

Income sources as underlying business models' attributes for scholarly journals: preliminary findings from analysing open access journals data

Panayiota Polydoratou and Ralf Schimmer

{polydoratou, schimmer}@mpdl.mpg.de

On behalf of the SOAP project¹

Abstract

The Study for Open Access Publishing (SOAP) project is one of the initiatives undertaken to explore the risks and opportunities of the transition to open access publishing. Some of the early analyses of open access journals listed in the Directory of Open Access Journals (DOAJ) show that more than half of the open access publishing initiatives (56%) were undertaken by smaller publishers associated with a small number of journals. The study differentiates between 14 large publishers and other publishers. The 14 large publishers publish more than 50 journals each or 1000 per year (data as per 2007 or 2008). Regarding income sources as means for sustaining a journal's functions, "article processing charges", "membership fee" and "advertisement" are the predominant options for the large publishers (publishers associated with more than 50 journals or 1000 articles); "subscription to the print version of the journal", "sponsorship" and somewhat less the "article processing charges" have the highest incidences for all other publishers.

Keywords: Open access publishing, business models, study of open access publishing

¹ Important notice: The research results of this Project are co-funded by the European Commission under the FP7 Research Infrastructures Grant Agreement Nr. 230220. This document contains material, which is the copyright of certain SOAP beneficiaries, and may not be reproduced or copied without permission. The information herein does only reflect the views of its authors and not those of the European Commission. The European Commission and the beneficiaries do not warrant that the information contained herein is capable of use, or that use of the information is free from risk, and they are not responsible for any use that might be made of data appearing herein.

1. Introduction

Open access literature is online, free of charge for all readers, and permits its distribution and further use for research, education and other purposes.

Activities around scholarly publication and dissemination of research from all stakeholders involved in the scholarly communication process demonstrate the increased recognition of open access advantages. Funding bodies are requiring more and more that research outputs funded with public money are made openly accessible; academic institutions mandate open access and encourage scholars to actively practice it; libraries develop policies in support and publishers experiment with new models for open access publishing.

In Europe, the European Commission recognised the need to examine the potential for change in the scholarly publishing arena² and explore initiatives that would make suggestions at policy level for a smooth transition to Open Access. The SOAP project is one of the initiatives undertaken to explore the risks and opportunities of the transition to full open access publishing.

The Study of Open Access Publishing (SOAP) project

The Study of Open Access Publishing (SOAP) is a two-year project, funded by the European Commission under FP7 (Seventh Framework Programme). The project is co-ordinated by CERN, the European Organization for Nuclear Research and the SOAP consortium represents key stakeholders such as publishers (BioMed Central Ltd (BMC), SAGE (SAGE Publications Ltd) and Springer Science+Business Media Deutschland GmbH (SSBM), funding agencies (Science and Technology Facilities Council (STFC) UK), libraries (Max Planck Digital Library) and a broad spectrum of research disciplines.

The SOAP project aims to deliver to the European Commission, publishers, libraries and research communities a description and analysis of models of open access publishing, so that these key players may ascertain which model, or combination and variation of models, will enable them to make a smooth transition to open access publishing.

The project objectives are stipulated in the following:

- The SOAP project will describe and analyse open access publishing. It will compare and contrast business models. Such an approach will allow for a better understanding of the marketplace as well as the opportunities and risks associated with open access publishing.

² http://ec.europa.eu/research/science-society/document_library/pdf_06/scientific-info-resultscrest-final-090609_en.pdf

*Scholarly journals and underlying business models' attributes: preliminary findings
from analysing DOAJ journal level metadata*

- The SOAP project will conduct a large scale survey that will investigate the European Research Area (ERA) scholars' requirements for scientific publishing. It is anticipated that the survey's findings will uncover what researchers as authors are willing to trade off (and what they are not) in the transition to open access publishing.

This paper presents preliminary results pertaining to the former of our objectives and specifically concentrating on income sources as one of business models' attributes for sustaining scholarly journals' operations. Income sources are discussed in association with a) publisher size³ and b) journal's subject coverage.

2. Scope, definitions and information sources

This paper concerns open access journals or otherwise known as "gold" open access. Business models for open access publishing have been in the spotlight of scholarly publishing for many years. The notion of a business model for open access publishing carries a different weight for publishers, researchers and libraries. It commonly includes aspects such as "client" segment (author, reader, funder, library, etc.), income sources (subscription, advertisement, grant, etc.), structure/hierarchies related to meeting costs and value proposition for these different "clients". The indicators for assessing the business models from the various standpoints are many⁴. This paper presents aspects of the "income sources" as one of the key differences from the traditional "pay for access" models. Definitions pertaining to income sources used in this paper are defined as follows:

- (a) article processing charge: a charge applied for the processing of an article. It might be requested at various stages of the publishing process, e.g. at submission or on acceptance. There was no differentiation for these variations.
- (b) membership fee: journal income via a membership option.

