

## Some limitations of DOAJ metadata for research purposes

Posted by [Dr. Heather Morrison](#)

*by: Xuan Zhao, Luan Borges, & Heather Morrison*

### Abstract

The Directory of Open Access Journals <http://doaj.org> is an excellent service that fulfills many important functions, in particular facilitating access to a vetted collection of over 15,000 freely available peer-reviewed journals. The DOAJ search services and metadata download are very useful for researchers as well. The purpose of this post is to alert researchers to some of the limitations of the DOAJ metadata that researchers need to take into account to avoid drawing erroneous conclusions. First, when downloading DOAJ metadata, it is necessary to open the .csv file in Unicode in order to retain non-English characters. We open in Open Office for this reason, then save as an excel file. The nature of the metadata means that some data is inserted in the wrong column; clean-up, as discussed below, is necessary before data analysis. When journal editors or others working on their behalf enter metadata into DOAJ, research is not the primary purpose of this exercise; for this reason, in-depth assessment and corrections may be necessary before analysis. Below, we present publisher size analysis as an example of what researchers may encounter. Finally, because the main purpose of DOAJ is connecting readers with content, the metadata of interest to a particular research project may not be up to date. As demonstrated below, as of Jan. 5, 2021, only 30% of DOAJ journals have a “last update” date within the previous year (2020). We do not know whether the “last update” date reflects a full or partial metadata review. We illustrate the potential impact on research results with the example of the SKC longitudinal APC study. Of the 4,292 DOAJ journals that responded “yes” to the APC question, only 30% have a last update date of 2020 or 2021. Even with this 30% of journals, we have no way of knowing whether the APC status and/or amount per se was updated, or only other unrelated metadata. This means that if we compare 2019 prices obtained from publisher websites in 2019 with 2021 DOAJ APC metadata, we will almost certainly get incorrect results, for example falsely assuming that matching APC amounts means no change in the prices. DOAJ provides rich and useful metadata for the researcher and the research question “is this journal listed in DOAJ?” is of value in and of itself. For this reason, we intend to continue using DOAJ metadata in addition to data derived from other sources, particularly data derived directly from publisher websites. See below to a link to an open data version of the DOAJ metadata reflecting the corrections explained in this post.

### Details

#### Correcting for displaced observations

As previously mentioned, the first step to confidently use the DOAJ metadata for analysis and research is identifying and correcting data inserted in the wrong column, herein also called displaced observations.

Below we can see an example of a displaced observation from the DOAJ metadata. Column BB has no assigned variable while containing some observations, apparently displaced one column to the right.

ACME					
AY	AZ	BA	BB	BC	BD
Last updated Date	Number of Article Records	Most Recent Article Added			
2017-01-04T14:19:54Z	2649	2020-06-10T21:49:11Z			
2020-02-24T09:07:42Z	166	2020-06-19T09:42:34Z			
2017-02-22T11:14:48Z	1049	2021-01-05T08:28:16Z			
2020-12-22T11:08:24Z	1207	2015-12-08T15:06:43Z			
2020-06-26T13:24:42Z	1140	2020-07-06T19:01:13Z			
2020-06-26T13:19:44Z	3656	2020-12-04T12:42:12Z			
2017-04-15T15:55:43Z	228	2020-08-04T10:39:40Z			
2019-08-06T14:46:13Z	280	2020-11-19T08:33:55Z			
2018-04-14T14:57:59Z	413	2020-03-18T00:59:08Z			
2015-11-18T14:06:52Z	1244	2019-10-09T21:18:35Z			
2020-06-01T13:34:49Z	39	2020-12-23T19:07:48Z			
2019-11-02T13:55:53Z	289	2020-11-20T08:49:18Z			
2017-04-03T14:55:55Z	381	2013-09-05T23:44:06Z			
2017-02-02T14:02:08Z	0				
2019-12-18T05:36:29Z	354	2021-01-04T12:46:13Z			
2019-11-04T18:04:20Z	1109	2020-11-23T20:36:30Z			
2020-12-04T18:45:02Z	0				
2016-04-01T16:37:30Z	0				
2019-11-02T14:12:31Z	156	2020-11-20T08:49:50Z			
2016-07-21T18:43:52Z	3240	2020-06-10T21:08:18Z			
2016-07-21T18:47:12Z	1172	2020-06-10T20:49:47Z			
2016-07-22T15:36:16Z	8036	2020-06-10T22:08:57Z			
2018-06-12T11:37:35Z	94	2019-05-12T11:05:04Z			
2018-07-30T15:46:30Z	169	2021-01-04T15:00:01Z			
2020-01-29T16:05:51Z	0				
2016-06-08T12:29:44Z	1803	2013-01-23T12:26:33Z			
2019-03-28T02:36:32Z	27	2020-07-04T12:35:10Z			
2016-03-14T11:19:15Z	7200	2018-12-28T13:29:46Z			
2020-09-10T08:43:45Z	1979	2019-11-12T11:41:40Z			
2016-07-22T15:40:28Z	3747	2020-06-10T21:21:18Z			
2017-11-24T16:48:46Z	144	2020-08-17T11:57:26Z			
2017-06-08T11:41:59Z	1668	2020-10-22T16:55:05Z			
2020-03-03T15:05:36Z	70	2020-08-03T14:53:54Z			
2018-03-14T21:10:12Z	88	2020-06-22T09:55:36Z			
2016-12-08T09:33:00Z	2017-09-06T17:24:27Z	25	2017-09-17T17:04:02Z		
2017-11-08T13:48:26Z	21	2017-09-01T06:59:13Z			
2020-07-13T08:29:22Z	111	2020-12-31T13:58:22Z			
2018-06-06T09:30:16Z	0				
2020-10-15T08:05:22Z	91	2020-11-01T22:53:08Z			

