in earlier chapters, and also notes segment-specific brackets for volume and APC levels. After that comes one chapter for each subject in the segment, somewhat similar to the blog posts at *Walt at Random* and chapters in *The OA Landscape 2011-2014: An Interim Subject View* but including many more journals. How many more? Tables later in this chapter show the increases—noting again that Megajournals in the earlier report has been split between Other Sciences and Miscellany, with *PLOS One* removed altogether, and that Miscellany (with some journals switched to Other Sciences) has been added to the Humanities and Social Sciences segment.

First, five pie charts—a graphic form I usually avoid because it takes up so much space and can only deal effectively with a few figures, but one that seems to make sense in this instance.

Figure 7.1 shows the percentage of gold OA journals (graded A and B) in each segment.

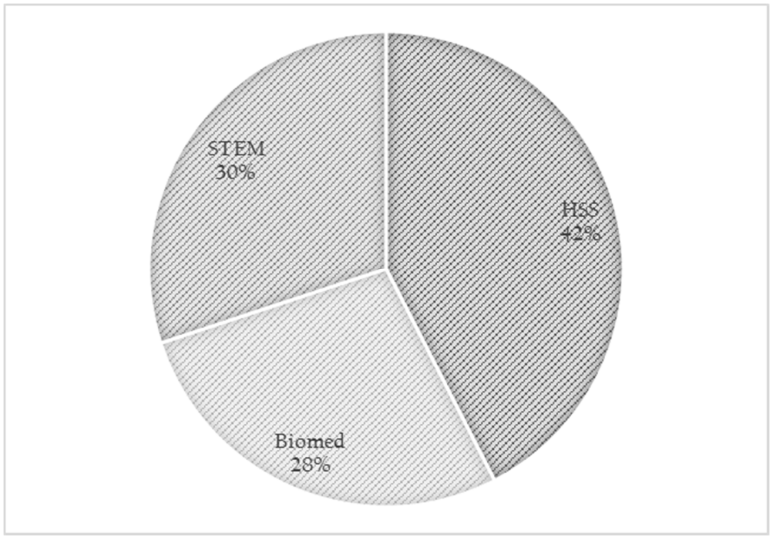

Figure 7.1. Gold OA journals by segment

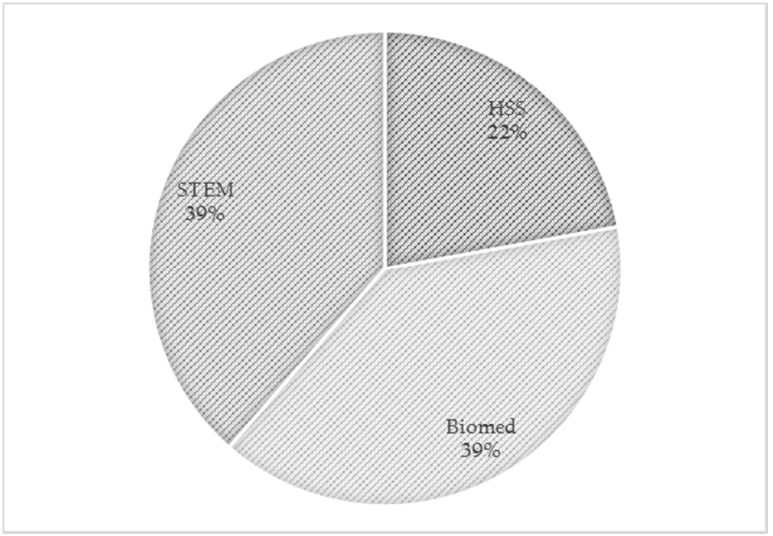

Figure 7.2, 2014 articles in gold OA journals by segment

Figure 7.2 shows the percentage of 2014 gold OA articles in each segment (excluding *PLOS One*, as all of these charts do).

