Journal of Informetrics 10 (2016) 919–932

Contents lists available at ScienceDirect

Journal of Informetrics

journal homepage: www.elsevier.com/locate/joi

Hybrid open access—A longitudinal study

Mikael Laakso*, Bo-Christer Björk

Information Systems Science, Department of Management and Organisation, Hanken School of Economics, P.O. Box 479 (Arkadiankatu 22), FI-00101 Helsinki, Finland

ARTICLE INFO

Article history:
Received 26 May 2016
Received in revised form 5 August 2016
Accepted 5 August 2016
Available online 13 September 2016

Keywords:
Open access
Scientific publishing
Business models
Science policy

ABSTRACT

This study estimates the development of hybrid open access (OA), i.e. articles published openly on the web within subscription-access journals. Included in the study are the five largest publishers of scholarly journals; Elsevier, Springer, Wiley-Blackwell, Taylor & Francis, and Sage. Since no central indexing or standardized metadata exists for identifying hybrid OA an explorative bottom-up methodological approach was developed. The individual search and filtering features of each publisher website and a-priori availability of data were leveraged to the extent possible. The results indicate a strong sustained growth in the volume of articles published as hybrid OA during 2007 (666 articles) to 2013 (13 994 articles). The share of hybrid articles was at 3.8% of total published articles for the period of 2011–2013 for journals with at least one identified hybrid OA article. Journals within the Scopus discipline categorization of Health and Life Sciences, in particular the field of Medicine, were found to be among the most frequent publishers of hybrid OA content. The study surfaces the many methodological challenges involved in obtaining metrics regarding hybrid OA, a growing business for journal publishers as science policy pressures for reduced access barriers to research publications.

1. Introduction

Open Access (OA) as a phenomenon has existed since the earliest days of the internet, although the term itself was formally established around the time when the Budapest Open Access Initiative was signed in 2002 (BOAI, 2002). Suber's (2012:4) definition of OA conveys the essence of most official definitions: "Open access (OA) literature is digital, online, free of charge, and free of most copyright and licensing restrictions.". Essentially we talk about either the journal publisher making the article available directly (gold OA), or alternatively manuscript versions being uploaded to the web by authors (green OA) which can act as substitutes for readers lacking access rights to a subscription-journal. OA is also relevant for other forms of scientific reporting, such as books and data sets, but the border conditions are different for these and they are outside the focus of this study.

Since the early 1990s full OA journals have been launched in increasing numbers, using a number of alternative business models to secure the finances or resources needed to operate them (Laakso & Björk, 2012). For the year 2014 the number of full OA journals exceeded 9500, collectively publishing more than 482000 articles during the year (Crawford, 2015). Since the majority of scholarly journals and articles are still only available for subscribers an increasing number of institutional

* Corresponding author.
E-mail address: mikael.laakso@hanken.fi (M. Laakso).

http://dx.doi.org/10.1016/j.joi.2016.08.002

and subject repositories have also been created, in which authors can upload and preserve manuscript versions of their articles (Björk, Laakso, Welling & Paetau 2014).

A solution in-between full OA journals and subscription journals is being offered by most of the leading publishers. In so-called hybrid journals, authors can free their individual articles for anybody to read by making an optional payment to the publisher, while the rest of the journal's content remains reserved for subscribers only (Weber 2009). Information about such options are marketed to authors, especially at the stage when a manuscript has been accepted for publication. The common benefit enabled by payment is that the authors usually retain full copyright of the final published article and the article is labeled with a Creative Commons license, which explicitly outlines what readers can do with the article. Choosing the hybrid option is for many authors an easy option for complying with OA mandates set by funders and universities, policies which are increasingly common (Swan, Gargouri, Hunt, & Harnad, 2015). Hybrid OA has also been discussed in the context of acting as a transition mechanism for journals eventually converting to full OA publishing, whereby a journal can gradually shift over to full OA as uptake grows (Prosser, 2003).

Initial discussions and experiments around hybrid OA started as early as 1998–1999 (Walker, 1998), but the concept was tested on a wider scale with Springer's "Open Choice" programme launched in 2004 (Springer, 2004). Springer set the pricing at 3000 USD per article (Velterop, 2007), which has since become more or less of a de facto industry standard. Hybrid OA has shown signs of escalating rapidly based on reported publication outlays of universities and research funders (e.g. Björk & Solomon, 2014; Pinfield, Salter, & Bath, 2015). Research funders, in particular in the UK, have signaled a readiness to remunerate the charges to authors and their universities. There has been an ongoing debate about the possible consequences of a potential rapid uptake of the hybrid OA option in the overall publishing and subscription costs of research intensive universities (Finch, 2012). Recent experiences of funders like Wellcome Trust (Björk & Solomon, 2014) and the Austrian Science fund (Reckling & Kenzian, 2015), show that the majority of their earmarked APC funding has gone to paying the charges of hybrid journals rather than the charges of full OA journals, which are comparatively less expensive to publish in.

One of the problems in the on-going debate about hybrid OA has been the lack of exact information on uptake of hybrid OA. It is difficult to distinguish, and in particular make exact counts of hybrid OA articles. Publishers have widely differing ways of tagging hybrid OA articles in their tables of content, and there is so far no uniform universally adopted standard. The few studies conducted so far have had to rely on partial data made available by individual publishers, or partial sampling of a wider population of publications. Their methodologies and results are summarized in the following section.

2. Earlier studies

Hardly any bibliometric studies have been conducted concerning the prevalence of hybrid OA alone, but different aspects of hybrid OA have been partially covered as part of broader and coarser studies.

In the EU-funded SOAP study report one chapter is devoted to hybrid OA (Dallmeier-Tiessen et al., 2010). The project team looked at the overall hybrid OA offering (number of journals) from 12 leading publishers. The actual number of articles published was determined by asking the publishers to supply figures. The number of journals was in 2009 1991 representing around 25% of the journal portfolios of the publishers in question, and the number of articles found was 4582.

In a study based on aggregating information made available by publishers and searches in PubMed Central, Björk (2012) identified 4381 journals and estimated the number of published articles to be 12089 for 2011. In a later study, Björk & Solomon (2014) found 8003 hybrid journals in 2013, almost double the amount reported for 2011 in Björk (2012).

Other studies have indirectly incorporated hybrid articles in the overall numbers of OA articles, but usually without trying to distinguish them from free manuscript versions (green OA) or promotionally free articles (Archambault et al., 2014; Gargouri, Larivieṙe, Gingras, Carr, & Harnad, 2012). Such studies usually start from a sample of scholarly articles from indexing services like Web of Science or Scopus and then try to determine if the full text is available freely. OA articles in full OA journals can be rather easily distinguished due to the indexing of OA journals in the Directory of Open Access Journals (DOAJ), but usually all other hits (whether in delayed OA journals, hybrid journals, promotionally free issues or any variant of green OA) are bundled as one category, unless they are classified by manual inspection.