Table 1 – An example of misplaced data from 2021 DOAJ metadata

Users may follow different steps to correct for displaced data. Here we explain in more detail how we have identified these displacements and corrected them.

Before proceeding with any analysis, it is important to get familiarized with the DOAJ metadata first. We recommend users to read the *DOAJ Guide to applying*, [available online](#), because the metadata reflects responses to questions asked in the application process. The DOAJ metadata, as of 5 Jan. 2021, possesses 53 variables ranging from *Journal Title* to *Country* to *Most recent article added*. It may be helpful to start correcting observations from variables with easily identifiable responses, such as « Country » or « Country of Publisher », or variables that allow only two types of answers (i.e Yes or No), such as *Author holds copyright without restrictions* and *APC*. It is recommended to create a pivot table to identify displaced observations, repeating this process until no observations are identified in a wrong column.

When cleaning-up the DOAJ metadata, users will notice that in some cases only one observation was displaced; in other cases, an entire row was displaced beginning on a specific variable. In the example highlighted in yellow below, all observations beginning at variable *Publisher* were displaced one column to the right.

	H	I	J	K	
1	Languages in which the journal accepts manuscripts	Publisher	Country of publisher	Society or institution	Count
2	English	Academia Brasileira de Ciências	Brazil		
3	Italian	Università degli Studi di Milano	Italy		
4	English	Society for Publication of Acta Dermato-Venerologica	Sweden		
5	English, Spanish	Colegio de Médicos y Cirujanos de Costa Rica	Costa Rica		
6	English	Polish Botanical Society	Poland		
7	English	Polish Botanical Society	Poland		
8	Croatian, English	University of Zagreb, School of Dental Medicine	Croatia	Croatian dental society, Croatian medical association	
9	English	SAGE Publishing	United Kingdom	GIGA German Institute of Global and Area Studies	
10	Spanish	Colegio Mexicano de Inmunología Clínica y Alergia, A.C.	Mexico	Colegio Mexicano de Inmunología Clínica y Alergia, A.C.	
11	English, French, Italian, Portuguese, Spanish	Universidade Estadual Paulista Júlio de Mesquita Filho	Brazil	Universidade Estadual Paulista	
12	English	American Society for Horticultural Science (ASHS)	United States		
13	English	University of Belgrade, Faculty of Law, Belgrade, Serbia	Serbia		
14	English, Portuguese	Instituto de Ciências Sociais da Universidade de Lisboa	Portugal	Instituto de Ciências Sociais da Universidade de Lisboa	
15	English	E.H. Angle Education and Research Foundation	United States	Edward H. Angle Society of Orthodontists	
16	English	Copernicus Publications	Germany	Leibniz Institute for Farm Animal Biology	
17	English, Spanish	Consejo Superior de Investigaciones Científicas	Spain		
18	Spanish	Sociedad Uruguaya de Pediatría	Uruguay	Sociedad Uruguaya de Pediatría	Urug
19	English, French, Italian, Portuguese, Spanish	Universidad de Córdoba	Spain		
20	Serbian	Pharmaceutical Association of Serbia, Belgrade, Serbia	Serbia		
21	English	Conselho Brasileiro de Oftalmologia	Brazil	órgão oficial do Conselho Brasileiro de Oftalmologia	
22	English	Instituto Brasileiro de Estudos e Pesquisas de Gastroenterologia (IBGE)	Brazil		
23	English	Academia Brasileira de Neurologia (ABNEURO)	Brazil	Academia Brasileira de Neurologia (ABN)	
24	English, Spanish	Facultad de Farmacia, Universidad de Granada	Spain	Universidad de Granada	
25	English	Taylor & Francis Group	United Kingdom	KoREMA - Croatian Society for Communications, Computing, Electronics, I	
26	English, German	Senckenberg Deutsches Entomologisches Institut	Germany	Deutsche Gesellschaft für allgemeine und angewandte Entomologie	
27	Dutch, English	BRILL	Netherlands	Vereeniging KTLV (NL)	
28	English, German	Sciencio	Austria	University of Natural Resources and Life Sciences, Vienna (BOKU)	
29	Spanish	Banco de la República	Colombia	Banco de la República	
30	English	South African National Biodiversity Institute	South Africa	South African National Biodiversity Institute	
31	English, Portuguese	Instituto Agronômico de Campinas	Brazil		
32	Croatian, English	Faculty of Mechanical Engineering and Naval Architecture	Croatia	University of Zagreb	
33	English	University of California Agriculture and Natural Resources	United States	University of California Agriculture and Natural Resources	
34	English	Firenze University Press	Italy	University of Florence	
35	Czech, English	Univerzita Karlova, Filozofická fakulta	Czechia		
36	English	Univerzita Karlova, Filozofická fakulta	United Kingdom	Chartered Institute of Library and Information Professionals (CILIP)	Catal
37	Spanish	Universidad de Granada	Spain	Departamento de Filosofía del Derecho – Facultad de Derecho	
38	English, French, German	Swiss Chemical Society	Switzerland	Swiss Chemical Society	
39	English, French, German, Italian, Spanish	Università degli Studi di Torino	Italy	Société Internationale des Amis de Cicéron	
40	Spanish	Universidad Nacional de La Plata	Argentina	Universidad Nacional de La Plata, Facultad de Ciencias Económicas	
41	Catalan, English, French, Portuguese, Spanish	Consejo Superior de Investigaciones Científicas	Spain		
42	English	Association of College and Research Libraries	United States		
43	English	Elsevier	United States		
44	English	Croatian Chemical Society	Croatia		
45	English	Columbia University Libraries	United States	Columbia University	
46	English	Elsevier	Netherlands		
47	English, French, Portuguese, Spanish	Universidade do Estado do Rio de Janeiro	Brazil		

