

PERSONAL VERSION

This is a so-called personal version (author's manuscript as accepted for publishing after the review process but prior to final layout and copyediting) of the article: Laakso, M. (2014). Green open access policies of scholarly journal publishers: a study of what, when, and where self-archiving is allowed. *Scientometrics*. 99 (2) 475-494.

<http://dx.doi.org/10.1007/s11192-013-1205-3>

This version is stored in the Institutional Repository of the Hanken School of Economics, DHANKEN. Readers are asked to use the official publication in references.

Green open access policies of scholarly journal publishers: a study of what, when, and where self-archiving is allowed

Mikael Laakso

Information Systems Science, Department of Management and Organisation

Hanken School of Economics, Arkadiankatu 22, 00101 Helsinki, Finland

mikael.laakso@hanken.fi

Abstract The degree to which scholarly journal articles published in subscription-based journals could be provided open access (OA) through publisher-permitted uploading to freely accessible web locations, so called green OA, is an underexplored area of research. This study combines article volume data originating from the Scopus bibliographic database with manually coded publisher policies of the 100 largest journal publishers measured by article output volume for the year 2010. Of the 1,1 million articles included in the analysis, 80.4% could be uploaded either as an accepted manuscript or publisher version to an institutional or subject repository after one year of publication. Publishers were found to be substantially more permissive with allowing accepted manuscripts on personal webpages (78.1% of articles) or in institutional repositories (79.9%) compared to subject repositories (32.8%). With previous studies suggesting realized green OA to be around 12% of total annual articles the results highlight the substantial unused potential for green OA.

Keywords Open access, self-archiving, scientific publishing, science policy

Introduction

Open access (OA), a term that refers to free and unrestricted access to scholarly journal articles. Since journal articles can end up being openly available on the web as the results of various processes, the terminology describing OA commonly splits OA into two main categories depending on who acts as provider for the free content. Journal-mediated OA is referred to as Gold OA and is commonly handled through the rapidly growing number of journals which publish articles OA directly on the journal website (Laakso and Björk 2012). Green OA refers to indirect free access to an article or an earlier version of the manuscript that is available on the web at a location other than the website of the journal publisher (Harnad et al 2004). “Green” in this context comes from the notion of publishers giving a “green light” for uploading openly available copies of the article contents. The terms green OA and self-archiving have often been used interchangeably, however, the term green OA will primarily be used in this study since it does not imply that the author is the only one who can be involved in the process of making the contents of a journal article openly available. Through various policies between publishers, authors, institutions and research funders the uploading of an OA version of an article might not involve any author involvement, so called mediated deposit into a repository. This study is exclusively focused on green OA and more specifically on exploring the extent to which publisher policies allow uploading of manuscripts and articles to web locations with free unrestricted access.

A large part of the quantitative research concerning green OA has focused on measurement of the realized proportion of green OA journal articles as a share of all published journals articles during a specific year. However, the fundamental definitions and research methodologies for green OA prevalence studies have varied, making longitudinal developments and cross-comparisons of studies hard to perform. A recent review and synthesis of green OA research including realized green OA studies is provided in Björk et al (2014) where the overall share of articles provided as green OA in recent years is estimated to be around 12% of all annual articles published in subscription-based journals. In studies enquiring into realized green OA article volumes there is usually the explicit or implicit assumption that the potential maximum is equal to the total number of annual journal articles published during a specific year. However, it is more of common that publishers place embargos or completely prohibit some or all forms of green OA, as such the assumption of 100% theoretical green OA implies breaches of publisher policies. The purpose of this study is to explore what that degree of publisher-permitted green OA actually is.

Before low-cost, systematic distribution of electronic documents was enabled by the Internet the purpose and need for agreements between publishers and article authors were different in nature than they are with today's information infrastructure. Going from a time when dissemination of research was most effectively handled within the covers of printed journals to realizing the benefits open digital distribution can have for all aspects research, the scholarly publishing industry is still in a state of intense transformation and exploration. There is tension between making research available as openly as possible, in an effort to expand the potential audience and increase potential citations in

the best interest of authors and publishers, and publishers striving to retain their current subscription-based business models. For a long time self-archiving was the only way to make research published in subscription-based journals freely available online. However, within the last few years subscription-based journal publishers have increasingly become involved in this process and realized the business opportunities for “selling” authors the rights to make their research available OA either directly through the journal or indirectly through self-archiving. One development is the founding of new OA journals based on the author-pays business model into the publisher journal portfolio. A recent example of this is Elsevier’s newly founded BBA Clinical journal which was introduced as an OA outlet part of the BBA portfolio of journals. Another development is the growing number of subscription-based journals which offer authors the paid option to make their individual article OA on the journal website, so called hybrid OA (e.g. Springer Open Choice, Wiley OnlineOpen). There are currently over 7000 subscription-based journals offering a hybrid OA option, most of them belonging to the largest commercial publishers. Both of these developments can be seen as a slowly creeping threat to liberal, unpaid, self-archiving rights.

