

Article

The DOAJ spring cleaning 2016 and what was removed – tragic loss or good riddance?

Jan Erik Frantsvåg^{1,*}

¹ UiT The Arctic University of Norway; jan.e.frantsvag@uit.no

* Correspondence: jan.e.frantsvag@uit.no;

Abstract: In December 2012, DOAJ's 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 – and that journals already included, would need to reapply in order to be kept in the registry. My 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, permanent article identifiers, journal licenses, article level metadata deposited with DOAJ, archiving policy/solutions and/or having a policy in SHERPA/RoMEO. The analysis shows my concerns to be correct.

Keywords: Open access; DOAJ; publisher size

1. Introduction

In December 2012, The Directory of Open Access Journals' (DOAJ) 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 – and that journals already included, would need to reapply in order to be kept in the registry. Those new criteria were launched on 19th March 2014 [2] and the deadline for re-application was set to December 31st 2015 [3], later extended to April 1st 2016 [4].

We had been through the re-application procedure for our own journals¹, and I experienced that this entailed quite some work and would need some understanding of Open Access (OA) and of various OA and publishing infrastructures, services and standards. Having earlier investigated the size-distribution of OA journal publishers listed in DOAJ [5] and having become skeptical to various aspects of small publishers, I immediately came to suspect that DOAJ would lose some small, stand-alone journals – rather than journals published by larger publishers. This would not be because they would lack scholarly quality, but because they would lack the necessary competence and resources concerning the technicalities of OA publishing.

My 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 metadata information would show 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, permanent article identifiers, journal licenses, article level metadata deposited with DOAJ, archiving policy/solutions and/or having a policy in SHERPA/RoMEO.

Why is it important whether a journal is listed in DOAJ, or not? DOAJ is the authoritative database over which journals are OA, and which not. 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

¹ I was at the time responsible for our institutional publishing service Septentrio Academic Publishing <http://septentrio.uit.no/> servicing seven DOAJ-listed journals.

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.

2. Materials and Methods

DOAJ continuously publishes journal level metadata [6] that can be downloaded in a CSV format, this are the data used for this analysis. I downloaded one file immediately before the removing of journals started (file time-stamped May 9th 2016 12:00) and one immediately after the process was finished (file time-stamped May 10th 2016 06:30). Both files are publicly available [7] 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 [5] and will not be repeated here. Note the discussion there of what the term “publisher” means in DOAJ and various pitfalls concerning this.

The post clean-up file contains 8,791 journals; the pre clean-up contains 11,644, indicating a removal of 2,853 journals, or 24.5 per cent of all journals in DOAJ at that time. However, the post clean-up file contains 6 journals added during the removal process, so 2,859 journals have been removed (and 8,785 kept), increasing the loss to 24.6 per cent. This tallies well with numbers from DOAJ’s own list of removed journals [8]. That file contains 2,861 entries, but two of these are errors, with two journals still listed in DOAJ being listed as removed in the file². An important point is that journals kept has not necessarily been re-accredited – but they are either re-accredited or have applied to be re-accredited. The re-accrediting work went on for quite some time after May 9th 2016. The re-accreditation project was officially ended December 13th, 2017 [9]. Interestingly, the announcement notes that during the re-application process 2,058 re-applications were rejected. The reason for reject 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 [5] 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. 2,589 journals removed from DOAJ because they did not send in a reapplication to DOAJ within the time limit. These are our focus.
2. 3,862 journals added to DOAJ after March 2014 and journals re-accredited after March 2014.
3. 4,923 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.

² 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.

In the discussions below I assume that publishers know they need to apply for re-accreditation. This is not necessarily so³. 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 note they manually updated more than 1,000 user accounts during the process.

3. Results

3.1. Publisher size

The pre clean-up file contained 11,644 journals, published by 6,081 different publishers. A summary of publisher size, as measured by the number of journals published by that publisher, is given in the table below.

Table 1 Pre clean-up publisher size distribution

Publisher size	Number of publishers	Number of journals	Share of publishers	Share of journals
1	5,097	5,097	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	1,647	0.1%	14.1%
Total	6,081	11,644	100%	100%

This is not very different from what I found previously [5], even if the journal share of the single journal publishers has been reduced from 55.0 per cent to 43.8 per cent, and the share of the largest publishers has increased from 10.8 per cent to 14.1 per cent.

