Cites & Insights

Crawford at Large

Libraries • Policy • Technology • Media

Volume 14, Number 10: October/November 2014 ISSN 1534-0937 Walt Crawford

Intersections

Journals and "Journals": Taking a Deeper Look

When I started working on JOURNALS, "JOURNALS" AND WANNABES: INVESTIGATING THE LIST (*Cites & Insights* 14:7, July 2014), I was counting *all* the articles published in a journal in 2014, 2013 and 2012—but gave up the total counts fairly quickly.

I gave up not because I didn't think it was valuable information but because *some* journal websites made it difficult to get to lists of articles—sometimes *very* difficult. Since it was already clear that the project was huge, I decided to use the "20 or more" breakpoint and leave it at that.

That worked—but it meant I could only guess at some fairly powerful numbers. Specifically, for a given group of journals (e.g., "A" from Beall's lists, "B" from OASPA), how many articles were published over a given period?

Those numbers are powerful directly and indirectly. Directly, because they give a sense of how the journal's doing; indirectly for journals with APCs, because they can be used to calculate the maximum revenue the journal could have generated.

This new study builds on the earlier study. It's broader in some ways and narrower in others:

- Where feasible, I counted (or calculated a close approximation of) *all* peer-reviewed articles in each journal for 2011, 2012, 2013 and the first half of 2014 (through June 30).
- I rechecked APCs in all cases and PDFs in many cases. Given the rechecked APCs, it's feasible to determine maximum revenue for a journal. (Why "maximum"? We'll get to that later, but think waivers and changing APCs.)
- There's a new third set of journals, representing roughly half of the journals in the *Directory of Open Access Journals* (DOAJ) that aren't in the other two sets, more specifically representing journals *not* in the fields of human medicine, biology and biomedicine. That includes humanities, social sciences, mathematics and a range of hard sciences.
- The new study is narrower in that portions of the first two sets (the Beall lists and journals from OASPA members) are omitted or dealt with summarily, and in some cases the omissions are somewhat arbitrary. More about that as we go along. (Small portions of the third set are also omitted, rarely for arbitrary reasons.)

This study does not directly replace the July report. In "The Overall Picture" I do include tables that are somewhat comparable to the overall tables in that report, but only somewhat: I changed some definitions along the way and made a major change at the end of the study.

This report is in two parts. Other than "The Overall Picture" and some prefatory material, the third set of journals, in some ways the most interesting set, shows up in the second part. The first part primarily concerns the first two sets. The second part, the December 2014 *Cites & Insights*, will also offer some comparisons, consider possible *in*efficiencies of scale, and offer a few comments on the new (and generally much improved) DOAJ criteria.

A formatting note: For these two reports—this issue and the December 2014 issue—if you're not planning to print out the issue, you're better off with the single-column 6x9" version. Tables are wider and sometimes have larger and easier-to-read type in the single-column version. I won't attempt to balance column lengths in either issue, since tables need to appear within a single column.

The study itself was done between July 1, 2014 and September 26, 2014: all checking of journal websites took place during that period.

Early Prefatory Material

The sections below were written in late June 2014—after I'd established the initial sets of journals to be studied, but before actually doing the study and making refinements along the way. Some of what's here is a little misleading—e.g., the F group disappeared entirely and D was expanded and subdivided. You may want to skip right to "The Overall Picture," since I'll try to define everything as it worked out. In a couple of cases where I find it essential to add a caveat or update to what I wrote in late June 2014, that caveat or update appears as quoted material (indented on both sides and in slightly smaller type), beginning with either "**Caveat**:" or "**Update**:" in boldface.

A lamppost study

I decided to do a followup study—but what I originally thought of as a "lamppost study," from the old joke about the cop finding a drunk on his hands and knees under a lamppost. "What'cha doin?" the cop says. "Looking for my wallet," the drunk says. "I dropped it somewhere over there." "Then why are you looking here?" the cop says. "Because this is where the light is."

Admittedly badly told, but you get the gist. I would look at journals that make it easy to get to lists of articles—and, correspondingly, easy to get to articles—and admit up front that it wasn't an entirely valid universal study.

Opacity

The more I thought about it, the more I thought that this was *not* a lamppost study. Instead, it was a study of transparency and quantity—and transparency should be a goal of any journal, and specifically of a Gold OA journal that doesn't already have a huge reputation.

Think about it. If you're a researcher with a paper ready to submit and you're looking at a journal you're not already familiar with, you'll consider a number of factors (gone over in previous articles). Whatever else you do, you will *certainly* look at the contents for the last year or three—article titles, how much is being published, who's writing published articles and the quality of papers.

If a journal makes that difficult to do, it's opaque, and that's not a positive quality.

A transparent journal will at best have a reverse chronological (newest first) list of all accepted articles, possibly with breakpoints for years, possibly with breakpoints for "issues" (although for e-only refereed journals that post articles as soon as they're accepted, which should be the norm for OA, I increasingly wonder what the point of an issue is).

An opaque journal makes you go through several steps to get to a list of articles for one issue and you have to repeat those steps for each issue and for each volume. In the most opaque cases, there *is* no list of articles: you have to download complete issues in PDF form and browse through them or look at issue tables of contents in PDF form.

There are stages in between. I'd generally say that transparent journals require no more than two clicks to get from an overall archival list to the papers for an issue.

So I'm not calling this a lamppost study. I'm calling it a study that filters based on transparency or, if you prefer, opacity.

> **Caveat**: Some journals—nearly all in the Beall set—were excluded because my spreadsheet didn't have URLs for the journals, almost always because the publisher didn't offer a list of journal titles with hyperlinks. That's an issue at the publisher level, not necessarily the journal level.

It's not exactly a repetition of the earlier study. In part, it works from that study, but with entirely new numbers. It changes in several ways: the groups and criteria for each journal, the information collected and the universes studied.

Preliminary changes at the Journal Level

If I'd already checked a journal, its group (what I called "grade" in the July article) didn't change *unless* the factors noted below came into play. For journals I hadn't checked before, the group is newly assigned (and in general publisher pages were *not* checked this time around).

Working backwards alphabetically:

- **X: No change—**unreachable or unworkable.
- **O: Opaque**: Assigned if it was difficult to find out how many articles the journal published in each recent year.
- **N**: **Not a refereed OA journal**: No change in definition.

  > **Caveat**: That may not be entirely true. I do break down the "N" cases in each set by reason, including some reasons that may not have applied the first time around.

- **H: Hybrid**: No change.

None of the above has article counts, and a fairly large number of journals may wind up as Opaque.

- **F: Few articles**. Refined criteria: At least *five* articles in each year 2011-2013 (and at least two in the first half of 2014), fewer than 30

articles in 2012-2013 combined and fewer than 20 articles in any given year.

- **E**: **Empty**. No peer-reviewed articles between 2011 and June 30, 2014.
- **D**: **Nearly empty**. This combines the old E2 and D but with slightly different criteria, based on the current *DOAJ* rules for listings. To wit: A journal is flagged as D if it has at least one article but has fewer than five articles in any year 2011-2013 (that it's been publishing) or fewer than two in the first half of 2014.

  Update: That final clause, "or fewer than two in the first half of 2014," does not apply in cases where an issue-oriented journal publishes either one or two issues a year. In such cases, lack of 2014 articles doesn't automatically flag a journal as D.

- **C**, **B**, **A**: No changes, except that missing or hidden APCs cause an immediate C (for journals not previously checked).

One key change: Some journals that were C, B, or A (and a few E2 or D) are now O because they're difficult to quantify.

Oh, there's also the inclusion of 2011 (if that doesn't turn out to be too much work).

Update: It didn't.

Preliminary changes in journal sets

This time around, there are three sample sets: a trimmed version of Beall's lists, a trimmed version of OASPA journals, and a *vastly* trimmed version of journals in *DOAJ* as of May 7, 2014.

Trimming Beall's set

I deleted all journals in groups X, N and H.

I deleted journals published by MDPI, as these are included in the OASPI set (I'm not convinced MDPI is questionable).

I deleted journals from opaque *publishers*—that is, cases where I was never able to download or generate a hyperlinked list of journal titles, and had to key the titles in and link to them individually from the publisher page.

Update: Just to clarify, the paragraph above describes the most arbitrary exclusion of journals. It eliminated about 1,100 journals, including 320 that were empty in any case but also several hundred journals that would otherwise be included.

That left 6,800 journals as a starting point (which may be reduced by journals that are underlined but don't actually have hyperlinks).

Update: There were very few of those. There was also at least one journal that's a duplicate within the Beall lists. The final count is 6,775.

Trimming the OASPA set

I deleted journals in groups X, N or H. That left 1,401 journals as a starting point.

Update: Cases where journals turned out not to have working hyperlinks reduced the final count to 1,378.

Creating the DOAJ set

I wanted a reasonably large sample but not one so overwhelming that I'd give up. Starting with the downloadable spreadsheet (actually CSV file) from May 7, I began by deleting all journals that didn't have English as the first-named language. That left something over 6,300 journals.

Then I thought about subjects and realized that the Beall test, the OASPA list (I believe) and most commentaries I've seen about Gold OA journals are all heavy on biology, biochemistry and medicine—which isn't surprising, especially for APC-charging journals, since that's where the money is.

For a change of pace, I *removed* all journals with primary subject headings for biology, biochemistry and medicine (medicine being *by far* the most numerous, with its scores of narrower subjects).

That left 4,280 journals.

Then I removed all journals with publishers on Beall's list, which brought it down to 3,935, and journals with OASPA publishers, which brought it down to 3,519. Finally, I normalized the "independent" Beall titles that were left and used Excel lookup tools to find matches. That brought the count down to 3,465.

Update: You can read that as saying that Beall's list has about 409 journals that are in *DOAJ* and aren't in the biology, biochem and medicine fields—and that the OASPA set has about 416 such journals. I believe those are reasonably good estimates, in each case representing a bit less than half of the journals from the set that are in *DOAJ*.

That's 3,465 journals with English-language interfaces available in the humanities, social sciences, mathematics, hard sciences *except* for biology, and pretty much everything outside of biology, biochem and medicine. (I left veterinary medicine in the set.)

This is the only set where I hadn't already seen most of the journals. It may turn out that the full grading process is too laborious, in which case I'll use some other method to reduce the sample size.

Update: The full grading process wasn't too laborious in the context of this large project, but 127 journals disappeared for other reasons: because I found that the actual publisher *was* in one of the other two sets (publisher names aren't always used consistently); because the journal title was included elsewhere; the most common, because although the

journal's DOAJ language codes began with English, there wasn't enough English content in the journal's interface for me to be able to understand such elements as whether there's an APC or where the archives were. The final set includes 3,338 journals.

I'm writing this after preparing the sets but before beginning the actual survey (which won't begin until at least July 1, 2014, since I want to include the first half of 2014). We'll see what we see…

Update: That's the end of the prefatory material, written before I started checking journals. As you'll see later, a few things changed along the way.