Agreements between authors and publishers come in different varieties and the terminology used to describe them is as diverse. Copyright transfer agreement, exclusive license, non-exclusive license are some of the common formal agreement types (Gadd et al 2003). But regardless of term and agreement type, the agreement is intended to define what the author is able to do with the manuscript on which the article is based once the agreement is accepted. The specific legal terms for different types of publishing agreements is of less importance for this study since the focus is on practical implications for the possibilities of green OA, the author-retained rights to disseminate the article or an earlier manuscript version. Throughout the rest of this study any such agreements are commonly referred to with the generic term of publisher policy.

This study takes a quantitative and strict perspective on something that involves human intervention to actually realize its potential, i.e. green OA copies of manuscripts or articles becoming available on the web. While the contents of publisher policies are important in defining explicit author-retained rights for articles and earlier manuscript versions, the policies’ impact on author behavior are not direct and straightforward; that is the fundamental reason for this study to exist. One factor complicating the relationship is awareness (Swan & Brown, 2005; Covey 2009). Authors might not actually read the terms of the publishing agreement. Another factor is attitude (Rowlands & Nicholas 2005; Swan & Brown, 2005; Creaser 2010; Spezi et al 2013). The terms in the publishing agreement might consciously be breached for a variety of reasons. Research looking into the relationship between publisher policies and green OA prevalence has shown weak or even inverse relationship between what publisher copyright policies allow and the likelihood of finding a green OA copy on the web (Covey 2009; Antelman 2006). To summarize the focus of this study, Figure 1 presents a diagram suggesting that (publisher) *agreements*, (author) *awareness*, and (author) *attitude* are moderating factors for determining *what*, *when* and *where* publisher agreement-compliant green OA copies are made available. This study explores the relationship between *agreements* and *what*, *where* and *when* green OA copies could be made available within the agreement limitations.

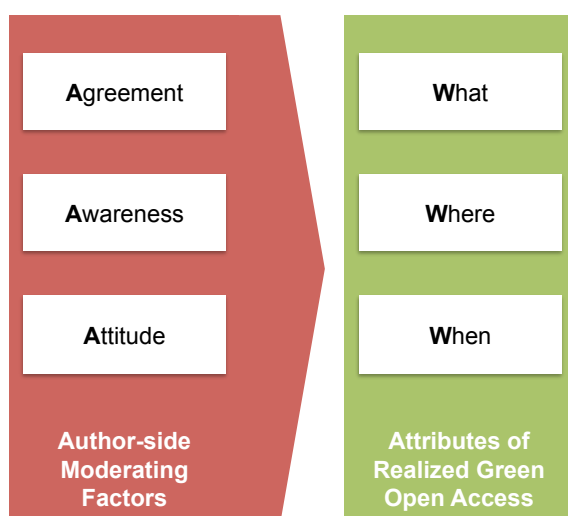


Fig. 1 Suggested relationship for interpreting author-side moderating factors and their combined influence on attributes of realized green OA

Previous studies

Publisher copyright policies have been analyzed in different types of quantitative studies. Most earlier studies have been limited to the publisher or journal level, looking at the number of publishers or journals allowing or prohibiting uploading of green OA copies, or have not included green OA embargoes in the analysis.

Publisher and journal-level studies

The earliest large-scale analysis of publisher policies with particular focus on author-retained green OA rights was performed by Gadd, Oppenheim & Proberts (2003). The authors contacted large publishers of highly ranked journals in the WoK (Web of Knowledge) (Web of Knowledge 2013) and Ulrichs (UlrichsWeb 2013) databases, asking for specific details regarding their publisher policies. Only 48 of 84 publishers provided usable answers, however, the dataset was further complemented with publisher policy data received through other sources. All in all the authors were able to analyze the policies for 80 publishers, collectively publishing 7,302 journals in the year 2002. As a result of the analysis only 49.1% of the journal titles allowed green OA. The authors also explored the specific arguments publishers use for why a publishing agreement is needed in the first place, with “protection from copyright infringement”, “effective third-party permissions”, and “wide dissemination of article” being the most frequent. Though the analysis is limited to the journal-level and the publisher dataset used had limitations (incomplete survey response rate, publishers drawn from various sources, list of included publishers not provided hindering replication) the results contribute valuable initial insight into a broader landscape of journal publisher copyright policies.