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

Table 2 Publisher size and loss of journals

Number of journals Publisher size group	Status journals		Total journals	Per cent lost
	Kept	Lost		

³ The following information is from a private communication with Dominic Mitchell of DOAJ, on July 14th 2016.

1	3,441	1,656	5,097	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	1,625	22	1,647	1.3 %
Total	8,785	2,859	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 per cent of journals, on average 27.6 per cent, the next category 21–50 lose 18.8 per cent while the larger (>50) publishers lose a negligible fraction of their journals. The losses among the larger (>50) publishers are as follows:

Table 3 Number of journals lost from publishers >50 journals

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

Of the 7 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 [10]. 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 7 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	Status publishers		Total publishers	Per cent lost
	Kept	Lost		
1	3,441	1,656	5,097	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	4,307	1,774	6,081	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 (risk of less than 1 per cent⁴) that many publishers with more than three journals would disappear. 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 distribution of publishers 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.

Table 5 Publisher size after the spring clean-up

Publisher size	Number of publishers	Number of journals	Share of publishers	Share of journals
1	3,590	3,590	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	1,625	0.2 %	18.5 %
Total	4,307	8,785	100.0 %	100.0 %

The 151 publishers with 6 or more journals are 3.5 per cent of publishers, but control 42.2 per cent of journals.

The numbers seem to support my 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, e.g. [11] discusses the spring clean-up

⁴ The average risk is 29.2 per cent. The risk of losing n journals if losses are random, is 0.292^n , The risk of losing 3 journals is 0.292^3 , which is 0.025 (2.5 per cent), the risk of losing 4 journals is 0.292^4 , equal to 0.007 or 0.7 per cent.

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.

3.2 Re-accreditation

Not only whether, but also how quickly and efficiently a publisher has re-applied for accreditation might be of interest.

The table below shows that the bigger the publisher, the larger a fraction of journals have been (re-)accepted after March 2014, when the new application procedures were put in force. Other journals that have not been removed, have a re-application pending, but not yet accepted.

Table 6 Publisher size and re-application status

Publisher size group	Application pending	Accepted after March 2014	Total	Per cent accepted after March 2014
1	2,311	1,130	3,441	33 %
2	376	258	634	41 %
3	223	183	406	45 %
4	142	136	278	49 %
5	105	80	185	43 %
6-10	277	261	538	49 %
11-20	371	300	671	45 %
21-50	280	346	626	55 %
51-100	192	189	381	50 %
>100	646	979	1,625	60 %
Total	4,923	3,862	8,785	44 %

This could of course indicate that DOAJ growth after March 2014 has had a tendency to come from larger publishers. Another explanation is that larger publishers have been more efficient in getting re-applications in early, or with better quality of information, resulting in their journals having finished their re-application process earlier.

3.3 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 kept⁵, Table 7 will give a good picture.

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 per cent (2,074) have a CC license, while among those lost only 19 per cent (550) had such a license. We could also note that nearly 97 per cent (5,802) of journals accepted after the new criteria were put in force have a CC license.

⁵ The original content of the license field has been grouped to get a better overview. E.g. all CC-enabled journals are grouped together here.

Table 7 Use of licenses among kept or lost journals

	Journal license	Application pending	Accepted after March 2014	Total
<i>Kept</i>	CC license	2,072	3,730	5,802
	Various journal-specific licenses	7	131	138
	None	2,844	1	2,845
<i>Kept Total</i>		4,923	3,862	8,785
<i>Lost</i>	CC license	550		550
	Various journal-specific licenses	2		2
	None	2,307		2,307
<i>Lost Total</i>		2,859		2,859
Grand total		7,782	3,862	11,644

The table 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, so the 2,859 removed journals are part of the numbers.

Table 8 Licenses and publisher size

Journal license	Publisher size group										
	1	2	3	4	5	6-10	11-20	21-50	51-100	>100	Total
CC license	2,037	413	300	199	143	390	471	541	304	1,554	6,352
Various journal-specific licenses	78	15	11	2	1	14	18	1			140
None	2,982	460	247	183	101	349	429	229	79	93	5,152
Total	5,097	888	558	384	245	753	918	771	383	1,647	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 %

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

3.4 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 – to me – 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 2 of 7782 journals accepted in DOAJ before March 2014 has any meaningful information about this. So instead, I checked the DOAJ journal data against a data file from SHERPA/RoMEO (RoMEO for short). This

file was from January 29th 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.