The Overall Picture

The six tables that follow are probably too terse to be very useful, as some groups or categories require a fair amount of explanation. If you attempt to compare these tables to those in the July 2014 article, you need to be aware of the following:

- This time around, if it seemed likely that a journal had an APC but I couldn't find the amount (or if the instructions mentioned an APC but not the amount), the journal received an automatic C. There's really no excuse for failing to state these charges clearly. (There were a few exceptions to this rule.)
- Group E2 disappeared, with journals generally merged into the expanded D group.
- E journals where the publisher has marked the title as ceased or cancelled were also moved into the expanded D group.
- The new D includes not only dying, dead and dormant journals but also "diminutive" journals—ones that wouldn't qualify for inclusion in the new *DOAJ* (as I understand the criteria) because there are fewer than five articles in a year other than the first or last (or fewer than two in the first half of 2014), but that may be workable journals in *very* narrow categories, mostly in the humanities. The new requirement of five or more articles per year is more stringent than the old one of four or more articles; this moved some journals from F to D. More about diminutive journals when we get to the DOAJ set.
- The first table includes F as previously defined: no year with at least 20 articles and no two-year period with at least 30 articles.
- As I was doing the study, I added a "subgroup" of A, B or C to each F journal—what the journal would get if it had more articles. Looking at many of the journals and thinking about it, I concluded that F was arbitrary. The second table in each set, and the discussion throughout, replaces F with the subgroup: that is, an A journal with 10 articles a year is still an A journal. The full discussions will, of course, include notes on article frequency.

With those minimal notes (and noting that an APC of $1,000 or more is still an automatic B at best), here are the overall numbers and percentages for the three sets of journals.

Beall's lists

Group	Journals	Percent
A: Apparently good	211	3.1%
B: May need investigation	579	8.5%
C: Highly questionable	521	7.7%
D: Dormant or diminutive	1,340	19.8%
E: Empty	2,052	30.3%
F: Few articles	1,226	18.1%
H: Hybrid	1	0.0%
N: Not OA peer-reviewed	46	0.7%
O: Opaque or obscure	372	5.5%
X: Unreachable	411	6.1%
Total	6,775	100.0%

Table 1: Beall set including group F

If you attempt to compare the numbers in Table 1 directly to those in the July 2014 report, note that all H, N and X journals were removed from this set even before I started, but also that MDPI's 124 journals, all of them A, B, E2 or F in July, moved from this set to OASPA.

Group	Journals	Percent
A: Apparently good	350	5.2%
B: May need investigation	1,130	16.7%
C: Highly questionable	1,057	15.6%
D: Dormant or diminutive	1,339	19.8%
E: Empty	2,046	30.2%
H: Hybrid	1	0.0%
N: Not OA peer-reviewed	50	0.7%
O: Opaque or obscure	375	5.5%
X: Unreachable	427	6.3%
Total	6,775	100.0%

Table 2: Beall set using current groups

The changes in D, E, N, O and X came about because of some rechecking while distributing F journals.

OASPA set

Group	Journals	Percent
A: Apparently good	386	28.0%
B: May need investigation	394	28.6%
C: Highly questionable	2	0.1%
D: Dormant or diminutive	185	13.4%
E: Empty	85	6.2%
F: Few articles	298	21.6%
H: Hybrid	0	0.0%
N: Not OA peer-reviewed	0	0.0%
O: Opaque or obscure	26	1.9%
X: Unreachable	2	0.1%
Total	1,378	100.0%

Table 3: OASPA set including group F

There were also some OASPA publishers whose sites didn't include hyperlinked journal lists and, thus, aren't included here. There were no *new* instances of hybrid or non-OA journals in addition to the 46 discarded from the earlier study, but there were two journals that now failed to show up, in addition to the single case in July 2014.

Group	Journals	Percent
A: Apparently good	611	44.3%
B: May need investigation	460	33.4%
C: Highly questionable	8	0.6%
D: Dormant or diminutive	229	16.6%
E: Empty	42	3.0%
O: Opaque or obscure	26	1.9%
X: Unreachable	2	0.1%
Total	1,378	100.0%

Table 4: OASPA set using current groups

Changes in D and E involve rechecking while distributing F journals. (For one thing, in all sets, if a journal had no articles through June 30, 2014 but *did* have later 2014 articles, it moved from E to D.) Empty groups (H, N) are omitted.

The obvious comparisons here are between Table 2 and Table 4. More than three-quarters of the OASPA journals qualify as A or B (and many of the "B" are there because of high APCs), while less than one-quarter of the Beall set do—and while there are only a handful of empty OASPA titles, there are more than two thousand such "journals" in the Beall set.

It may be worth noting that *all* of the opaque OASPA journals come from a single publisher that maintains article archives in no order I could grasp.

DOAJ set

Group	Journals	Percent
A: Apparently good	1,368	41.0%
B: May need investigation	230	6.9%
C: Highly questionable	49	1.5%
D: Dormant or diminutive	558	16.7%
E: Empty	8	0.2%
F: Few articles	638	19.1%
H: Hybrid	0	0.0%
N: Not OA peer-reviewed	140	4.2%
O: Opaque or obscure	175	5.2%
X: Unreachable	172	5.2%
Total	3,338	100.0%

Table 5: DOAJ set including group F

There's no July table to compare these to, which may be a good thing. None of these was a hybrid and almost none were empty. Most of these journals are singletons: the 3,338 journals come from more than 2,500 entities (in many cases universities, libraries and societies). The largest single-publisher clusters are 79, 59 and 35 journals respectively, although that may be misleading.

Group	Journals	Percent
A: Apparently good	1,942	58.2%
B: May need investigation	274	8.2%
C: Highly questionable	69	2.1%
D: Dormant or diminutive	558	16.7%
E: Empty	8	0.2%
N: Not OA peer-reviewed	140	4.2%
O: Opaque or obscure	175	5.2%
X: Unreachable	172	5.2%
Total	3,338	100.0%

Table 6: DOAJ set using current groups

Possibly because the DOAJ set was generally checked more recently, there were no odd changes in other categories when distributing F journals—that is, they all became A, B and C. Possibly worth noting: after completing the general sweep but before splitting out F journals, I revisited all of the X journals that weren't 404s. That set of revisits yielded 16 journals with spotty availability, moved to other groups as appropriate.

Publisher counts

There's one other area in which the new study and the old one might be compared: the number of publishers represented in each set.

The Beall set began with 501 distinct publishers, of which only 432 had any journals or "journals" to test. Once the Beall list was trimmed to the 6,775 journals and "journals" in this study, only 224 publishers—slightly more than half—are included. Furthermore, of those 224 publishers, only 170 show journals with published articles. In other words, 170 of the 501 "publishers" are actually publishers. "Hundreds" used to describe the number of questionable or predatory publishers is at best misleading; as of this study, there were fewer than 200 actually publishing journals, and I'll argue that a fair number of those aren't actually predatory or questionable (or at least no more so than all the subscription-based publishers that charge author-side fees).

The OASPA set began with 62 publishers. Various exclusions (including publishers that only publish books, not journals) reduced that to 40 for this study. Of those 40, 38 had at least one journal with published articles. One single-journal publisher was unreachable and another single-journal publisher showed no content from 2011 through mid-2014.

Fleshing Out the Groups

Let's look at each of the current groups in a little more detail.

A: Apparently good

These journals seem like good places for authors and readers, based on what I could see in a quick examination—and they publish at least five peer-reviewed articles a year (except possibly the first year and 2014). For 2014, unless the journal is issue-oriented and appears one or two times a year, there were at least two articles in the first half year.

If there's an article processing charge (APC) or similar charge, it's stated clearly (either as its own tab or link or as an easily visible part of journal description or author instructions) and it's less than $1,000. It must be an actual charge, not a statement that the author will be told what the charge will be—that's just not reasonable.

I didn't find enough questionable English, apparent lack of competence in site design or oddities in article titles checked to downgrade the journal from A to B or C.

There could be more to it that I can't determine. A journal could come from a publisher that engages in wholesale email solicitations, it could have plagiarized articles it could have faulty peer review. There are probably cases where an A journal isn't a good place for authors because it only accepts submissions from, say, grad students at one university or scholars within an institution, but that's an entirely different issue. Since those limits are stated, they don't disqualify the journal. "An apparently good place to submit an article" carries what may be an obvious but unstated additional clause: "assuming that you're eligible to submit to the journal and your paper is appropriate for it."

B: May need investigation

These journals publish at least five peer-reviewed articles a year (with the same exceptions as for A, above) and have clearly stated APCs (if they have such charges), but there's something about them that's troublesome enough to suggest that an author might want to know more. I think of these as yellow-flag journals, where some caution is appropriate. For example:

- Any journal charging $1,000 or more as an APC automatically gets a B or lower, simply because I believe such high charges require some justification.
- Journals published in English where the site itself has clearly defective English earn B or lower, because the lack of language proficiency raises issues about publisher competence.
- Publishers and journals using America, Canada or Europe as part of the name or an important element of their website, where the nature of the editors, editorial boards and websites make it *highly* likely that they aren't American, Canadian or European, drop down to B at best.
- Journals where the site is busy or peculiar tend to earn B or lower. I think the average author or reader would want to know more before trusting them. That can include boastful language (that doesn't quite lie) and other issues.

To a great extent, the line between A and B is arbitrary; I'd consider all A and B journals to be plausible resources.

C: Highly questionable

If B journals raise yellow flags, these raise red ones: Proceed with caution if at all. For most readers and authors, I believe these journals should *not* be regarded as appropriate resources. Why not?

- Some clearly or probably have article processing charges but don't say what they are. That's just not appropriate: Saying "We'll tell you how much you owe us later" is nonsense. (Saying "you'll owe us *no more than* X" is fine, and that didn't cause downgrading.)

- Some sites appear so incompetent (either at the journal or publisher level) that one has to wonder whether they're doing anything right.
- Some publishers make claims on their sites that are clearly false; that's an automatic C.

There are *very* few C journals in the OASPA and DOAJ sets: A total of 77 out of 4,716 journals or just over 1.6%. Of the 1,055 C journals in the Beall set, 810 come from 26 publishers each having at least 10 C journals (with 210 C journals in one case). Checking my notes on some of those publishers, I see things like "false statements," "lack of editorial boards," "APC not stated" (in eight of the 26 cases), the interesting statement "we do not reject any paper," and more. Those 26 publishers have a total of 150 B journals and no A journals at all. If you're doing the math, you'll find that eliminating those 26 publishers reduces the total number of C journals to 245—not zero, but a relatively small number.

D: Dormant, diminutive, dying or dead

These journals didn't publish much—in some cases they didn't publish anything at all. For a journal to be D, it has either been explicitly cancelled (or merged with another journal) or it has fewer than five articles in 2011, 2012 or 2013 (when the year isn't the first of the journal's existence) or fewer than two articles in the first half of 2014 (when the journal appears more than twice a year or isn't issue-oriented).