One of the most recent journal-level studies is Miguel et al (2011) which combines data from the Scopus database, Directory of Open Access Journals (DOAJ 2013) and the SHERPA/RoMEO database (SHERPA/RoMEO 2013) to compare the profiles of gold OA journals, journals that allow green OA uploading and journals that do not allow it. A problem in interpreting the results is the bundling of journals into a single group that explicitly forbid green OA posting and those for which no information could be found or which lack a policy. The study found that only 32 % of journals had an explicit policy allowing green OA posting. Geographical differences in journal policies were also explored based on the region of the world the journals were based in. The study found higher percentages of journals allowing green OA in Europe (37.4 %) and North America (34.7 %), while smaller percentages were measured for Asia, Africa and Oceania (ranging from 15.5% to 16.2%), and Latin America (0.4%). The low figures for the last four regions can partly be explained by proportionally higher shares of gold OA journals, and by publishers from such regions usually being small and lacking explicit policies altogether.

While not a study in the traditional sense, the continuously updated statistics for the green OA policies of publishers included in the SHERPA/RoMEO database makes the service a very important source of summarized policy information by itself (SHERPA/RoMEO statistics 2013). The data is aggregated by a mixture of information coded from publisher policies and publisher self-reported information. At the time of writing, May 2013, the service covered 1250 publishers, and provides summarized data about the publisher policies both in chart and table form. Table 1 contains data replicated directly from the website and provides a comprehensive summary of the publisher-level policies in the database. In an effort to make the essential information easily accessible for a wide audience SHERPA/RoMEO categorizes publishers by different colors based on the policy restrictions. A summary of the publisher policies in the SHERPA/RoMEO database is provided in Table 1.

RoMEO colour	Archiving policy	Publishers	%
Green	Can archive pre-print and post-print	369	30
Blue	Can archive post-print (ie final draft post-refereeing)	397	32
Yellow	Can archive pre-print (ie pre-refereeing)	97	8
White	Archiving not formally supported	387	31

Table 1 Summarized publisher policy information from the SHERPA/RoMEO database as of May 2013 (SHERPA/RoMEO statistics 2013)

Based on the SHERPA/RoMEO summary 62% of publishers allow post-prints, which are also referred to as accepted manuscripts, to be disseminated as green OA. There are benefits and drawbacks to consider in using the SHERPA/RoMEO database as a source of empirical data. The viability of utilizing this database as a foundation for further policy research is discussed in the Methods section.

Article-level studies

John Cox and Laura Cox present a comprehensive article-level analysis of green OA in Morris (2009). Their survey addresses publisher policies from 181 WoK publishers and was augmented with data concerning 2007 year article volumes for the respective publishers. The data accounted for 753,037 (or 74.7%) of all the articles covered by WoK that year. Publishers allowed posting 80% of the accepted manuscripts to personal or department websites, 70% to an institutional repository and over 40% to a subject repository. Publishers were considerably more restrictive with the published versions of their articles, Less than 20% of the published articles were allowed to be disseminated by means other than the publisher's website.

Hansen (2012) also focused on journal articles included in the WoK index in an article-level study on the author-retained rights of journal articles output by five major U.S. research universities. Hansen (2012) queried the WoK index for articles published during 2011 by authors affiliated with the pre-defined institutions and matched the article citation data by individual ISSNs to publisher policy information contained in the SHERPA/RoMEO database. Of the 29,322 unique articles included in the analysis 15.4% could archive the publisher version immediately upon publication and an additional 5.37% within embargos ranging from 6 to 24 months. Posting the publisher version was explicitly prohibited for 58.5% of the articles. The corresponding figures for accepted manuscripts were 55.46% immediately without restrictions and an additional 19.25% within a maximum embargo of 24 months. For 4.65% of the articles archival of the accepted manuscript was explicitly prohibited. The study is limited by use of the exclusive WoK index and reliance on SHERPA/RoMEO for copyright data. The study found 7.4% of the articles had unclear or unknown rights regarding accepted manuscripts and 12.67% was unclear or unknown for the publisher version. It would have been interesting to see the share of articles which are either allowed an accepted manuscript or publisher version to be uploaded as the two categories are treated without measuring overlap. It should be noted that the study did not identify or single out OA journals from the analysis, nor did the study look at the type of web locations which allowed or prohibited for green OA copies.

