

Article

The DOAJ Spring Cleaning 2016 and What Was Removed—Tragic Loss or Good Riddance?

Jan Erik Frantsevåg 

The University Library of Tromsø, UiT The Arctic University of Norway, NO-9037 TROMSØ, Norway;
jan.e.frantsevag@uit.no

Received: 22 January 2019; Accepted: 28 May 2019; Published: 27 June 2019



Abstract: In December 2012, DOAJ’s (The Directory of Open Access Journals) parent company, IS4OA, announced they would introduce new criteria for inclusion in DOAJ and that DOAJ would collect vastly more information from journals as part of the accreditation process—journals already included would need to reapply in order to be kept in the registry. My working hypothesis was that the journals removed from DOAJ on May 9th 2016 would chiefly be journals from small publishers (mostly single journal publishers) and that DOAJ journal metadata information would reveal that they were journals with a lower level of publishing competence than those that would remain in the DOAJ. Among indicators of publishing competence could be the use of APCs (Article Processing Charges), permanent article identifiers, journal licenses, article level metadata deposited with DOAJ, archiving policy/solutions and/or having a policy in SHERPA/RoMEO, the database containing self-archiving policies for more than 30,000 journals. The analysis shows my concerns to be correct.

Keywords: open access; DOAJ; publisher size

1. Introduction

In December 2012, DOAJ’s (The Directory of Open Access Journals) parent company, IS4OA, announced they would introduce new criteria for inclusion in DOAJ [1] and that DOAJ would collect vastly more information from journals as part of the accreditation process—journals already included would need to reapply in order to be kept in the registry. Those new criteria were launched on 19 March 2014 [2] and the deadline for re-application was set to 31 December 2015 [3], later extended to 1 April 2016 [4]. On 9 May 2016, journals in DOAJ that had not re-applied for inclusion were removed from DOAJ.

The working hypothesis was that the journals removed from DOAJ on 9 May 2016 would chiefly be journals from small publishers (mostly single journal publishers) and that metadata information would show that they were journals with a lower level of publishing competence than those that would remain in the DOAJ. Publishing competence is not easy to measure, but by looking at how the journals “score” when it comes to some indicators, one can get a picture. This study only looks at the technical side of publishing, not at the editorial quality and the scholarly quality of the content. The technical quality concerns the journals ability to produce articles that satisfy general norms for articles when it comes to layout and design, and to disseminate these widely and efficiently. Among indicators of publishing competence could be the use of APCs (Article Processing Charges), permanent article identifiers, journal licenses, article level metadata deposited with DOAJ, archiving policy/solutions and/or having a policy in SHERPA/RoMEO, the database containing self-archiving policies for more than 30,000 journals. Data on a number of such aspects of publishing quality are only available in the data-set for journals accepted after March 2014; thus, they cannot easily be used for analysis in this context. An example of this is the DOAJ Seal, which is a good indicator of the technical quality of a journal.

2. DOAJ as a Whitelist

Why is it important whether a journal is listed in DOAJ, or not? DOAJ is the authoritative database over which journals are OA (Open Access), and which not. It became the most authoritative of such services early after its start in 2003, due to its quality and coverage, see Morrison [5] and Bi [6]. For authors or administrators, DOAJ is a source of information on whether journals are truly OA or merely hybrid, and also on journal quality as DOAJ screens applicants for quality before admitting them. For journals, DOAJ is a tool to becoming visible, as various services—including library ones—regularly harvest DOAJ for journal and article metadata. A bona fide OA journal not listed in DOAJ will be markedly less visible to readers and authors than one that is listed, and will hence be less well-suited for publishing in for an author interested in having his/her work efficiently disseminated. At least in Norway, listing in DOAJ is normally necessary if an author wants to have costs for APCs in that journal refunded from a publication fund. In short, not being listed in DOAJ is a bad idea for an OA journal that wants readers and authors.

The implementation plan for Plan S [7] requires a listing in DOAJ to be an acceptable publishing venue for authors funded by Plan S participants. For journals wanting to attract manuscripts from such authors, being listed in DOAJ has become a *sine qua non*. A listing in DOAJ does not in itself make a journal Plan S compliant, but it is one of the mandatory criteria in the implementation plan. Furthermore, an inclusion in DOAJ, which requires the journal to follow some minimum editorial quality standards, is one of only two requirements that point to editorial quality. Other requirements address other, more technical aspects of the journals.

DOAJ functions as a whitelist, i.e., a list of journals that are “acceptable” according to certain norms. For DOAJ the important norms are that the journals fulfill certain criteria for being accepted as OA journals, and that they fulfill quality criteria that merit scholarly journal status. Journals not fulfilling these criteria satisfactorily will not be accepted. DOAJ also looks for accepted journals that over time have changed their policies so they do not fulfill the criteria any longer.

To function well as a whitelist, a list needs to be inclusive, i.e., contain as many acceptable journals as possible, and to be exclusive in the sense that it does not include journals not fulfilling the criteria. This is a difficult balancing act, and a strong focus of avoiding the one error will likely increase the risk of making the other error. Admitting a journal that does not really satisfy the criteria could be called a false positive, rejecting a journal that in reality satisfy the criteria could be termed a false negative. These and other important aspects are widely discussed in [8].

Having strict criteria for inclusion—(which I think one can agree that DOAJ has [9] also given that a listing in DOAJ is a requirement for Plan S eligibility [7]) and looking for policy changes or changes in practices that conflict with inclusion criteria should ensure that the number of false positives, i.e., journals included without meriting it, is kept low. DOAJ publishes a list of journals removed for various reasons, see <https://doaj.org/faq#list>.

