



---

## The size distribution of open access publishers: A problem for open access?

by Jan Erik Frantsvåg

---

### Abstract

I stumbled across the question of publisher size while preparing for an earlier article. From the viewpoint of an economist, the size distribution of open access publishers looked inefficient. In this article I first explore reasons to be sceptical to a situation with a large number of small publishers. Then I go through the numbers from the *Directory of Open Access Journals*, also discussing problems inherent in the material. The results are then compared to similar data about toll access publishing. A conclusion is that, even though numbers may lack in exactitude, there seems to be a need for institutions to look at how they organize their publishing activities.

### Contents

[Introduction](#)

[Method](#)

[Results](#)

[Consequences](#)

[Toll access publishing](#)

[Conclusion](#)

---

### Introduction

Analyzing data for a previous article (Frantsvåg, 2010), I stumbled across the question of the size distribution of open access (OA) publishers. A quick analysis of *Directory of Open Access Journals (DOAJ)* data at that time suggested that the distribution was extremely skewed, with single journal publishers dominating at nearly 90 percent of all publishers, publishing the majority of OA journals. Larger publishers were few, and accounted for about 25 percent of all journals. OA publishing seems, in other words, to be a small-scale industry.

Why should this concern us? In economics, we learn that size is important in various ways. Economics is — in this context — not to be confused with profit maximization, it is the science of how one household's resources in order to maximize one's goals. For a commercial publisher, this could mean maximizing profits. For an institution-based scholarly publisher, this goal to be maximized could be dissemination of scientific content. In this article, I will take dissemination of scientific content to be the primary goal of small OA publishers. Any resource use that is inefficient will be detrimental to this goal — increased efficiency will bring more dissemination at the same cost, or the same level of dissemination at a lower cost. Both situations will be more advantageous to science than a less efficient one.

How does size influence economic efficiency? There are two major concepts that we should look at here, "(average) fixed costs" and "economies of scale". These concepts are treated in most general introductory economics textbooks [1]. In the discussion, I will look on a journal as an institution ("factory") that produces quality-assured scientific articles.

#### Fixed costs

Fixed costs are costs that are the same whatever the size of the output (here: the number of articles

published). In OA journal publishing this could be annual service payments to some external or internal service provider, annual fees for some subscription, etc. It also covers the annual depreciation of initial investments that has been made, *i.e.*, the investment cost is divided over the expected life of the investment. For example, an editor in a small journal often has to invest a lot of time in learning the technicalities of the publishing system, the hows and whys of getting the journal listed and embedded in various Internet services, etc. These costs are seen to be “used” over the time the editor works for the journal.

The sum of these annual costs divided by the number of articles published becomes the average fixed cost of an article. Clearly, the higher the number of articles, the lower the average fixed cost.

### Economies of scale

Economies of scale (not to be confused with “returns to scale”, which is a very theoretical concept not to be discussed here) is about the efficiency of the production. Economic theory (based on observation of reality) holds that as production increases, the effort of producing another unit (in this discussion: another article) could become lower. This is because the persons involved in the production gains knowledge and experience, and also because increased production gradually will allow specialization among the persons involved. People will increasingly specialize and concentrate on the elements of the production process that they are good at, instead of having to do everything. This also means that the average variable cost (the costs that are not fixed) of producing an article will decrease as the volume of articles in a journal, or in a publishing venture, increases.

### Other size-related matters

Polydoratou, *et al.* 2010) and Polydoratou and Schimmer (2010) found that there are important differences in financing between smaller and larger publishers, with the smaller ones being more reliant on sponsorship for their operations. These findings are further explored in Dallmeier-Tiessen, *et al.* (2010).

A quick glance at the *DOAJ* Web site (<http://www.doaj.org/>, as of 21 October 2010) indicates that less than half (2,361 of 5,542) of the journals are searchable at article level. Delivering article metadata is an important function in order to have content widely disseminated. Larger publishers generally make sure their content metadata is to be found in *DOAJ*. It is the smaller publishers who do not use this functionality, which is free to use and — provided you have the right software — simple to use. Smaller publishers do not have the right software or minimal resources needed to exploit this service — or they lack the necessary competence to see the need to use it. Whatever the reason, the result is less efficient dissemination of their content.

Hicks and Wang (in press) shows that journals from smaller publishers suffer disadvantages in being recognized as scholarly and in being efficiently distributed.

### Consequences

All these elements suggest that small-scale operation of OA publishing is economically inefficient, and that OA publishing best be organized in larger publishing institutions.

