

Delayed Open Access – an overlooked high-impact category of openly available scientific literature

Mikael Laakso

Hanken School of Economics, P. B. 479 00101 Helsinki, Finland.

E-mail: Mikael.Laakso@Hanken.fi

Bo-Christer Björk

Hanken School of Economics, P. B. 479 00101 Helsinki, Finland.

E-mail: Bo-Christer.Bjork@Hanken.fi

Delayed open access (OA) refers to scholarly articles in subscription journals made available openly on the web directly through the publisher at the expiry of a set embargo period. Though a substantial number of journals have practiced delayed OA since they started publishing e-versions, empirical studies concerning open access have often overlooked this body of literature. This study provides comprehensive quantitative measurements by identifying delayed OA journals, collecting data concerning their publication volumes, embargo lengths, and citation rates. Altogether 492 journals were identified, publishing a combined total of 111 312 articles in 2011. 77,8 % of these articles were made open access within 12 months from publication, with 85,4 % becoming available within 24 months. A journal impact factor analysis revealed that delayed OA journals have on average twice as high average citation rates compared to closed subscription journals, and three times as high as immediate OA journals. Overall the results demonstrate that delayed OA journals constitute an important segment of the openly available scholarly journal literature, both by their sheer article volume as well as by including a substantial proportion of high impact journals.

Introduction

Open access (OA) is a term widely used to refer to unrestricted online access to articles published in scholarly journals. Scientists started experimenting with open access journals in the early 1990s but the term itself was coined around the year 2000. There are a number of semi-official definitions of OA, but the most widely quoted is the one included in the declaration of the Budapest Open Access Initiative meeting in 2001 (BOAI 2001):

"By "open access" to this literature, we mean its free availability on the public internet, permitting any users to read, download, copy, distribute, print, search, or link to the full texts of these articles, crawl them for indexing, pass them as data to software, or use them for any other lawful purpose, without financial, legal, or technical barriers other than those inseparable from gaining access to the internet itself. The only constraint on reproduction and distribution, and the only role for copyright in this domain, should be to give authors control over the integrity of their work and the right to be properly acknowledged and cited."

The BOAI definition is rather liberal, as it grants users a lot of freedom to do what they want with the published content. OA is, however, not a simple on-off phenomenon where a publication either is OA or not. Ideally an article is open from day one, directly through the publisher's own web site and provided with extensive and well-defined usage rights (known as libre OA) which are often defined by referring to a Creative Commons license. Less ideal

forms restrict specific usage of the published content (e.g. no re-distribution, human reading only), limit openly available copies to non-final manuscript versions, or delay the open availability through an embargo period. In an attempt to provide an overview of the key variables involved, Table 1 lists some criteria according to which OA can be classified. The categories applicable to this study are presented in bolded italics.

Availability	Prior to publication	Immediately at publication	Up to 12 months after publication	In excess of 12 months after publication	
Version	Version Original publication		Preprint version		
Usage rights	Libre	Gratis			
Provider of OA copy	Publisher	Funder or employer	Author	Third-party	
Setting	In fully free journal	Individually opened			
Location	Publisher's site	Subject- based repository	Institutional repository	Homepage or other	
Permanency	Permanently OA	Temporarily OA			
Legal status	Legal copy	Illegal copy			

Table 1 - OA classification table, scope of study in bolded italics

Earlier efforts at identifying and labeling different aspects of OA have been made, recognizing the complexity of the phenomenon is nothing new. In his seminal book Willinsky (2005) describes ten flavors of Open Access, including delayed OA as one of them. The classification table above includes within its scope both what in the OA debate is commonly called gold OA (provided by the publisher) and green OA (manuscript copies provided by the author and other parties). Interestingly gold OA is by definition always immediate, whereas green OA includes delayed articles, due to publisher embargoes or delays in self-archiving. Thus gold and green OA are not two opposite concepts, rather gold + delayed OA should be contrasted with green OA.

Going by the strict BOAI definition the term 'delayed OA' is something of an paradoxical oxymoron. On his FAQ pages concerning the Budapest Open Access Initiative leading OA advocate Peter Suber answers the question "Is open access compatible with an embargo period?" in the following way: "No. Open access is barrier-free access, and embargo periods are barriers to access. Many of the benefits of open access are not achieved when embargoes are in place. However, while delayed free access does not serve all the goals of the BOAI, it does serve some of them. Just as open access is better than delayed access, delayed free access is better than permanently priced access" (Suber 2011). Since this study is not bound by the BOAI definition and adopts a more inclusive perspective, the term delayed OA will be used to refer to this category of journals.