One big change in *my* thinking from the July 2014 report: I now believe that a fair number of D journals, mostly in the humanities, are perfectly reasonable journals that just don't publish much. A lot more about that in Part 2.

In individual set discussions, I break D down into six categories:

- **C: Ceased**—either an explicit statement from the publisher or a journal that has had no articles since 2012. Some of these journals *never* had any articles and were categorized as E in the earlier study.
- **D: Dying**—a pattern of article counts suggesting that a journal's disappearing.
- **E: Erratic**—a journal that has more than five articles in some years (sometimes a *lot* more) and fewer in others, but doesn't appear to be in a dying pattern.
- **H: Hiatus**—a journal that appears to be on a declared or undeclared hiatus, usually the latter.
- **N: New**—a journal that only had articles in one year and either fewer than five in a year before 2014 or only one in 2014.
- **S: Small**—a journal that has had fewer than ten articles each year and has a publishing pattern suggesting that it's a journal with a very narrow focus (or simply unable to attract many articles). You'll see some examples in Part 2 of this report.

The distinctions between some of these categories are as subjective as the distinction between A and B.

E: Empty

These journals have not published *any* articles, at least between January 1, 2011 and June 30, 2014, but they also haven't been marked as cancelled or ceased by their publisher.

Most "journals" fall into this category—web pages frequently filled using templates (sometimes in painfully obvious ways), often with no ISSN or editorial board, waiting to either go away or somehow attract an article. There are only 50 E "journals" in the OASPA and DOAJ sets combined, compared with more than two thousand in the Beall set.

H: Hybrid

Journals—or, rather, *journal* (since at this point there's only one in all three sets combined)—that charges subscription prices but offers OA at an article level for a usually-high APC. Based on all the evidence I've seen to date, this is double dipping (the journal's subscription price isn't going to be reduced based on the few OA articles).

N: Not open access peer-reviewed articles

A journal can be flagged as N for several reasons, some of which are tougher than DOAJ criteria and may be too tough. I note some of them in individual set discussions, but here are a few:

- Journals that aren't OA and in some cases don't claim to be.
- Journals that require reader registration before articles can be read—the most common situation. If you can't read anonymously, I don't think it's really open access.
- Journals that consist entirely or almost entirely of conference proceedings rather than peer-reviewed articles.
- Journals that claim to be OA but don't provide access to the latest issue(s). Embargoed OA isn't OA, especially for gold OA.
- Journals that don't appear to consist of peer-reviewed articles, including "journals" that are really monographic series and websites that are

in DOAJ (but may not stay there) but don't appear to have peer-reviewed articles as such. (Post-publication peer review still counts as peer review, but open discussions and essays posted without any peer review don't.)

This is the only category beyond A-E where most instances are in the DOAJ set rather than the Beall set. In most DOAJ cases, the websites are neither deceptive nor bad—they're just something other than collections of peer-reviewed articles. More about that in Part 2.

O: Obscure or opaque

Most journals in this group, cases where I found it too cumbersome to try to count articles over date ranges, disappeared *before* I started the analysis—journals from publishers where there's no list of hyperlinked journal titles on the publisher website.

The rest fall into several categories and I mention some of those in individual set discussions, such as:

- **Whole-issue PDFs**, where the only way to see article titles is to download journal issues.
- **Undated archives**, either issue-oriented archives where neither the volumes nor the issues contain years or archives where there doesn't seem to be any chronological order to articles. (If there's a search function that can take publication date as a search, those journals are *not* O: I used the search function.)
- **Combined archives**, where archives for several journals are all in one intermingled lump.
- **Multilayer/complex archives**, where it took too long to try to figure out what was what.

I revisited many O journals near the end of the study and was able to move some into other groups, but most were just too hard to cope with. Fortunately, these account for only about 5% of the Beall and DOAJ sets and less than 3% of the OASPA set: this turns out not to be a lamppost study in any major sense even without the disadvantages of opacity.

X: Unreachable or unworkable

These journals just weren't there (in almost all cases) or were so defective that I couldn't analyze them (e.g., a journal in which all archive and author-instruction links yield 404s or blank pages).

If the journals showed as 404s, I gave up. If they showed other error messages—or, in a few cases, "just kept spinning" for five minutes or more—I'd try to give the journal a second shot a few days or a few weeks later. That sometimes worked.

Why F went away

On one hand, I could make a case for a slightly tougher version of F: As I understand it, a journal won't be considered for Impact Factor by Thomson Reuters unless it publishes at least 20 articles a year. (I may be wrong on that and can't currently verify it.) That's not a bad measure for most medical, biomedical and broad science journals.

On the other hand, the more I looked at these journals, *especially* in the DOAJ set (largely humanities and social sciences), the more it looked like an arbitrary distinction.

Approximations

How did I arrive at the article counts in the set discussions that follow? I counted the articles—but that's too simple.

In most cases, I did exactly that: opened either a date-oriented archive or individual issues and counted what appeared to be peer-reviewed articles (based on the journal's own definitions).

In some cases, I took shortcuts, mostly for journals with a lot of articles. For example, if there was a date search function based on publication date, I'd use that, filtering for peer-reviewed articles if that seemed sensible. If the archive was 25 articles to a page, I'd zip through pages to look for annual breakpoints, not attempting to remove a handful of editorials or short notes. If articles within issues were numbered or DOIs were assigned sequentially within issues or volumes, I used those numbers (but if every issue had an editorial, I subtracted those).

In a few cases, mostly journals with a lot of issues per year or a lot of articles per issue, that became too cumbersome. I used various forms of approximation to come up with ballpark figures. Since discussions rarely mention individual journal titles, such approximations shouldn't affect the discussion, although those approximations are one reason the raw datasets will not be made freely available (except in stripped form). Approximation techniques, from what I regard as the most reliable to the least reliable, include these:

- For a journal with many issues and quite a few articles per issue but also continuous pagination within a volume, I counted the articles in the final issue of a volume (or the last June issue for 2014), calculated the average pages per article in that issue, and calculated the probable articles per volume based on that average length. So, for example, if the

last issue covered pages 7,200 to 7,400 and had 20 articles, I calculated 10 pages per article and 740 articles for the volume.

- For a journal that did *not* have continuous pagination, either because there was no overall pagination or because pagination was on an issue-by-issue basis, and that also had quite a few articles per issue, I counted the number of articles on the first full browser page (after moving past overhead), then multiplied that number by the number of page-downs required to get to the end of the issue. (Thus, if a journal had eight articles on the first page and 25 pages, I did *not* manually count all 200 articles; I assumed 200.)
- **Worst case**: Many issues (typically *more* than 12) and no volume pagination. I counted the articles in each of two to four issues throughout the year, then multiplied the average by the number of issues. If the issues varied widely in number of articles, I'd increase the sampling. For some erratic journals, this is the most unreliable approximation technique.

How many journals required approximation? As far as I can tell, taking approximation as a percentage of those journals that *had* countable articles (that is, categories A-D), less than 1% of those in the Beall set, about 1.3% of those in the OASPA set, and about 2.2% of those in the DOAJ set—or, to put it another way, 115 journals in all out of 8,027 with one or more articles. I revisited most approximations at the end of the data analysis to see whether I could do a better job, and in a few cases was able to refine the numbers.

One other mild caveat: In a few cases, I may have included a few 2014 articles that appeared after June 30, 2014, but I generally controlled for this (explicitly if publication dates were part of the displayed metadata, approximately if not).

As an aside, it may be obvious to readers that I did all of this *manually*—that I didn't use programming tools, APIs and what have you to automate this enormous task. I don't believe most of it could have been automated, and if it could, I don't know enough to do so. (Maybe it's a reaction to a career spent doing computer programming, analysis and design: I find that I really don't *want* to be a programmer in retirement, at least not for free.) I also found that seeing article titles—at least some of them—was useful, as was seeing journal sites. I probably encountered two or three dozen journals where I may return at leisure to read some of the articles. In any case, while I've done a lot of Excel massaging and could not have done this project were it not for Excel's ability to open browser tabs based on hyperlinks in cells, data gathering was manual.

New Calculations

I believe in transparent methodology. This interminable introduction probably makes that clear. Before getting on to actual set analysis and discussion, I should note some of the new calculations in this broader report.

Peak articles and total volume

Each journal row in the spreadsheets has peak and total cells. The peak count is the highest of any of the individual article counts (2011, 2012, 2013, first half of 2014). The total volume is the sum of the individual article counts.

Article counts

The biggest problem with the July 2014 report, I believe, is that A-C article totals, either by APC range or overall, were so far removed from reality because I stopped counting at 20.

This time around, allowing for some small variation because of approximations, the article counts should be more meaningful. They cannot, of course, be compared to the July 2014 counts.

Maximum revenue, annual and overall

When I discuss potential revenues for a journal or a publisher, I'm taking the *current* article processing charge (for a 10-page article if pagination counts, for a non-member if that's an issue, for a "foreigner" when that's an issue, the lowest membership cost if publication requires membership) multiplied by the *peak* article count. Thus, publisher totals are likely to be high because one journal may have had its peak in 2012 and another in 2013.

It's also a maximum, almost certainly wrong in that respect for many publishers. Some publishers automatically waive or reduce APCs for authors from some nations; some waive APCs on request; some change APCs quite frequently. (In that last case, the calculated figure could as easily be low as high.)

I believe maximum peak-year revenues and overall revenues are useful general indicators, not specific to-the-dollar data. It is useful to know, for example, that only 12% of journals that actually publish articles within the Beall set have maximum annual revenues of $25,000 or more or that 3.6% of journals in the OASPA set have maximum overall revenues of one million dollars or more (both real figures).

This ain't no scholarly research

This report appears in *Cites & Insights*, which is *not* a peer-reviewed journal: it's a periodical. The report has not been reviewed by my peers. Further, it is not standard scholarly research: I did not begin with a thesis and attempt to prove or disprove it.

It's descriptive research. I went out to see what was actually happening. I didn't attempt to draw a statistically valid sample (and will suggest that these journals are too heterogeneous for that to be meaningful); I looked at everything that was out there.

What basis do you have for believing I actually did all of this work, reasonably meticulously, and didn't just make it up? Only my name and reputation: More than three decades of transparent research and publication, much of it in formal venues from regular publishers. There's also the fact that there's no conceivable reason I would go to such lengths without having done the work: Why bother?

Now, finally, on to the journal sets, beyond the overall tables you've already seen (but repeating those for good measure).

The Beall Set

If you haven't already done so, you should read three other *Cites & Insights* essays before reading this report—and especially before looking at what's left of the Beall set: two ETHICS AND ACCESS essays in the April and May 2014 issues and the entirety of the July 2014 issue. Those essays provide useful background for this discussion.