A recent study [10] indicates less than half of scholarly OA journals in the Nordic countries are listed in DOAJ, without analyzing the reasons for not registering with DOAJ. If these numbers hold for the rest of the world, DOAJ need to do something about their coverage in order to serve efficiently as a whitelist. As Plan S requires a DOAJ listing for a journal in order to be an acceptable publishing venue, this becomes increasingly important [7]. At least one national funder has earlier implemented a DOAJ requirement in order to finance APCs for authors [11]. A journal satisfying the criteria, but not included in a whitelist like DOAJ is a situation that can be remedied through action from the publisher.

A blacklist is a list of journals that are not acceptable according to certain norms. A well-known blacklist has been Beall’s list [12], which is now defunct. This list aimed to publish information about publishers and journals with questionable business practices, not about publishers not being OA in the way DOAJ defines OA, even if it all was about OA publishers and journals. Blacklists may also have false positives (unmerited inclusions in the list) or negatives (journals that should have been included, but are not). False positives in a blacklist creates problems for legitimate publishers/journals, while false negatives may result in authors submitting manuscripts to journals they should have avoided,

trusting that the blacklist was complete. Walter Crawford has a very critical view of Beall's list [13], especially regarding a lack of clear criteria for inclusion. Both whitelists and blacklists need clear criteria for inclusion and exclusion to function well.

3. The DOAJ Re-Accreditation

Others have also to some extent discussed the DOAJ re-accreditation process. An interesting paper is [14], which has a number of analysis parallel to the present manuscript. Perspectives and parts of the discussion differ from the present one, and some analysis are different. I have chosen to retain also the analyses that could seem to be duplicates for the sake of retaining the discussion of the findings, and of the somewhat different perspective here. Both papers look at the geographical distribution of removed journals, but [14] gives a more extensive overview of this. It also looks at publishers with many journals removed and removal across subject categories. It should be noted that [14] discusses all removals from DOAJ between March 2014 and May 2016, not only this specific removal on 9 May 2016 which is the subject of this study. Moreover, it discusses aspects not discussed here: the overlap between DOAJ and Beall's list; the overlap between DOAJ and Scopus and JCR; the DOAJ Seal and overlap between DOAJ and journals published by OASPA (Open Access Scholarly Publishers Association) members. The present manuscript discusses aspects not covered by [14] like publisher size, licenses, self-archiving policies present in SHERPA/RoMEO and use of APC or not. There has been research [15] that discusses the challenges facing DOAJ in the context of both the re-accreditation process and of the "Bohannon Sting". Bohannon published an article in *Science* [16] after having sent a spoof paper to a number of OA journals, some of which were listed in DOAJ. It is well worth noting that the foundations of the re-accreditation process were made public in December 2012, while Bohannon's article was published in October 2013, despite wordings in [15] that could lead one to believe it was Bohannon's activities that started it all. It is, however, reasonable to assume that the re-accreditation initiative was informed by the activities of Jeffrey Beall, starting much earlier than the decision to have a re-accreditation process, see [17], and that the Bohannon sting informed the re-accreditation requirements and made such a process urgent.

DOAJ continuously publishes journal level metadata [18] that can be downloaded in a CSV format; this is the data used for this analysis. I downloaded one file immediately before the removing of journals started (file time-stamped 9 May 2016 12:00) and one immediately after the process was finished (file time-stamped 10 May 2016 06:30). Both files are publicly available [19] in their original format. By comparing these files (after converting to Excel format to enable use of Excel tools), I could ascertain which journals in the older file were kept and which were removed during the clean-up. The older file was then used to identify publisher size and various characteristics of the journals kept and removed. Various technical aspects of the DOAJ data that could be relevant for this study are discussed in detail in [20] and will not be repeated here; how publisher size is constructed is also treated in detail there. Note the discussion there of what the term "publisher" means in DOAJ and various pitfalls concerning this. It is important to note that size is here measured in terms of journal titles published, not in terms of articles published—the latter could be a better measure, but is much more difficult to find reliable data for.

The post clean-up file contains 8791 journals; the pre clean-up contains 11,644, indicating a removal of 2853 journals, or 24.5 percent of all journals in DOAJ at that time. However, the post clean-up file contains six journals added during the removal process; thus, 2859 journals have been removed (and 8785 kept), increasing the loss to 24.6 percent. This tallies well with numbers from DOAJ's own list of removed journals [21]. That file contains 2861 entries; however, two of these are errors, with two journals still listed in DOAJ being listed as removed in the file. (I informed DOAJ of these two erroneous entries in the list of removed journals. The two journals were subsequently removed from the list of removed journals.) An important point is that journals kept have not necessarily been re-accredited—they are either re-accredited or have applied to be re-accredited. The re-accrediting work went on for quite some time after 9 May 2016. The re-accreditation project was officially ended

13 December 2017 [22]. Interestingly, the announcement notes that during the re-application process 2058 re-applications were rejected. The reason for rejection is not stated, but one can suspect editorial quality issues to be one important explanation.

The extent to which DOAJ gives a complete picture of all OA journals is also discussed in [20], where the number of journals not listed in DOAJ was found to be of minor importance. An independent survey of whether this is still the case has not been attempted in conjunction with the present study, but the removal of one quarter of all journals listed will certainly influence the reliability of DOAJ as a comprehensive source for future studies.

It should be noted that in the following we are actually discussing three different groups of journals:

- (1) 2859 journals removed from DOAJ because they did not send in a reapplication to DOAJ within the time limit. These are our focus.
- (2) 3862 journals added to DOAJ after March 2014 and journals re-accredited after March 2014.
- (3) 4923 journals that have applied for re-accreditation but have not yet had their application processed.

Journals in group (2) and (3) are below generally grouped together as journals kept in DOAJ, as opposed to (1), which are lost from DOAJ. Looking at some aspects one needs to note that published metadata are much richer and up to date for (2) than for (3), making it necessary to look at them separately in some cases.