In this article I will try to establish some facts about the size distribution of OA publishers in terms of journals published, and discuss what consequences the findings could have. This will also lead to a discussion of similar traits in TA publishing.



## Method

While it is the number of articles that really says something about the size of publishers in terms of efficiency, in the following discussion I will use the number of journals per publisher as an indication of size. This is because these data are more readily available than the number of articles published. The *Directory of Open Access Journals (DOAJ)*, considered to be the best source of information about OA journals) has no data on the annual output in terms of articles published of the journals they list. They have information on the number of articles publishers have uploaded metadata for, but these numbers haven't been made available and they could also be misleading as there is much uploading of metadata for old articles.

On 5 August 2010, I downloaded a file of all *DOAJ* journals from the *DOAJ* Web site, using the link <http://www.doaj.org/doaj?func=csv> on the *DOAJ* FAQ Web page (<http://www.doaj.org/doaj?func=loadTempl&templ=faq>). An Excel file containing both the downloaded raw data and the various tables and numbers constructed on the basis of these data is available as supplementary files to accompany this article.

The raw data file contained 5,256 titles. 95 titles that were listed with a value in the column “End Year” were excluded from further analysis, leaving a total of 5,161 titles.

Using the pivot functionality in MS Excel, a table listing all unique publishers in the raw data file and giving the number of journals listed per publisher was constructed. In this table, 3,316 unique publishers were listed. However, on inspection a number of publishers were listed twice, and this turned out to be due to extraneous blanks in the Publisher field for some records. Removing these extraneous blanks, using Excel's "trim" function, reduced the number of publishers to 3,231.

### Publishers — sources of error

The sorting and counting by publisher is based on the content of the field "Publisher" in the *DOAJ* file. There are a number of problems with this approach.

One problem is the trivial one mentioned earlier, that the occurrence of some extraneous blanks makes Excel distinguish between publishers that are really one and the same. This source of error is easily corrected — if it is detected.

A source of error closely connected to this is spelling variants — *e.g.*, the use of "and" and "&" — and different names of the same publisher in different languages. There is no reason to believe that "Igitur, Utrecht Publishing & Archiving Services" and "Igitur, Utrecht Publishing and Archiving Services" actually are two different publishers. There are a number of such "near misses" in the data, but as I have no secure method of discerning between intended and non-intended small differences in spelling, I have chosen not to correct even the errors that seem self-evident errors. This will mean that the number of publishers in the analysis is larger than what is really true, bringing the average number of journals per publisher lower than what is correct [2].

A source of error that is more difficult to handle and to have some idea of the size of, is what is meant by publisher in the *DOAJ* data set. Is this the editorial organization, or is it the publishing institution — and if institution, on what level? And will similarities or dissimilarities indicate whether the technical production — which is the most important when it comes to production cost — is organized as larger units common to a number of journals, or is organized by the individual journals? One could imagine that a number of journals have the same publisher listed in *DOAJ* (*e.g.*, the publisher could be a university), but the production could still be done by individual journals. On the other hand, one could imagine institutions where production is centralized but where the publisher listed is the institute the journal belongs to, etc. I suspect that we will find errors both ways in the data, but I find it probable that this kind of error on the whole also will tend to exaggerate the number of small or single journal publishers.

In *DOAJ*, we only have information about OA journals. We know, however, that some OA journals are published by publishers that also publish toll access (TA) journals. The efficiency of a publisher — as we discuss it here — is not dependent upon whether their journals are published OA or TA, but on their total size. In order to gain some idea of this (and other aspects of what is discussed here) we bought a data set from *Ulrich's Periodicals Directory* (received on 22 February 2010) containing active, academic, refereed journals from their publication database. Unfortunately, *Ulrich's* forbid me to post these data with this paper. In the Ulrich file, we found 9,970 publishers publishing 24,263 journals. Of these publishers, 229 publish both OA and TA journals: 996 OA and 3,016 TA, making a total of 4,012 journals.

Table 1: Publishers with both OA and TA journals, by number of journals in each category.																
		Number of TA journals														
		1	2	3	4	5	6	7	8	9	10	11–20	21–50	51–100	101–	Total
Number of TA journals	1															
	1	93	24	9	8	3	2	6	2	3	1	8	3	2	4	168
	2	4	3	3	1	1	1	1	1	1			3	3	1	23
	3	2	3	2	1	1		1		1						11
	4	1	1			1	1	1				2			1	8
	5	3				1										4
	6–10	1	1	1	1	1						1				6
	11–20	1	1													2
	21–50			1	1		1									3
	51–100						1									1
	100+												3			3

Total	105	33	16	12	8	6	9	3	5	1	11	9	4	6	229
-------	-----	----	----	----	---	---	---	---	---	---	----	---	---	---	-----

Note: N=229.