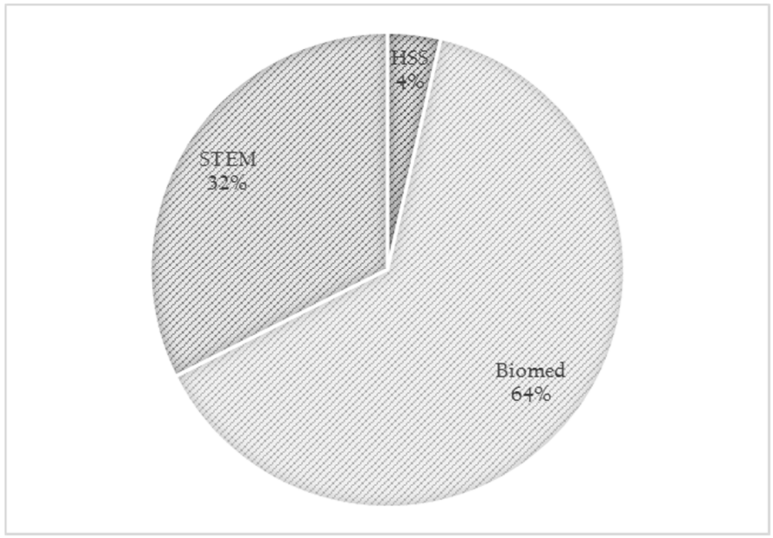

Figure 7.3. Potential 2014 revenue by segment

Figure 7.3 shows the startlingly different picture for total revenue in each segment (again excluding *PLOS One*).

Now, two more pie charts showing 2014 articles by free-vs.-pay percentage—noting that, rounded to the nearest percentage (and excluding *PLOS One*), the percentages for STEM and biomed are identical.

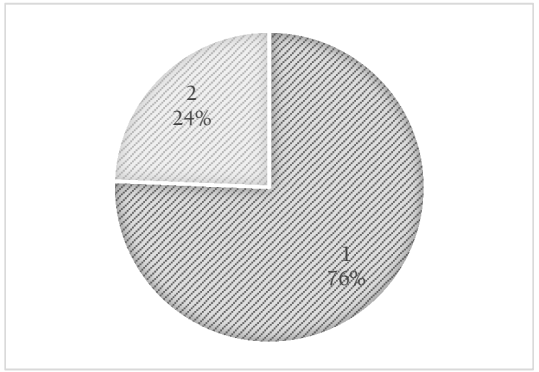

Figure 7.4. HSS 2014 journal articles

Figure 7.4 shows the percentages for HSS: just over three-quarters free.

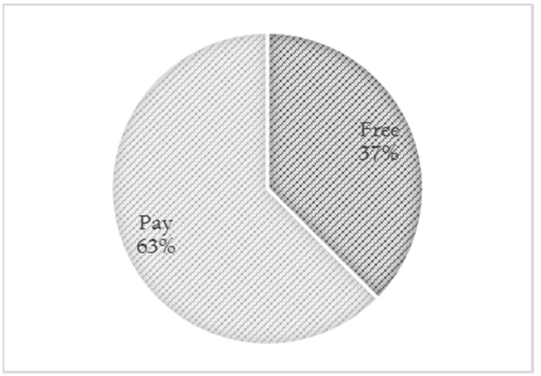

Figure 7.5. STEM and biomed 2014 articles

Figure 7.5 shows the percentages for STEM articles; the percentages for biomed are the same (rounded to the nearest whole percent).

Changes in Journal and Article Numbers

These last two tables show the differences in journal coverage between *Open-Access Journals: Idealism and Opportunism* and this report and in article coverage between the interim report (which covered the same journals) and this report. (If you're doing direct comparisons, note that the old numbers included journals graded C, which I've removed from the new discussion.)

Because megajournals moved to either Miscellany or Other Sciences and a number of journals were moved between the two, those rows are omitted from these tables as being meaningless in terms of growth.

There's quite a wide growth range, generally higher (in percentage terms) in the humanities and social sciences. While noting that this report should be far more inclusive of the global south and of nations with strong linguistic preferences (e.g., France, Spain and Portugal), I wouldn't attempt to draw firm conclusions from those patterns.