In the discussions below, it is assumed that publishers know they need to apply for re-accreditation. This is not necessarily so. (The following information is from a private communication with Dominic Mitchell of DOAJ in the form of a comment in a manuscript version of this article, on July 14th 2016. Available from this author upon request.) There are many reasons for that; the major one being the need of having correct contact information (i.e., e-mail address) for publishers. Every publisher gets an account, and DOAJ tries to assign all journals from a publisher to that same account. That means fewer places to keep information up to date, but also means the information in such an account is important to more journals. DOAJ's impression is also that consolidating journals into a single account is easier with larger publishers than with smaller ones—single journal publishers of course will have a single account. It is the responsibility of the publisher to keep contact details updated; this does not happen to the extent one could want for—and the smaller the publisher, the greater the risk of such information not being updated. DOAJ assumes a relatively high risk that they were not able to reach out to all publishers due to this, and that the smaller publishers were the ones more likely not to be reached during the process. DOAJ notes they manually updated more than 1000 user accounts during the process.

4. Results

4.1. General Discussion

There are two types of information discussed and analyzed as to whether journals are kept or lost here, background information and information about technical aspects of the journals. The background variables looked at are publisher size, geography, language and subject, while licenses and self-archiving policy are technical aspects. APC is neither, but it has been shown to be connected to publisher competence.

4.2. Publisher Size

The pre clean-up file contained 11,644 journals, published by 6081 different publishers. A summary of publisher size, as measured by the number of journals published by that publisher, is given in the table below. Publisher size is not in itself a sign of quality and competence, but analyses show that there is a connection. Both in this study and in [23], one sees clear connections between publisher size and the ability to fulfill technical demands. Hence, it is of interest to look at publisher size and various indicators of technical competence, and—in this case—whether journals were kept or lost.

What we see in Table 1 is not very different from what was found previously [20], even if the journal share of the single journal publishers has been reduced from 55.0 percent to 43.8 percent, and the share of the largest publishers has increased from 10.8 percent to 14.1 percent.

Table 1. Pre clean-up publisher size distribution.

Publisher Size	Number of Publishers	Number of Journals	Share of Publishers	Share of Journals
1	5097	5097	83.8%	43.8%
2	444	888	7.3%	7.6%
3	186	558	3.1%	4.8%
4	96	384	1.6%	3.3%
5	49	245	0.8%	2.1%
6–10	104	753	1.7%	6.5%
11–20	64	918	1.1%	7.9%
21–50	28	771	0.5%	6.6%
51–100	6	383	0.1%	3.3%
>100	7	1647	0.1%	14.1%
Total	6081	11,644	100%	100%

If we turn to Table 2, showing the loss of journals over publisher size (size before removals), we get the following picture:

Table 2. Publisher size and loss of journals.

Number of Journals Publisher Size Group	Status Journals		Total Journals	Percent Lost
	Kept	Lost		
1	3441	1656	5097	32.5%
2	634	254	888	28.6%
3	406	152	558	27.2%
4	278	106	384	27.6%
5	185	60	245	24.5%
6–10	538	215	753	28.6%
11–20	671	247	918	26.9%
21–50	626	145	771	18.8%
51–100	381	2	383	0.5%
>100	1625	22	1647	1.3%
Total	8785	2859	11,644	24.6%

We see that the smallest publishers—the single journal publishers—lose nearly one third of journals. The publishers with a size between 2 and 20 journals lose between 20 and 30 percent of journals, on average 27.6 percent, the next category 21–50 lose 18.8 percent while the larger (>50) publishers lose a negligible fraction of their journals. The losses among the larger (>50) publishers are as follows in Table 3:

Table 3. Number of journals lost from publishers >50 journals.

Publisher	Journals Lost
Hindawi Publishing Corporation	8
De Gruyter Open	6
BioMed Central	6
Springer	2
PAGEPress Publications	1
Libertas Academica	1
Total	24

Of the seven largest publishers (>100) MDPI AG, Elsevier and Dove Medical Press lost no journal during the process.

A “partial preliminary” list of cuts by publisher has been prepared by Walt Crawford [24]. We note that Internet Scientific Publications, LLC, by having all their 46 journals removed from DOAJ is the publisher that lost the largest number of journals, and the only publisher with more than 20 journals to be removed entirely from DOAJ through this process—a fate shared with seven of 64 publishers in the 11–20 category.

A total picture of publishers that have lost all their journals through the clean-up, and thus disappear as publishers, is given in Table 4.

Table 4. Original publisher size and loss of publishers.

Number of Publishers Publisher Size Group	Status Publishers		Total Publishers	Percent Lost
	Kept	Lost		
1	3441	1656	5097	32.5%
2	374	70	444	15.8%
3	169	17	186	9.1%
4	86	10	96	10.4%
5	45	4	49	8.2%
6–10	95	9	104	8.7%
11–20	57	7	64	10.9%
21–50	27	1	28	3.6%
51–100	6		6	0.0%
>100	7		7	0.0%
Total	4307	1774	6081	29.2%

Except for the smallest category, where the removal of the one journal published also results in the removal of the publisher, percentages are smaller than in Table 2. We can clearly see that here, as with journal losses, there is a tendency that smaller publishers are more likely to disappear than larger ones. One reason is, of course, that larger publishers are more robust in the sense that they only need to retain one journal to stay in DOAJ. If journals had been removed at random, it is rather unlikely that many publishers with more than three journals would disappear, with a risk of less than 1 percent for any given publisher (The average risk of losing a journal is 24.6 percent, see Table 2. The risk of losing n journals if losses are random, is 0.246^n . The risk of losing three journals is 0.246^3 , which is 0.015 (1.5 percent), the risk of losing four journals is 0.246^4 , equal to 0.004 or 0.4 percent.) For all size groups smaller than 51, the percentage of publishers removed is much higher than follows from a random removal of journals. Hence, the removal of journals is skewed towards specific publishers, and it seems reasonable to conclude that removals are related to some aspect of the publisher.