The usefulness of different types of OA varies, but it can be argued that any kind of access is better for prospective readers than restricting access to only subscribers or pay-per-view

customers. If a web search is done for the full texts of a large number of randomly selected scholarly articles in a setting with no subscription access, the copies found are most likely a mixture of different OA categories. However, OA background variables such as usage rights, provision mechanisms, type of copy and when the article was made openly available are not likely to be a key concern for a reader just wishing to read the article, to be able to obtain a full-text copy of the article is most likely the highest priority.

In a broader perspective the value of different types of OA can also be seen as a function of the typical readership patterns and citing behaviors of academics. Recently published literature is mostly read in what can be called "current awareness" reading where scholars rapidly scan the tables of contents of particularly interesting journals that they follow, often prompted by table-of-contents e-mails. Various web tools have also become increasingly important sources for tracking when relevant articles matching specific interest profiles have been published. Articles published in earlier years are more likely to be encountered via keyword searches, citations or by recommendations from colleagues. For such cases delayed OA is as good as immediate OA. For instance King et al (2009) found that roughly half of all scientific articles read are at least one year old. One could argue that the value of reading an article tracked down through a reference is higher in terms of impact on scholarship than the average current awareness reading.

The benefit of delayed OA (as being defined in his study excluding delayed green OA) being provided in the original version directly through the publisher is also of relevance. In a survey distributed among UK scientific societies 52 % of the survey respondents said they would not opt to access green manuscript copies even when they do not have access to the published version (Morris & Thorn, 2009). Some answers suggested that the integrity of self-archived author manuscripts is questioned, whether the content is fully identical to the actual published version. In some fields of science it is also important to be able to cite specific pages in publications, which can be problematic with manuscript copies.

Previous research

As the introduction delineated, this study focuses on publisher-provided delayed OA of original subscription publications. Most studies providing quantitative measures of the prevalence of OA have overlooked this part of the OA spectrum. Partly this is due the fact that there is no comprehensive index of such journals similar to the service the Directory of Open Access Journals (DOAJ) provides for immediate OA journals, partly by the divided opinions concerning delayed OA as a valid form of OA. In a discussion about the lack of an aggregating index Jacsó (2011) suggested that delayed OA journals should be added to the DOAJ as a separate category, motivated by the fact that the population is made up of many widely read and cited journals with a combined historical archive exceeding at least 2 million articles. To emphasize this point it is worth mentioning that five of the ten most cited journals in the Web of Knowledge during 1999-2009 provide delayed OA (Sciencewatch 2009).

Of the few quantitative studies available, Björk, Roos and Lauri (2009) is the first to explore delayed OA article volumes and embargo lengths as part of a larger OA study. By sampling article volumes for delayed OA journals listed in HighWire Press, the authors came to the result that 234 delayed OA journals published an estimated 47 499 articles during 2006. The embargo distribution among the articles was 11,5 % 2-6 months, 81,2 % 12 months, and 7,3 % with an embargo of 24 months or longer. Based on an approximation for the total number of articles published during 2006, these delayed OA articles made up 3,5 % of all published articles.

In contrast to Björk, Roos and Lauri (2009) who used journal annual article volume data as the base data for analysis, Björk et al (2010) used a bottom-up methodology where the metadata of randomly selected articles was used to estimate the open availability of the total body of scholarly articles published during 2008. Through a manual web search for the individual articles included in the sample and classification of found full text hits into specific OA types, the authors were also able to estimate the prevalence of delayed OA, which ended up being 1,2 % out of the total OA rate of 20,4 %. The low result compared with the earlier study highlights the challenge to correctly identify delayed OA for random individual articles as many journals do not explicitly state their delayed OA policy on their webpages, much less directly within the full-text articles.

In a study commissioned by the Association of Learned and Professional Society Publishers a survey was sent out to the main contact persons of a large number of scholarly journals, including delayed OA journals webhosted by Highwire Press (Kaufmann-Wills Group 2005). An interesting question in the survey focused on which groups had influenced the decision to offer delayed OA. The most frequently reported groups were full open access proponents, society members, authors, publishing oversight bodies, and the journal publisher. The study also included 22 in-depth case studies of journals, of which some offered delayed OA. A general finding was that delayed OA was relatively common among society journals. Society journals seemed to have found a working recipe for continuing as subscription journals while not fearing too much loss of revenue even though they open up the e-versions after an embargo period. Society publishers, while recognizing the importance of protecting their revenue and continuing high-class publishing services, also might value OA as an important part of their "mission".