Statistical data derived from *Ulrich's Periodicals Directory*, © 2010 ProQuest LLC. All rights reserved.

The picture we get is that publishing both OA and TA only occurs with a minority of OA publishers. Of 1,633 OA publishers in *Ulrich's*, 229 also publish TA. Nearly half of these OA publishers (105) publish one OA journal, 93 of these publish one TA journal additionally. The general picture is that of small OA publishers publishing few additional TA journals, and that additional TA publishing doesn't really alter the publisher size of OA publishers in any substantial way. Few small OA publishers would become significantly larger by adding TA publishing. Of course, there are exceptions — most notably Routledge, with one OA journal and 837 TA journals.

A fourth source of errors is the existence of OA journals that are not listed in *DOAJ*. This is by definition an unknown quantity. We know that OJS (Open Journals Systems at <http://pkp.sfu.ca/ojs/>), the most common platform for OA journals, has some 6,600 installations. *DOAJ* lists 5,258 journals, many of which do not use OJS. One installation of OJS can cover a virtually unlimited number of journals. On the other hand, OJS is also used for non-OA journals. Comparing the file from *Ulrich's*, we find that of the 2,639 OA journals listed there, 131 journals are not in *DOAJ* if we do a lookup in the *DOAJ* file based on the ISSN in *Ulrich's* file (journals without an ISSN in the *Ulrich's* file being excluded) [3]. We can probably conclude that there are a number of OA journals out there that are not included in our numbers. As listing in *DOAJ* actually is useful for OA journals, we can also conclude that a vast majority of these unlisted journals are published by single journal publishers that are not currently listed in *DOAJ*. This source of error will under-estimate the number of single journal publishers. Whether this source of error is larger than the other sources of error mentioned, is impossible to say, but this last error will at least to some extent even out the effect of the other sources of error.

The possible errors, seen as a whole, indicate that the numbers found in *DOAJ* — and used here — will give a fairly good picture of the current realities when it comes to actual size of OA publishers measured in terms of the number of journals they publish.

## Results

Ordering the publishers by their size in terms of the number of journals they publish and counting the number of publishers that have a given size, we arrive at [Table 2](#) after grouping the larger publishers in size groups. The grouping is somewhat arbitrary, but is chosen so that the numbers could be useful in later analysis.

Table 2: OA publisher size by number of journals.				
	Number		Percentage of total	
Publisher size	Publisher	Journals	Publisher	Journals
1	2,839	2,839	87.9%	55.0%
2	212	424	6.6%	8.2%
3	50	150	1.5%	2.9%
4	30	120	0.9%	2.6%
5	16	80	0.5%	1.6%
6	16	96	0.5%	1.9%
7	13	91	0.4%	1.8%
8	9	72	0.3%	1.4%
9	7	63	0.2%	1.2%
10	2	20	0.1%	0.4%
11–20	26	357	0.8%	6.9%
21–50	6	183	0.2%	3.5%
51–100	2	132	0.1%	2.6%
100+	3	534	0.1%	10.3%

	3,231	5,161	100.0%	100.0%
Note: N=3,231.				

As we see, 87.9 percent of all publishers publish only a single journal (amounting to 55 percent of all journals) while the larger publishers (with more than 10 journals) [4] total 1.1 percent of all publishers and publish 23.3 percent of all journals. We also note that the average number of journals per publisher is near to 1.6, while the median and mode of number of journals per publisher are both 1.

## Consequences

It could seem reasonable, based on what economics tells us, and the data we have, to conclude that open access publishing, as it is organized today, is vastly inefficient compared to traditional publishing and that it either should be abolished or strongly re-organized. The data seem to indicate that OA publishing, as a whole, is ready to take advantage of any diseconomy of scale, and every inefficiency, available. A natural conclusion could be that OA publishing in its present form should be abolished.

## Toll access publishing

But are things really that different in traditional publishing? If we, once again, turn to *Ulrich's* and perform the same analysis, we get the following picture. (In this analysis we exclude publishers that only publish OA, in order to create the maximum contrast between the two sets of data. For TA publishers, the OA journals they publish are counted.)

*Ulrich's* list 9,970 publishers publishing 24,263 journals. Of these, 1404 publishers publish only OA journals, with a total publishing volume of 1,643 journals. The remaining 8,566 publishers publish 22,620 journals (996 OA and 21,624 TA).