This creates a new distribution of journals over publishers of different sizes, with an increased degree of concentration. The new distribution of publishers in Table 5 is rather similar to Table 1, but the smallest publishers have an even smaller share of journals, and the largest ones an even larger share.

The 151 publishers with six or more journals are 3.5 percent of publishers, but control 42.2 percent of journals.

The numbers seem to support the hypothesis that the journals lost were mainly published by small publishers, publishers that seem to lack the competence or the resources necessary either to understand why re-application was important, or to go through the re-application procedure. There is nothing in this process to indicate the journals lost were of a lower quality content-wise than journals kept. That the spring clean-up has little to do with the scholarly quality of the journals de-listed is however not well understood in the OA community, one example is [25], which discusses the spring clean-up with this wording: “After excluding 3000 dubious open access journals from its index [. . .]”. This totally misses the point that the journals removed were not scrutinized and found lacking in scholarly quality—they were removed because they had not asked to be re-scrutinized.

Table 5. Publisher size after the spring clean-up.

Publisher Size	Number of Publishers	Number of Journals	Share of Publishers	Share of Journals
1	3590	3590	83.4%	40.9%
2	341	682	7.9%	7.8%
3	131	393	3.0%	4.5%
4	57	228	1.3%	2.6%
5	37	185	0.9%	2.1%
6–10	70	523	1.6%	6.0%
11–20	53	732	1.2%	8.3%
21–50	15	446	0.3%	5.1%
51–100	6	381	0.1%	4.3%
>100	7	1625	0.2%	18.5%
Total	4307	8785	100.0%	100.0%

4.3. Licenses

One important point about the publishing and distribution quality of an OA journal is that it has a readily available and comprehensible user license, so that a reader knows to what extent content may or may not be re-used to various purposes.

Looking at which licenses were used by the journals lost or kept, Table 6 will give a good picture. The original content of the license field has been grouped to get a better overview. All CC-enabled journals are grouped together here, as are all journals with journal specific licenses.

Table 6. Use of licenses among kept or lost journals.

	Journal License	Application Pending	Accepted after March 2014	Total
Kept	CC license	2072	3730	5802
	Various journal-specific licenses	7	131	138
	None	2844	1	2845
Kept Total		4923	3862	8785
Lost	CC license	550		550
	Various journal-specific licenses	2		2
	None	2307		2307
Lost Total		2859		2859
Grand total		7782	3862	11,644

Now, journals accepted after March 2014 were more or less forced to give information about their licenses as part of the re-accreditation process, so we get a more relevant picture if we compare the lost journals and the journals that have an application pending, and exclude the journals accepted after March 2014. We see that among the journals that have an application pending, 42 percent (2074) have a CC license, while among those lost only 19 percent (550) had such a license. We could also note that nearly 97 percent (5802) of journals accepted after the new criteria were put in force have a CC (Creative Commons) license.

Table 7 below shows that the share of journals having a CC license or no license is closely connected to publisher size. Numbers are from before the removal, thus, the 2859 removed journals are part of the numbers.

“Various journal-specific licenses” totals 140 journals, mainly from stand-alone journals. It is difficult to imagine having such licenses is meaningful, and they will probably be harmful to distribution of content. Having no license is also harmful to distribution.

Table 7. Licenses and publisher size.

Journal License	Publisher Size Group										Total
	1	2	3	4	5	6–10	11–20	21–50	51–100	>100	
CC license	2037	413	300	199	143	390	471	541	304	1554	6352
Various journal-specific licenses	78	15	11	2	1	14	18	1			140
None	2982	460	247	183	101	349	429	229	79	93	5152
Total	5097	888	558	384	245	753	918	771	383	1647	11,644
Percentage “None”	59%	52%	44%	48%	41%	46%	47%	30%	21%	6%	44%
Percentage CC license	40%	47%	54%	52%	58%	52%	51%	70%	79%	94%	55%

4.4. Having a Self-Archiving Policy in SHERPA/RoMEO

Another aspect of publishing and distributional quality of a journal is that it has published a self-archiving policy in SHERPA/RoMEO, enabling authors and administrators to find out to which extent self-archiving is permitted, and with which restrictions. Increasingly, this is information authors need in order to ascertain if a given journal is a journal that enables them to meet various OA mandates or contractual requirements. Not having a policy in SHERPA/RoMEO is a sign of low publishing quality, as it makes life harder for the users of a journal. Not having a policy there is probably more a question of competence or resources than an active resistance to the idea. Having a policy there is, after all, free.

DOAJ data for this (there is info in the DOAJ journal metadata) are not reliable—only two out of 7782 journals accepted in DOAJ before March 2014 has any meaningful information about this. Thus, instead, DOAJ journal data was checked against a data file from SHERPA/RoMEO (RoMEO for short). This file was from 29 January 2016 so there is some risk of missing data; however, new journals and publishers are not extremely frequent in this database—one could wish for a stronger influx of such information.

We see from Table 8 that the percentage of journals remaining in DOAJ having a policy in RoMEO is 37, while among those removed only 13 percent had such a policy listing.

Table 8. Policies in SHERPA/RoMEO.

In RoMEO?	Found Post Clean-Up?		
	Kept	Lost	Total
No	5575	2496	8071
Yes	3210	363	3573
Total	8785	2859	11,644
Percentages			
No	63%	87%	69%
Yes	37%	13%	31%

As the above table shows, the situation is generally not good when it comes to publishing self-archiving policies. It should be noted that DOAJ accepts publishing such policies in other services than RoMEO. This is a debatable position—efficient OA demands centralized services—but as these other services combined only have data for a small portion of DOAJ-listed journals, I have concentrated on RoMEO.