Table 2 – Line 36 illustrates an example of an entire row with displaced observations

## Data entry inconsistencies

When correcting for displaced observations, we have also identified some inconsistencies in the way observations are registered in the DOAJ metadata. The table below lists the main visible inconsistencies found for some variables. In the majority of instances, the inconsistencies will not impact DOAJ users looking up information for a particular journal. However, it is important to take into account these inconsistencies before proceeding to any automated statistical analysis. For example, DOAJ metadata as is can be used to identify the number of journals with persistent article identifiers, but automated counting of DOI v. ARK or other approaches would require some advance data manipulation.

### Variable Example

**Alternative title** Some journals alternative titles may be registered as a number. Some examples are “2300-6633” and “0”.

Some observations have some special characters as follows:

6. rheology, tribology, hydrodynamics, thermodynamics, mechanics of structures, mechatronics.

**Keywords** ♣ water cycles, water environment, water treatment and reuse, water resource, water quality, hydrology

• natural sciences, • environmental sciences, • social sciences, agricultural sciences, veterinary medicine, medical sciences

**Copyright information URL** Some URLs lack a letter « h » at the beginning or the end. The example below illustrates this small error. There should be an “h” at the beginning and an “l” at the end of the link.  
<http://www.emeraldgrouppublishing.com/services/publishing/jiuc/authors.htm>

**Plagiarism** Some URLs lack a letter « h » at the beginning or the end. The example below illustrates this small error. There should be an « h » at the beginning and an « l » at the end of the link.  
<http://www.emeraldgrouppublishing.com/services/publishing/jiuc/authors.htm>