Group	Count	%All	%A-E
A: Apparently good	350	5.2%	5.9%
B: May need investigation	1,130	16.7%	19.1%
C: Highly questionable	1,057	15.6%	17.8%
D: Dormant or diminutive	1,339	19.8%	22.6%
E: Empty	2,046	30.2%	34.5%
H: Hybrid	1	0.0%	
N: Not OA peer-reviewed	50	0.7%	
O: Opaque or obscure	375	5.5%	
X: Unreachable	427	6.3%	
Total	6,775		5,922

Table 7: Journals in Beall set

Table 7 is *almost* identical to Table 2, but adds a third column: percentages of those journals and "journals" that are actually studied here. Groups H, N, O and X represent journals that really *aren't* part of the studied set for various reasons.

Before offering overall tables, let's look at those excluded groups in a little more detail.

H: Hybrid journals

The initial Beall set had 200 of these, all omitted before this study began. One more—a journal that either became hybrid quite recently or where I missed its true nature the first time—emerged during this study. I don't believe hybrid journals are really OA at all; they're double dipping. This particular journal has an $1,800 fee for OA.

There's a huge difference between a hybrid journal and a journal that doesn't offer open access to everything that subscribers see. That difference is peer review. It's entirely appropriate for a journal to offer subscribers (or only print subscribers) access to editorials, news reports, conference reports, book reviews and other non-peer-reviewed material that isn't accessible to everybody.

N: Not an OA peer-reviewed article journal

As noted earlier, my version of OA here is fairly stringent. It doesn't allow for embargoes and it doesn't allow publishers to require readers to register before gaining access to articles (I can see no plausible reason to do that and very good ones *not* to, including the desire to read without being identified). I also exclude sites that allow access to everything but don't appear to represent peer-reviewed articles, at least not primarily.

Among the fifty Beall-set journals in this group, there are nearly three dozen that require registration (almost all from two publishers, one of which also requires APCs but doesn't state them), half a dozen that aren't OA at all (and don't claim to be), another half-dozen that appear to be entirely conference-oriented and a few true oddballs, including one journal that has all the trappings of a standard scholarly journal but, as far as I can tell, is entirely written by one man; another that calls itself a journal but seems to be an anti-HMO pressure group; one that consists of images rather than articles; one that shows abstracts but not full papers; and one that doesn't seem to have any way to get at articles.

As for APCs, five of the N journals charge $500 to $600; one charges $200; 23 charge $33; and nine clearly have APCs but don't state them.

O: Opaque or obscure

My goal was to count articles by year in as many journals as possible. In 377 cases in the Beall set, I gave up. Reasons include more than 115 where there were no dates at either the volume or issue level; at least 100 where it was so difficult to navigate the ar-

chive that I gave up; 15 (almost all from one publisher) that kept popping up ad windows whenever I attempted to do *anything*; at least nine (I think more) where the archives consisted entirely of whole-issue PDF downloads; more than a dozen where all articles from several different journals were dumped into a single archive; and more than a dozen where the archives were too complex to figure out. Some journals simply lacked any visible archive link.

APC	Journals	Percent
$2,000+	0	
$1,000-$1,999	1	0.3%
$600-$999	0	
$300-$599	31	8.2%
$200-$299	134	35.5%
$100-$199	29	7.7%
$50-$99	57	15.1%
$1-$49	2	0.5%
None	4	1.1%
Unknown	119	31.6%
Total	377	

Table 8. APCs for Beall journals in group O

Table 8 shows APCs for these journals, using the same ranges used throughout this report (but with zero-journal lines included). Note the high percentage of unknown APCs; nearly all of these either *definitely* or *probably* have APCs, but I couldn't find them or they were stated as "we'll let you know." The sweet spot appears to be $200-$299.

X: Unreachable or unworkable

Between 525 unreachable journals in the original Beall set (omitted before starting this study) and 427 that were unreachable when rechecked this summer, that's some 952 "journals" that at this point simply don't exist. In some cases, the publisher's disappeared (or, in one case, now denies access to its site—an odd situation for an OA publisher!). One "publisher" has exactly the same five "papers" in all of its "journals" and finds that it needs to protect the names of its "editors." (In another case, all of a publisher's so-called journals are pointers to a single stream of articles.) The journal URLs for one publisher are all flagged by my browser as security risks: I chose not to ignore that warning. At least one publisher apparently changed all the URLs for its journals, within a four-month period, and *left no links* at the old URLs: that level of incompetence makes the journals essentially nonexistent.

At the journal level, 79 titles yielded 404 errors; four yielded a "reported web forgery" message; 11 showed as "suspended page"; 25 showed "access denied"; 37 resolve to a single useless publisher page; more than 50 have archives so broken or unreachable as to make the journal worthless (in addition to similar cases for O journals); at least 65 are now parking pages; more than 61 were unreachable (but didn't yield 404s); and there were other cases.

In all of these cases, if there were actual worthwhile articles submitted by actual authors, that's a shame: they're effectively gone. If the publishers or journals do re-emerge, I'd regard them as wholly untrustworthy.

That's the cruft. Let's proceed with the rest.

Peak article count, all groups with articles

Table 9 shows the number of articles in the peak year in each journal in groups A-D—3,876 of them in all, including ten with no articles that were explicitly cancelled, putting them in D rather than E. It also shows the total volume of articles for each group of journals between 2011 and June 30, 2014 (Volume).

If you accept the estimates that there are roughly 28,000 peer-reviewed journals and roughly two million articles per year (neither estimate solidly proved, and these numbers may or may not include humanities and social sciences), then the "average journal" publishes 71.4 articles per year. By that reckoning, and excluding entirely empty journals, more than 85% of the Beall journals have fewer articles than average.

Peak	Journal	Percent	Volume	Percent
1,000+	14	0.4%	50,615	15.6%
600-999	22	0.6%	40,347	12.5%
300-599	39	1.0%	36,486	11.3%
100-299	183	4.7%	66,362	20.5%
75-99	104	2.7%	20,800	6.4%
50-74	211	5.4%	28,659	8.9%
35-49	240	6.2%	21,090	6.5%
20-34	533	13.8%	28,067	8.7%
10-19	796	20.5%	19,088	5.9%
5-9	848	21.9%	8,989	2.8%
1-4	876	22.6%	2,988	0.9%
None	10	0.3%	0	
Total	3,876		323,491	

Table 9. Peak articles in Beall journals, groups A-D

Nearly two-thirds of the journals published fewer than 20 articles per year and less than 10% published more than 100 articles in the busiest year.

Two of the top 14 journals published somewhat more than 2,000 articles in their busiest year. The handful of journals publishing more than 1,000 articles in their busiest year account for roughly one out of six of all articles published—and the 6.7% publishing 100 articles or more per year account for more than two-thirds of all articles published.

Article Processing Charges

APC	Journals	Percent	Volume	Percent
$2,000+	32	0.8%	1,670	0.5%
$1,000-$1,999	99	2.6%	7,507	2.3%
$600-$999	708	18.3%	58,019	17.9%
$450-$599	487	12.6%	30,570	9.5%
$300-$449	735	19.0%	31,901	9.9%
$200-$299	625	16.1%	23,738	7.3%
$100-$199	341	8.8%	65,316	20.2%
$50-$99	192	5.0%	34,580	10.7%
$1-$49	46	1.2%	6,701	2.1%
None	162	4.2%	5,290	1.6%
Unknown	449	11.6%	58,199	18.0%
Total	3,876		323,491	

Table 10. APCs for Beall journals, groups A-D

Table 10 shows how many journals in the Beall set charge article processing charges (APCs) or equivalent charges—and what volume of overall articles are accounted for by those journals. Remember that all journals in this set have been accused of being possibly predatory, pretty much in it for the money—even though 162 of them have no APC at all. I find the volume figures here particularly interesting: apparently journals charging $100 to $199 are particularly successful at attracting articles, with those charging $50 to $99 not far behind, while those charging $200 to $449 are relatively unsuccessful. I admit to some surprise at the success of journals that don't clearly state their APCs in attracting articles.

Maximum revenue

Revenue	Journals	%t	Volume	%
$1 million +	2	0.1%	10,177	3.9%
$250K-$999K	11	0.3%	29,803	11.5%
$100K-$249K	47	1.4%	37,738	14.5%
$50K-$99K	98	3.0%	35,672	13.7%
$25K-$49K	218	6.7%	44,275	17.0%
$15K-$24K	243	7.5%	26,100	10.0%
$10K-$14K	260	8.0%	22,989	8.8%
$5K-$9K	550	16.9%	27,708	10.7%
$2,500-$4,999	598	18.4%	14,291	5.5%
$1,000-$2,499	645	19.8%	8,615	3.3%
$1-$999	586	18.0%	2,634	1.0%
Subtotal	3,258		260,002	

Table 11: Maximum annual revenue, Beall journals A-D

Table 11 is speculative at best. It represents the maximum amount that a journal (with explicit APCs) *could* have yielded in any year 2011 and beyond, assuming that the journal's current APC was in force and that there were no waivers whatsoever (and that, where pages affect APCs, every paper was 10 pages long). Actual revenue for any journal would probably be considerably smaller.

The subtotal for journal count isn't the same as what you'd get by adding up all journals with explicit APCs in Table 10, because some journals with APCs didn't publish any articles.

No journal in Beall's set brought in more than $2 million in a year; the two very high earners both had APCs in the $600-$699 range.

I won't draw conclusions as to the profitability of "predatory" publishing, although I personally can't imagine bothering if the annual take was under $25,000, as it was for 88% of these journals.

Ah, but publishers in the Beall set tend to publish more than one journal—sometimes a *lot* more. So let's look at maximum annual revenues by publisher, even more speculative because the peak year is for each journal and may vary within a publisher. (A publisher may also have some journals without stated APCs and some with no APCs, in which case the volume may be higher than it should be.)

Revenue	Pubs	Volume	Percent
$7-9 mill.	3	75,436	40.0%
$1-1.7 mill.	2	14,835	7.9%
$500-$999K	7	31,667	16.8%
$250-$499K	11	23,046	12.2%
$100-$249K	9	17,505	9.3%
$50-$99K	19	12,497	6.6%
$25-$49K	18	6,816	3.6%
$10-$24K	30	5,114	2.7%
$1-$9K	25	1,518	0.8%
$66-$999	8	139	0.1%
Subtotal	132	188,573	

Table 12: Maximum annual revenue by publisher, Beall

In the first version of this table, there was a second row between $4 and $5 million, representing an astonishing 93,000+ articles—but that turned out to be the aggregate of all the journals Beall calls "independent," 276 of them in this study.

What I take from this table—speculative as it is—is that very few publishers on Beall's list are making a fortune at this, but a few may be handling fairly large sums of money while also processing fairly large numbers of articles. (For those not familiar with the usage, "K" stands for "thousand" and "$50-$99K" is short for "$50,000 to $99,999.")