As a summary to the reviewed previous research it can be concluded that there is a lack of studies incorporating journal literature which could be considered broad enough in either sample or population to represent the wide landscape of journal publishing.

Methodology

Research aim

The aim of this study is to explore the degree to which green OA could provide free access to articles published in subscription-based scholarly journals if all authors would exercise the rights granted to them by journal publishers.

There are variations on *what* can be uploaded (preprint, accepted version, publisher version), *when* a content is allowed to be uploaded (upon publication, some specified time after publication), as well as *where* (author website, institutional repository, subject repository, other websites). This study aims to cover all of the above variables in its scope; what versions can be uploaded, the locations permitted for each version, and any potential delay enforced in providing free access to such full-text copies. The Scopus bibliographic database was used as the main source of data due to it being far more inclusive than the WoK (Larsen 2010), thus arguably being a better representation of the scholarly journal publishing landscape.

Publishers, journals and article volumes

Since the main Scopus database requires a subscription and limits the amount of data that can be exported with each single query, this study utilized a data acquisition method documented in Solomon et al (2013) which merges data from two freely available resources to reconstruct the Scopus data. Elsevier, the owner of Scopus, provides a frequently updated spreadsheet containing all indexed journal titles along with a rich set of metadata for each entry. The spreadsheet along with other metadata includes ISSN (International Standard Serial Number), publisher, subject categorization (Elsevier 2013a). The second resource which was merged with the Elsevier metadata via ISSN was Scopus data available from the SCIMago Journal and Country Rank website (SCIMago, 2007). Through a direct collaboration agreement with Elsevier, SCIMago provides delayed and aggregated free access to journal level information contained in the Scopus database. Journal citation averages, article counts and other statistics for individual journals starting from 1999 are all available by year. After obtaining the ISSN for the journals in Scopus from the downloadable spreadsheet, we developed software that automated the process of searching the SCIMago database by

ISSN and capturing a digital copy for each journal of the full set of information displayed on the site for the years 1999–2011. The software extracted the statistics for each year including article counts. The data were captured on July 26 and 27, 2012. The data from 2011 appeared to be incomplete in the SCImago database at the time of query and was not used in the study.

Of the 6038 unique publisher names acquired 98 were identified as duplicate entries and merged either due to small variations in spelling and explicit imprint ownership with identical policies. The 100 largest publishers in terms of the number of articles published annually were identified using data for over 18,000 journals. The publisher size ranged from Elsevier with 1,887 journals and 329,747 articles in 2010 to the IEEE Computer Society with 11 journals and 1,208 articles. As such small publishers were not included by design. Even if the limitations of manual data collection were removed the main problem is that many small publishers have no explicit policies (as observed by Miguel et al (2011)). 10 dominantly immediate gold OA publishers part of the initial top 100 were excluded from the analysis and in order to focus on opening access to otherwise restricted content through green OA. For the record these gold OA publishers accounted for a combined 535 journals and 51,819 articles during 2010. Five publishers were also discarded and replaced due to the publisher site being unavailable and SHERPA/RoMEO also missing data for said publishers. The publishers were categorized into one of six different publisher types according to the same grouping criteria as in Laakso & Björk (2012; 2013): Commercial publisher; Scientific society or professional association; University Press; Professional non-commercial publisher; University, University department or research institute; Individual scientist or group of scientists.

Publisher policies

Initially the use of the SHERPA/RoMEO publisher policy database was considered as the main source of data like similar studies have done in the past. However, while the database is the most comprehensive service of its kind it, it is currently too limited for obtaining accurate results considering the level detail needed for the present study. For example, permitted locations and embargo lengths are not stored in a structured way and are generally missing from the metadata. Specific information for different versions of the article are also commonly missing. The database might also be outdated compared to the publisher website, with information for some publisher being several years out of date. Thus, to ensure the quality of the study and the reliability of the results the decision was made to primarily consult each of the 100 publisher websites directly.

As it is impossible to retrospectively review the publisher policies exactly as they were when published in 2010 the aim of this study was to create a reliable snapshot of the policies as they were formulated during Spring 2013. The year-to-year variations in relative publisher volume can be assumed to be fairly stable.