informat	
ion URL	
URL for	
journal's	Some URLs lack a letter « h » at the beginning or the end. The example below illustrates this
instructi	small error. There should be an « h » at the beginning of the URL
ons for	<a href="https://revistas.unasp.edu.br/LifestyleJournal/about/submissions">https://revistas.unasp.edu.br/LifestyleJournal/about/submissions</a>
authors	
	Some URLs have extra letters. The example below, for instance, has a letter « i » at the
Other	beginning of the URL
submissi	<a href="https://journals.univie.ac.at/index.php/voebm/m/index">https://journals.univie.ac.at/index.php/voebm/m/index</a>
on fees	Some URLs lack a letter « h » at the beginning or the end. The example below illustrates this
informat	small error. There should be an « h » at the beginning of the URL
ion URL	<a href="http://psr.ui.ac.id/index.php/journal/about/submissions#authorGuidelines">http://psr.ui.ac.id/index.php/journal/about/submissions#authorGuidelines</a> <a href="https://www.karger.com/Journal/Guidelines/261897#sec62">https://www.karger.com/Journal/Guidelines/261897#sec62</a>
Preserva	
tion	<i>Preservation services</i> can be registered as a name or a website
Services	
Preserva	
tion	
Service:	<i>Preservation services – national library</i> can be registered as a name or a website
national	
library	
Preserva	Some URLs lack a letter « h » at the beginning or the end. The example below, for instance,
tion	has a small error. There should be an « h » at the beginning of the URL
informat	<a href="https://periodicos.uff.br/revistagenero/about/editorialPolicies#focusAndScope">https://periodicos.uff.br/revistagenero/about/editorialPolicies#focusAndScope</a> <a href="http://ejournal.st">http://ejournal.st</a>
ion URL	<a href="http://kip-pgri-sumbar.ac.id/index.php/economica">kip-pgri-sumbar.ac.id/index.php/economica</a>
Deposit	
policy	
director	<i>Deposit policy directory</i> can be registered as a name or a website
y	
Persiste	<i>Persistent article identifiers</i> can be registered as an acronym (UDC, DOI, ARK), but also as a
nt article	website, such as <a href="http://dc.identifier.uri">dc.identifier.uri</a> (DSpaceUnipr) or
identifie	NBN <a href="http://www.depositolegale.it/national-bibliography-number/">http://www.depositolegale.it/national-bibliography-number/</a> .
rs	Another example is the occurrences UDC and UDC (Universal decimal Classification), which
	are equivalents but were registered differently
URL for	
journal's	Some URLs lack a letter « h » at the beginning or at the end, or they have an extra h at the
Open	beginning of the URL. The example below has an extra letter « h » at the beginning of the
Access	URL.
stateme	<a href="http://www.revistas.usp.br/gestaodeprojetos/about">http://www.revistas.usp.br/gestaodeprojetos/about</a>
nt	

Table 3 – Visible inconsistencies identified in the DOAJ metadata

## Publisher's names duplicates investigation and clean-up

The purpose of this project is preparation to develop a rough picture of publisher size to compare with Solomon & Björk's findings (2012). In order to better perform publisher size analysis, we have specifically investigated the publisher duplicates and corrected most of the obvious errors, such as small differences in punctuation and/or characters, extra spaces at the beginning and/or at the end, and minor differences in entering the publisher name when it is the same, etc. (Please see examples in *Table 4 – Investigative Strategies – Publisher Names Duplicates*).

The process of clean-up was divided into three stages. Firstly, we created a pivot table for the publisher column to identify the entries in rows which were slightly different but weren't gathered. Secondly, when potential duplicates were found, we conducted an investigation to confirm duplicates and/or to decide which name to keep (in priority order: use the name with the most journal entries; correct name with obvious typo; use the first name listed). Please see the investigative strategies below:

IF	Publisher names are identical, e.g., extra spaces at the beginning or at the end of names, or short of spaces ("Abant İzzet Baysal University" vs. "Abant İzzet Baysal University"; "Universidade de São Paulo" vs. "Universidade de São Paulo"; "Australian International Academic Centre PTY.LTD." vs. "Australian International Academic Centre PTY.LTD.")	THEN	Copy one instance of the name and replace the duplicate(s)	THEN	Re-run the pivot table to make sure this worked.
	Publisher names are almost identical, except for one very minor detail, e.g., non-english character in one but not the other, or minor difference in punctuation and/or characters, or abbreviation in one but not the other ("Abant İzzet Baysal Üniversitesi" vs. "Abant İzzet Baysal University"; "Academica Brâncuși" vs. "Academica Brâncuși"; "Area de Innovación y Desarrollo S.L." vs. "Área de Innovación y Desarrollo, S.L."; "Asociación Interuniversitaria de Investigación Pedagógica" vs. "Asociación Interuniversitaria de Investigación Pedagógica (AIDIPE)")		Assume a typo in one. Select one version and copy / paste over the other		
	There is a potentially meaningful difference between similar publisher names ("University of Ottawa Press" vs. "University of Ottawa" vs. "University of Ottawa department X")		Keep the original variations.		
	Two or more publisher names appear to be identical but there is more than one minor variation ("Academica Brâncuși" vs. "Editura Academica Brâncuși"; "Anadolu University" vs. "Anadolu University, Eskisehir")		Check to see if the journal URLs are partially the same (common server), other information similar (e.g. country of publisher, language). Make a judgment call as to whether the publishers are the same or not. If same, copy / paste one name over duplicates If not same, keep the original variations.		Re-run pivot table to make sure this worked.