That completes the overall tables. Let's look at the remaining groups, A through E, with tables and discussion as appropriate.

A: Apparently good

APC	Journals	Peak	Volume	Percent
$600-$999	11	5,317	13,527	24.5%
$450-$599	70	6,628	15,984	29.0%
$300-$449	55	3,557	9,145	16.6%
$200-$299	107	2,509	4,845	8.8%
$100-$199	61	3,471	8,054	14.6%
$50-$99	9	800	2,051	3.7%
$1-$49	1	131	292	0.5%
None	36	639	1,271	2.3%
Total	350	23,052	55,169	

Table 13: Beall A, journals and articles by APC

Table 13 shows the distribution of journals, peak articles and total article volume by APC level. The percentage in this case is of total volume. Since it's not possible to be A with an APC higher than $999 or an unstated APC, those rows don't appear. It is interesting that most articles appear in journals with moderately high APCs.

Revenue	Journals	Peak	Volume	Percent
$1 million +	1	2,530	5,905	11.0%
$250K-$999K	6	5,606	14,490	26.9%
$100K-$249K	5	2,581	7,006	13.0%
$50K-$99K	11	1,848	4,631	8.6%
$25K-$49K	34	2,718	6,410	11.9%
$15K-$24K	35	2,132	5,265	9.8%
$10K-$14K	31	1,204	2,923	5.4%
$5K-$9K	54	1,841	3,826	7.1%
$2,500-$4,999	57	1,204	2,127	3.9%
$1,000-$2,499	66	677	1,190	2.2%
$1-$999	14	72	125	0.2%
$0	36	639	1,271	2.4%
Total	350	23,052	55,169	

Table 14: Beall A journals and articles by revenue

The same caveat applies to Table 14 as to any other tables involving revenue: these are maximum figures, assuming today's APC and no waivers at all.

Peak	Jrnls	2014	2013	2012	2011
1,000+	4	828	2,777	5,171	5,778
600-999	4	983	2,569	2,839	2,604
300-599	5	523	1,477	1,572	1,665
100-299	17	1,136	2,245	1,976	1,354
75-99	15	363	986	900	1,020
50-74	41	888	1,945	1,555	1,471
35-49	41	693	1,103	940	937
20-34	82	942	1,684	1,062	729
10-19	68	540	825	273	163
5-9	55	262	298	15	0
1-4	18	55	23	0	0
Total	350	7,213	15,932	16,303	15,721

Table 15: Beall A journals, article distribution by peak

Table 15 shows approximate article count by year, arranged by peak article count. (There's no room for percentages in this already-crowded table.) Several items may be worth noting. For example, although four journals had at least 1,000 articles in the peak year, the four put together only exceeded 4,000 articles in 2011 and 2012—and fell considerably below the 2,000 mark for the first half of 2014 (which may or may not mean much). It's also mildly interesting that *none* of the journals with fewer than 10 articles in the peak year had any articles in 2011 (many of them didn't exist that far back). As for those with fewer than five, the lack of articles in 2011 and 2012 is definitional: It's only possible for such journals to

be in group A, B or C if they only have articles in 2013 and 2014 (since the "five or more" rule doesn't apply to the first year or the first half of 2014).

B: May need investigation

APC	Journals	Peak	Volume	Percent
$1,000-$1,999	4	177	381	0.3%
$600-$999	226	10,621	27,648	21.1%
$450-$599	93	2,695	4,960	3.8%
$300-$449	265	6,496	13,347	10.2%
$200-$299	196	5,109	9,939	7.6%
$100-$199	178	20,591	44,858	34.3%
$50-$99	96	10,244	22,012	16.8%
$1-$49	18	2,152	4,623	3.5%
None	54	1,351	3,105	2.4%
Total	1,130	59,436	130,873	

Table 16: Beall B, journals and articles by APC

These journals show a distinctly different pattern from the A group, with the sweet spot being in the $50-$199 APC range, including more than half of all articles. (You may find a small discrepancy in adding up totals in these tables and the overall table: a handful of titles with unclear APCs seemed to merit B rather than C groups, and were assigned a $0 APC.)

Revenue	Journals	Peak	Volume	Percent
$250K-$999K	4	4,973	11,890	9.3%
$100K-$249K	24	9,468	23,621	18.5%
$50K-$99K	42	9,228	21,183	16.6%
$25K-$49K	96	10,561	23,549	18.4%
$15K-$24K	92	5,953	12,305	9.6%
$10K-$14K	90	4,621	9,965	7.8%
$5K-$9K	214	7,150	14,121	11.1%
$2,500-$4,999	188	3,437	6,490	5.1%
$1,000-$2,499	198	1,967	3,560	2.8%
$1-$999	128	727	1,084	0.8%
$0	54	1,351	3,105	2.4%
Total	1,130	59,436	130,873	

Table 17: Beall B journals and articles by revenue

No B journal earned $1 million or more (with the usual caveats) in its peak year, and most journals earned very little. (As in Table 16, the $0 line includes a few journals with unclear APCs.) Note that, as in Table 14, "Percent" is of article volume.

Peak	Jrnls	2014	2013	2012	2011
1,000+	6	5,118	7,970	5,992	1,556
600-999	10	4,684	6,330	3,571	2,035
300-599	17	3,708	5,308	2,884	1,420
100-299	88	6,860	11,545	8,504	5,107
75-99	40	1,726	3,095	2,170	1,062
50-74	94	2,580	4,732	3,792	1,795
35-49	104	1,983	3,696	2,451	901
20-34	184	2,108	4,089	2,680	905
10-19	245	2,101	2,758	899	146
5-9	214	1,091	855	160	19
1-4	128	354	133	0	0
Total	1130	32,313	50,511	33,103	14,946

Table 18: Beall B journals, article distribution by peak

Table 18 shows (approximate) annual article distribution arranged by peak year. This time around, 2013 is the only year in which all six journals with peaks of at least 1,000 could have published at least 1,000 articles. Among this group, a few of the journals with five to nine articles did appear in both 2011 and 2012; as with A journals, that's not possible for those with one to four, as it would drop them into D.

Remember that "2014" really means the first half of the year—worth noting since, for B journals as grouped here, there are always more articles in the first half of 2014 than in all of 2011.

Looking at journal titles for journals with very few articles—within the Beall set, that is—it's easy to assert that most of these journals just aren't making it. A journal with a scope as broad as archaeology, agriculture, statistics and probability, social science or surgery should certainly be able to attract 20 or more articles in its second or third year unless it's perceived as damaged goods—and it's fair to assume that a great many of these journals *are* perceived as damaged goods, with or without Beall's "predatory" label. When we get to the DOAJ set, we'll see examples where publishing as few as five to nine articles per year—or even fewer—may be more reasonable.

C: Highly questionable

In my opinion, these journals *are* pretty clearly damaged goods—places where most knowledgeable authors would not want their work to appear. (That may be true for some of the A, B and D journals as well; I tend to be an easy grader.) Clearly, thousands of scholars think otherwise or are willingly submitting articles to questionable journals.

APC	Jrnls	Peak	Volume	Percent
$2,000+	18	630	1,572	1.3%
$1,000-$1,999	84	2,930	6,770	5.7%
$600-$999	223	6,326	13,600	11.4%
$450-$599	125	3,361	7,107	6.0%
$300-$449	127	3,481	7,215	6.0%
$200-$299	118	3,105	7,081	5.9%
$100-$199	24	5,376	10,516	8.8%
$50-$99	62	4,905	9,344	7.8%
$1-$49	7	1,050	1,657	1.4%
None	27	977	2,236	1.9%
Unknown	242	20,875	52,195	43.8%
Total	1,057	53,016	119,293	

Table 20: Beall C, journals and articles by APC

I find Table 20 sad in several respects. First, the 18 journals charging $2,000 and more—although it's noteworthy that even at peak, these journals averaged a *lot* fewer than 100 articles per year. (Indeed, the only APC brackets where journals averaged more than 100 articles per year at peak are $100-$199 and $1-$49.) Second, there's that huge group of journals and even huger group of articles—nearly 44% of the total—where authors probably didn't know how much they were going to be charged.

Revenue	Journals	Peak	Volume	Percent
$1 mill +	1	1,768	4,272	6.4%
$250K-$999K	1	1,477	3,423	5.1%
$100K-$249K	18	2,963	7,111	10.6%
$50K-$99K	39	4,442	9,427	14.0%
$25K-$49K	79	6,330	13,440	20.0%
$15K-$24K	95	3,269	7,726	11.5%
$10K-$14K	92	4,272	8,151	12.1%
$5K-$9K	141	3,348	6,081	9.1%
$2,500-$4,999	157	2,078	3,399	5.1%
$1,000-$2,499	109	931	1,465	2.2%
$1-$999	56	286	367	0.5%
$0	27	977	2,236	3.3%
Subtotal	815	32,141	67,098	

Table 21: Beall C journals and articles by revenue

The last line of Table 21 is labeled "Subtotal" rather than "Total" because the numbers don't include journals with unknown APCs, since even the unlikely maximum revenue couldn't be calculated for those. It may be noteworthy that the two journals with the highest potential revenues (one at more than $1.1 million, the other at more than $400,000) do *not* have extremely high APCs; instead, they published a *lot* of articles. Very high APCs show up in the next group. But even here, the largest group of articles overall were in journals where maximum annual revenue was between $25,000 and $49,999—a significant amount, but hardly a fortune.

Peak	Jrnls	2014	2013	2012	2011
1,000+	4	2,536	5,385	4,819	2,685
600-999	8	3,383	5,639	3,848	1,862
300-599	17	3,573	5,901	4,989	3,466
100-299	71	5,523	8,781	6,447	4,769
75-99	45	1,740	3,186	2,276	1,704
50-74	66	1,953	3,076	2,208	1,590
35-49	80	1,659	2,795	1,715	998
20-34	202	2,719	4,274	2,674	1,417
10-19	274	2,371	3,092	1,144	409
5-9	207	1,108	1,003	226	45
1-4	83	234	71	0	0
Total	1,057	26,799	43,203	30,346	18,945

Table 22: Beall C journals, article distribution by peak

I'm not sure there's anything especially noteworthy in Table 22, although the patterns are interesting. Note that there's only one case in which there were more articles in 2011 than in the first half of 2014—four journals publishing more than 1,000 articles each.