Waltham (2005) provides a rare insight into the financial aspects involved in implementing delayed OA for previously subscriber-only journals. The study reports figures for delayed OA influence on journal subscriptions for ten Oxford University Press journals. Compared to the subscriber figures prior to providing delayed OA, two journals with an embargo period of 6-months saw an average subscriber decrease of 6 % and eight journals with a 12-month delay a reduction of 2 %. These results show some degree of reduction in subscription income, however, with a sample so small the conclusions should not be extensively generalized.

Aim of the study

The aim of this study was to identify scholarly peer-reviewed journals continuously providing OA to full issues after a set embargo period, and collect data regarding embargo lengths and articles published during 2011. By analyzing this data the total volume of delayed OA can be estimated as well as an analysis of article volumes per embargo length.

A retrospective longitudinal study of the development over time of delayed OA would require accurate records for when individual journals have started or ceased publishing the OA volumes. This was found to be excessively labor-intensive and thus the study was limited to articles published during 2011.

In addition to the volume of delayed OA articles another aim was to measure the scientific impact, in terms of average citation rates compared to closed journals and immediate open access journals. The aim was not attempt an estimation for the influence of the open availability on the number of citations (the so-called citation effect of OA), but to measure the effect of delayed OA on the availability of cited articles. The methodology for this has been

developed in a recent study (Björk and Solomon 2012) and some data from that study could be reused.

Method

Data collection was set up as a two-stage process. First, a list of delayed OA journals was composed by combining data from multiple sources. In addition to journal title, publisher, and embargo length, the ISSN/eISSN codes were considered essential data for enabling cross analysis with major publication indexes during later stages of data analysis. Second, the annual article volumes for the identified journals were collected by retrieving data available in major publication indexes or by manually visiting the web pages for individual journals. Data was collected during May and June 2012. Details for how both steps in the process were conducted is provided in the remainder of this section.

The first step of the process, compiling the aggregate list of delayed OA journals, was largely explorative since there is currently no comprehensive pre-existing listing or database that could be used for the purpose. The goal of the data collection was to identify journals publishing content openly on the web with a delay from the original publication date, thus excluding full immediate OA journals from the scope. In the end journals were identified from six main sources: HighWire Press, PMC, Elsevier, the Wikipedia page for delayed OA journals, delay OA journals discovered during a previous study (Björk 2010), and additional individual journals known to the authors.

HighWire Press, a division of the Stanford University Libraries, provides online access to a large number of journals as well as other scholarly literature. While the majority of the 1349 journals listed in HighWire Press at the time of the study were subscription or pay-per-view only, 38 were self-reported as full immediate OA, and 240 journals as offering delayed OA with the added criterion of having published at least one volume between the years 2005 and 2011 (HighWire Press 2012). Though exact historical records for the increase of delayed OA journals in HighWire Press are not available, it can be derived from chart data provided in Waltham (2005) that there were around 85 such journals in 2001, around 170 journals in 2005, and this study pegs the number at 240 for 2011.

PMC, a free full-text archive of biomedical and life sciences journal articles, provides access to full journal volumes as well as individual articles or manuscripts of articles published in subscription journals. As this study focuses on delayed OA of complete journal volumes, individual articles made available in PMC were not included in the calculations. Some publishers deposit articles stemming from NIH grants directly into PMC, usually with a delay, rather than have the authors self-archive the manuscript versions, but such articles were not included in the study. The total journal count of PMC at the time of the study was 1564 (PMC 2012), however, after removing journals registered as depositing only selected articles, full immediate OA journals listed in the DOAJ, journals which have not published anything during 2005-2011, and journals which are listed as not depositing any new content to the repository, 186 delayed OA journals remained.

Elsevier, a major publisher of scientific journals, provides a convenient list of journals for which they provide delayed OA. At the time of the study this list contained 73 titles (Elsevier 2012). The Wikipedia page for the category of delayed OA journals contained 60 journals for which the webpages were active and available for article counting at the point in time of the study (Wikipedia 2012). Through a previous study where the availability of randomly selected articles as well as the OA model through which they were delivered were studied,

(Björk 2010) the authors had previously collected information about eight individual journals providing delayed OA. Five other journals of which the authors were aware were also added.