*Table 4 – Investigative Strategies – Publisher Names Duplicates*

Thirdly, after identifying inconsistencies in publisher names, we created a table (please see *Table 5 – Corrections Gathering – Publisher Names Duplicates*) to register all the corrections on the variable *Publisher*. About 500 inconsistencies were corrected. Thus, the number of publishers in the pivot table has decreased from 7218 entries (data resource: pivot table based on DOAJ metadata) to 6804 entries (data resource: pivot table based on the cleaned-up version of database).

	A	B	C	D
1	Variable	Original	New	Notes
2	Publisher	Dirección General de Investigación, Universidad Peruana Unión	Dirección General de Investigación, Universidad Peruana Unión	blank space beginning
3	Publisher	Institute for Animal Husbandry, Belgrade	Institute for Animal Husbandry, Belgrade	blank space beginning
4	Publisher	Razavi Hospital, Mashhad, Iran	Razavi Hospital, Mashhad, Iran	blank space beginning
5	Publisher	Universidade de São Paulo	Universidade de São Paulo	blank space beginning
6	Publisher	"Ștefan cel Mare" University of Suceava, Faculty of Forestry and "Marin Drăcea" National Research-Development Institute in Forestry - Station Cămpulung Moldovenesc	"Ștefan cel Mare" University of Suceava, Faculty of Forestry and "Marin Drăcea" National Research-Development Institute in Forestry - Station Cămpulung Moldovenesc	blank space beginning
7	Publisher	Academy Publishing Center	Academy Publishing Center	blank space beginning
8	Publisher	Advancements in Agricultural Development Inc	Advancements in Agricultural Development Inc	blank space beginning
9	Publisher	Al-Khwarizmi College of Engineering – University of Baghdad	Al-Khwarizmi College of Engineering – University of Baghdad	blank space beginning
10	Publisher	AO Research Institute Davos	AO Research Institute Davos	blank space beginning
11	Publisher	Aras Part Medical International Press	Aras Part Medical International Press	blank space beginning
12	Publisher	Arkat USA, Inc.	Arkat USA, Inc.	blank space beginning
13	Publisher	Asociación Española de Geografía	Asociación Española de Geografía	blank space beginning
14	Publisher	Associação Brasileira de Educação a Distância (ABED)	Associação Brasileira de Educação a Distância (ABED)	blank space beginning
15	Publisher	Associação Latino-americana de Estudos do Discurso	Associação Latino-americana de Estudos do Discurso	blank space beginning
16	Publisher	Association "Professional analytics of autopoiesis systems"	Association "Professional analytics of autopoiesis systems"	blank space beginning
17	Publisher	Association Française d'Etudes Américaines	Association Française d'Etudes Américaines	blank space beginning
18	Publisher	Association of Naval Architects and Marine Engineers	Association of Naval Architects and Marine Engineers	blank space beginning
19	Publisher	Association of Surveyors of Slovenia (Zveza geodetov Slovenije)	Association of Surveyors of Slovenia (Zveza geodetov Slovenije)	blank space beginning
20	Publisher	Babol University of Medical Sciences	Babol University of Medical Sciences	blank space beginning
21	Publisher	Bagher al-Olum University	Bagher al-Olum University	blank space beginning
22	Publisher	Bila Tserkva National Agrarian University	Bila Tserkva National Agrarian University	blank space beginning
23	Publisher	Borys Grinchenko Kyiv University	Borys Grinchenko Kyiv University	blank space beginning
24	Publisher	Çanakkale Onsekiz Mart University	Çanakkale Onsekiz Mart University	blank space beginning
25	Publisher	Center for Quality, Faculty of Engineering, University of Kragujevac, Serbi	Center for Quality, Faculty of Engineering, University of Kragujevac, Serbi	blank space beginning
26	Publisher	Dnipro National University of Railway Transport named after Academician	Dnipro National University of Railway Transport named after Academician	blank space beginning
27	Publisher	Dnipro State Agrarian and Economic University	Dnipro State Agrarian and Economic University	blank space beginning

Table 5 – Corrections Gathering – Publisher Names Duplicates

As illustrated in the two tables above, there were different types of data inconsistencies. In order to respect metadata to the greatest extent, we acted prudently when making decisions. In some minor variation cases, we tried to click on the URLs to check publisher websites and to collect convincing evidence. However, we met some intricate complex challenges.