D: Dormant, diminutive, dying or dead

Category	Jrnls	%	Peak	Sum	%
C: Ceased	218	16%	1,745	2,607	14%
D: Dying	121	9%	1,714	2,814	15%
E: Erratic	74	6%	1,537	2,706	15%
H: Hiatus	127	9%	2,896	5,248	29%
N: New	389	29%	1,043	1,054	6%
S: Small	410	31%	2,020	3,727	21%
Total	1,339		10,955	18,156	

Table 23: Beall D journals by category

What all of these journals have in common—indeed, the *only* thing they have in common—is that each published fewer than five articles in 2011, 2012, or 2013, or fewer than two in the first half of 2014 (the last enforced only for journals that either publish as continuous article streams or have more than two issues per year). Otherwise, they're a hodgepodge. A few notes:

- **C: Ceased**. 75 of these were formally discontinued by their publishers, merged into an-

other journal or stopped accepting submissions. With a couple of exceptions, "publishers" should be "publisher": they're all from a single publisher. The rest have no articles in 2013 or 2014, leading me to conclude that they're goners. Note that this category also includes a few explicitly discontinued journals that *never* published an article, and would otherwise be in E.

- **D**: **Dying**. Based entirely on publication pattern and I was gentle in applying this category (I could have included at least a couple of dozen more). To give some sense of that pattern, here are the total articles for these 121 journals: 715 in 2011; 1,196 in 2012; 776 in 2013—and only 127 in the first half of 2014.
- E: **Erratic**. I can't think of any other way to describe these journals. Annual article numbers bounce all over the place, but bounced down below the limit in at least one year. So, for example, one journal that started in 2011 had 232 articles in the first half of 2014, 95 in 2013, 22 in 2011…and only four in 2012. Another, with seven in 2014, 31 in 2012 and 18 in 2011, didn't have any articles at all in 2013. Perhaps the most mysterious: a journal that began in 2009, published 106 articles in 2011 and a reasonable 12 in the first half of 2014—but nothing in either 2012 or 2013.
- **H**: **Hiatus**. None of these journals has any articles in the first half of 2014—but the pattern in prior years isn't the usual "dying" pattern, although, realistically, all of these could be lumped into D. Some may have suddenly collapsed; others may need new editors or editorial boards. This category is a way of suggesting that some of these *might* come back.
- **N**: **New**. With a few exceptions, these are journals that have either only a single article in the first half of 2014 or a handful in 2013 and one or none in 2014.
- **S**: **Small**. Never more than nine articles in a year and a publishing pattern that suggests it's either a niche journal—or it's not making it. For journals in the Beall set, the latter is more likely to be the case: it's hard to see how journals on preventive medicine, molecular science, veterinary science or psychiatry could fail to attract at least 10 articles a year unless something's very wrong. It's fair to note that 89 of these come from the same notorious publisher as 70-odd of the Ceased journals (in all, 202 of that publisher's journals fall into the D group).

APC	Journals	Peak	Volume	Percent
$2,000+	14	98	98	0.5%
$1,000-$1,999	11	213	356	2.0%
$600-$999	248	1,932	3,244	17.9%
$450-$599	199	1,763	2,519	13.9%
$300-$449	288	1,418	2,194	12.1%
$200-$299	204	1,221	1,873	10.3%
$100-$199	78	947	1,888	10.4%
$50-$99	25	617	1,173	6.5%
$1-$49	20	71	129	0.7%
None	53	432	691	3.8%
Unknown	199	2,243	3,991	22.0%
Total	1,339	10,955	18,156	

Table 24: Beall D, journals and articles by APC

Possibly worth noting in Table 24: All 14 of the journals with more than $2,000 APC are from the same publisher, and all have articles in 2013—but *only* in 2013.

Revenue	Journals	Peak	Volume	Percent
$50K-$99K	6	381	431	3.0%
$25K-$49K	9	483	876	6.2%
$15K-$24K	21	450	804	5.7%
$10K-$14K	47	1,035	1,950	13.8%
$5K-$9K	141	2,180	3,680	26.0%
$2,500-$4,999	196	1,468	2,275	16.1%
$1,000-$2,499	272	1,524	2,400	16.9%
$1-$999	388	759	1,058	7.5%
$0	60	432	691	4.9%
Subtotal	1,140	8,712	14,165	

Table 25: Beall D, journals and articles by revenue

Not at all surprisingly, none of these journals could have gained $100,000 or more in potential revenue, even though a few had reasonably high article counts in some years. The biggest category is journals that might have earned at least $1 (actually at least $20) but no more than $999, a level that's barely worth mentioning.

Peak	Jrnls	2014	2013	2012	2011
100-299	7	244	391	751	729
75-99	4	0	109	307	156
50-74	10	19	285	403	367
35-49	15	23	295	461	440
20-34	65	143	801	965	875
10-19	209	286	1,690	1,330	1,061
5-9	372	319	1,531	1,149	908
0-4	657	288	950	483	397
Total	1,339	1,322	6,052	5,849	4,933

Table 26: Beall D journals, article distribution by peak

Table 26, not surprisingly, lacks the top three rows—none of these journals published 300 or more articles in any year, although that's certainly possible for erratic, dying and hiatus cases. (Two journals had more than 200 articles in at least one year, one erratic and one quite possibly shutting down.)

E: Empty

There's not a lot to say about this astonishingly large set of "journals," more than half of which come from four "publishers." They have no articles later than 2010, in almost all cases none at all, so groupings by revenue or peak articles are ridiculous. In practice, when a publisher clearly had loads of empty journals, I didn't go looking for an APC for each one, so the table that follows includes only those cases where I *did* see an APC or an explicit "no APC" statement.

APC	Journals	Percent
$2,000+	24	2.3%
$1,000-$1,999	33	3.2%
$600-$999	38	3.7%
$450-$599	255	24.9%
$300-$449	473	46.1%
$200-$299	70	6.8%
$100-$199	27	2.6%
$50-$99	20	1.9%
$1-$49	53	5.2%
None	33	3.2%
Subtotal	1,026	

Table 27: Beall E journals, distribution by APC (partial)

There it is. Just under half of the "journals" with no articles at all wanted $300-$449 per article, with the next largest group asking $450-$599. As you might expect, most of those "journals" come from a handful of "publishers," with six "publishers" representing a majority of the "journals."

Sorry for all the scare quotes—but in most cases, these never were journals. They're web pages (frequently not even separate sites, just pages on the publisher's domain), frequently populated with template text, tossed out in the hope that some author might actually submit a paper, although in some cases the web pages are so defective that I can't imagine how that could happen. The pages hadn't disappeared by the time I checked them during July-September 2014; I wouldn't be surprised if several hundred were gone by this time next year.

Predatory journals with outrageous charges

I keep seeing phrases similar to that, frequently as part of an overall attack on gold OA (or OA in general), and it leads me to add one more table for the Beall set. Not that I agree that all Beall journals are predatory (I don't), but if they are, just how effective are they? I address that to some extent in Tables 11 and 12, but let's look at it another way: To what extent are robust journals with high APCs ripping off authors?

I can't answer that question, but I can provide some estimates. Considering all the journals that published more than 100 articles in some year from 2011 through 2014 as being robust, let's see where they stand in terms of APCs and *total* maximum revenue from 2011 through mid-2014 (not the same as peak-year revenue in any other table).

APC	Jrnls	Volume	Revenue
$900-$999	11	5,735.00	$1,915,802
$800	12	5,344.00	$1,539,200
$600-$700	11	19,532.00	$4,870,120
$500-$550	16	14,276.00	$3,372,900
$300-$450	17	12,576.00	$1,685,298
$200-$285	24	8,981.00	$812,861
$160-$194	21	17,643.00	$1,310,907
$120-$155	19	17,473.00	$1,052,917
$100-$110	20	16,828.00	$855,315
$80-$95	17	11,102.00	$438,580
$55-$78	20	13,930.00	$424,211
$16-$50	10	8,131.00	$152,475
$0	1	422.00	
Unknown	50	41,837.00	
Subtotal	249	193,810.00	$16,514,784

Table 28: Journals with peak volume 100 or more

Table 28 shows the results, breaking down APCs in more detail than usual (but only showing

APCs that actually exist for such journals—e.g., *none* of them charges $1,000 or more, and all of the ones between $701 and $899 are at $800 exactly). I will note that the biggest chunk of 3.5-year maximum revenue among known APCs, $600-$700, is two journals that, between them, published approximately 10,100 articles during this period.

And that's it for the Beall set, unquestionably heavier on phantom "journals" and questionable publishers and journals than the open access field in general. It includes a few hundred journals that seem OK; slightly more than a thousand that might also be OK but need investigation; another thousand or so that look pretty sketchy—and more than 1,300 that are either dying, erratic or otherwise odd. Relatively few of these are in the current *DOAJ*; even fewer, I suspect, will be in *DOAJ* once the new rules are fully in place.

The OASPA Set

Here's what I had to say about the set of journals from members of the Open Access Serial Publishers Association (OASPA) in July 2014:

> The OASPA membership list serves as a *convenient* control group of OA publishers and independent journals, but I don't believe it's especially representative of Gold OA in general or of what's in *DOAJ*. It is, however, the only control group I could think of, so it will have to do.
>
> Why isn't it representative? Mostly because OASPA includes a number of publishers who've chosen another route to making OA profitable: performing the necessary duties and charging very high APCs. Some of those publishers also publish subscription journals. It's interesting that one such publisher calls its $1,750-$1,950 fees "competitive" and that Frontiers, with fees as high as 2000 Euros, finds it advantageous to at least indirectly denigrate OA publishers that *don't* charge high fees: "Like most other serious open-access publishers, Frontiers maintains our high quality of service through an 'author-pay' model." Since most OA publishers do *not* charge APCs, Frontiers is saying most OA publishers aren't "serious."

The last sentence in the first paragraph is no longer true, and the second part of this report will—in part—focus on that new "control group," more than 3,300 journals listed in the *Directory of Open Access Journals* (in May 2014) that aren't in one of these other two sets, have enough English so I could determine APCs and count articles—and, perhaps crucially, aren't in the big "money subjects" for APC-based OA journals, namely medicine, biology, biochem and biomedicine.

Meanwhile, the OASPA set is interesting to compare and contrast with the Beall set. This time around, the OASPA set includes journals from MDPI, which were included with the Beall set in the July article. The same exclusions were done as for Beall: namely, all journals marked as hybrid, not open access or unreachable—but there weren't many of those.

Group	Count	%All	%A-E
A	611	44.3%	45.3%
B	460	33.4%	34.1%
C	8	0.6%	0.6%
D	229	16.6%	17.0%
E	42	3.0%	3.1%
O	26	1.9%	
X	2	0.1%	
Total	1,378		1,350

Table 29. Journals in OASPA set by group

Table 29 is similar to Table 4, but with one extra column, showing journals in groups A-E as percentages of journals actually studied here (that is, excluding groups H, N, O and X).

It's easy to think of the OASPA set as being just over one-fifth as large as the Beall set—but I think it's more realistic to say that it's just over *one-third* as large: there are 1,308 journals with articles included in this study, as compared to 3,876 in groups A-D for Beall. It's also realistic to say that OASPA is a bigger force in OA publishing than the Beall set: as you'll see comparing Table 31 to Table 9, the OASPA journals published *more* articles during the study period than did the Beall journals.