Table 9 Policies in SHERPA/RoMEO

In RoMEO?	Found post clean-up?		Total
	Kept	Lost	
No	5,575	2,496	8,071
Yes	3,210	363	3,573
<i>Total</i>	8,785	2,859	11,644
Percentages			
No	63 %	87 %	69 %
Yes	37 %	13 %	31 %

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

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.

3.5 Geography

The journals lost were not evenly distributed over countries, quite the opposite. 7 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 1 journal, 7 published 2 journals while 7 published from 3 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 per cent. The country losing the most journals is the United States, with a loss of 403 journals out of 1070 – a 37.7 per cent loss.

Table 10 Status for Japanese journals

Publisher size	Clean-up status			Percentage lost
	Kept	Lost	Total	
1	16	69	85	81 %
2	5	3	8	38 %
11-20	2		2	0 %
21-50	2		2	0 %
>100	1		1	0 %
<i>Total</i>	26	72	98	73 %

The Japanese journals are mainly published as stand-alone journals, and these obviously have had problems with the re-accreditation process. The 5 journals published by larger publishers are all published by international publishers – and have managed to re-apply for accreditation. Publisher size obviously plays a role.

Table 11 Status for US journals

Publisher size	Clean-up status		Total	Percentage lost
	Kept	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	1,070	38 %

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 3 publishers now entirely removed from DOAJ.

Numbers in [12] 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 1 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 5 journals, hence themselves as a publisher, removed; University of California UCL/UCLA losing 8 journals; Centers for Disease Control and Prevention losing 3 out of 4 journals, etc.

3.6 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 must 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.

The current metadata file does not contain reliable information about APCs for journals added before March 2014, only 8 of those journals indicate they use APCs. Using these metadata often, I have an archive of files from various dates and have found that a file from February 7th 2014 contains such information. The file contains information about 9,804 journals, matching the current metadata file with the old one leaves information about 8,112 journals that were in DOAJ on February 7th 2014, that also were in DOAJ on May 9th 2016. After grouping the data (N for No and NY which probably

means No⁶, into No; and CON for conditional and Y for Yes into Yes, leaving 4 where data is lacking) we find the following:

Table 12 APC-information for journals in DOAJ both in February 2014 and in May 2016

Journals		APCs February 2014						Grouped			
Publishe size group	Empty	N	NY	CON	Y	Total	No	Yes	Rest	Per cent APCs	
1	3	3,199	86	182	477	3,947	3,285	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 %	
<i>Total</i>	<i>4</i>	<i>5,625</i>	<i>128</i>	<i>435</i>	<i>1,920</i>	<i>8,112</i>	<i>5,753</i>	<i>2,355</i>	<i>4</i>	<i>29 %</i>	
<i>Percentage</i>							<i>71 %</i>	<i>29 %</i>			

Table 13 APC information for journals in DOAJ after the clean-up

	APCs after the clean-up						Grouped			
Publisher size group	Empty	N	NY	CON	Y	Total	No	Yes	Rest	Per cent APCs
1	1	1,926	17	105	250	2,299	1,943	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 %
<i>Total</i>	<i>1</i>	<i>3,568</i>	<i>29</i>	<i>252</i>	<i>1,426</i>	<i>5,276</i>	<i>3,597</i>	<i>1,678</i>	<i>1</i>	<i>32 %</i>
<i>Percentage</i>							<i>68 %</i>	<i>32 %</i>		

- Of the 8112 journals still in DOAJ before the clean-up, 71 per cent had no APC, while 29 per cent had.
- After the clean-up we are left with 5276 journals, of which 68 per cent have no APC while 32 per cent have.

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 find that in total the percentage of journals having APCs has grown as a result

⁶ Private communication with DOAJ on July 19th 2016

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.

3.7 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, 1,971 of 7,782 (25.3 per cent) lack this information. A bit baffling, of the journals lacking this information, 29 per cent were lost in the clean-up, while 39 per cent 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.

Table 14 Language and removals

Main language	Journals			Per cent lost
	Kept	Lost	Total	
English	2,567	1,670	4,237	39 %
Other than English	964	610	1,574	39 %
Info lacking	1,392	579	1,971	29 %
<i>Total</i>	4,923	2,859	7,782	37 %

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 above table shows no difference between English or another first language for journals lost – both categories show a loss of 39 per cent.

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.