³ Size is defined by the number of journals or articles that are associated with a given publisher in DOAJ

⁴ Some examples include the following reports from various professional bodies:

SPARC/ARL, <http://www.arl.org/resources/pubs/rli/archive/rli266.shtml> and http://www.arl.org/sparc/bm-doc/incomemodels_v1.pdf;

ALPSP, http://www.alpsp.org/ngen_public/article.asp?id=200&did=47&aid=270&st=&oid=-1;

STM, http://www.stm-assoc.org/2009_10_13_MWC_STM_Report.pdf?PHPSESSID=dcd8480886aa0a262a4751e315910863;

- (c) advertisement: journal income through accepting and hosting advertisements.
- (d) sponsorship: journal income through sponsorship, by single or multiple institutions/organisations or at an individual level.
- (e) subsidy: financial assistance by an organisation hosting the publishing activity or by a funding agency concerned by assuring that the publishing activity remains ongoing.
- (f) subscription: income from subscription to the print version of the journal.
- (g) hard copy: income from hard copy sales, either individual volumes or the archives of the journal with some given periodicity.
- (h) consortium: income from the fact the journal was offered as part of a library consortium agreement.
- (x) other: groups some of the other sources for income such as: article page charge, colour page charge, off-prints and reprints sales, print on demand, income via conference fees, donations, services to authors (copy-editing, proof reading, etc.), sales in other formats than hard-copy (e.g. CD-ROM with digital archives).

Directories of scholarly journals have long been a means to increase the visibility and use of journals. The reference directory for open access journals is the DOAJ⁵. It was set up in 2003 by the Lund University Libraries with the support of the Open Society Institute. It lists more than 4,000 Open Access journals published by over 2,000 publishers, often via different platforms and in different languages.

The DOAJ was selected as the primary source of data for this study owing to the following criteria:

- Reputation and visibility as the most comprehensive registry of open access scholarly journals.
- Quality control of open access journals as being either peer-reviewed or having other forms of editorial assurance⁶.
- Availability of an initial sample of descriptive metadata on publisher and journal information with ease of access.
- Permission to locally ingest and further enrich the data.

⁵ <http://www.doaj.org/doi/func=suggest>.

⁶ Description of the DOAJ quality control criteria: <http://www.doaj.org/doi/func=loadTempl&templ=about#criteria>.

3. Methodology

A data file of journal-level metadata was downloaded from the DOAJ website in the form of a spreadsheet during July 2009. The data retrieved were parsed and assigned to a relational database structure using PHP and SQL. Final entities comprised what are later referred to as: "journal title", "publisher" and "subject heading".

The data was enriched with additional information such as the number of published articles per year, the publication end date and the journal impact factor. This information was extracted from the following data sources, through an ISSN matching at the journal level with the DOAJ record:

- The Electronic Journals Library (EZB)⁷; data as of year 2009.
- SCOPUS⁸; data as of year 2009.
- Journal Citation Reports (ISI-JCR)⁹; data for year 2008; retrieved in 2009.
- SCImago Journal & Country Rank (SCImago)¹⁰; data for year 2008; retrieved in 2009.

Additional information on the journals and publishers was manually collected between September 2009 and January 2010, with some subsequent double-checking and corrections during the analysis phase. The information was investigated and collected from the websites of the journals and publishers.

4. Results

Publisher information –size

The DOAJ data file listed 4,032 unique journal records corresponding to 2,588 publisher names.

More than half of the publishing activity (56% of the journals) is conducted by small publishers associated with one journal only. Less than a quarter (21%) of the journals are produced by publishers who own between 2 and 9 journals and 9% own between 10 and 49 journals. There are only five

⁷ <http://rzblx1.uni-regensburg.de>.

⁸ <http://www.scopus.com>.

⁹ http://thomsonreuters.com/products_services/science/science_products/a-z/journal_citation_reports.

¹⁰ <http://www.scimagojr.com/journalrank.php>.

publishers with more than 50 journals titles each (14%). Those publishers are: Bentham open, BioMed Central, Hindawi Publishing Corporation, Internet Scientific Publications – LLC and Medknow Publications (Table 1).