Subject	Journals	Growth
Agriculture	418	46%
Anthropology	263	104%
Arts & Architecture	226	52%
Biology	399	28%
Chemistry	155	24%
Computer Science	369	16%
Earth Sciences	302	63%
Ecology	246	70%
Economics	541	64%
Education	549	74%
Engineering	302	34%
History	275	102%
Language & Literature	524	100%
Law	218	106%
Library Science	131	72%
Mathematics	274	25%
Media & Communications	166	113%
Medicine	2,217	39%
Philosophy	175	82%
Physics	160	32%
Political Science	212	64%
Psychology	167	139%
Religion	114	78%
Sociology	404	74%
Technology	201	51%
Zoology	213	27%

Table 7.1. Growth in OA journals from previous (less complete) study

Subject	Articles	Growth
Agriculture	19,861	33%
Anthropology	5,703	119%
Arts & Architecture	4,139	41%
Biology	30,844	6%
Chemistry	13,400	14%
Computer Science	21,517	9%
Earth Sciences	10,433	40%
Ecology	11,705	40%
Economics	14,979	57%
Education	13,314	92%
Engineering	23,520	17%
History	5,883	90%
Language & Literature	10,711	84%
Law	4,394	129%
Library Science	2,542	82%
Mathematics	13,907	8%
Media & Communications	3,902	99%
Medicine	146,054	23%
Philosophy	3,035	119%
Physics	13,558	22%
Political Science	4,261	65%
Psychology	5,798	76%
Religion	2,784	67%
Sociology	13,338	73%
Technology	12,138	27%
Zoology	9,677	17%

Table 7.2. Growth in articles in gold OA journals from previous (less complete) study

8. Biology and Medicine

Biomed—subjects related to human biology and the many subjects related to medicine, including pharmacies, some aspects of nutrition and most aspects of sports and sports medicine—is distinctly where the money is.

Of the three segments, this one has the fewest journals. It's roughly tied with STEM for most number of articles and percentage of articles in APC-charging journals (although that's without *PLOS One*). But it has *by far* the most revenue, possibly as much as $167.8 million in 2014, nearly two-thirds of all OA APC revenue. This overview adds tables and graphs not already included in other chapters, then looks at APC and volume brackets based on this segment rather than on the overall field.

Cost per Article

	2014	2013	2012	2011
Revenue	$167,813,590	$122,361,263	$98,525,112	$80,723,564
Pay Articles	110,841	85,518	70,595	57,084
$/article	$1,514.00	$1,430.82	$1,395.64	$1,414.12
Tot. Articles	176,898	150,253	134,006	114,711
$/article	$948.65	$814.37	$735.23	$703.71

Table 8.1. Possible revenues* and cost per article, biomed, 2011-2014

Table 8.1 shows the possible revenues and cost per particle (for pay articles and for all articles) with two huge assumptions: that there were no waivers or discounts and that APCs for each journal were the same

throughout the four years. Assumptions or no assumptions, these are much higher figures than for other segments.

Journal and Article Volume per Year

	2014	2013	2012	2011
Journals	2,478	2,469	2,275	2,069
Free%	55%	56%	58%	59%
Articles	176,898	150,253	134,006	114,711
Free%	37%	43%	47%	50%

Table 8.2. Biomed journal and article volume and free% per year

Table 8.2 includes only those journals that actually published articles each year, a figure that grows close to 10% in 2012 and 2013 but essentially stays unchanged for 2014—while article volume continues to grow at a healthy rate. Meanwhile, a literature that was half-free (on the author side) as recently as 2011 is now 63% based on APCs.