Mueller-Langer and Watt (2014) conducted a study on the citation effect of hybrid OA among articles published in 15 economics journals offering a hybrid OA option. The authors included 14 journals from Springer and 1 from Oxford University press, with a total of 1329 articles published from December 2006 to December 2011. Based on manual identification 208 articles were found to be available hybrid OA. Hybrid OA shares of articles published in the 15 journals ranged from 3.02% to 18.06%, with a total hybrid OA share of 6.5% for all included articles. The authors note that the uptake figures for hybrid OA among the journals included in the sample were influenced by the pilot hybrid OA agreements that e.g. the Dutch Consortium of University Libraries, University of California, Max Planck Society (MPG), University of Goettingen, and the University of Hong Kong had made with publishers that commonly enable all affiliated authors from the organizations to get their articles published as hybrid OA without paying individual fees. Appendix A of Mueller-Langer and Watt (2014) contains the identified institutions and agreement time periods which are likely useful for interpretation of data on the growth of hybrid OA. The study concludes that, controlling for institution quality and citations to RePEc OA pre-prints, there was not a significant relationship between hybrid OA status of articles and received citations to said articles.

The theoretical potential and realized uptake of various OA publishing models was recently studied by Jubb et al. (2015), with a particular focus on research output by UK-affiliated authors but also providing comparative metrics on uptake levels globally. The part of the broad study most relevant to this article was based on disciplinary stratified random sampling (divided as per the four main UK Research Excellence Framework panels) of Scopus-indexed articles. The study utilized manual data collection of observations for 9400 articles in with UK-affiliated authors, and 5100 articles from authors globally. The study estimated that 52.3% of all articles globally for 2014 could have been published as hybrid OA, while the actual uptake for 2014 was estimated at 2.4% of such articles. For the articles with UK-affiliated authors the hybrid OA option was estimated to have been available for 67.4% of all articles, and the realized hybrid OA uptake for UK-affiliated authors was estimated at 6.5%. Strong growth in uptake could be observed when comparing to a previous sampling of hybrid OA for articles published in 2012 (78.9% relative increase globally, and 61.7% relative increase for the UK). The study provides strong evidence for the globally accelerating uptake of hybrid OA, and that certain countries like the UK have a higher proportional share of articles output as hybrid OA seen globally, which is likely influenced by science policy and research funding supporting the option.

Sotudeh, Ghasempour, and Yaghtin (2015) studied the relationship between OA and citations by analyzing OA articles published by Elsevier and Springer during the years 2007–2011. The authors used the publisher websites to collect data about articles published OA in both subscription journals and full OA journals. A total figure of 15656 hybrid OA articles were reported for 2007–2011, where over 90% were articles published by Springer and only a small portion by Elsevier. The study provides an exclusively article-level and discipline-level analysis, omitting the measurement level of journals which within the articles are published. As such no insight is provided on the share of papers being published as hybrid OA within journals, nor were the years separated so as to provide annual figures for OA articles.

The lack of an index for OA articles published in hybrid journals can be assumed to be a major obstacle for why no extensive earlier studies have been conducted. No study so far has been based around both journal and article-level metrics, going from observations of individual articles to aggregate figures for the journal level of analysis. So far studies have been individual snapshots of specific points in time put together by disparate bits and pieces of data to paint partial pictures when it comes both to scope and chronology. There has been no longitudinal study based on a standardized bibliometric data collection methodology to study the uptake of hybrid OA publishing.

3. Aim & methods

The aim of the study was to produce an article-level measurement of the uptake of hybrid OA over the years 2007–2013, as well as to study the relationship between the uptake level with other factors such as subject field and impact of the journals in question.

Study of full OA journals is commonly facilitated by data retrievable from the Directory of Open Access Journals and study of OA uptake in general can be attempted by conducting a search for instance using Web of Science or Scopus indexed articles as a basis (see for instance Archambault et al., 2014). No such index exists for hybrid journals and searches are made difficult because many subscription journals in addition to hybrid OA articles also include other articles made open for promotional purposes, often on a temporary basis using a moving wall technique. Alternatively, publishers might incorporate a delayed OA policy where all articles become freely available after a set timeframe from publishing.

The study is informed by earlier studies of the number of hybrid journals offered by leading publishers (Björk, 2012; Björk and Solomon, 2014; Dallmeier-Tiessen et al., 2010; Sotudeh et al., 2015). Even though there are likely hundreds of publishers offering a hybrid OA option, the vast majority of individual journals can be assumed to be published by the leading commercial publishers. Hence the scope of the study was restricted to them.

Research question 1: What has the longitudinal uptake for hybrid OA been for the five largest scholarly journal publishers (Elsevier, Springer, Wiley-Blackwell, Taylor & Francis, Sage) during the time period of 2007–2013?

Research question 2: How have hybrid OA articles been distributed across scientific disciplines and journal impact metrics?

Research question 3: What is the relationship between hybrid OA uptake for a journal, its subject field and, and journal impact?

Journal impact will be operationalized by comparing the field-normalized, citation-based Source Normalized Impact per Paper (SNIP) value calculated annually by Scopus (Journalmetrics, 2016).

3.1. Article data

The foundation of the study is a dataset concerning hybrid OA articles and metadata collected from the open web. Data collection could not be conducted with a fully automated nor uniformly identical methodology for all five publishers since labeling concerning which articles have been published OA in a subscription journal is not standardized across publishers.

One central principle of the study was to leverage publicly available resources to the extent possible to facilitate replicability, a choice which also leads to highlighting the most problematic aspects of hybrid OA identification and measurement. Where publishers had made hybrid OA journal listings available they were used to narrow down the population of studied journals. Article data was collected between February 2014–December 2014. In the following the data retrieval and filtering methods for each publisher are described. All data was collected outside of any institutional network or login which could lead to subscription content being available for download. However, even so there is a lot of non-hybrid OA content that is available for download from subscription journal websites. In order to improve the signal-to-noise ratio in detecting actual hybrid OA articles among other journal content the following exclusion criteria were consistently enforced, omitting any entries that fulfilled the following criteria:

Single page documents

2-page documents which contained the word "editorial" somewhere in the full-text

Max 3 page documents which contained the word "errata", "erratum", "corrigendum", or "corrigenda" somewhere in the full-text

Articles published prior to 2007 or after 2013

However, these steps were only a preliminary weeding for relevant content. Specific inclusion criteria are outlined in the following sections describing how each publisher sample was constructed.

3.1.1. *Elsevier*

As a starting point a list of journals providing the hybrid OA option was accessed at Elsevier's website through the following URL https://web.archive.org/web/20130516113934/http://cdn.elsevier.com/assets/pdf_file/0008/109448/journal_list.pdf (April 2014). Upon investigating the URL structure of Elsevier-hosted journals it became evident that journal websites can be accessed through "http://www.sciencedirect.com/science/journal/XXXXXXX/" where XXXXXXXX represents the journal ISSN. Furthermore, Elsevier's publisher-wide web platform has a convenient link displayed on the page of each journal which displays all OA articles in said journal, including among them all articles assumed to be hybrid OA. This specific page could be accessed for every journal by adding the suffix "open-access/" to the above URL structure.