Publisher size	Number		Percentage of	
	Publishers	Journals published	Publishers	Journals
1	7,168	7,168	83.7%	31.7%
2	671	1,342	7.8%	5.9%
3	186	558	2.2%	2.5%
4	116	464	1.4%	2.1%
5	76	380	0.9%	1.7%
6	46	276	0.5%	1.2%
7	28	196	0.3%	0.9%
8	29	232	0.3%	1.0%
9	21	189	0.2%	0.8%
10	25	250	0.3%	1.1%
11–20	85	1,221	1.0%	5.4%
21–50	66	1,984	0.8%	8.8%
51–100	22	1,632	0.3%	7.2%
100+	27	6,728	0.3%	29.7%
	8,566	22,620	100.0%	100.0%

\* The number of journals include OA journals, but publishers that only publish OA journals are excluded.

N=8,566.

Statistical data derived from *Ulrich's Periodicals Directory*, © 2010 ProQuest LLC. All rights reserved.

**Table 3**, though different from **Table 2**, shows a remarkable likeness to it in that the single journal publishers dominate, being 83.7 percent of the total number of publishers. For OA publishers (Table 2) the percentage is 87.9.

If we combine Tables 2 and 3, we arrive at Table 4 which makes it easier to compare the two groups of publishers.

Table 4: TA and OA publishers, combined table.								
	TA publishers with or without OA (Ulrich's*)				OA publishers (DOAJ)			
	Number		Percentage of		Number		Percentage of	
Publisher size	Publishers	Journals published	Publishers	Journals	Publishers	Journals	Publishers	Journals
1	7,168	7,168	83.7%	31.7%	2,839	2,839	87.9%	55.0%
2	671	1,342	7.8%	5.9%	212	424	6.6%	8.2%
3	186	558	2.2%	2.5%	50	150	1.5%	2.9%
4	116	464	1.4%	2.1%	30	120	0.9%	2.3%
5	76	380	0.9%	1.7%	16	80	0.5%	1.6%
6	46	276	0.5%	1.2%	16	96	0.5%	1.9%
7	28	196	0.3%	0.9%	13	91	0.4%	1.8%
8	29	232	0.3%	1.0%	9	72	0.3%	1.4%
9	21	189	0.2%	0.8%	7	63	0.2%	1.2%
10	25	250	0.3%	1.1%	2	20	0.1%	0.4%
11-20	85	1,221	1.0%	5.4%	26	357	0.8%	6.9%
21-50	66	1,984	0.8%	0.8%	6	183	0.2%	3.5%
51-100	22	1,632	0.3%	7.2%	2	132	0.1%	2.6%
100+	27	6,728	0.3%	29.7%	23	534	0.1%	10.3%
	8,566	22,620	100.0%	100.0%	3,231	5,161	100.0%	100.0%
Average	2.64				1.60			
Mean	1				1			
Mode	1				1			

\* Statistical data derived from *Ulrich's Periodicals Directory*, © 2010 ProQuest LLC. All rights reserved.

The typical publisher in both TA and OA is a single-journal publisher, and such publishers comprise more than 80 percent of all publishers. In TA publishing there are a larger number and fraction of publishers publishing more than 20 journals than in OA, so that the average number of journals per publisher is higher for TA than for OA. We also see that the fraction of TA journals being published by the largest publishers is greater than for OA journals.

We know that TA publishing over the last decades has undergone restructuring, where medium-sized publishers have been acquired by larger publishers, creating a number of very large publishers. This concentration has not yet started in OA publishing. The difference in average number of journals per publishers could be ascribed to this concentration process, so that we could expect the numbers in OA publishing to become more like TA publishing numbers in the future. The TA publishing business is, after all, centuries old, while OA publishing started about a decade ago.

Both in TA and OA publishing, a general picture — looking at the raw data — is that the professional, commercial publishers are among the larger publishers, while institutional publishers and publishers representing professional societies are small. This is corroborated by Polydoratou, *et al.* (2010).

## Conclusion

This analysis is not without flaws. However, both OA and TA publishing — covering major parts of publishing output — are highly inefficient. Professional publishers are organized in large corporations, having a large number of journals and producing their output in a relatively efficient way. Other publishers — mainly professional societies and academic institutions — have organized their publishing — be it OA or TA — in inefficient ways.