O: Opaque or obscure

As with other sets, I wanted to count articles by year where that seemed feasible. There were 26 journals in the Beall set where it didn't seem feasible, all from the same publisher, all with archives that appear to be in random order or some order I couldn't figure out. For what it's worth, Table 30 shows the APCs for those 26 journals.

APC	Journals	Percent
$620-$940	9	34.6%
$410-$570	6	23.1%
None	11	42.3%
Total	26	

Table 30. APCs for OASPA journals in group O

Oddly enough, this isn't a case where—given a single publisher—I could use a couple more lines and give a single APC on each line: in all, these 26 journals have 13 different APCs (including $0).

X: Unreachable

Two journals, two publishers (one with only the one title), both yielding 404s. Nothing more to say, except to express pleasure at the very low level of H, N, O and X among the OASPA set.

Note that, as with Beall, group E (which is, fortunately, only about 2% as large as in the Beall set) is included among journals actually studied—but doesn't show up in any of the next four tables, since there were no articles to count.

Peak article count, groups A-D

Peak	Journal	Percent	Volume	Percent
1,000+	10	0.8%	124,654	37.3%
600-999	13	1.0%	28,590	8.6%
300-599	40	3.1%	46,072	13.8%
100-299	141	10.8%	62,881	18.8%
75-99	64	4.9%	13,603	4.1%
50-74	134	10.2%	20,659	6.2%
35-49	149	11.4%	14,986	4.5%
20-34	239	18.3%	14,235	4.3%
10-19	203	15.5%	5,786	1.7%
5-9	155	11.9%	1,934	0.6%
1-4	116	8.9%	467	0.1%
None	44	3.4%	0	0.0%
Total	1,308		333,867	

Table 31. Peak articles in OASPA journals, groups A-D

If you're at all familiar with OASPA, you know what's behind that anomalous "volume" figure in the 1,000+ row: *PLOS One*, which not only published more than 1,000 articles in each of those years, it published more than *10,000*—including 16,000 in the first half of 2014—and, according to the journal's own search engine, 31,509 in 2013. That megajournal alone accounts for 84,718 of the 124,654 articles from those 14 journals. (Two other journals published more than 2,000 articles in the peak year, in both cases actually in the first half of 2014.) The ten busiest journals account for more than a third of all articles published by OASPA publishers, but if you remove *PLOS One* that comes down to just under one-fifth, not a lot higher than for the Beall set.

The large "none" count is easy to explain: one OASPA publisher, BioMed Central, has been diligent in ceasing attempted journals that weren't getting any traction. (There's one other journal in that group, a case where it was reborn as new OA journals.)

By way of comparison with the Beall set, more than 15% of the OASPA journals published 100 articles or more in their best years and just over half of those publishing any articles published fewer than 20 in their busiest years. The group publishing at least 100 articles per year accounts for more than three-quarters of all articles published, but again that's heavily influenced by *PLOS One*.

Article Processing Charges

APC	Jrnls	Percent	Volume	Percent
$2,000+	251	19.8%	101,212	30.3%
$1,000-$1,999	252	19.9%	173,991	52.1%
$600-$999	115	9.1%	14,674	4.4%
$450-$599	23	1.8%	2,359	0.7%
$300-$449	69	5.5%	9,993	3.0%
$200-$299	8	0.6%	514	0.2%
$100-$199	5	0.4%	1,663	0.5%
$50-$99	3	0.2%	1,075	0.3%
None	532	42.0%	27,817	8.3%
Unknown	8	0.6%	569	0.2%
Total	1,266		333,867	

Table 32. APCs for OASPA journals, groups A-D

It may *not* be surprising that so many OASPA journals have very high APCs, given the nature of OASPA itself, but it's nonetheless startling that nearly four out of ten journals and fully 82% of articles are in journals charging $1,000 or more, with roughly one out of five charging $2,000 or more. (220 of those 251 are from two publishers, BioMed Central and Frontiers.)

On the other hand, given the nature of this group, it's gratifying to see that more than 42% of the journals actually publishing articles *don't* charge APCs—oddly enough, precisely ten times as high a percentage as for the Beall set (4.2%). While 325 of those 532 freebies (where "free" must usually be said with "for now" added) come from a single publisher (Hindawi), there are quite a few others—including all of those from the University of Pittsburgh and quite a few from MDPI. In general, those journals have fewer articles than most, and for some commercial publishers there's probably a correlation between "no APC" and "not very many articles." Still, OASPA journals with no APCs published more than five times as many articles as Beall journals with no APCS between 2011 and June 30, 2014—a reasonable comparison, since OASPA journals as a whole pub-

lished slightly more articles than Beall journals did during that period.

Revenue	Jrnls	Percent	Volume	Percent
$1 million +	26	3.6%	154,748	50.7%
$250K-$999K	113	15.7%	83,983	27.5%
$100K-$249K	114	15.8%	29,216	9.6%
$50K-$99K	133	18.4%	16,709	5.5%
$25K-$49K	112	15.5%	11,080	3.6%
$15K-$24K	74	10.3%	5,292	1.7%
$10K-$14K	50	6.9%	2,622	0.9%
$5K-$9K	48	6.7%	1,199	0.4%
$2,500-$4,999	32	4.4%	483	0.2%
$1,000-$2,499	10	1.4%	130	0.0%
$1-$999	9	1.2%	29	0.0%
Subtotal	721		305,491	

Table 33. Maximum annual revenue, OASPA A-D

As everywhere else mentioning revenue, these brackets represent speculative maximum amounts: the amount that a journal would have received if, during its peak year (or half-year) of publishing, it used its current APC and had *no* waivers. (And every paper was ten pages long, where pagination matters.) Actual revenues may be considerably smaller.

You already know the name of the journal with revenues potentially exceeding $4 million and, indeed, $40 million in its peak year. Two others had potential maximum annual revenue between $3 and $4 million, with seven more in the $2 to $2.99 million range. I leave you to draw your own comparisons between Table 33 and Table 11.

Revenue	Publishers	Volume	Percent
$28-$60 mill.	3	256,203	76.6%
$10-$19 mill.	2	46,605	13.9%
$2-$6 mill.	2	8,923	2.7%
$1-$1.5 mill	3	7,651	2.3%
$500-$999K	1	1,877	0.6%
$250-$499K	2	2,840	0.8%
$100-$249K	3	2,082	0.6%
$50-$99K	4	1,866	0.6%
$25-$49K	1	789	0.2%
$10-$24K	1	464	0.1%
$1-$9K	1	49	0.0%
Subtotal	38	329,349	

Table 34. Maximum annual revenue by publisher, OASPA

Table 34 doesn't include publishers with no known revenue from APCs during this period, and as always it's highly speculative. It's also so wildly different from Table 12 that it's almost pointless to discuss the differences. (I kept the lower rows as in Table 12, which is why there are so many single-publisher rows.) It is interesting, at the very least, that the three biggest OA publishers by potential revenue published more than three-quarters of all articles published with APCs during the 3.5-year period.

A: Apparently good

APC	Journals	Peak	Volume	Percent
$600-$999	92	6,233	14,270	27.5%
$450-$599	21	982	2,283	4.4%
$300-$449	65	3,973	9,914	19.1%
$200-$299	6	192	485	0.9%
$100-$199	3	670	1,643	3.2%
$50-$99	3	552	1,075	2.1%
None	421	10,377	22,231	42.8%
Total	611	22,979	51,901	

Table 35. OASPA A, journals and articles by APC

The most obvious thing that stands out in Table 35 is the high percentage of journals (nearly two-thirds) and articles (more than 40%) without current APCs. It's sort of a barbell-shaped table, as the second highest number of journals and titles is the highest APC range for A journals—as many journals as the next three ranges combined (and more articles than the next three ranges combined). Since there were no OASPA journals with APCs in the $1-$49 range, that row is omitted.

Revenue	Jrnls	Peak	Volume	Percent
$100K-$249K	11	2,829	6,994	13.5%
$50K-$99K	30	2,758	6,660	12.8%
$25K-$49K	51	3,639	8,231	15.9%
$15K-$24K	36	1,846	4,250	8.2%
$10K-$14K	30	1,035	2,313	4.5%
$5K-$9K	14	315	818	1.6%
$2,500-$4,999	15	155	346	0.7%
$1,000-$2,499	2	23	56	0.1%
$1-$999	1	2	2	0.0%
$0	421	10,377	22,231	42.8%
Total	611	22,979	51,901	

Table 36. OASPA A journals and articles by revenue

The most interesting thing in Table 36 may be what's missing: the top two rows (max. revenue $250K-$999K and $1 million+). (How does a journal with only two articles end up in group A? The journal's brand-new, with both articles in the first half of 2014.)

Peak	Jrnls	2014	2013	2012	2011
600-999	1	640	483	281	159
300-599	3	722	1,091	912	632
100-299	35	1,626	3,027	3,127	2,310
75-99	29	760	1,390	2,323	1,280
50-74	71	1,551	2,757	3,800	2,276
35-49	86	1,340	2,347	2,551	1,897
20-34	134	1,637	2,746	2,204	1,396
10-19	128	1,016	1,356	730	503
5-9	76	369	334	96	68
1-4	48	125	39	0	0
Total	611	9,786	15,570	16,024	10,521

Table 37. OASPA A journals, article distribution by peak

It's a bit striking that *none* of these journals published more than 1,000 articles in any year from 2011 through 2013 (and, as should be clear, the single journal with more than 600 [640, to be precise] did that in the first half of 2014 rather than in any full year).

To some extent, these tables may be warped by my assertion that APCs of $1,000 or more need more explanation. Many (perhaps most) of the journals in group B appear to be good journals other than the high APCs.

B: May need investigation

APC	Jrnls	Peak	Volume	Percent
$2,000+	247	36,151	101,088	36.6%
$1,000-$1,999	191	63,518	173,012	62.6%
$600-$999	4	117	320	0.1%
$450-$599	1	42	63	0.0%
$100-$199	1	5	7	0.0%
None	16	736	1,722	0.6%
Total	460	100,569	276,212	

Table 38. OASPA B, journals and articles by APC

Did I say "perhaps most" above? How about "nearly all"? More than 95% of the B journals, including more than 99% of the articles, have APCs of $1,000 or more. (Most of the "None" are explicitly sponsored or society/university publications that have other mild issues.)

Perhaps noteworthy: while there are more journals charging more than $2,000 than there are charging $1,000 to $1,999, the journals with (slightly) lower APCs publish *considerably* more articles. (Or perhaps not: Take out *PLOS One* and the most expensive journals show a larger total volume than the next group.)