4.5. Article Processing Charges (APC)

While it will be wrong to proclaim that the fact that a journal charges an APC is a sign of quality, the two are not quite separated. Charging APCs gives a journal a business model that allows it to pay for the use of resources, this may make it possible both to operate more efficiently and to buy, or develop internally, publishing competence. In this sense, charging APCs could be positively correlated with quality. On the other hand, dubious publishers charge APCs; their business model is to make authors part with their money without the journal actually delivering the quality assurance services

paid for—a “predatory” publisher not charging an APC would be meaningless. Our recent article [23] clearly indicates that charging an APC gives a journal a better chance of satisfying technical criteria.

The current metadata file does not contain reliable information about APCs for journals added before March 2014; only eight of those journals indicate that they use APCs. Using these metadata often, I have an archive of files from various dates and have found that a file from 7 February 2014 contains such information. Morrison et al. [26] analyses APCs but based on manually collected data from DOAJ, because the APC information was purged from the metadata files because of dubious quality. This purge must have happened soon after my having downloaded the file containing APC information, Morrison et al. describes their data gathering process as performed in May 2014. This indicates one should be cautious about the validity of the data and findings about APC here.

The file contains information about 9804 journals, matching the current metadata file with the old one leaves information about 8112 journals that were in DOAJ on 7 February 2014, which also were in DOAJ on 9 May 2016. After grouping the data (N for No and NY which probably means No [27], into No; and CON for conditional and Y for Yes into Yes, leaving four where data is lacking) we found the following, shown in Tables 9 and 10.

- Of the 8112 journals still in DOAJ before the clean-up, 71 percent had no APC, while 29 percent did.
- After the clean-up, we are left with 5276 journals, of which 68 percent have no APC while 32 percent did.
- The majority of journals removed did not charge an APC—2156 did not, 677 did.

Table 9. APC-information for journals in DOAJ both in February 2014 and in May 2016.

Journals		APCs February 2014					Grouped			
Publisher Size Group	Empty	N	NY	CON	Y	Total	No	Yes	Rest	Percent APCs
1	3	3199	86	182	477	3947	3285	659	3	17%
2		511	8	31	88	638	519	119		19%
3		274	2	17	74	367	276	91		25%
4		219	6	3	27	255	225	30		12%
5		112	5	7	43	167	117	50		30%
6–10		345	7	43	130	525	352	173		33%
11–20		435	7	63	130	635	442	193		30%
21–50	1	342	5	37	136	521	347	173	1	33%
51–100		98		1	96	195	98	97		50%
>100		90	2	51	719	862	92	770		89%
Total	4	5625	128	435	1920	8112	5753	2355	4	29%
Percentage							71%	29%		

Table 10. APC information for journals in DOAJ after the clean-up.

Journals		APCs after the Clean-Up					Grouped			
Publisher Size Group	Empty	N	NY	CON	Y	Total	No	Yes	Rest	Percent APCs
1	1	1926	17	105	250	2299	1943	355	1	15%
2	0	323	4	17	41	385	327	58	0	15%
3	0	169	0	8	40	217	169	48	0	22%
4	0	129	2	1	19	151	131	20	0	13%
5	0	70	0	1	36	107	70	37	0	35%
6–10	0	217	2	20	73	312	219	93	0	30%
11–20	0	291	1	18	79	389	292	97	0	25%
21–50	0	259	1	30	86	376	260	116	0	31%
51–100	0	97	0	1	96	194	97	97	0	50%
>100	0	87	2	51	706	846	89	757	0	89%
Total	1	3568	29	252	1426	5276	3597	1678	1	32%
Percentage							68%	32%		

In most size groups, there is a tendency towards a lower percentage of journals with APCs remaining, but as the larger publishers have more APCs and are also the size groups with a larger percentage of journals being kept, we found that in total the percentage of journals having APCs has grown as a result of the clean-up. The higher occurrence of APC-charging journals among those lost in

the clean-up in most size groups can lead one to speculate that we have lost some dubious publishers with an APC business model that were weak on quality. This could be sheer incompetence, but could also be an indication of “predatory-ness”. The numbers being small, and without looking closer into details about those lost APC-charging publishers, this is only speculation—but may be something to be looked further into at some future date.

4.6. Geography

The journals lost were not evenly distributed over countries, quite the opposite. Seven countries publishing a total of 11 journals lost them all. At the other end of the scale, 31 countries kept all their journals in DOAJ. 17 of these published only one journal, seven published two journals while seven published from three to 15 journals.

The major country (having more than 50 journals pre clean-up) that loses the largest percentage of its journals is Japan who loses 72 of 98 journals, a loss of 73.5 percent. The country losing the most journals is the United States, with a loss of 403 journals out of 1070—a 37.7 percent loss. These findings correspond well with findings in [14].

The Japanese journals are mainly published as stand-alone journals, and as we see from Table 11 these obviously have had problems with the re-accreditation process. The five journals published by larger publishers are all published by international publishers—they have managed to re-apply for accreditation. Publisher size obviously plays a role.

Table 11. Status for Japanese journals.

Publisher Size	Clean-Up Status		Total	Percentage Lost
	Kept	Lost		
1	16	69	85	81%
2	5	3	8	38%
11–20	2		2	0%
21–50	2		2	0%
>100	1		1	0%
Total	26	72	98	73%

We see from Table 12 that smaller publishers are also dominant in the US, but not over-represented among the journals lost. A closer inspection of the raw data reveals that one reason for the high losses among larger US publishers (except for the very largest ones) is that a number of mid-sized publishers are removed in their entirety. An example of this are the 46 journals lost in the 21–50 category—that is the result of one publisher with 46 journals being removed, the largest publisher to be removed entirely, irrespective of country. Of the 41 journals lost in the 11–20 category, 39 are published by three publishers now entirely removed from DOAJ.