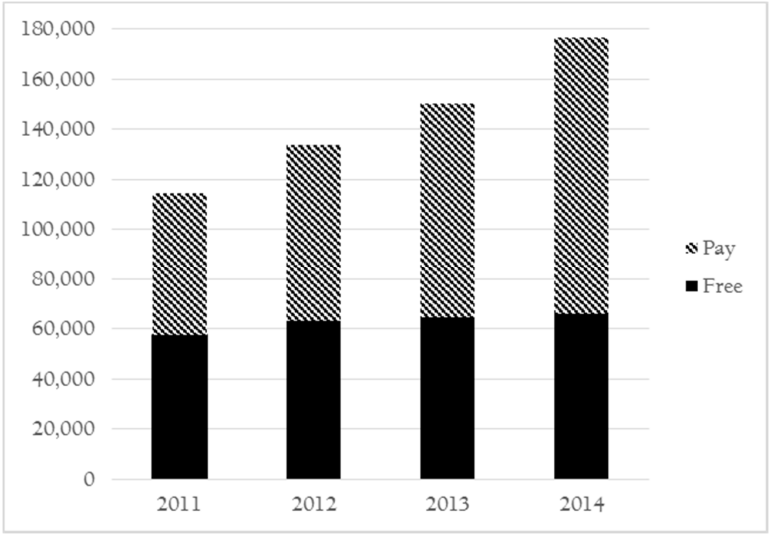

Figure 8.1. Biomed articles per year

Figure 8.1 makes the trend clearer: while article volume in no-fee OA biomed journals has grown slowly (but it has grown every year), volume in APC-charging journals has grown rapidly, and is 94% higher in 2014 than in 2011.

Of the 2,616 biomed gold OA journals, 1,349 (52%) published more articles in 2014 than in 2013; 162 (6%) published the same number; and 1,105 (42%) published fewer articles. Looking at significant changes, 1,118 journals (43%) published at least 10% more articles in 2014; 615 (24%) stayed about the same; and 883 (34%) shrank by at least 10%.

Revenue Brackets

Revenue	Journals	Cum J	Articles	Art/J
$2 million +	8		11,563	1,445
$1 to $1.94 million	26	34	18,890	727
$500,000-$999,999	52	86	17,134	330
$300,000-$499,999	59	145	11,741	199
$200,000-$299,999	48	193	6,990	146
$100,000-$199,999	92	285	9,575	104
$50,000-$99,999	151	436	10,401	69
$30,000-$49,999	120	556	6,196	52
$20,000-$30,000	117	673	5,739	49
$10,000-$19,999	159	832	6,112	38
$5,000-$9,999	114	946	2,804	25
$2,500 to $4,999	78	1,024	2,348	30
$1 to $2,499	88	1,112	1,307	15
$0	33	1,145		

Table 8.3. Journals by revenue bracket

If OA journals in general aren't easy get-rich-quick schemes, are biomed journals better? Yes, but not all that much, as shown in Table 8.3. Only 34 journals had at least $1 million potential revenue in 2014, and only 285 had at least $100,000. The steady decline in average articles per journal as revenues decline (except for an anomaly in the $2,500 to $4,999 bracket) may seem entirely reasonable.

New Volume and APC Brackets

Before moving on to subject chapters (only two in this case), it seems plausible to arrive at different article volume and APC-level brackets based on biomed rather than on all gold OA journals.

Article Volume Quintiles

Sorting by descending 2014 volume and looking at cumulative totals, the breakpoints for one-fifth of the articles (around 35,380, breaking between numbers) are 519 articles and up; 216-518 articles; 105-215 articles; 49-104 articles; and 0-48 articles. Note that these are lower numbers in the top three quintiles than the overall numbers, with a much narrower central quintile. Table 8.4 shows the breakdown.

	Journals	%Free	Articles	%Free
Largest: 519+	36	11%	35,521	10%
Large: 216-518	110	22%	35,361	20%
Med.: 105-215	242	39%	35,408	37%
Small: 49-104	508	57%	35,617	57%
Smallest: 0-48	1,720	62%	34,991	63%

Table 8.4. Article volume quintiles based on biomed cumulative articles

Even in biomed, most journals are quite small—and, as usual, the free percentage falls as the volume rises.