Table 1: “Size” of publishers by number of open access journals a) for all DOAJ records b) for records selected for the SOAP study

size of publisher by number of DOAJ journals	DOAJ publishers		DOAJ journals *	
1	2,271	88 %	2,271	56 %
2 to 9	287	11 %	849	21 %
10 to 49	25	1 %	358	9 %
≥ 50	5	0 %	554	14 %
Total	2,588		4,032	

*The DOAJ journals columns list the number of journals (and their relative value) associated with the different publishers by size. For example, there are 2,271 journals published by 2,271 publishers, 849 journals associated with 287 publishers that publish between 2-9 journals, 358 journals associated with 25 publishers that publish between 10-49 journals and 554 journals that are published by 5 publishers.

size of publisher by number of selected journals	publishers		Journals **		estimated articles per year	
1	1,621	90 %	1,621	57 %	63,887	55 %
2 to 9	171	9 %	491	17 %	25,442	22 %
10 to 49	12	1 %	190	7 %	12,623	11 %
≥ 50	5	0 %	536	19 %	14,931	13 %
Total	1,809		2,838		116,883	

** The journals column lists the number of journals (and their relative value) associated with the different publishers by size for the selected DOAJ sample that was analysed in this study.

The total number of articles per publisher and year is also considerably skewed as presented in Table 1 and

Table 2. Most of the publishers selected (~90%) publish less than 100 articles per year and altogether contribute approximately one third of the total articles estimated. The remaining two thirds of the articles are published by only 10% of the publishers selected. Only 13 publishers (1%) publish more than 1,000 articles per year and account for 30% of the annual articles appearing in the journals selected for this study.

*Scholarly journals and underlying business models' attributes: preliminary findings
from analysing DOAJ journal level metadata*

Table 2: "Size" of publishers by number of articles per year

size of publisher by number of articles	publishers		selected journals		estimated articles per year	
missing	91	5 %	94	3 %		
0 to 9	318	18 %	326	11 %	1,852	2 %
10 to 99	1,212	67 %	1,357	48 %	40,004	34 %
100 to 999	175	10 %	507	18 %	39,588	34 %
≥ 1000	13	1 %	554	20 %	35,439	30 %
Total	1,809		2,838		116,883	

The uneven distribution suggests it is valid to aggregate publishers into two categories: large publishers and other publishers. A publisher is a "large publisher" if either of two criteria is fulfilled: they published more than 50 journals or more than 1,000 articles in 2007 or 2008. These criteria selected 14 large publishers, which are listed in Table 3.

Table 3: The 14 large publishers identified in this study, ordered by number of articles per year

shorthand	full name	number of journals	articles per year
bmc	BioMed Central	176	8,993
iucr	International Union of Crystallography	1	5,165
plos	Public Library of Science	7	4,368
ansi	Asian Network for Scientific Information	13	2,514
hindawi	Hindawi Publishing Corporation	85	2,044
copernicus	Copernicus Publications	18	2,012
osa	Optical Society of America	1	1,961
waset	World Academy of Science, Engineering and Technology	18	1,960
bentham	Bentham Open	154	1,663
medknow	Medknow Publications	59	1,574
ias	Indian Academy of Sciences	10	1,152
oup	Oxford University Press	2	1,032
acadj	Academic Journals	10	1,001
ispub	Internet Scientific Publications	62	657

Income funds

Between October 2009 and January 2010, the project partners manually collected information about visible income funds of the journals from their websites.

The following table lists the seven income sources that were investigated and gives their relative share [%] at the level of journal title. The selection of income sources allowed for multiple responses. "Article processing charges", "membership fee" and "advertisement" are the predominant options for the large publishers, whereas "subscription", "sponsorship" and somewhat less the "article processing charges" have the highest incidences for all other publishers. However one should take into consideration that these findings differ at article level as compared to journal level which is discussed here.

Information on income sources was available for almost all of the 620 journals of the large publishers but retrievable only for 1,338 (60%) of the journals from the other publishers. The results are presented in

Table 4 and Figure 1.