Numbers in [28] could seem to indicate that problematic publishers are overrepresented in the US, i.e., they often have US addresses. The large losses inflicted on the mid-sized category of US publishers could be a sign that some problematic publishers that have been accredited in the DOAJ have seen reason not to try to get re-accredited. Assessing the content quality of publishers is difficult, so this is only speculation—others might try to delve further into this aspect of the losses.

A tendency I see among the smaller publishers with more than one journal is that there are many universities or comparable institutions there that have not managed to get into the re-accreditation process. I note e.g., Duke University School of Law having all their five journals, hence themselves as a publisher, removed; University of California UCL/UCLA losing eight journals; Centers for Disease Control and Prevention losing three out of four journals, etc.

Table 12. Status for US journals.

Clean-Up Status				
Publisher Size	Kept	Lost	Total	Percentage Lost
1	329	214	543	39%
2	70	44	114	39%
3	22	12	34	35%
4	33	19	52	37%
5	5	10	15	67%
6–10	38	16	54	30%
11–20	42	41	83	49%
21–50	38	46	84	55%
51–100	73		73	0%
>100	17	1	18	6%
Total	667	403	1070	38%

4.7. Language

Was there anything noteworthy about the publishing languages of the journals kept or lost? DOAJ asks journals to list languages they publish full-text in. The languages should be listed in order of importance, i.e., publishing volume. Among journals (re-)accepted into DOAJ after March 2014, only one lacks this information. Among those admitted before March 2014, 1971 out of 7782 (25.3 percent) lack this information. A bit baffling is that of the journals lacking this information, 29 percent were lost in the clean-up, while 39 percent of journals having this information were lost. Looking at other aspects, it seems that journals having information in DOAJ on various aspects of their work fared better than those lacking such information, but not when it comes to language.

Another question is whether having English as the most important language has any influence upon the risk of being removed. A more detailed study of the numbers behind the below table shows no difference between English or another first language for journals lost—both categories show a loss of 39 percent, as seen in Table 13.

Table 13. Language and removals.

Journals				
Main Language	Kept	Lost	Total	Percent Lost
English	2567	1670	4237	39%
Other than English	964	610	1574	39%
Info lacking	1392	579	1971	29%
Total	4923	2859	7782	37%

Having tried to look at other aspects regarding language, I cannot find any clear tendencies in any direction, only minor differences that are too small to merit discussion—and pointing in different directions.

4.8. Subject

Is there any connection between the scholarly field in which a journal is active, and the risk of being removed in the clean-up?

DOAJ has assigned a subject classification to each journal. Unfortunately, this is quite detailed and with a number of options, resulting in 1421 different values in the “Subjects” field in the metadata. I have tried to group these into a smaller number of broader categories, generally built on the first element in the subject. For example, I categorized the category “Philosophy. Psychology. Religion:” as Philosophy, even though Psychology could be the more important word. The appropriate category is impossible to define without a detailed analysis of the individual journals.

Looking at only the 7782 journals either removed because they did not re-apply, or that applied but was not yet re-evaluated after March 2014, we find the following in Table 14:

Table 14. Journals not re-evaluated after March 2014 by subject.

Subject	Journals			
	Kept	Lost	Total	Percent Lost
Agriculture	212	131	343	38%
Education	326	183	509	36%
Fine Arts	75	48	123	39%
General Works	267	149	416	36%
Geography	235	134	369	36%
History	151	66	217	30%
Language and Literature	333	190	523	36%
Law	75	80	155	52%
Library science	73	41	114	36%
Medicine	1068	614	1682	37%
Military Science	7	3	10	30%
Music	23	12	35	34%
Naval Science		1	1	100%
Philosophy	220	140	360	39%
Political science	110	49	159	31%
Science	831	490	1321	37%
Social Sciences	546	325	871	37%
Technology	371	202	573	35%
“Missing”		1	1	100%
All journals	4923	2859	7782	37%

Of the subjects with more than 100 journals, Law with a loss of 52 percent and History and Political Science with losses of “only” 30 and 31 percent respectively, stand out. The numbers here reflects the same tendencies as in [14], though the absolute numbers differ. The differences might be due to different sets of journals ([14] discussing a larger set) or different ways of assigning a subject category.

4.9. Other Aspects

Initially, I listed a number of aspects that could be interesting to have a look at to see if journals/publishers performing well would be more likely to remain in DOAJ, with the following so far not looked further into:

- Permanent article identifiers
- Archiving policy/solutions
- Article level metadata deposited with DOAJ

Using permanent article identifiers, such as DOI (Digital Object Identifier, a permanent and unique identifier), is a sign of publisher competence, partly because using them in a journal requires some competence. Information about permanent article identifiers (not necessarily using them) must be given in the (re-)application form; however, this information only exists for journals having been (re-)admitted since March 2014.

The same goes for archiving solution/policy, i.e., information on whether the journal has a long-time archiving solution in place, such as LOCKSS, a service providing the necessary storing and other functionality.

No information was found in the journal metadata file regarding whether a journal has deposited article level metadata with DOAJ. Depositing article level metadata with DOAJ is a cheap instrument to make content visible, as this is information harvested by and re-used in various other services. This is one of the services provided by DOAJ that is more important to smaller publishers than to the larger

ones—they have other and more mechanisms available to achieve the same effects. Despite this, the larger ones seem to be better at using these possibilities in DOAJ.

DOAJ does, however, display information about the total volume of this on their front page, and I took some screen dumps around the time of the removal.

Tabulating data from these screen dumps and adding other data documented earlier, I found the following data, detailed in Table 15.

Table 15. Article level metadata deposit and removal from DOAJ.

Date and Hour	Data from Screen Shots				Other Data	
	Journals	Searchable at Article Level	Articles	Publishers	Articles Per Journal	Articles Per Publisher
09.05.2016 11:30	11,650	7290	2,296,024	6081	197	378
10.05.2016 06:30	8795	6095	1,960,409	4307	223	455
Lost	2855	1195	335,615	1774	118	189
Percent lost	24.5%	16.4%	14.6%	29.2%		

We see that journals depositing article level metadata have a lower risk of being removed during the clean-up process, 16.4 percent compared to 24.5 percent. Furthermore, we see that those removed despite having deposited such metadata had, on average, deposited information about fewer articles than those remaining.