Listing all the identified journals together and removing duplicates resulted in 492 unique delayed OA journals. It should be pointed out that this list is conservatively composed, in cases where there was doubt about the delayed OA status of the journal it was not added from the source.

For the second step of the data collection process data was gathered from three different sources. Web of Knowledge provided article volume data was used for journals included in that index, Scopus data for journals not indexed by Web of Knowledge but indexed by Scopus, and lastly manual article counting was performed for the remainder of the journals not covered by either index. Since 409 of the journals are included in the Web of Knowledge index most of the data could be gathered by querying the service with ISSN/eISSN codes. It is worthwhile to note that journals often include editorials, news, book reviews, obituaries and other non-research content. Such material was excluded from all data, both data retrieved from indexes as well as manually gathered data.

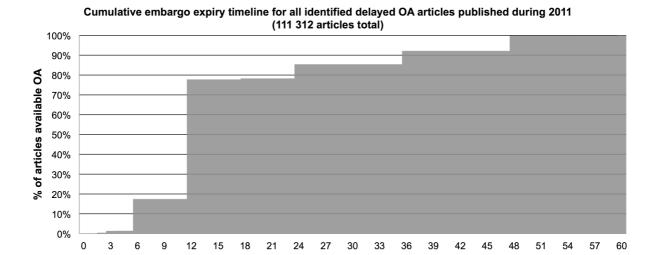
Doing the analysis by article numbers, rather than based on the number journals is an important aspect of the study and similar to methods used in a number of earlier studies from our research group. In this connection it was particularly important since the average article volume of the delayed OA journals is over 200, compared to an average of about 110 for all Web of Knowledge journals (Björk et al 2009), and around 50 for immediate OA journals (Forthcoming, Laakso and Björk 2012). Hence the impact of these articles eventually becoming OA is far bigger than the number of journals might initially suggest.

Analysis and interpretation of the data was very straightforward, sampling was not needed, and the data could be processed unweighted since everything is based on absolute numbers. However, one caveat to doing this type of time-sensitive study retrospectively, even for just a single year backwards, is the lack of data on when identified journals have started providing delayed OA. Where such information was unobtainable from the originating data source the assumption was that a journal currently providing delayed OA did so for the full 2011 calendar year as well.

Though almost all identified journals provide unrestricted access to articles once the embargo period has expired, there are journals that require readers to register for a free account before granting access to full text content. The example we came across is Science which has en embargo of 12 months after which registered users can access the journal content. Ultimately we decided to include Science as well. One argument for this is the rarity of this requirement, another is that interested readers are unlikely to be hindered by having to sign up if that enables them to access the requested content for free and that the articles, despite the registration barrier, are indexed in most indexing services and findable through web search engines.

Results

Altogether 492 journals were identified publishing a combined total of 111 312 articles in 2011. 77,8 % (86 599) of these articles are made open access within 12 months from publication, with 85,4 % (95 109) available within 24 months. Very few journals have embargos exceeding 48 months. Figure 1 shows the embargo distribution among the articles.



Months since publication

Figure 1 – Embargo expiry timeline for delayed OA articles

During the data collection all delayed OA publishers were categorized into six major categories based on what kind of organization is publishing the journal. The article volumes output by the six publisher categories is provided in Figure 2, together with a comparison to article volumes of immediate OA journal publishers (Laakso & Björk 2012). The publisher analysis revealed that 98 % of the 111 312 delayed OA articles were published by three of the six publisher categories: 52 % by scientific societies or professional associations, 33 % by commercial publishers, and 13 % by University press publishers. The relative distribution among the publisher categories is more skewed towards the three aforementioned categories compared to publishers of immediate OA articles. The distribution among publisher types is also markedly different compared to scientific journals in general. In 2009 the overall distribution of article output by publisher type of journals included in the Web of Knowledge index was reported to be 64 % commercial (including some proportion of publishing for societies), 30 % society publishers, 4 % university presses and 2 % by other types of publishers (Ware and Mabe 2009), i.e. delayed OA is composed of a higher proportion of society publisher output and a lower proportion of commercial publisher output than journals in general.