From the original list of 1527 journals offering hybrid OA all publicly available PDF files were downloaded and put through the data refinement process outlined below. To minimize the inclusion of non-hybrid OA content only articles that included any of the following phrases where included:

'creative commons'

"This is an open access article"

"This is an open-access article"

From random sampling among the freely available articles not explicitly marked as CC or OA it became apparent that they had been labeled as such in updated versions of the PDF files available from the publishers website (May 2015), for what reason this retrospective license addition had occurred remained unclear but in order to avoid re-downloading the content these articles were included in the sample.

3.1.2. *Wiley-Blackwell*

Wiley-Blackwell (referred to as only Wiley from here onwards) had published a convenient list of 1399 journals which provide the hybrid OA option at https://authorservices.wiley.com/bauthor/onlineopen_order_articleaccepted.asp (April 2014). By investigating the HTML source of the page a list of 1399 hybrid OA journals and their website URLs could be compiled. 8 of the journal entries on the original list were on closer inspection found to be either no longer published by Wiley, trade magazines or book series and were as discarded from further investigation. Wiley provides no mechanism for filtering OA articles per journal so the only option for the completion an exhaustive article-level study was to open the web page for each volume between 2007 and 2013 and query PDF links. To minimize the inclusion of non-hybrid OA content only articles that included any of the following phrases where included:

"This is an open access article"

"creative commons"

"onlineopen"

"© 20** The Authors ".

3.1.3. *Springer*

At the time of the study Springer had not published any exhaustive list of all hybrid OA journals on the web. To avoid scouring through all journals in Springers portfolio, which would have been very time-consuming considering the size of the publisher, we made an exception and reached out to Springer's Open Access Manager in order to obtain a list of all eligible titles. As a reply we got a list of dated May 2014 with 1569 journal titles and associated URLs to journal websites. Based on this list it was possible to extract the "journal ID" for each journal and create individual search queries for Springer's publisher-level search function found at http://link.springer.com/search to find out what content in these journals is not restricted to preview only. An example URL: http://link.springer.com/search?showAll=false&facet-journal-id=40521 &sortOrder=newestFirst&facet-content-type=Article&date-facet-mode=between&facet-start-year=2007&facet-end-year=2014. To minimize the inclusion of non-hybrid OA content only articles that included any of the following phrases where included:

"This article is published with open access at Springerlink.com"

"creative commons"

"This is a 'Springer Open Choice' article"

3.1.4. *Sage*

Sage provides a listing of full immediate OA journals and participating hybrid OA journals at http://www.uk.sagepub.com/repository/binaries/pdf/SAGE-Choice-Participating-Title-List.pdf (April 2014). After removing full OA journals and one journal which was no longer published by Sage, 706 journals remained in the population of hybrid OA journals published by Sage. Since Sage also uses a standardized URL scheme for all their journals it was possible to generate hyperlinks to for every volume of every journal between 2010 and 2013 based on the participating hybrid OA list. The convention was the following http://XXX.sagepub.com/content/by/year/YYYY where XXX is the three letter abbreviation for the individual journal and YYYY the volume of interest. To minimize the inclusion of non-hybrid OA content only articles that included any of the following phrases where included:

"creative commons"

"© The Author"

"The Author(s)"

"Sage Choice"

Searching for "open access" did not result in meaningful results

3.1.5. Taylor & Francis

For Taylor & Francis (T&F) there was no predefined list of hybrid OA journals to utilize as a starting point so a more labor-intensive approach was necessary. T&F offers a publisher-wide search facility at http://www.tandfonline.com/action/doSearch? where it is possible to narrow down the article search query to "Journals", "Only content I have full access to", and being published during the time of "2007–2013". To minimize the inclusion of non-hybrid OA content only articles that included any of the following phrase where included:

"Open Select"

Searching for "open access" did not result in meaningful results. While the methodology differed from how data was collected for the other publishers, the functionality of T&Fís search tool, the manageable journal portfolio (relative to some others), and clear marking of hybrid OA content made it viable rely on this method. Based on observing the collected material the reliability of the method was comparable to that of the other publishers.

3.1.6. Combining and refining the data

After each article matching the above criteria had been identified and retrieved, a bibliometric database containing the full reference information for each article was constructed containing each articles journal name, journal ISSN/E-ISSN, article title, publication year, author(s), volume, issue, page number, and DOI. DOI matching, i.e. matching each article with its unique DOI in order to retrieve associated reference information, was handled through the Papers 2 reference management software. Papers 2 has a robust matching functionality which is capable of identifying DOI information embedded in article metadata or within the article itself, and look up associated reference data from sources like Crossref, Scopus, Web of Science, and Google Scholar (Papersapp, 2014). Most articles were successfully matched without manual intervention, a small minority required manual lookup. Once a complete reference database of all identified articles had been created, a standard. bib file containing all reference data was exported from Papers 2 to JabRef (Jabref 2014) which enabled a simple conversion of the .bib file to a .csv file of the full bibliometric dataset. This file was then imported into IBM SPSS and Microsoft Excel for analysis. This description has so far covered how the unique data over hybrid OA articles was retrieved. At the stage of analysis we matched that manually-created data to journal metadata freely downloadable from SCImago (2014) and Scopus (2014), containing e.g. journal subject categories, annual article publication counts, and SNIP values.

Further filtering out likely noise from the data the following procedures where performed:

- Only journals included in Scopus were included, to ensure that focus was on peer-reviewed academic journals, and standardized metadata was available for all journals.
- Journals publishing under a total of 20 articles during 2011–2013 which OA share was over 1/3 of all content were excluded. Such outliers were assumed to be hybrid OA false positives, where a large share of articles in very small journals were available OA.
- Journals registered in the DOAJ were excluded, i.e. had at some point during the observation period flipped from subscription-access to full OA or had been full OA from the start. This is a weakness that is hard to avoid for a semi-automated retrospective study. Also journals which on closer inspection were openly marketed as OA journals but not registered to the DOAJ were also excluded.
- An existing listing of journals previously identified to incorporate delayed OA, stemming from Laakso and Björk (2013), was cross-matched with the sample to exclude journals having a policy of systematically granting access to all content after a set embargo. Again, this is something that was done due to the retrospective nature of the study.

4. Results

This section presents the results of the study, first focusing on journal level and later on the article-level. The number of journals reported throughout are journals that contained at least one assumed hybrid OA article during the 2007–2013. Hybrid OA journals without any assumed hybrid OA articles are not included as such journals could not be identified with the selected approach to data collection. However, data regarding the total number of hybrid OA journals of the same five publishers is reported for the years 2009, 2012, and 2013 in Björk and Solomon (2014) which will be used as a point of reference to gauge uptake between potential and actual uptake.