Again, this points to removal being connected to both publishing competence and journal/publisher size. Publishing competence as journals not depositing article level meta-data are more prone to be removed; thus, size was important, as the journals removed on average have metadata about fewer articles than those journals that were kept.

5. Summing Up and Discussion

We see from the data presented that many journals were removed as a result of not having submitted an application for re-accreditation may have problems with their publishing competence. An analysis of metadata cannot say anything about their scholarly quality and competence; however, it is important to point out that there is nothing about the process that indicates that the journals that were removed were of inferior scholarly quality compared to those remaining—those removed at this particular point of the process were not evaluated at all. Some data relating to journals with APCs could make it reasonable to speculate that some journals of doubtful quality may have been removed. On the other hand, a majority of journals removed did not charge an APC, and hence cannot be predatory journals. However, not being predatory is not synonymous with being of satisfactory quality. Re-applications from 6359 journals were evaluated, 2058 journals were rejected during the re-application process. [22]. This means more than 2/3 of all journals re-evaluated were accepted. Some journals among those removed may be assumed to be “dead” journals that have ceased publishing. The research in [21] indicates that generally about 15 percent of journals being removed are removed because they have ceased publishing, or because their websites do not work any longer. As we have no reason to believe the scholarly quality of the journals removed due to not re-applying was different from the quality of journals that re-applied, we may assume that roughly the same percentage of journals still functioning would have been re-accepted, meaning we have lost a sizable number of bona fide OA journals. We know nothing about whether any individual journal of the 2859 met the re-application criteria, and can only assume they were not significantly different from other DOAJ journals in this respect. Thus, it seems reasonable to conclude that the majority of the de-listed journals that still were functioning journals were journals that actually merit a listing in DOAJ. These journals will now become less visible and less useful for their authors as their dissemination to the potential readership will be made more difficult. Science and scholarship would probably have been better served if they had not been removed.

On the other hand, being an open access journal necessitates acquiring the necessary competence to function as one. You are not really competent to operate an open access journal if you are not able to answer the DOAJ questionnaire. Data here point to many of the problems being associated with publishers being small. Many such small publishers are part of, or associated with, larger institutions.

In Norway (and probably more countries), larger institutions have set up publishing infrastructures to help editors of small open access journals [29]. The institutions provide an OJS (Open Journal Systems, a wide-spread journal publication platform) installation and keep it (somewhat) upgraded and help editors with the more technical aspects of publishing. Norway lost eight of 54 journals, 15 percent—far below the average. As far as I can see, none of the journals lost are published by the publishing services set up by the larger institutions—they obviously have had a mission in the context of the re-application process.

Over time, we might expect a sizeable fraction of the journals lost to be back in DOAJ. This fraction might be increased if institutions engaged in or associated with journals see their responsibility to create an environment where editors get the financial, technical and publishing support needed to be able to operate open access journals with a sufficient level of competence. For an institution, being involved in a journal without ensuring it can comply also with the more technical publishing norms seems meaningless.

The findings in [23] show the same pattern that small publishers publishing non-APC journals are the ones most likely to have problems, as this study does. As commented on in [26] most journals are published either by very small, or very large, publishers—with little in-between. With processes like the present one, which removes many small publishers from the playing field by making them disappear from DOAJ, or the Plan S process, which seem to make many small publishers useless as providers of legitimate publishing venues under Plan S, the small scholar-led journals seem to be in danger. Two aspects of a solution to this emerge:

- The need to create larger publishers, by creating publishing services that can be a center of competence for a larger number of journals, helping them overcome some of the drawbacks of being small. This should probably be done on a level beyond the individual institution, unless these are really large.
- The clear indications that APC helps on the technical quality indicates that other forms of financing often leads to underfinancing. In other words, organizations having some kind of ownership to journals need to strengthen their financing to the extent that journals become able to develop or buy the necessary competence.

The removal of many journals that should have been kept, because they did not reapply, lowers the quality of DOAJ as a whitelist. It is possible—although impossible to document through this analysis and the data available—that a large number of the journals removed due to their not having re-applied, would have been kept had they re-applied. The author in [10] show another source of missing journals, that of journals never having applied for inclusion. Action is needed to make DOAJ a reliable whitelist again, aiming to contain the whole population of acceptable OA journals. The Plan S process does not make journals disappear from DOAJ, but create an inferior class of journals, not suited for publishing in if you have Plan S financing. In Europe, especially in some countries, Plan S-financed research will be an important part of future research undertaken.

Looking back to the title of this article: we probably have no reason to lament all journals lost; however, many journals lost are journals we would be better off having kept.

Funding: This research received no external funding.

Acknowledgments: Thanks to Dominic Mitchell of DOAJ for reading through an early version of this manuscript, helping me avoid some pitfalls. And thanks to my colleague Aysa Ekanger for commenting on both content and language. Errors and omissions of fact and of grammar and spelling remaining are of my own making.

Conflicts of Interest: The author is a member of the DOAJ Advisory Board, and works with the university's publishing service Septentrio Academic Publishing, which publishes seven journals that went through the re-accreditation process. The author is a member of *Publications'* editorial board.