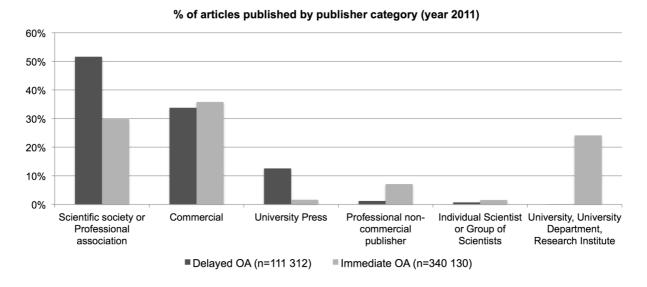


Figure 2 – Publisher category analysis

Previous studies have indicated that delayed OA journals include many widely cited publications. To study this a comparison between closed journals, immediate open access journals, and delayed OA journals was performed. Comparing the scientific quality of journals is a challenging task and the only feasible method in our case was to rely on readily available data on average journal citation levels (impact factors). The analysis of impact factors for journals included in the Web of Knowledge index (Table 2) shows that delayed OA journals exceed comparable scientific impact averages for closed journals as well as immediate OA journals, both calculated on the journal and the article level.

Journal Type	Journals (n)	Journal Level		Article Level	
Closed ¹	7 609	Mean 1.97	SD 2.95	Mean 2.81	SD 3.31
Immediate Open Access ¹	610	1.50	4.02	2.04	2.28
Delayed Open Access	409	4.42	5.05	5.93	6.39

¹ = Results from Björk & Solomon (2012)

Table 2 – 2-year citation averages for Web of Knowledge-indexed immediate open access journals, delayed open access journals, and closed journals. Calculated using 2010 impact factors.

The article volume estimations of this and a previous study (Forthcoming, Laakso and Björk 2012) can be combined with the average impact factors presented in Table 2 in order to analyze to accessibility of articles cited in the Web of Knowledge. The analysis is restricted to articles in journals tracked by the Web of Knowledge and of citations in journals included in the index. The focus is on the citations, not the articles behind them, hence an article cited 100 times get as much weight as the number of citations. The analysis shows that:

- 80 % of the citations point to articles in closed subscription journals. Some of these can nevertheless be found as self-archived manuscript copies (green OA) and around 1 % as individually opened articles in hybrid OA journals.
- 6 % point to articles in immediate OA journals.
- 14 % point to articles published in delayed OA journal with embargo periods of 12 months or less.

The distribution is of importance since citations in a sense constitute the votes of the global scientific community about which articles, from a scientific viewpoint, provide the most valuable reading.

Due to space limitations the full list of delayed OA journals can not be listed within this article. However, in order to illustrate the high-quality nature of many delayed OA journals, Table 3 lists some attributes of the 19 delayed OA journals which are among the top 50 journals in Google Scholar's list of top publications. The Google Scholar ranking is based on a Hirsch index for citations to a journal during the last five years.

Google Scholar Journal Rank	Google Scholar h5- index	ISI Impact Factor 2010	Nr of Articles 2011	Embargo Length (months)	Journal Title
2	274	53.5	687	6	New England Journal of Medicine
3	265	31.4	1718	12	Science
5	195	32.4	441	12	Cell
6	189	9.8	3342	6	Proceedings of the National Academy of Sciences
11	159	14.4	703	12	Circulation
13	141	10.6	1592	12	Blood
14	140	7.4	2472	12	The Astrophysical Journal
15	139	14.3	469	12	Journal of the American College of Cardiology
25	120	8.2	744	12	Cancer Research
28	117	14.0	410	12	Neuron
31	115	24.2	205	12	Immunity
32	114	14.8	216	6	The Journal of Experimental Medicine
35	113	16.7	246	6	Annals of Internal Medicine
41	110	5.3	4291	12	Journal of Biological Chemistry
42	110	7.3	1821	6	The Journal of Neuroscience
45	106	12.9	237	6	Genes & Development
47	104	5.8	1406	12	The Journal of Immunology
48	103	14.2	367	12	Molecular Cell
49	103	26.9	165	12	Cancer Cell

Google Scholar 9th July 2012 (http://scholar.google.com/citations?view_op=top_venues&hl=en) Non-journal entries removed from list and ranking adjusted accordingly (RePEc, arXiv, Social Science Research Network, NBER Working Paper Series, Cochrane database of systematic reviews).

h5-index is the h-index for articles published in the last 5 complete years. It is the largest number h such that h articles published in 2007-2011 have at least h citations each.