Table 1
Overview.

	2007	2008	2009	2010	2011	2012	2013	Total
Journals with at least one hybrid article	239	807	1082	1382	1788	2102	2714	3483 (unique)
Total hybrid articles	666	3325	5481	7774	9414	10802	13994	51456
Average nr of hybrid articles/journal	2.79	4.12	5.07	4.35	5.27	5.14	5.16	5.09
Total number of hybrid journals available by the same five major publishers (Björk and Solomon, 2014).			1522			4000	6740	
% share of hybrid journals with uptake of at least one hybrid article			71%			53%	40%	

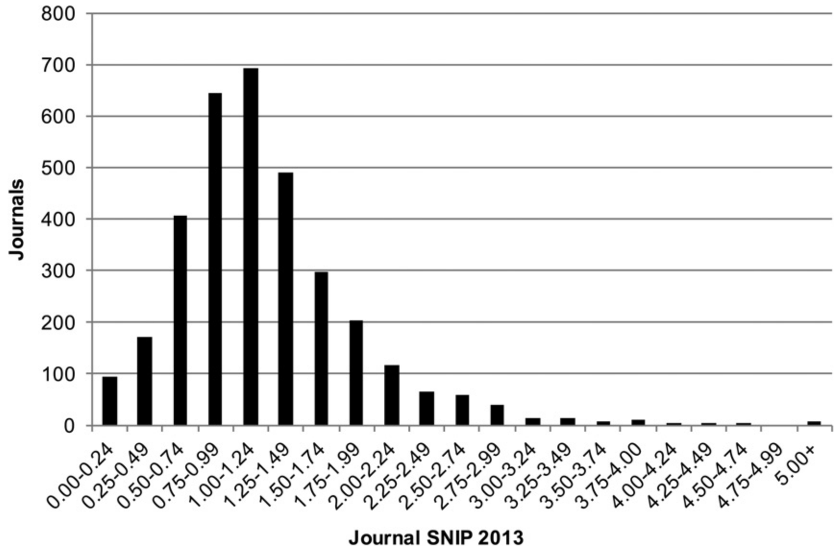

Fig. 1. SNIP 2013 values of journals with >0 hybrid article uptake 2007–2013.

Table 1 presents the high-level findings, reporting a growth from 239 (2007) to 2714 (2013) journals publishing at least one hybrid OA article. While this over ten-fold increase is substantial, the total hybrid OA article count in said journals has increased over twenty-fold in total, from 666 articles in 2007–13994 in 2013. There has been almost a doubling in the average number of hybrid OA articles published per journal per year, going from 2.79 in 2007–5.16 in 2013. Table 1 also contains compatible data points from Björk and Solomon (2014), which suggests that the% uptake of hybrid OA has dropped from 71% to 40%, much due to the rapid expansion of the hybrid OA option among the five publishers.

Figs. 1 and 2 provide a high-level visualization of the distribution of journals with >0 hybrid OA articles during 2007–2013. Fig. 1 is based on categorization by journal SNIP (year 2013), and Fig. 2 by relating the share of hybrid OA articles to the total share of published citable items in said journals during 2011–2013. The SNIP category of 1.00-1.24 is the most populated, with the neighboring categories coming 2nd (SNIP 0.75–0.99) and 3rd (SNIP 1.25–1.49).

Fig. 2 reveals that hybrid OA content accounted for 0.1–4% of all citable documents during 2011–2013 for the majority of journals where more than one hybrid OA article could be found. The distribution follows a constant decrease across the categories with a long tail ending with one journal at hybrid OA content between 34.1%–36%.

Scopus categorizes indexed journals into one or several of five major categories: Life Sciences, Social Sciences, Health Sciences, Physical Sciences, and General. Many journals are categorized into several of these categories which means that analysis needs to account for this, analyzing the categories individually would lead to a lot of articles being counted more than once. Accounting for all permutations in categorization, Table 2 presents the longitudinal development by each discipline category variant individually for every year 2011–2013. For 2013 the Physical Sciences (820 journals), Social Sciences (736 journals) Health Sciences (543 journals) were the most numerous in journals publishing at least one hybrid OA article during the year. Calculations based on SNIP (year 2013) are also provided in Table 2, both as averages for journals by year and for each subject category individually. Longitudinally it is only for year 2007 that the article-volume weighted SNIP value is below that of the simple average of all journals active in hybrid OA publishing in the year. This is an indication that journals above the average SNIP of all journals are publishing more hybrid OA articles and thus increasing the article-weighted number. The

Table 2
Journal-level overview. Annual figures report the number of journals with >0 hybrid OA articles published that year. SNIP values are provided both weighted and un-weighted, where the former takes into account the number of articles published in each journal, and the latter is a simple average of all journals with >0 hybrid OA articles in the category or year.

Scopus Journal Classification	2007	2008	2009	2010	2011	2012	2013	Total (unique)	Average SNIP 2013 for all journals	Average article-volume weighted SNIP 2013
Life	66	154	188	224	299	310	369	428	1.12	1.13
Life + Social	6	7	12	15	18	19	26	31	1.20	1.35
Life + Physical	13	53	71	90	131	150	178	205	1.24	1.25
Life + Health	41	101	122	141	207	222	262	304	1.18	1.26
Social	18	109	155	214	225	314	497	736	1.13	1.41
Social + Physical	8	48	62	75	83	104	140	202	1.40	1.49
Social + Health	8	25	47	52	52	73	83	130	1.12	1.35
Physical	33	176	257	315	475	518	646	820	1.52	1.51
Physical + Health	0	3	4	6	7	14	14	21	1.32	1.53
Health	42	118	151	228	263	346	449	543	1.07	1.30
Life + Social + Physical	0	4	3	6	7	10	15	19	1.36	1.12
Life + Social + Health	2	3	3	4	7	5	10	13	0.92	1.09
Social + Physical + Health	0	0	0	2	3	2	3	5	1.13	1.08
Life + Physical + Health	2	6	7	10	10	13	19	22	1.19	1.22
General	0	0	0	0	1	2	2	3	1.71	0.74
Life + Health + Physical + Social	0	0	0	0	0	0	1	1	0.60	0.60
Total	239	807	1082	1382	1788	2102	2714	3483	1.24	1.32
Average SNIP 2013 for all journals	1.27	1.21	1.22	1.21	1.27	1.28	1.27	1.24		
Average article-volume weighted SNIP 2013	1.07	1.29	1.29	1.28	1.32	1.34	1.35	1.32		