One of the challenges was the language. Due to the massiveness and the wide-range of publishers (124 countries, 80 languages, DOAJ, 7 Feb. 2021) [https://doaj.org/], we were unable to identify all of the sources of information. Besides, when there were invalid URLs or unmatched information, it was difficult to seek out any precision. What's more, among 7218 entries of publisher names, some of the potential duplicates weren't gathered because of their different beginning words. For example, "Editora da Universidade Estadual de Maringá (Eduem)" vs. "Eduem – Editora da Universidade Estadual de Maringá" and "Academica Brâncuși" vs. "Editura Academica Brâncuși". They were usually far apart and hard to be detected. More details can be found in the Table 6 below:

Different beginning words (examples)	"Academica Brâncuși" vs. "Editura Academica Brâncuși";
	"Alexandru Ioan Cuza University of Iași" vs. "Editura Universităţii 'Alexandru Ioan Cuza' Iași";
	"Editora da Universidade Estadual de Maringá (Eduem)" vs. "Eduem – Editora da Universidade Estadual de Maringá"

Table 6 – (1)

Unmatched publisher names (examples):

Original publisher names	Possible correct names	URLs
Canadian Society for the Study of Education.	The Canadian Association for Curriculum Studies	<a href="https://jcacs.journals.yorku.ca/index.php/jcacs/index">https://jcacs.journals.yorku.ca/index.php/jcacs/index</a>



Badan	URL directs to a new web link:	