Table 3 – Delayed OA journals (n=19) within the Google Scholar top 50 ranking

Conclusions

OA is not a straightforward on-off phenomenon, rather there are many degrees and shades of OA. While immediate publisher-provided OA with a liberal reuse license is the ideal form, any form of OA is better than only closed access. What this study demonstrates is that delayed OA, which is mainly provided by society publishers and university presses, is important both in volume and in granting access to highly cited scholarly articles. It is interesting to note that the yearly volume of delayed OA articles (111 000) is almost ten times as big as the number of individually freed article in so-called hybrid journals (12 000) (Björk 2012), and that total number of citations to delayed OA article in the Web of Knowledge exceeds the number of citations to articles in full OA articles.

A concrete suggestion following from this study would thus be for the DOAJ to also start indexing delayed OA journals and their embargo periods. Most of the journals belonging to this category are mature with technologically advanced webpages facilitating automated indexing even on article level. Compared to the more heterogeneous lot of webpages for immediate OA journals complete article-level coverage would be more easily realized. Perhaps the list assembled for this study and made available with this article can offer a starting point for improved coverage for delayed OA journals in the future.

Having such a centralized list of delayed OA journals would also facilitate OA prevalence studies using robotized tools, that usually perform the searches with meta-data of articles

published at least a year old (i.e. Gargouri et al 2012). Using such a list with ISSN numbers the found copies could more easily be split into immediate OA, delayed OA and green copies.

References

Björk, B-C., Solomon, D. (2012). Open access versus subscription journals: a comparison of scientific impact. *BMC Medicine*, 10 (73)

Björk, B-C., Roos, A., Lauri, M. (2009). Scientific journal publishing: yearly volume and open access availability. *Information Research*, *14*(1), 391.

Björk, B-C., Welling, P., Laakso, M., Majlender, P., Hedlund, T., & Guðnason, G. (2010). Open Access to the Scientific Journal Literature: Situation 2009. *PLoS ONE*, 5(6), e11273

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*, Volume 63, Issue 8, pages 1496–1504, August 2012

BOAI (2011). Budapest Open Access Initiative, http://www.soros.org/openaccess/read doi:10.1087/2009208

Elsevier (2012). Elsevier Delay OA Journal Listing http://www.elsevier.com/wps/find/authorsview.authors/open_archives

Gargouri, Y., Lariviere, V., Gingras, Y., Carr, L., Harnad, S. (2012). Green and Gold Open Access Percentages and Growth, by Discipline. 17th International Conference on Science and Technology Indicators (STI), 5-8 September, 2012, Montreal, Quebec, Canada, 11pp., http://eprints.soton.ac.uk/340294/

HighWire Press (2012). HighWire Press Journal Listing http://highwire.stanford.edu/librarians/AtoZList.xls

Jacsó, P. (2011). Péter's Digital Reference Shelf – DOAJ http://www.gale.cengage.com/reference/peter/200912/doaj.html

Kaufmann-Wills Group (2005). The facts about Open Access - A study of the financial and non-financial effects of alternative business models for scholarly journals. Research report, The Association of Learned and Professional Society Publishers

King, D., Tenopir, C., Choemprayongchoe, S., Wu, L. (2009). Scholarly journal information-seeking and reading patterns of faculty at five US universities, *Learned Publishing*, 22, 126–144

Laakso, M., Björk, B-C. (2012) Anatomy of open access publishing - a study of longitudinal development and internal structure. Accepted for publication in BMC Medicine

Morris, S., Thorn, S. (2009). Learned society members and open access. *Learned Publishing*, 22(3), 221–239.

PubMed Central (2012). PubMed Central Journal Listing http://www.ncbi.nlm.nih.gov/pmc/journals/#csvfile

Sciencewatch (2009) Top Ten Most-Cited Journals (All Fields), 1999-2009 http://sciencewatch.com/dr/sci/09/aug2-09 2/

Suber, P. (2011). BOAI FAQ http://www.earlham.edu/~peters/fos/boaifaq.htm#openaccess, last revised 16.12.2011

Waltham, M. (2005). JISC: Learned Society Open Access Business Models, Report, MaryWaltham.Com, Princeton, NJ, USA, www.marywaltham.com/JISCReport.pdf

Ware, M., Mabe, M. (2009). The stm report - An overview of scientific and scholarly journals publishing. International Association of Scientific, Technical and Medical Publishers (p. 68).

Wikipedia (2012). Wikipedia Delay OA Journal Listing http://en.wikipedia.org/wiki/Category:Delayed open access journals

Willinsky, J. (2005). The Access Principle. MIT Press.