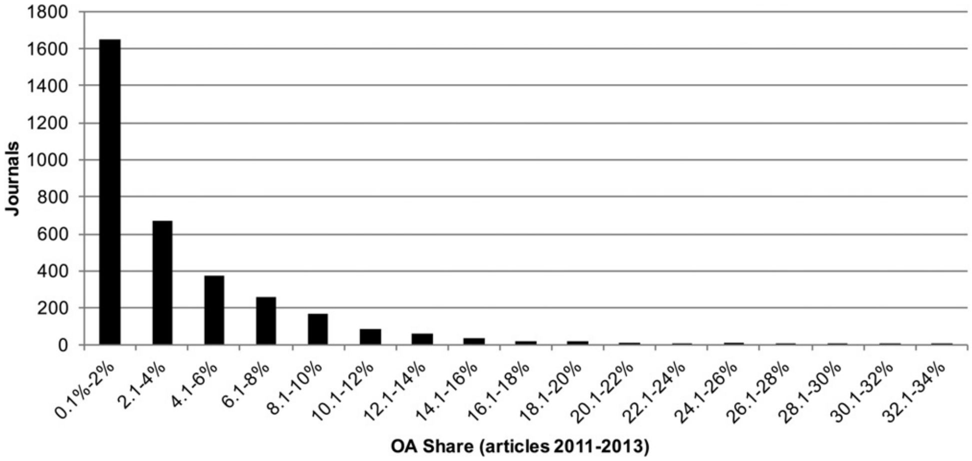

Fig. 2. Journals categorized by% of articles published 2011–2013 being hybrid OA articles (only includes journals with at least one published hybrid OA article during 2011–2013).

SNIP averages for subject categories reveal that most categories have a higher article-weighted average than one calculated without taking into account article volumes, the exception being: Physical, Life + Social + Physical, Social + Physical + Health, General, Life + Health + Physical + Social. Some of the categories have a small population size which likely causes the inverse trend to other categories, however, the Physical Sciences category is not small (N = 820) and the fact that the other outliers contain partial Physical Sciences classification suggests that the relationship between hybrid OA and SNIP might be different within this discipline since SNIP is designed to be a field-normalized indicator enabling cross-discipline comparisons. The article-level analysis will shed more light on potential explanations for the finding.

Scopus also provides a lower-level of classification of journals into 27 categories, again where each journal commonly belongs to multiple sub-categories. Wang and Waltman (2016) recently evaluated the strength of the connection between journals categorizations in Scopus and Web of Science based on inter-disciplinary citations of published papers, with results suggesting that the Scopus classification is lenient in including journals into multiple categories to which the connection through citations within the discipline are low. However, in absence of better options and in order to be able to make article-level calculations for the distribution of articles across these 27 categories, the article volume of each journal was divided and assigned equally to all subject categories which it was a member of. The results of this analysis can be found in Table 3. Due to the nature of division and adding up category totals small discrepancy from absolute numbers due to rounding of fractions (Total number of articles adds up to 14094 while actual observations through data collection was 13994, as presented earlier in Table 1).

Focusing on the journal discipline categories with a sufficient number of observations (>35), Fig. 3 visualizes the relationship between average OA% and journal average SNIP (year 2013), further highlighting some of the already indicated disparities between disciplines. Physical sciences have on average the highest relative SNIP values compared to Life, Health and Social Sciences, however, at the same time the% share of OA articles in journals is relatively the lowest in the Physical sciences. See Table 4 for excluded categories not visualized due to small population size.

By the results presented so far it is clear that the various disciplines have discrepancies when it comes not only absolute hybrid OA article output, but perhaps more interestingly the percentage of hybrid OA articles when compared to the total output of the associated journals during the timespan of 2011–2013. The total average is 5.2 hybrid OA articles for 2013, and 3.8% of all output between 2011 and 2013 being hybrid OA. The hybrid OA percentage ranges from the lowest of 1.6% within Materials Science (despite an average hybrid OA article count of 4.5 which is close to the overall total) and highest of 5.2% within the Arts and Humanities sub-category (while the average hybrid OA article count of the category was only 3 which is below the total average). What can be discerned from this overall is that the sizes of journals differ between the disciplines leading to varying results depending on whether one observes the situation based on article count or as a percentage article output, as such generalizations of results across all categories obscures some of the important insight that can be gained. This also supports the choice of using percentages as an additional measure to absolute counts for evaluating and comparing hybrid OA uptake between journals and disciplines. Counting only absolute articles skews the analysis in different ways due to the size differences in journals between subject categories.

Correlation between journal SNIP 2013 value and OA ratio (as% of all articles 2011–2013 found OA) was tested for with Pearsons Correlation, which found a weak but significant correlation between the two variables, $r = -0.068$ with a 95% confidence interval of $[-0.1, -0.035]$. In other words, as also the scatterplot in Fig. 4 visually suggests, there is overall not a strong linear relationship between journal impact and the OA ratio of articles in the journals, however, a slight negative relationship can be discerned based on the collected data (which has been collected and constructed with the bias of the journals having >0 hybrid OA articles published in them during 2007–2013). What also factors in when considering journal citation-based indicators and measurements influenced by output volume (such as the OA% here) is the inherent relationship

Table 3
Article-level analysis (article volume of journals categorized into multiple subject categories split evenly among included categories. Small discrepancy from absolute numbers due to rounding of fractions).

Scopus Journal Classification	2007	2008	2009	2010	2011	2012	2013	avg. open articles per journal 2013	Standard deviation	avg. open articles as share of all articles 2011–2013	Standard deviation
Life Sciences	330	1091	1622	2131	3026	3737	4136	6.5	4.95	4.1%	0.03
Agricultural and Biological Sciences	25	233	412	609	674	764	836	4.9	2.04	3.6%	0.02
Biochemistry. Genetics and Molecular Biology	87	410	665	840	1205	1497	1879	7.4	3.95	4.3%	0.02
Immunology and Microbiology	32	81	107	180	282	340	375	6.8	1.70	3.8%	0.01
Neuroscience	44	123	231	297	470	515	586	7.6	2.33	4.6%	0.01
Pharmacology. Toxicology and Pharmaceutics	142	244	206	204	395	621	460	7.1	1.89	4.1%	0.01
Health Sciences	221	1084	1679	2525	2963	3263	3894	6.1	4.59	4.2%	0.03
Medicine	188	1020	1600	2403	2794	2963	3602	6.2	4.53	4.3%	0.03
Nursing	26	37	34	55	53	119	110	3.5	0.59	4.0%	0.01
Veterinary	6	8	12	18	40	76	71	8.0	1.05	1.7%	0.00
Dentistry	0	4	5	10	42	49	30	3.5	0.29	3.7%	0.00
Health Professions	1	14	27	39	34	56	80	3.9	0.61	4.1%	0.01
Physical Sciences	57	733	1457	2100	2548	2590	3828	4.6	4.34	3.0%	0.03
Chemical Engineering	3	26	40	73	120	93	163	5.0	1.17	2.1%	0.01
Chemistry	5	65	106	193	379	349	471	6.3	2.28	2.0%	0.01
Computer Science	3	109	176	223	259	251	284	2.9	1.30	3.3%	0.01
Earth and Planetary Sciences	15	105	187	258	308	350	595	5.1	1.93	3.7%	0.01
Energy	0	5	4	18	42	37	102	4.4	0.83	1.6%	0.00
Engineering	5	79	178	319	260	256	349	3.6	1.32	2.1%	0.01
Environmental Science	7	100	198	317	346	445	650	5.3	2.06	3.5%	0.02
Materials Science	7	55	124	203	202	205	336	4.5	1.52	1.6%	0.01
Mathematics	8	108	228	271	284	256	333	3.2	1.33	3.2%	0.01
Physics and Astronomy	4	81	216	224	348	347	546	5.7	2.60	2.2%	0.01
Social Sciences	59	417	760	1042	1012	1335	2225	3.5	2.57	4.3%	0.03
Arts and Humanities	6	33	74	104	106	112	250	3.0	0.81	5.2%	0.02
Business. Management and Accounting	4	41	51	55	70	91	190	3.3	0.82	4.0%	0.01
Decision Sciences	3	26	50	53	62	43	54	2.5	0.36	3.3%	0.01
Economics. Econometrics and Finance	6	69	102	103	88	87	151	3.0	0.77	3.8%	0.01
Psychology	19	91	191	256	280	318	471	3.8	1.38	3.5%	0.01
Social Sciences	22	158	291	471	406	684	1109	3.4	2.06	4.6%	0.03
General	0	0	0	0	3	6	12	4.0	0.17	1.9%	0.00
General	0	0	0	0	3	6	12	4.0	0.17	1.9%	0.00
Total	666	3325	5518	7797	9551	10930	14094	5.2	6.90	3.8%	0.04