APC Quartiles

Using the same two methods as in Chapter 4 yields very different results if biomed is taken as the universe. Table 8.5 shows the journal and article counts breaking down the APC-charging journals to as close as possible to one-quarter in each group, by descending APC (as usual, it's not possible to get exact quarters: in this case, quite a few journals charge $1,960.) I think it's odd to call a $491 APC "nominal" or a

$1,049 charge "low," but within the high-priced biomed universe, that's where things fall out.

	Journals	Articles
High: $1,960+	302	49,695
Medium: $1,050-$1,959	275	23,554
Low: $492-$1,049	281	14,807
Nominal: $1-$491	287	22,785

Table 8.5. APC-charging biomed journals by APC level

Table 8.6 is fairly startling: breaking down quarters of total revenue, sorting by decreasing APC, then by journal revenue. Note the incredibly narrow second bracket ($2,187 to $2,249) and the very high brackets in general. This breakdown won't be used for subject chapters.

	Journals	Articles
Top: $2,250+	103	14,212
Next: $2,187-$2,249	93	18,262
Mid: $1,750-$2,186	219	22,593
Bottom: $1-$1,749	730	55,774

Table 8.6. APC-charging biomed journals by cumulative revenue levels

9. Biology

Biology includes most everything with "bio" in the title and more specifically all aspects of human biology, biochemistry and the like. Some areas such as marine biology are included in zoology. This subject includes 399 journals, which published a total of 24,710 articles in 2013 and a considerably higher 30,844 in 2014.

	2014	2013	2012	2011
Journals	374	378	339	297
%Free	43%	44%	44%	45%
Articles	30,844	24,710	23,183	20,605
%Free	20%	26%	25%	26%

Table 9.1. Biology journals and articles by year

Table 9.1 shows journals that actually published articles each year, the number of articles published, and the free percentages. The always-low free percentage of journals (considerably lower than medicine) declined only slightly over these four years, but the percentage of articles in free journals—always much lower than in most fields, including medicine-dropped sharply in 2014. (It's not that there were fewer articles in no-fee journals—that number has grown each year—but that there were *many* more articles in APC-charging journals, a jump of 34%.)

On a journal-by-journal basis, 182 journals (46%) published more articles in 2014 than in 2013; 17 (4%) published the same number; and 200 (50%) published fewer articles. Looking at significant changes, 153 journals (38%) published at least 10% more articles; 77 (19%) stayed about the same; and 169 (42%) published at least 10% fewer articles.

Figure 9.1 shows pay-vs.-free articles over the years graphically.

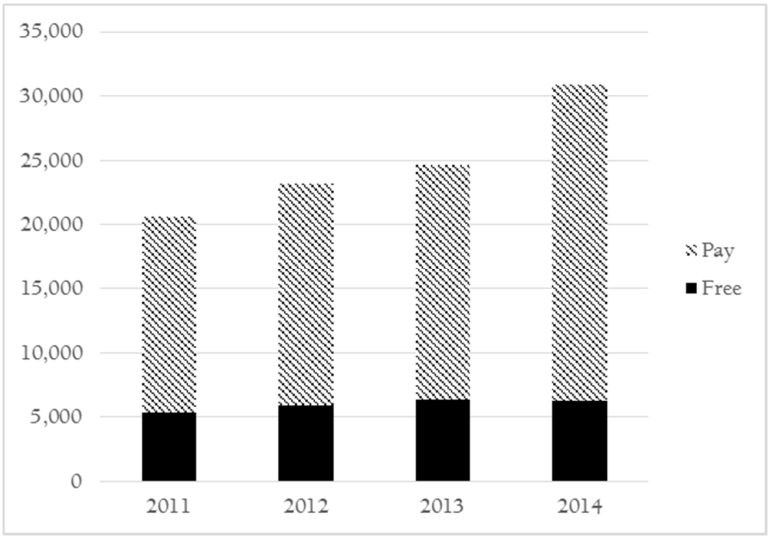

Figure 9.1. Biology articles per year

Other Details

	Journals	%Free	Articles	%Free
Largest: 519+	11	9%	11,740	5%
Large: 216-518	17	6%	5,529	5%
Med.: 105-215	30	30%	4,634	32%
Small: 49-104	58	38%	4,005	39%
Smallest: 1-48	283	50%	4,936	46%

Table 9.2. Biology journals by article volume

Table 9.2 shows the number of journals in each size category as adjusted for biomed journals, 2014 articles for those journals, and free percentages. The biggest journals—all but one of them with APCs—especially dominate biology, with more than twice as many articles as any other group. The second largest journals (also with only one free) also publish more articles than might be expected—and in both cases almost none of the articles appear in free journals.