*Scholarly journals and underlying business models' attributes: preliminary findings
from analysing DOAJ journal level metadata*

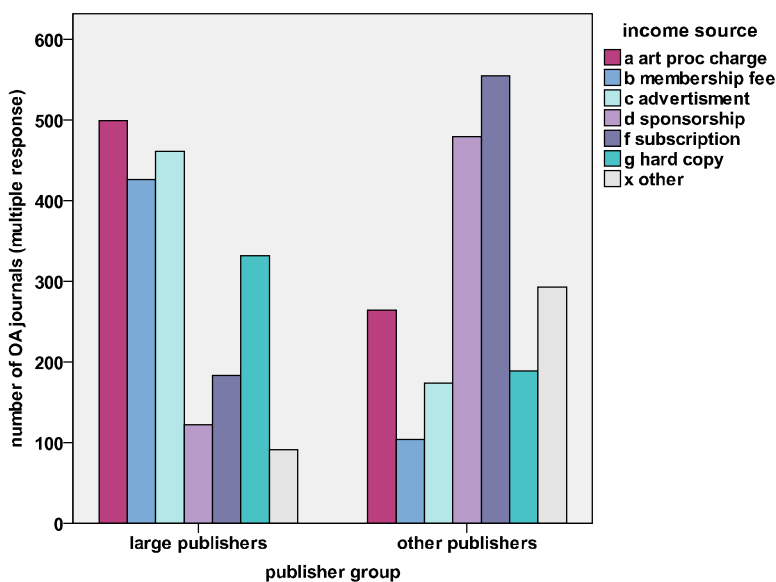
The first column in

Table 4 represents the publisher codes for the large publisher, the second column the number of journals. The following two columns show the total number of journals for which the information was retrieved. The last seven columns represent the percentage of journals published by the publisher which appear to have such an income stream (a-article processing charge, b-membership fee, c-advertisement, d-sponsorship, f-subscription, g-hard copy x-other). The last two rows represent total figures. Given the use of multiple possibilities, the percentages in the last seven columns exceed 100%.

There is no substantial prevalence of any of the eight specified income options. Their relative importance changes depending on publisher size. Large publishers do have a considerably higher incidence of "article processing charge", "membership fees" and "advertisements" as income sources than the other publishers. For the latter "article processing charges" still appear, but this is rarely the case for "membership fees" and "advertisements". "Sponsorship" and "print subscriptions" play a comparably smaller role for the large publishers, whereas these are the most frequent sources amongst the other publishers. Hard copy sales are at an intermediate position for both groups.

Table 4: Income sources for journals by publisher.

publisher	journals		info found	fraction of journals with income source						
				a	b	c	d	f	g	x
acadj	10	10	100%	all	-	all	-	-	-	all
ansi	13	13	100%	-	-	-	-	all	15 %	all
bentham	154	154	100%	all	all	all	-	-	99 %	1 %
bmc	176	176	100%	97 %	96 %	99 %	-	1 %	all	-
copernicus	18	18	100%	83 %	83 %	-	-	83 %	22 %	all
hindawi	85	85	100%	all	all	-	-	all	-	-
ias	10	10	100%	-	-	10 %	-	all	30 %	10 %
ispub	62	62	100%	all	-	all	all	-	-	all
medknow	59	59	100%	-	-	all	all	all	-	all
osa	1	1	100%	all	-	-	-	-	-	-
oup	2	2	100%	all	-	-	-	50 %	-	all
plos	7	7	100%	all	all	all	all	29 %	-	all
iucr	1	1	100%	all	-	all	-	-	-	all
waset	18	18	100%	-	-	-	-	-	-	all
large	616	616	100 %	82 %	70 %	76 %	21 %	30 %	55 %	31 %
other	2222	1338	60 %	20 %	8 %	13 %	36 %	42 %	14 %	22 %