Table 4
Journal OA Ratios 2010–2013 (in%) and Journal SNIP 2013 values for Scopus discipline classifications, and in total. 0% journal OA ratios due to journals publishing an hybrid OA article between 2007 and 2009 but not in the timespan of 2010–2013.

Scopus Subject Category		SNIP 2013	OA Ratio 2010–2013
Life	Minimum	0.181	0.000
	Maximum	3.941	0.343
	Mean	1.116	0.041
	Grouped Median	1.047	0.029
	N	428	428
	Std. Deviation	0.468	0.044
Life + Social	Minimum	0.298	0.005
	Maximum	3.344	0.303
	Mean	1.204	0.049
	Grouped Median	1.109	0.033
	N	31	31
	Std. Deviation	0.539	0.063
Life + Physical	Minimum	0.000	0.000
	Maximum	4.136	0.325
	Mean	1.238	0.033
	Grouped Median	1.109	0.019
	N	205	205
	Std. Deviation	0.537	0.042
Life + Health	Minimum	0.000	0.000
	Maximum	4.251	0.323
	Mean	1.181	0.048
	Grouped Median	1.066	0.033
	N	304	304
	Std. Deviation	0.628	0.051
Social	Minimum	0.000	0.000
	Maximum	6.751	0.308
	Mean	1.129	0.046
	Grouped Median	1.04	0.026
	N	736	736
	Std. Deviation	0.681	0.049
Social + Physical	Minimum	0.000	0.000
	Maximum	4.126	0.286
	Mean	1.405	0.036
	Grouped Median	1.222	0.022
	N	202	202
	Std. Deviation	0.738	0.043
Social + Health	Minimum	0.161	0.000
	Maximum	5.012	0.278
	Mean	1.125	0.045
	Grouped Median	1.018	0.027
	N	130	130
	Std. Deviation	0.680	0.049
Physical	Minimum	0.000	0.000
	Maximum	12.916	0.250
	Mean	1.525	0.027
	Grouped Median	1.34	0.014
	N	820	820
	Std. Deviation	0.925	0.035
Physical + Health	Minimum	0.464	0.002
	Maximum	3.658	0.076
	Mean	1.322	0.026
	Grouped Median	1.291	0.021
	N	21	21
	Std. Deviation	0.758	0.023
Health	Minimum	0.000	0.00000
	Maximum	4.609	0.317
	Mean	1.070	0.040
	Grouped Median	1.035	0.022
	N	543	543
	Std. Deviation	0.548	0.047

Table 4 (*Continued*)

Scopus Subject Category		SNIP 2013	OA Ratio 2010–2013
Life + Social + Physical	Minimum	0.390	0.000
	Maximum	2.693	0.228
	Mean	1.357	0.048
	Grouped Median	1.353	0.028
	N	19	19
	Std. Deviation	0.644	0.065
Life + Social + Health	Minimum	0.139	0.001
	Maximum	1.410	0.061
	Mean	0.916	0.018
	Grouped Median	0.984	0.014
	N	13	13
	Std. Deviation	0.360	0.015
Social + Physical + Health	Minimum	0.547	0.003
	Maximum	2.102	0.054
	Mean	1.129	0.020
	Grouped Median	0.893	0.006
	N	5	5
	Std. Deviation	0.634	0.022
Life + Physical + Health	Minimum	0.448	0.000
	Maximum	3.886	0.110
	Mean	1.186	0.026
	Grouped Median	1.03	0.017
	N	22	22
	Std. Deviation	0.732	0.028
General	Minimum	0.574	0.000
	Maximum	2.696	0.044
	Mean	1.713	0.019
	Grouped Median	1.868	0.012
	N	3	3
	Std. Deviation	1.069	0.023
Life + Social + Physical + Health	Minimum	0.601	0.006
	Maximum	0.601	0.006
	Mean	0.601	0.006
	Grouped Median	0.601	0.006
	N	1	1
	Std. Deviation	.	.
Total	Minimum	0.000	0.000
	Maximum	12.916	0.343
	Mean	1.241	0.038
	Grouped Median	1.117	0.022
	N	3483	3483
	Std. Deviation	0.719	0.045

between journals of different sizes and citation-based indicators. Huang (2016) recently presented evidence for the positive correlation between quantity and impact when it comes to scholarly journals, which in this context would influence by journals having a higher SNIP value also tending to be larger in publication volume and thus have a lower relative OA content as measured by%. However, the bottom line is that there is not overall a strong linear relationship between SNIP and hybrid OA uptake as percentage of published articles.

5. Discussion

The study provides an experimental methodology for taking a measurement of the state of hybrid OA through a bottom-up approach. The insight available previously has been fragmented and usually not longitudinal, and often limited to either only the journal-level or article-level of analysis. Through this study it was possible to discern the total growth for both journals and articles, and for each subject category individually. The relationship between hybrid OA uptake and journal impact (as operationalized by SNIP) proved to be weak on the total sample of journals. In the future, if hybrid OA metrics improve and become more readily available, conducting the analysis on a broader set of journals which also include hybrid OA journals without uptake of the hybrid OA option, could produce a different conclusion.