	Jour.	%APC	%All	Art.	%APC	%All
High	62	28%	16%	11,741	48%	38%
Medium	58	26%	15%	7,736	31%	25%
Low	49	22%	12%	1,998	8%	6%
Nominal	56	25%	14%	3,114	13%	10%
None	174		44%	6,255		20%

Table 9.3. Biology journals and articles by fee range

Table 9.3 shows the number of journals and articles in each fee range, with fee ranges based on overall biomed APC levels (that is, roughly a quarter of fee-charging journals in each level): high, $1,960 and up; medium: $1,080 to $1,959; low: $492 to $1,079; nominal: $1 to $491.

Nearly half of all articles in fee-charging journals are in the most expensive group of journals—and only 21% of the articles are in the two least expensive levels.

APCs could have totaled $43,116,330 in 2014, with no waivers or discounts. Average charge per article for all articles in fee-charging journals in 2014 is $1,753. Average charge per article for *all* articles is $1,398, an extremely high figure.

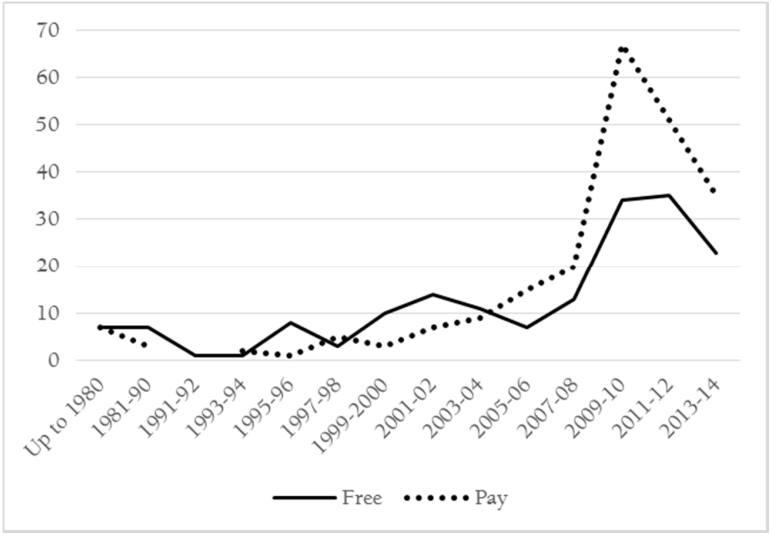

Figure 9.2. Biology OA journals by starting date

Biology OA journals are somewhat unusual in almost all respects, including the possibility of a "gold rush." While it's certainly true that many more APC-charging journals began in 2009-2012, such journals outpaced free journals in some earlier periods as well. Figure 9.2 shows the patterns.

Country	Articles
United Kingdom	8,549
United States	5,649
Egypt	4,886
Switzerland	3,014
Brazil	1,628
India	1,061
Bangladesh	800
Iran, Islamic Republic of	507
Netherlands	490
Japan	340
Turkey	305
Serbia	281
South Korea	252
Poland	249
Romania	221

Table 9.4. Articles by country of publication

Biology journals claiming to be published in more than 50 countries published articles in 2014; Table 9.4 shows the article counts for the fifteen countries with at least 200 articles.

Summing Up

Rapidly growing between 2013 and 2014, with nearly all that growth in fee-charging journals, which dominate this field more than most: That's the overall story for biology, where the fees are high and the biggest journals dominate.