Academic institutions need to look at ways to organize the technical side of publishing in new ways in order to exploit efficiencies of size, so that dissemination can be both less costly and more efficient. This is not an argument against OA publishing; this concerns TA publishing as well. But in a process of transitioning to OA, one should take care to organize OA in a way that will truly serve scholarship.

Much debate over the economics of OA turns around business models for large and medium-sized publishers. The numbers presented here should demonstrate a need to create debate and research that is more relevant to the small OA publishers — after all, they are the vast majority! 📰

## About the author

Jan Erik Frantsvåg is an Open Access Adviser at the University Library of Tromsø.  
E-mail: jan [dot] e [dot] frantsvag [at] uit [dot] no

## Acknowledgments

The work leading to this article has been funded in part by Nordbib, a funding programme under NordForsk, which is financed by the Nordic Council of Ministers, through the project Aiding Scientific Journals Towards Open Access Publishing (NOAP); partly by NORA — Norwegian Open Research Archives (financed by The Norwegian Ministry of Education and Research) and partly by the University of Tromsø.

## Notes

1. See *e.g.*, Sloman and Sutcliffe, 1991, pp. 140–178 or Samuelson and Nordhaus, 1989, pp. 472, 498–531.
2. Hopefully, this discussion could lead to publishers looking through their *DOAJ* data and correcting and standardizing their information there.
3. There is some reason for concern about data quality: Of the 2,639 OA journals in *Ulrich's*, 1,070 has some kind of difference in the journal name between *DOAJ* and *Ulrich's*. Most differences seem to be trivial, but still.
4. The definition of "large" in this context is arbitrary. Any publisher publishing more than four journals is among the 100 largest publishers in this data set.

## References

Suenje Dallmeier-Tiessen, Robert Darby, Bettina Goerner, Jenni Hypoelae, Peter Igo-Kemenes, Deborah Kahn, Simon Lambert, Anja Lengenfelder, Chris Leonard, Salvatore Mele, Panayiota Polydoratou, David Ross, Sergio Ruiz-Perez, Ralf Schimmer, Mark Swaisland, and Wim van der Stelt, 2010. "First results of the SOAP project. Open access publishing in 2010" (14 September), at <http://arxiv.org/ftp/arxiv/papers/1010/1010.0506.pdf>, accessed 28 November 2010.

Directory of Open Access Journals, at <http://www.doaj.org/>, accessed 21 October 2010.

Jan Erik Frantsvåg, 2010. "The role of advertising in financing open access journals," *First Monday*, volume 15, number 3, at <http://www.uic.edu/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/2777/2478>, accessed 28 November 2010.

Diana Hicks and Jian Wang, in press. "Coverage and overlap of the new social science and humanities journal lists," *Journal of the American Society for Information Science and Technology* (forthcoming), and at [http://works.bepress.com/diana\\_hicks/22/](http://works.bepress.com/diana_hicks/22/), accessed 28 November 2010.

Paniyoti Polydoratou and Ralf Schimmer, 2010. "Income sources as underlying business models' attributes for scholarly journals: Preliminary findings from analysing open access journals data," *Proceedings ELPUB2010, — Conference on Electronic Publishing (Helsinki)*, at [http://elpub.scix.net/data/works/att/999\\_elpub2010.content.pdf](http://elpub.scix.net/data/works/att/999_elpub2010.content.pdf), accessed 28 November 2010.

Panayiota Polydoratou, Margit Palzenberger, Ralf Schimmer, and Salvatore Mele. 2010. Open access publishing: An initial discussion of income sources, scholarly journals and publishers. In: Gobinda Chowdhury, Chris Koo and Jane Hunter (editors). *The role of digital libraries in a time of global change. Lecture Notes in Computer Science*, number 6102. Berlin: Springer-Verlag, pp. 250–253, and at <http://www.springerlink.com/content/627663m2q8320282/fulltext.pdf>, accessed 28 November 2010.

Paul A. Samuelson and William D. Nordhaus. 1989. *Economics*. Thirteenth edition. New York: McGraw-Hill.

John Sloman and Mark Sutcliffe. 1991. *Economics*. New York: Harvester Wheatsheaf.

---

## Editorial history

Received 21 October 2010; accepted 28 November 2010.

---



"The size distribution of open access publishers: A problem for open access?" by Jan Erik Frantsvåg is licensed under a [Creative Commons Attribution 3.0 Norway License](http://creativecommons.org/licenses/by/3.0/norway/).

The size distribution of open access publishers: A problem for open access?

by Jan Erik Frantsvåg.

*First Monday*, Volume 15, Number 12 - 6 December 2010

<http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/rt/printerFriendly/3208/2726>