This study was considerably more time-consuming than initially expected, which was largely due to the variations in the degree of clarity that publishers mark openly available content. And even after identifying and retrieving the OA content it is not always evident if the content has always been OA, will remain OA in the future, or just happens to be so for the time being due to promotional reasons or technical glitches. Since the data collection of this study was performed there has

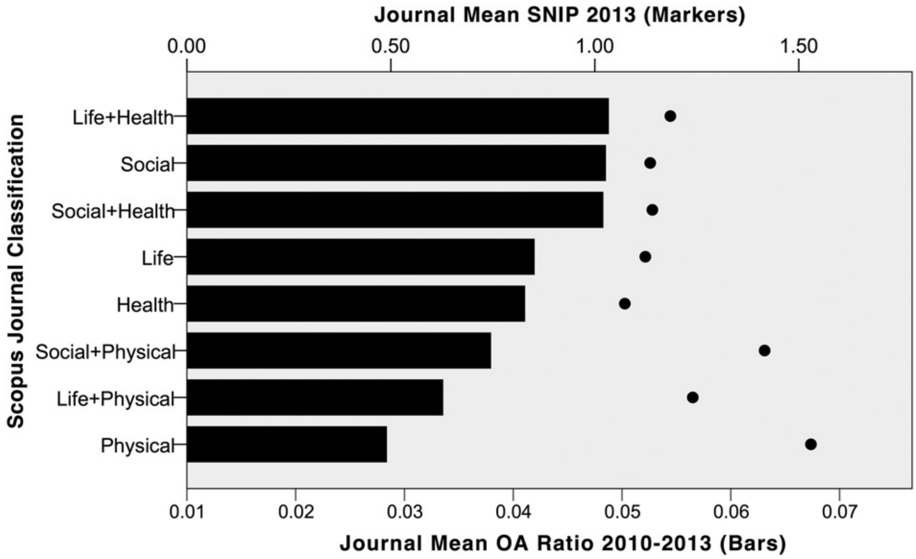

Fig. 3. Mean Journal OA Ratios 2010–2013 (in%) and Journal SNIP 2013 values across journals in Scopus journal discipline classifications with >35 observed journals.

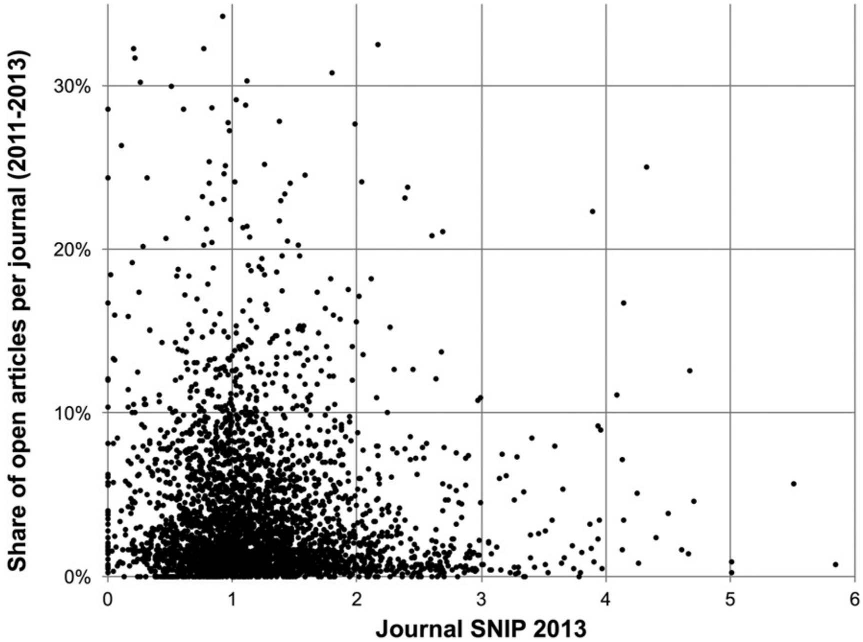

Fig. 4. Share of OA% per journal 2011–2013 contrasted to SNIP (2013) metric of journal. Each dot represents one unique journal, two extreme outliers outside of visualization to improve legibility (SNIP 11.18 <> 5%, and SNIP 12.92 <> 1.69%).

been slow and steady progress towards common metadata standards for identification of hybrid OA. A summary of the key developments is given by Chumbe, MacLeod, and Kelly (2015), who point out the key difficulties in getting publishers and discovery service providers to adopt common practices to discern between open and closed content. The authors created a promising web service, JournalTOCs, that could prove very valuable for the purpose of monitoring and studying hybrid OA (JournalTOCs, 2016). It is a web service that aggregates article and journal metadata feeds from scholarly journal publishers, including the five that are included in this study. The metadata currently contains info regarding hybrid OA status on the journal level, but comprehensively identifying individual articles as hybrid OA is still lacking, which is something that the metadata addition suggested by Chumbe et al. (2015) could rectify.

The major limitation of this study is that it makes the most of what is available – there are unavoidably uncertainties introduced when it comes to doing bottom-up data collection through the web where thousands of journals are scoured for marginally available openly accessible content. And even when content is found the task of determining what content is actually persistently meant to be out in the open is another step introducing non-perfect results. By documenting the experimental methodology as transparently as possible the intention has been to cater to those researchers who want to improve on any stage of the collection, identification, or analysis.

Future research into hybrid OA publishing could place focus on analyzing traits of individual articles, e.g. author affiliations or potential funding sources mentioned in article acknowledgements. A large-scale analysis of citation accrual to hybrid open access articles could also help shed light on what the relationship between hybrid open access and received citations is like, studies so far have been limited in scope. Mueller-Langer and Watt (2014) focused only on articles within economics, Sotudeh et al. (2015) only Elsevier and Springer exclusively at the article-level for years 2007–2011. Another study could focus on journal size, looking at the relationship between journal size on the uptake of hybrid OA. A comparison between hybrid OA uptake and publisher self-archiving policies for journals would also be valuable to better understand the current OA phenomenon, i.e. could authors of hybrid OA articles have uploaded an accepted manuscript of the article just as well? And if so, is there any relationship between strictness of self-archiving policies and hybrid OA uptake?

What has been reported here is only the inception of hybrid OA. In the beginning the alternative was sold as an optional service to authors, since then hybrid OA has become integrated into agreements and policies at different levels. The increasing offsetting deals with publishers, OA mandates, restrictions in journal self-archiving policies, and consortia similar to that of SCOAP3 (Romeu et al., 2016) will, or likely have already, fueled hybrid OA growth to much higher levels. Now that the content is increasingly open it should also be made as discoverable as possible.

Authors' contribution

Concieved and designed the analysis, collected the data, contributed data or analysis tool, performed the analysis, wrote the paper: Mikael Laakso.

Concieved and designed the analysis, collected the data, contributed data or analysis tool, performed the analysis, wrote the paper: Bo-Christer Björk.

References

Archambault, E., Amyot, D., Deschamps, P., Nicol, A., Provencher, F., Rebout, L., et al. (2014). Proportion of Open Access Papers Published in Peer-Reviewed Journals at the European and World Levels: 1996–2013 (41p.). Report produced for the European Commission DG Research & Innovation. http://www.science-metrix.com/pdf/SM_EC_OA_Availability_2004-2011.pdf.