Penelitian	<a href="https://ejournal2.litbang.kemkes.go.id/index.php/jki/index">https://ejournal2.litbang.kemkes.go.id/index.p</a>	
dan	<a href="http://ejournal.litbang.kemkes.go.id/index.php/jki">hp/jki/index</a>	<a href="http://ejournal.litbang.kemkes.go.id/index.php/jki">http://ejournal.litbang.kemkes.go.id/index.php/jki</a>
Pengembangan	whose publisher name is:	
gan	Pusat Penelitian dan Pengembangan Biomedis	
Kesehatan	dan Teknologi Dasar Kesehatan	
Shaheed		
Beheshti		
University		
of Medical	Kowsarmedical	<a href="http://journals.sbm.ac.ir/jme">http://journals.sbm.ac.ir/jme</a>
Sciences		
and Health		
Services		

Table 6 – (2)

Invalid URLs (examples):

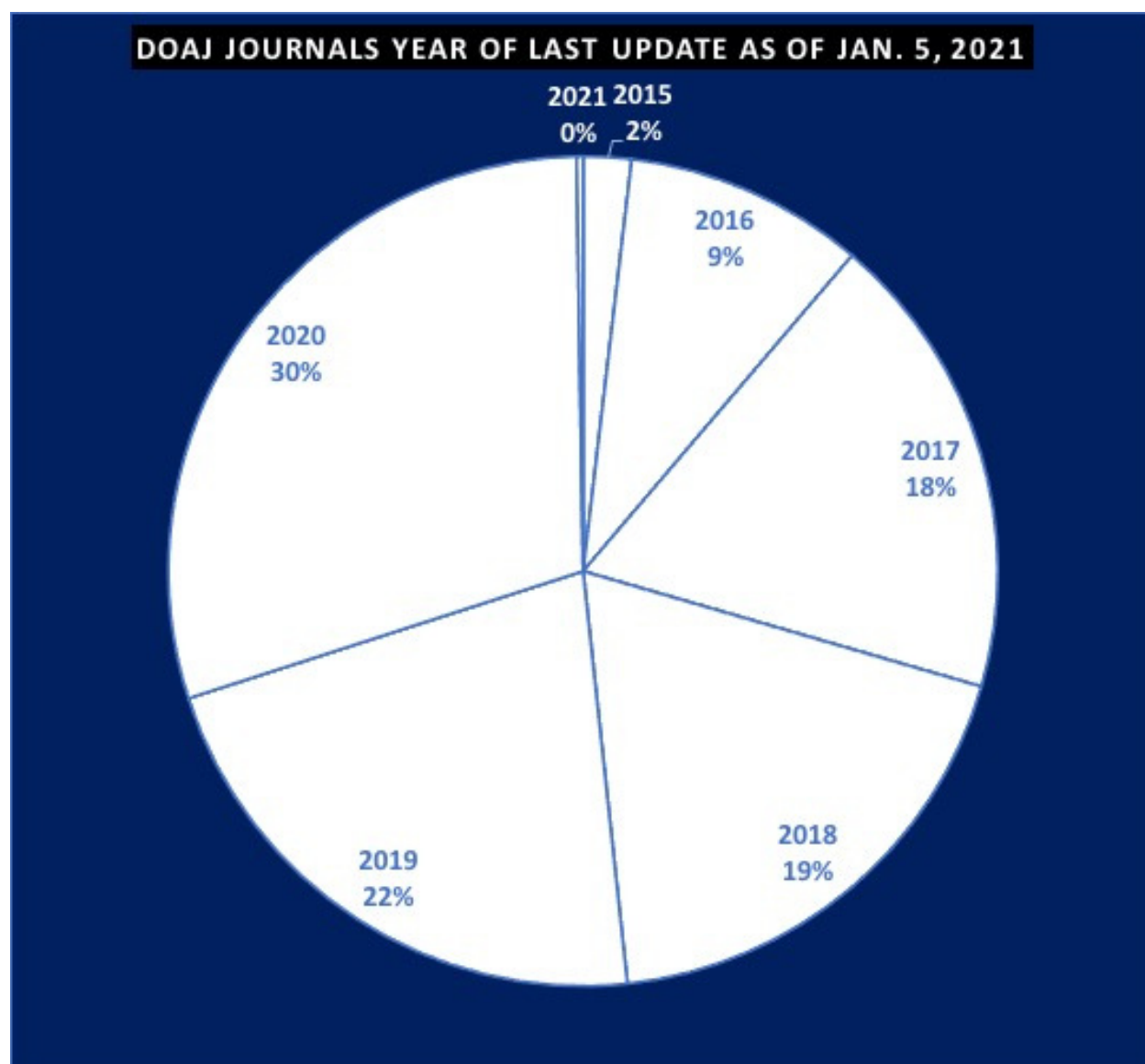
Original publisher names	Original URLs (invalid)
Alborz University of Medical Sciences (URLs wrongly directs to a website whose contents are meaningless; when we searched the journal title, we were directed to this website : <a href="https://enterpathog.abzums.ac.ir/">https://enterpathog.abzums.ac.ir/</a> )	<a href="http://enterpathog.com/?page=home">http://enterpathog.com/?page=home</a> ; <a href="https://jehe.abzums.ac.ir/index.php?slc_lang=en&amp;sid=1">https://jehe.abzums.ac.ir/index.php?slc_lang=en&amp;sid=1</a>
Instituto Nacional de Salud (INS)	<a href="http://revistas.ins.gov.py/index.php/rspp/">http://revistas.ins.gov.py/index.php/rspp/</a>
Instituto Superior de Ciências de Educação do Huambo	<a href="http://revista.isced-hbo.ed.ao/rop/index.php/ROP/index">http://revista.isced-hbo.ed.ao/rop/index.php/ROP/index</a>