References

- DOAJ. Future Plans for the Development of the DOAJ. Available online: <https://is4oa.org/2012/12/18/future-plans-for-the-development-of-the-doj/> (accessed on 15 July 2017).
- Bjørnshauge, L. *The New Application Form for Journals' Inclusion in DOAJ Has been Released*; DOAJ, 2014. Available online: <https://docs.google.com/document/d/1dr3jnOygvuDIONSBv8lho4McQsEPFd0a5gtxjCmKd9k/edit> (accessed on 22 January 2019).
- DOAJ. Reapplications are open. Have you submitted yours yet? *DOAJ News Service*; DOAJ, 2015. Available online: <https://blog.doaj.org/2015/09/30/reapplications-are-open-have-you-submitted-yours-yet/> (accessed on 22 January 2019).
- DOAJ. Reapplications deadline extended. *DOAJ News Service*; DOAJ, 2015. Available online: <https://blog.doaj.org/2015/12/30/reapplications-deadline-extended/> (accessed on 22 January 2019).
- Morrison, H. Directory of Open Access Journals (DOAJ). *Charlest. Advis.* **2008**, *9*, 19–26. [CrossRef]
- Bi, X. Quality open access publishing and registration to Directory of Open Access Journals. *Sci. Ed.* **2017**, *4*, 3–11. [CrossRef]
- Guidance on the Implementation of Plan S. Available online: <https://www.coalition-s.org/implementation/> (accessed on 10 April 2019).
- Teixeira da Silva, J.A.; Tsigaris, P. What Value Do Journal Whitelists and Blacklists Have in Academia? *J. Acad. Librariansh.* **2018**, *44*, 781–792. [CrossRef]
- DOAJ. Information for Publishers 3) Publishing Best Practice and Basic Standards for Inclusion. Available online: <https://doaj.org/publishers#advice> (accessed on 7 May 2019).
- Björk, B.-C. Open access journal publishing in the Nordic countries. *Learn. Publ.* **2019**. [CrossRef]
- Funding for Open Access Publication for the Fiscal Year 2015. Available online: <https://www.forskningsradet.no/en/Funding/STIMOA/1254015944566/p1254006410319?visAktive=false> (accessed on 12 April 2019).
- Beall, J. List of publishers Beall's list. *Scholarly Open Access. Critical Analys of Scholarly Open-Access Publishing*. 2017. Available online: <https://archive.fo/6EByy> (accessed on 12 April 2019).
- Crawford, W. "Trust Me": The Other Problem with 87% of Beall's Lists. *Walt at Random: The Library Voice of the Radical Middle*. 2016. Available online: <https://walt.lichost.org/2016/01/trust-me-the-other-problem-with-87-of-bealls-lists/> (accessed on 12 April 2019).
- Marchitelli, A.; Galimberti, P.; Bollini, A.; Mitchell, D. Helping journals to improve their publishing standards: A data analysis of DOAJ new criteria effects. *JLIS Ital. J. Libr. Arch. Inf. Sci.* **2017**, *8*, 1–21. [CrossRef]
- Teixeira da Silva, J.A.; Dobránski, J.; Al-Khatib, A.; Tsigaris, P. Challenges Facing the DOAJ (Directory of Open Access Journals) as a Reliable Source of Open Access Publishing Venues. *J. Educ. Media Libr. Sci.* **2018**, *55*, 349–358. [CrossRef]
- Bohannon, J. Who's Afraid of Peer Review? *Science* **2013**, *342*, 60–65. [CrossRef] [PubMed]
- Beall, J. "Predatory" Open-Access Scholarly Publishers. *Charlest. Advis.* **2010**, *11*, 10–17. [CrossRef]
- DOAJ. How Can I Get Journal Metadata from DOAJ? Available online: <https://doaj.org/faq#metadata> (accessed on 22 January 2019).
- Frantsvåg, J.E. *The DOAJ 2016 Spring Clean-Up*; UiT Open Research Data Dataverse, 2016. Available online: <http://dx.doi.org/10.18710/KCL56N> (accessed on 12 April 2019).
- Frantsvåg, J.E. The size distribution of open access publishers: A problem for open access? *First Monday* **2010**, *15*. [CrossRef]
- DOAJ. *DOAJ: Journals Added and Removed: Failed to Submit a Reapplication*. 2016. Available online: https://docs.google.com/spreadsheets/d/183mRBRqs2jOyP0qZWXN8dUd02D4vL0Mov_kgYF8HORM/edit#gid=1678073646 (accessed on 12 April 2019).
- DOAJ. The Reapplications project is officially complete. *DOAJ News Service*; DOAJ, 2017. Available online: <https://blog.doaj.org/2017/12/13/the-reapplications-project-is-officially-complete/> (accessed on 22 January 2019).
- Frantsvåg, J.E.; Strømme, T.E. Few Open Access Journals Are Compliant with Plan S. *Publications* **2019**, *7*, 26. [CrossRef]
- Crawford, W. (Ed.) DOAJ cuts by publisher: Partial preliminary list. *Walt at Random*. 2016. Available online: <https://walt.lichost.org/2016/05/doaj-cuts-by-publisher-partial-preliminary-list/> (accessed on 22 January 2019).

25. Koroso, N.H. *Directory of Open Access Journals Removes Thousands of Journals from Its Database United Academics Foundation*. 2016. Available online: <http://www.ua-foundation.org/directory-of-open-access-journals-removes-thousands-of-journals-from-its-database/> (accessed on 11 May 2019).
26. Morrison, H.; Salhab, J.; Calvé-Genest, A.; Horava, T. Open Access Article Processing Charges: DOAJ Survey May 2014. *Publications* **2015**, *3*, 1–16. [[CrossRef](#)]
27. Mitchell, D.; Frantsovåg, J.E.; Question about NY, New York, NY, USA. Personal communication, 2016.
28. Shen, C.; Björk, B.-C. 'Predatory' open access: A longitudinal study of article volumes and market characteristics. *BMC Med.* **2015**, *13*, 230. [[CrossRef](#)] [[PubMed](#)]
29. Nordic Journal Hosting Possibilities. Available online: https://www.ub.uit.no/wiki/noap/index.php/Nordic_Journal_Hosting_Possibilities (accessed on 7 May 2019).



© 2019 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).