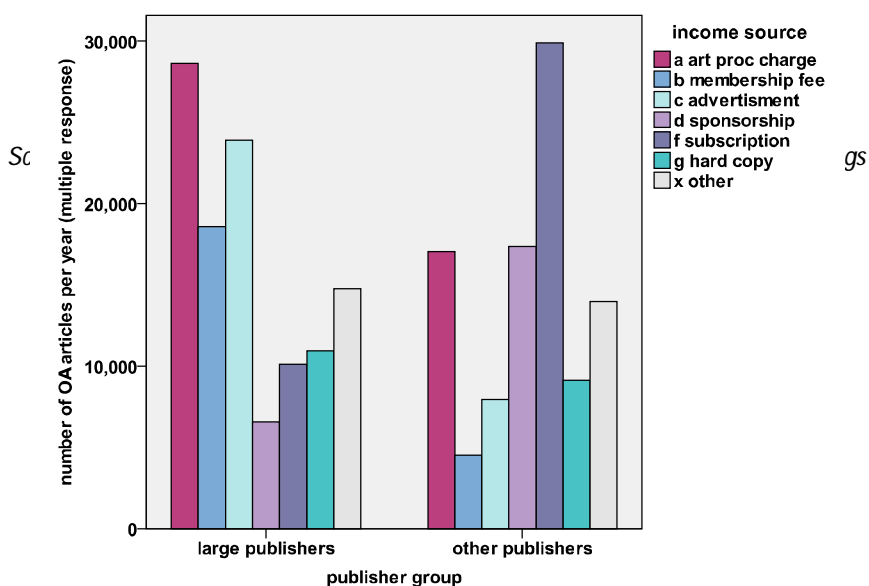


Figure 1: Number of journals and articles as a function of the income source of publishers, for the large publishers and the other publishers.

5. Discussion and future work

The preliminary findings presented above aim to further the understanding of the current existing open access offering from many publishers, using the DOAJ as an entry point. A similar approach has been followed in the past. For example, Kaufman-Wills (2005)¹¹, Dewatripont (2006)¹², Regazzi (2004)¹³, Morris (2006)¹⁴ used data from the DOAJ in their studies addressing open access journals, number of articles for journals indexed in ISI-JCR, frequency of use of an article processing fee. The results augment the existing body of knowledge for the following reasons:

- Article level information was not only collected for journals indexed in ISI-JCR or SCOPUS but for a wider set.
- Income sources as a means to sustain the functional operation were investigated in detail, beyond the article-processing-charge attribute, which was the focus of similar analyses.

While this approach brings new aspects and insight into the open access debate, it must be remembered that our sample did not cover the entire DOAJ sample, of 4,032 journals at the time of the data extraction. Some 1,200

11 http://www.alpssp.org/ngen_public/article.asp?id=200&did=47&aid=270&st=&oid=-1

12 http://ec.europa.eu/research/science.../scientific-publication-study_en.pdf

13 DOI: 10.1016/j.serrev.2004.09.010.

14 DOI: 10.1087/095315106775122565

journals were removed from the original DOAJ sample as not in the English language. This decision stems from the analysis of open access as a global issue, where journals have an offer beyond their national borders, which is in the remit of the SOAP studies. Some graphs and tables in this study would have looked different if all 4,032 DOAJ journals had been considered. Another known limitation in our approach is the fact that not all data fields were filled for the selected journals, given the impracticability of manually exploring thousands of web pages to extract the relevant information. Efforts were concentrated for the group of large publishers. Small systematic uncertainties arising from the manual harvesting and entry of information could be present in the data sample, but are not likely to alter any of the statistically significant findings of this study.

The main findings of the analysis discussed in this paper are summarised as follows:

- The distribution of journals per publishers is extremely skewed. A small number of large publishers appear on one side, with a large number of journals and/or articles. On the other side there is a vast majority of about 90% of all publishers with a single journal. The middle ground is hardly populated.
- Large publishers are more likely to rely on article processing charges (as well as membership fees and advertisement) as their income source, whereas the other (smaller) publishers base their operations more on sponsorship and subscriptions in addition to article processing charges, which they use as well. This information was collected from the journal websites; there might of course be other financial aspects of the journal incomes which are not made publicly available.

Work is ongoing and is focused on finalising the writing of results with respect to copyright/licensing options that are practiced, income options found in subject domains as well as a comparison of large publishers' experimentation with open access. Specifically, SOAP partners have reviewed the share of hybrid journals in the market, e.g. to analyse which open access share hybrid journals have and which open access share does the total article output of publishers have?

We are also currently conducting a large scale questionnaire survey looking into scholars' practices, attitudes and requirements when it comes to open access publishing. The outputs of the SOAP project will be made publically available via the project's website (<http://project-soap.eu>).

6. Acknowledgements

The project is funded by the European Commission under FP7 (Seventh Framework Programme, Capacities, Science in Society). Thanks go to people who contributed to the data entry process: Johannes Breimeier, Sylvia Cremer, Julia Graepel, Volker Kruppa, Anja Lengenfelder, Tina Planck, Teodora Todorova; Jenni Hyppoelae, Sergio Ruiz and Zanni Ali. The text in this paper has benefited from reviews conducted by members of the SOAP consortium, namely: Jenni Hyppoelae, Peter Igo-Kemenes, Salvatore Mele, Sergio Ruiz, Simon Lambert, Chris Leonard, David Ross, Bettina Goerner and Wim van der Steldt as well as others: Wolfgang Kurtz and Bo-Christer Björk. Thanks to Margit Palzenberger for conducting the statistical analysis.