Table 6 – (3)

Given the barriers and challenges mentioned above, we can draw a conclusion to the limitations of publisher names clean-up project. Precision is not possible in this project because the question “who is the publisher” is complex. Instead of making any definitive claims about publisher size, we are primarily interested in whether the long tail effect (a few big publishers, a few more middle-sized, most very small) reported by Solomon & Björk (2012) can still be observed in DOAJ in 2021.

### DOAJ metadata update analysis

The following analysis was conducted to determine whether DOAJ metadata on article processing charges (APCs) – charging status and amount – would be sufficient for SKC’s longitudinal study on APC trends over time. The answer is clearly no. The metadata for the vast majority of journals in DOAJ (overall and APC charging) has not been updated for more than a year, and it is unknown whether the most recent update would have included an update to APC or other metadata. We will continue to use DOAJ metadata as it is rich and the question “is this journal listed in DOAJ” is of value in and of itself, however for price comparisons we cannot rely on this data as it would likely result in erroneous conclusions.



DOAJ journals by year of last update.

This chart illustrates the percentage of DOAJ journals last update by year. Detailed figures are in the table below. Note that just under half the journals were last updated 2 or more years ago (2018 or earlier).

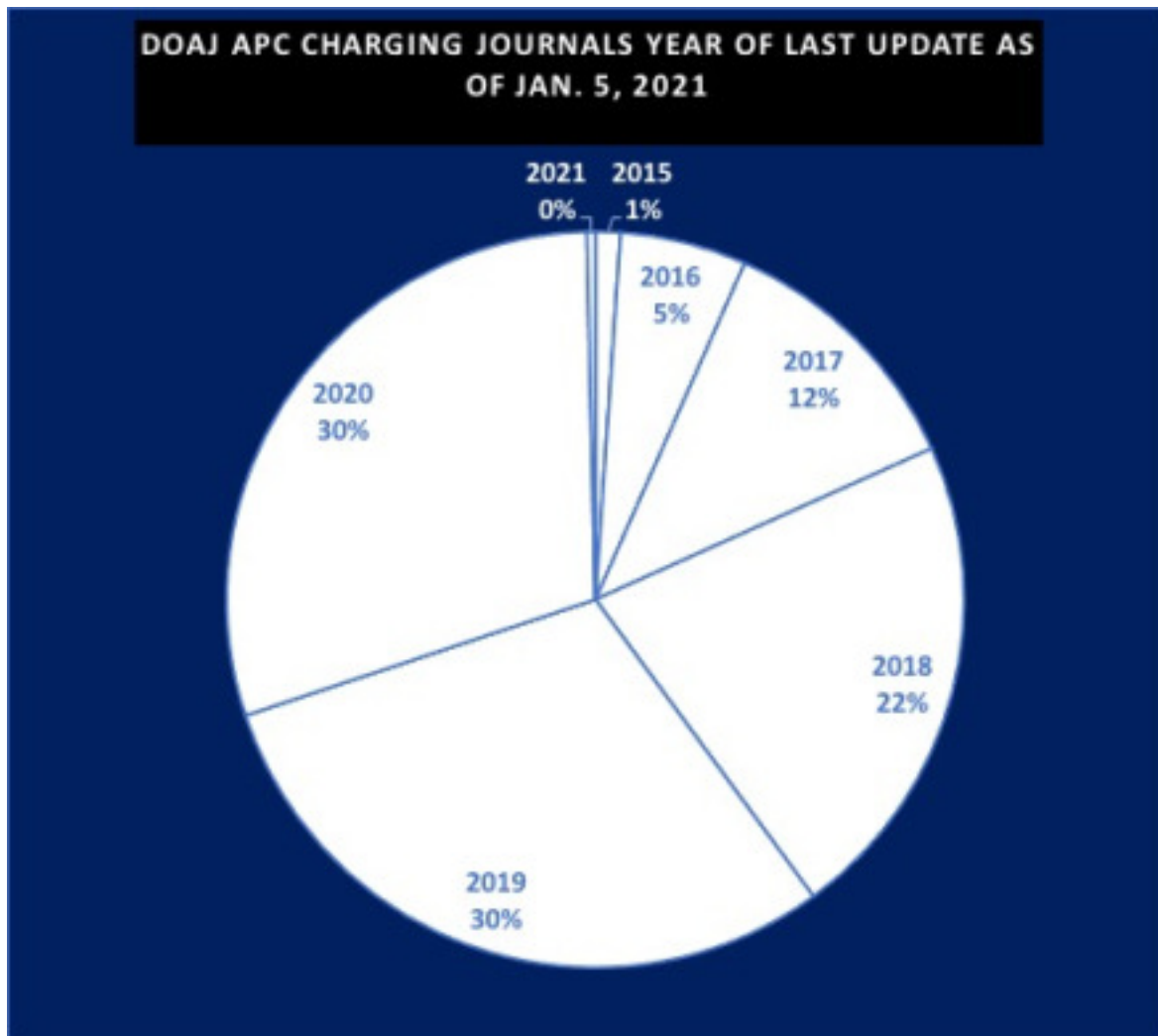
DOAJ last update as of Jan. 5,  
2021

Year	# journals last updated	% journals last updated
2015	294	2%
2016	1,469	9%
2017	2,864	18%
2018	2,951	19%



2019	3,412	22%
2020	4,662	30%
2021	39	0%
Total	15,691	100%

Table 7



DOAJ APC charging journals by year of last update

The chart above illustrates the percentage of journals that answered “yes” to the DOAJ question about charging APCs by year of last update. The table below provides the detailed figures. Note that only 30% of DOAJ journals that charge APCs were updated in the past year (2020 or 2021). It is also unknown whether in these cases the last update was a thorough review of the metadata, or might have been an update of non-APC data.

DOAJ last update APC journals only Jan. 5, 2021

Year of last update	# of journals last updated	% journals last updated
---------------------	----------------------------	-------------------------

2015	47	1%
2016	238	6%
2017	499	12%
2018	930	22%
2019	1,286	30%
2020	1,276	30%
2021	16	0%
Total	4,292	100%

Table 8

A version of the Jan. 5, 2021 DOAJ metadata file reflecting the corrections explained below is available as open data here:

Directory of Open Access Journals; Zhao, Xuan; Borges, Luan; Morrison, Heather, 2021, "DOAJ\_metadata\_2021\_01\_05\_with\_SKC\_clean\_up", <https://doi.org/10.5683/SP2/G5LEXG>, Scholars Portal Dataverse, V1

## References

The Directory of Open Access Journals (DOAJ) online: <https://doaj.org/>

Solomon, D. J., & Björk, B. (2012). A study of open access journals using article processing charges. *Journal of the American Society for Information Science and Technology*, 63(8), 1485–1495. <https://doi.org/10.1002/asi.22673>

Cite as: Zhao, X., Borges, L., & Morrison, H. (2021). Some limitations of DOAJ metadata for research purposes. *Sustaining the Knowledge Commons*. <https://sustainingknowledgecommons.org/2021/02/10/some-limitations-of-doaj-metadata-for-research-purposes/>