Revenue	Jrnls	Peak	Volume	Percent
$1 mill +	26	56,064	154,748	56.0%
$250K-$999K	113	30,099	83,983	30.4%
$100K-$249K	103	8,276	22,222	8.0%
$50K-$99K	101	3,822	9,921	3.6%
$25K-$49K	56	1,141	2,699	1.0%
$15K-$24K	24	296	676	0.2%
$10K-$14K	10	77	145	0.1%
$5K-$9K	7	41	65	0.0%
$2,500-$4,999	3	12	24	0.0%
$1-$999	1	5	7	0.0%
$0	16	689	1,652	0.6%
Total	460	100,522	276,142	

Table 39. OASPA B journals and articles by revenue

Setting aside *PLOS One* and its potential $42.5 million revenue in its peak year (so far!), there are two journals with potential revenue between $3 and $4 million and seven between $2 and $3 million. Most (but not all) of these are medical or biology journals—but there are also a couple of math journals and one physics journal in the million dollar club.

Peak	Jrnls	2014	2013	2012	2011
1,000+	10	27,296	42,830	33,168	21,360
600-999	12	5,112	8,968	7,334	5,613
300-599	37	8,914	14,216	11,494	8,091
100-299	106	10,947	17,476	13,536	10,832
75-99	32	1,438	2,312	1,863	1,557
50-74	55	1,946	2,901	2,259	2,017
35-49	54	1,158	1,965	1,568	1,342
20-34	78	1,122	1,622	1,261	849
10-19	44	399	455	314	237
5-9	28	165	81	56	51
1-4	4	9	8	0	0
Total	460	58,506	92,834	72,853	51,949

Table 40. OASPA B journals, article distribution by peak

Table 40 says that the sweet spot for these pricey OA journals is 100 to 299 articles per year—more journals than in any other bracket and, setting aside *PLOS One*'s domination of the top row, more articles than in any other bracket. It's mildly interesting that this group as a whole published more articles in the first half of 2014 than in all of 2011, and that's true not only for the group as a whole but for six other brackets (including 1-4, where that's necessarily true).

C: Highly questionable

This is going to be an odd section, since there are only eight journals in it—six because they appeared to have APCs but didn't identify them clearly, two because they seemed to have other problems (and might very well be B). Thus, the tables will have very few lines—only those that apply.

APC	Jrnls	Peak	Volume	Percent
$1,000-$1,999	1	23	66	10.4%
None	1	9	18	2.8%
Unknown	6	193	550	86.8%
Total	8	225	634	

Table 41. OASPA C, journals and articles by APC

The tiny number of C journals publish an even tinier number of articles—about two-tenths of one percent of the overall total.

Revenue	Jrnls	Peak	Volume	Percent
$15K-$24K	1	23	66	78.6%
$0	1	9	18	21.4%
Subtotal	2	32	84	

Table 42. OASPA C journals and articles by revenue

And, finally for this little group:

Peak	Jrnls	2014	2013	2012	2011
50-74	2	53	110	114	91
20-34	3	45	57	84	49
10-19	1	1	10		
5-9	2	2	9	9	
Total	8	101	186	207	140

Table 43. OASPA C journals, article distribution by peak

Table 43 is a little more interesting, but only a little. None of these journals is very active.

D: Dormant, diminutive, dying or dead

Category	Jrnls	%	Peak	Sum	%
C: Ceased	103	45%	1,560	3,468	67%
D: Dying	10	4%	164	280	5%
E: Erratic	17	7%	219	423	8%
H: Hiatus	4	2%	58	120	2%
N: New	18	8%	45	45	1%
S: Small	77	34%	389	854	16%
Total	229		2,435	5,190	

Table 44. OASPA D journals by category

Compared to the Beall D journals, this is a very clean group: other than ceased and small journals, there are so few as to be almost not worth noting.

C: Ceased. Most of the ceased journals—97 of the 103—are explicitly ceased, discontinued or, for 45 of them, merged into a single new OA journal. A few simply disappeared, with no articles after 2012 (and very few before then). BioMed Central is diligent in ceasing journals that never really published anything.

D: Dying. The ten journals flagged as dying all seem to be pretty clear cases. The total articles per year for that group: 69 in 2011, 131 in 2012, 72 in 2013—and eight in the first half of 2014.

E: Erratic. None of the erratic journals had the sheer drama of Beall's extreme cases; most had fewer than 20 articles even in the best years, with the biggest exception a journal going back to 2005 with three articles in 2011, 35 in 2012, 23 in 2013 and four in the first half of 2014.

H: Hiatus. Similarly, none of the journals that may be on hiatus published very many articles, with only one ever exceeding 20 articles in a year and one that can just as well be categorized as small (a mostly-student annual with four, three and two articles respectively in 2011, 2012 and 2013, and none yet for 2014).

N: New. Ten of the 18 journals categorized as New come from a single publisher, Hindawi, and all of them are free for now.

S: Small. Some of these may be legitimately very narrow niches, although most are in medical fields that would seem to justify more articles. There may be a reason: 46 of the 77 come from a single publisher (with another 14 journals in other D categories), Libertas. Some, such as three from Living Reviews, four from Ubiquity Press and seven from the University Library System at University of Pittsburgh, do indeed appear to be niche journals that may be healthy for their niches, with titles such as *Present Pasts*, *Pittsburgh Papers on the European Union* and *Living Reviews in European Governance*.

APC	Journals	Peak	Volume	Percent
$2,000+	3	37	61	1.2%
$1,000-$1,999	61	459	976	18.8%
$600-$999	19	52	84	1.6%
$450-$599	1	5	13	0.3%
$300-$449	4	39	79	1.5%
$200-$299	2	17	29	0.6%
$100-$199	1	9	13	0.3%
None	94	1,798	3,916	75.5%
Unknown	44	19	19	0.4%
Total	229	2,435	5,190	

Table 45. OASPA D, journals and articles by APC

Three-quarters of the articles appeared in journals with no APCs, far more per journal than the

second largest group of journals ($1,000-$1,999). All but two of the "Unknown" journals are ceased journals from BioMed Central; I may not have bothered to check for APCs with these journals, none of which had any articles later than 2010. The peak and volume numbers come from the two others, one of which should have an APC but doesn't say what it is, the other of which has a site that manages to hide the APC too well.

Revenue	Jrnls	Peak	Volume	Percent
$50K-$99K	2	62	128	2.5%
$25K-$49K	5	90	150	2.9%
$15K-$24K	13	150	300	5.8%
$10K-$14K	10	67	164	3.2%
$5K-$9K	27	141	316	6.1%
$2,500-$4,999	14	62	113	2.2%
$1,000-$2,499	11	39	74	1.4%
$1-$999	7	7	10	0.2%
$0	96	1,798	3,916	75.7%
Subtotal	185	2,416	5,171	

Table 46. OASPA D journals and articles by revenue

There's little in Table 46 that isn't fairly implicit from Table 45: relatively few of these journals had any maximum revenue at all, and almost none had enough to jump for joy over. (The difference in journal count between $0 here and None in Table 45 is easy to explain: two journals with APCs but with no articles—both formally ceased, putting them in D rather than E.)

Peak	Jrnls	2014	2013	2012	2011
75-99	3	91	160	226	203
50-74	6	99	192	234	259
35-49	9	72	230	315	201
20-34	24	133	430	452	148
10-19	30	90	220	255	200
5-9	50	120	254	178	144
0-4	107	57	99	61	67
Total	229	662	1,585	1,721	1,222

Table 47. OASPA D journals, article distribution by month

Table 47 lacks the top *four* rows—none of these journals published 100 or more articles in their busiest year. (All but two of those with 35 or more articles in the peak year are among the group merged into a single journal.)

E: Empty

Not surprisingly, given the shape of OASPA journals in general, there are very few journals with no articles since 2010 that haven't been shut down formally. Table 48 includes all but two of these journals; given the lack of articles, I didn't check the other two for APCs as carefully as I might have.

APC	Journals	Percent
$2,000+	11	27.5%
$1,000-$1,999	9	22.5%
$600-$999	4	10.0%
$300-$449	1	2.5%
None	15	37.5%
Subtotal	40	

Table 48. OASPA E journals, distribution by APC (partial)

The high and the low: most of these journals without content either asked very high APCs or didn't ask for an APC at all (*when checked:* these might have had APCs in the past).

I'm inclined to believe that all of these are legitimate attempts to start journals—either early ones that faded away by 2011 or newish ones that, so far, haven't attracted articles.

Next Time

That's it for Part 1 of this two-part story. You may be able to draw additional conclusions by looking at the tables; I didn't go out of my way to find meaning when there might not be any.

The role of no-fee OA journals in these two sets is quite different. Among the Beall set, it's almost nonexistent (perhaps not surprisingly given the hobbyhorse Beall's riding), with only 4.2% of those journals that published any articles and an even lower 1.6% of the articles.

Most (not all) OASPA members are commercial publishers, so it's not surprising that the percentages of no-fee journals and articles are still much lower than what's been calculated about gold OA in general: 42% of journals (with any articles) and 8.3% of articles. Significantly, I believe, no-fee OA journals tend to be smaller journals. Still: ten times the percentage of journals and five times the percentage of articles compared to the Beall set.

I believe Part 2, featuring a large subset of *DOAJ*-listed journals that aren't included here, will be revealing in this and other areas. It may also include new tables that look at other aspects of these journals (for example, article *and journal* distribution by year for various subsets), add more comparisons among the three sets and offer some additional commentary. As already noted, I also plan to discuss the possible *in*efficiencies of scale when it comes to

publishing some kinds of OA journals—and maybe offer some comments about what's good and what's possibly less than ideal about the new *DOAJ* rules.

If you believe this work is worthwhile, I encourage you to donate to *Cites & Insights*. Comments are, of course, welcome, to waltcrawford at gmail.com or on the various social media I (in)frequent.

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.

Yes, this issue has a half-blank last page. I could certainly find ways to add another few hundred words or so to fill it. I didn't choose to do so.

I will add one note for those surprised that there are no graphs in this report. I tried, and Excel will draw a graph with 3,600 or 6,700 data points (although it grumps at you when you change things on such a graph). But I ran into two problems, one of which convinced me *not* to use graphs:

- Word makes it difficult to place a full-page (42 pica) graph *with a caption* in a fixed position in a two-column document. Not impossible—I could create text boxes containing both the graph and the caption—but cumbersome. That didn't stop me: it would be a nuisance but not a major one.
- All of the graphs I tried turned out to be extreme cases of power law graphs—mostly curves that looked like somebody was trying to replicate the axes but got a little sloppy near the crossing. Even changing to logarithmic vertical scales didn't help much: it was still a steep vertical line just starting to move as it neared the bottom. I could not convince myself that any of the graphs offered any useful information. That *did* stop me: I couldn't see the point of adding graphs that didn't show much of anything.

I like meaningful graphs. So far, at least, I've been unable to determine what graphs might be both meaningful and workable for this set of data.

Masthead

Cites & Insights: Crawford at Large, Volume 14, Number 10, Whole # 178, 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/civ14i10.pdf