3.8 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. But e.g. the category “Philosophy. Psychology. Religion:” I have categorized as Philosophy, even if Psychology could be the more important word – this is impossible to say without a detailed analysis of the individual journals.

Looking at only the 7782 journals not yet re-evaluated after March 2014, we find the following:

Table 15 Journals not re-evaluated after March 2014 by subject

Subject	Journals			Per cent lost
	Kept	Lost	Total	
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	1,068	614	1,682	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	1,321	37 %
Social Sciences	546	325	871	37 %
Technology	371	202	573	35 %
"Missing"		1	1	100 %
<i>All journals</i>	4,923	2,859	7,782	37 %

Of the subjects with more than 100 journals, Law with a loss of 52 per cent and History and Political Science with losses of "only" 30 and 31 per cent respectively, stand out.

3.9 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, like DOI, are a sign of publisher competence, also 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, but 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, like LOCKSS.

I cannot find any information 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 find this:

Table 16 Article level metadata deposit and removal from DOAJ

	Data from sceen shots		Other data
--	-----------------------	--	------------

Date and hour	Journals	Searchable at article level	Articles	Publishers	Articles per journal	Articles per publisher
09.05.2016 11:30	11,650	7,290	2,296,024	6,081	197	378
10.05.2016 06:30	8,795	6,095	1,960,409	4,307	223	455
Lost	2,855	1,195	335,615	1,774	118	189
<i>Per cent lost</i>	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 per cent compared to 24.5 per cent. We also see that those removed despite having deposited such metadata, on average had deposited information about fewer articles than those remaining.

Again, this points to removal being connected to both publishing competence and journal/publisher size.

4. Summing up and discussion

We see from the data presented that many journals 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, but 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. Some data relating to journals with APCs could make it reasonable to speculate that some journals of doubtful quality may have been removed. But the general picture is that we have lost nearly 3,000 scholarly journals, that now will 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 Denmark, Sweden and Norway (and probably more countries) larger institutions have set up publishing infrastructures to help editors of small open access journals. The institutions provide an OJS installation and keep it (somewhat) upgraded and help editors with the more technical aspects of publishing. Norway lost 8 of 54 journals, 15 per cent – 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.

I have received information that a couple of the Norwegian journals lost that are published by rather competent organizations, have experienced traditional organizational “snafu” in that information about the re-accreditation process has not reached the right persons. These are now preparing new applications for listing in DOAJ. I trust a number of journals worldwide will, when they discover they have lost their listing, start processes to re-enlist.

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 is meaningless.

Looking back to the title of this article: We probably have no reason to lament all journals lost, but most 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 7 journals that went through the re-accreditation process. The author is a member of *Publications'* editorial board.

5. References

1. DOAJ. Future plans for the development of the DOAJ. Available online: <https://is4oa.org/2012/12/18/future-plans-for-the-development-of-the-doaj/> (accessed on 2017-07-15).
2. Bjørnshauge, L. The new application form for journals' inclusion in DOAJ has been released. DOAJ: 2014.
3. DOAJ. Reapplications are open. Have you submitted yours yet? . In *DOAJ News service*, DOAJ: 2015.
4. DOAJ. Reapplications deadline extended In *DOAJ News service*, DOAJ: 2015.
5. Frantsvåg, J.E. The size distribution of open access publishers: A problem for open access? *First Monday* **2010**, 15, doi:<https://doi.org/10.5210/fm.v15i12.3208>.
6. DOAJ. How can I get journal metadata from DOAJ? Available online: <https://doaj.org/faq#metadata> (accessed on 2019/01/22).
7. Frantsvåg, J.E. The DOAJ 2016 spring clean-up. UiT Open Research Data Dataverse, 2016; <http://dx.doi.org/10.18710/KCL56N>.
8. DOAJ. DOAJ: journals added and removed: Failed to submit a reapplication. 2016.
9. DOAJ. The Reapplications project is officially complete. In *DOAJ News service*, DOAJ: 2017.
10. Crawford, W. DOAJ cuts by publisher: partial preliminary list. In *Walt at Random*, Crawford, W., Ed. 2016.
11. Koroso, N.H. Directory of Open Access Journals removes thousands of journals from its database United Academics Foundation: 2016.
12. Shen, C.; Björk, B.-C. 'Predatory' open access: a longitudinal study of article volumes and market characteristics. *BMC Medicine* **2015**, 13, 15, doi:<https://doi.org/10.1186/s12916-015-0469-2>.