BOAI (2002). Read the Budapest Open Access Initiative. http://www.budapestopenaccessinitiative.org/read Accessed 10.02.14.

Björk, B.-C., & Solomon, D. (2014). *Developing an effective market for open access article processing charges*. London: Wellcome Trust. Retrieved from. http://www.wellcome.ac.uk/stellent/groups/corporatesite/@policy_communications/documents/web_document/wtp055910.pdf

Björk, B.-C. (2012). The hybrid model for open access publication of scholarly articles—A failed experiment? *Journal of the American Society of Information Sciences and Technology*, *63*(8), 1496–1504. http://dx.doi.org/10.1002/asi.22709

Björk, B.-C., Laakso, M., Welling, P., & Paetau, P. (2014). Anatomy of green open access. *Journal of the Association for Information Science and Technology*, *65*, 237–250. http://dx.doi.org/10.1002/asi.22963

Chumbe, S., MacLeod, R., & Kelly, B. (2015). We should not light an open access lamp and then hide it under a Bushel. In B. Schmidt, & M. Dobreva (Eds.). *Proceedings of the 19th international conference on electronic publishing. New avenues for electronic publishing in the age of infinite collections and citizen science: scale, openness and trust*, 102–112. http://dx.doi.org/10.3233/978-1-61499-562-3-102

Crawford, W., 2015. The Gold OA Landscape 2011–2014. ISBN: 9781329547131. Lulu Publishing. 219p.

Dallmeier-Tiessen, S., Goerner, B., Darby, R., Hyppoelae, J., Igo-Kemenes, P., Kahn, D., 2010. Open Access Publishing – Models and Attributes. Max Planck Digital Library/Informationsversorgung, 62p. http://edoc.mpg.de/478647

Finch, J. (2012). Accessibility, sustainability, excellence: how to expand access to research publications. Report of the Working Group on Expanding Access to Published Research Findings 140p. Retrieved from: http://www.researchinfonet.org/publish/finch/.

Gargouri, Y., Larivière, V., Gingras, Y., Carr, L., & Harnad, S. (2012). Green and Gold Open Access Percentages and Growth, by Discipline. arXiv.org.arXiv:1206.3664v1. http://arxiv.org/abs/1206.3664.

Huang, D.-W. (2016). Positive correlation between quality and quantity in academic journals. *Journal of Informetrics*, *10*(2), 329–335. http://dx.doi.org/10.1016/j.joi.2016.02.002

Jabref (2014). JabRef is an open source bibliography reference manager. http://www.jabref.org Accessed 10.02.14.

JournalTOCs (2016). http://www.journaltocs.hw.ac.uk Accessed 16.03.16.

Journalmetrics (2016). About Source Normalized Impact per Paper (SNIP), https://www.journalmetrics.com/snip.php Accessed 26.05.16.

Jubb, M., Goldstein, S., Amin, M., Plume, A., Oeben, S., Aisati, M., et al. (2015). *Monitoring the transition to open access: a report for Universities UK*. London: Research Information Network on behalf of Universities UK. Retrieved from. http://www.researchinfonet.org/oamonitoring/

Laakso, M., & Björk, B.-C. (2013). Delayed open access: An overlooked high-impact category of openly available scientific literature. *Journal of the American Society for Information Science and Technology*, *64*(7), 1323–1329. http://dx.doi.org/10.1002/asi.22856

Laakso, M., & Björk, B.-C. (2012). Anatomy of open access publishing: A study of longitudinal development and internal structure. *BMC Medicine*, *10*, 124. http://dx.doi.org/10.1186/1741-7015-10-124

Mueller-Langer, F., & Watt, R. (2014). The hybrid open access citation advantage: How many more cites is a $3,000 fee buying you? Max Planck Institute for Innovation & Competition Research Paper, (14–02), http://ssrn.com/abstract=2391692.

Papersapp (2014) Papers 2 reference manager. http://papersapp.com Accessed 10.02.14.

Pinfield, S., Salter, J., & Bath, P. A. (2015). The total cost of publication in a hybrid open-access environment: Institutional approaches to funding journal article-processing charges in combination with subscriptions. *Journal of the Association for Information Science and Technology*, http://dx.doi.org/10.1002/asi.23446

Prosser, D. C. (2003). From here to there: a proposed mechanism for transforming journals from closed to open access. *Learned Publishing*, *16*(3), 163–166. http://dx.doi.org/10.1087/095315103322110923

Reckling, F., & Kenzian, M. (2015). Austrian Science Fund (FWF) Publication Cost Data 2013, figshare. 10.6084/m9.figshare.988754.v4.

Romeu, C., Gentil-Beccot, A., Kohls, A., Mansuy, A., Mele, S., & Vesper, M. (2016). The SCOAP3 initiative and the Open Access Article-Processing-Charge market: global partnership and competition improve value in the dissemination of science. Report number:CERN-OPEN-2014-037. 15p. 10.2314/CERN/C26P.W9DT.

SCImago (2014). SCImago Journal & Country Rank. http://www.scimagojr.com/journalrank.php?category=0&area=0&year=2014&country=&order=sjr&page=0&min=0&min_type=cd&out=xls Accessed 10.02.14.
Scopus (2014). Journal Title List https://www.elsevier.com/__data/assets/excel_doc/0015/91122/title_list.xlsx Accessed 10.02.14.
Sotudeh, H., Ghasempour, Z., & Yaghtin, M. (2015). The citation advantage of author-pays model: The case of Springer and Elsevier OA journals. Scientometrics, 10.1007/s11192-015-1607-5.
Springer (2004). Open Choice: Springer adds new publication model. Press Release. 1 July 2004. http://www.springer.com/about+springer/media/pressreleases?SGWID=0-11002-2-803577-0 Accessed 10.02.14.
Suber, P. (2012). *Open access*. MIT Press., 230p. Available at:. http://mitpress.mit.edu/books/open-access
Swan, A., Gargouri, Y., Hunt, M., & Harnad, S. (2015). Open Access Policy: Numbers, Analysis, Effectiveness. Pasteur4OA Work Package 3 report: Open Access policies 2015. arXiv:1504.02261v1.
Velterop, J. (2007). *Open access is a choice. ZfBB 54 (4–5) 268–272*. http://edoc.hu-berlin.de/oa/articles/reml4evBkibI/PDF/26OGwKncskJbU.pdf
Walker, T. (1998). Free internet access to traditional journals. *American Scientist*, *86*(5), 463. http://dx.doi.org/10.1511/1998.5.463
Wang, Q., & Waltman, L. (2016). Large-scale analysis of the accuracy of the journal classification systems of Web of Science and Scopus. *Journal of Informetrics*, *10*(2), 347–364. http://dx.doi.org/10.1016/j.joi.2016.02.003
Weber, D. (2009). Hybrid OA journals: A progression or a destination. Preprint. http://eprints.rclis.org/13744/1/Hybrid_OA_Journals.pdf.