The publisher policies of each publisher were reviewed, looking specifically for information about if, when and where preprints, accepted manuscripts or publisher versions can be uploaded to the web. While most of the data points of interest were known a priori there was some evolution in the framework as data collection progressed to accommodate for all observed values for embargo length categories. To facilitate reliability and repeatability of the study the URLs to each publisher policy was recorded and paragraphs of text supporting key coding decisions for each of the three article versions were captured as part of the data collection. The final publisher policy-coding framework is available in Table 2.

Data Points	Coding
Preprint	
Explicit policy available	YES/NO
Preprint permitted	YES/NO
Location	
Personal website	YES/NO
Institutional repository	YES/NO
Subject repository	YES/NO
Elsewhere online with free access	YES/NO
Accepted manuscript	
Accepted manuscript permitted	YES/NO
Embargo	
6 months	YES/NO
12 months	YES/NO
18 months	YES/NO
24 months	YES/NO
Location	
Personal website	YES/NO
Institutional repository	YES/NO
Subject repository	YES/NO
Elsewhere online with free access	YES/NO
Publisher version	
Published article permitted	YES/NO
Embargo	
6 months	YES/NO
12 months	YES/NO
18 months	YES/NO
24 months	YES/NO
Location	
Personal website	YES/NO
Institutional repository	YES/NO
Subject repository	YES/NO
Elsewhere online with free access	YES/NO

Table 2 Publisher policy coding framework

This study simplified the coding of the timing to be aligned with time of publication of the original article unless an embargo was specified. Prior self-archiving as in pre-submission, pre-acceptance, pre-publication were decided to be left outside the scope of the study as the main aim was to calculate the potential for green OA for published journal articles and not explore the possibilities and limitations for pre-publication dissemination.

Assumptions and methodology summary

A summarizing methodology diagram is provided in Figure 2.

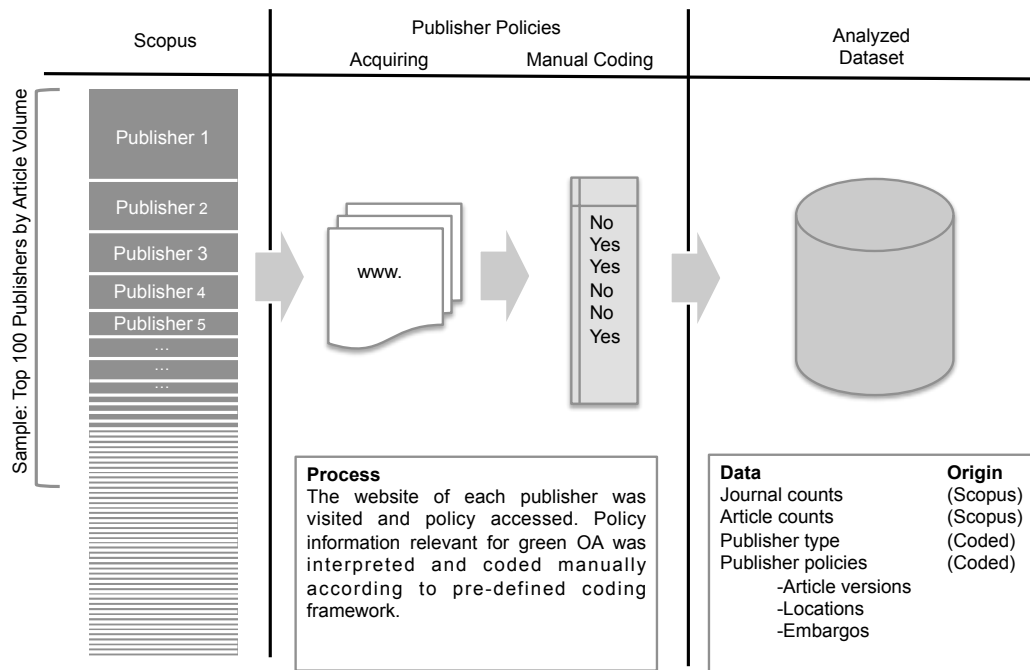


Fig. 2 Methodology visualized

Simplifying a wide variety of complex agreement terms into binary coding for facilitating quantitative analysis requires some guidelines to standardize the coding and interpretation of the terms. As such, the following assumptions were used throughout the data collection process:

- The author should not need to have a research funder or institution mandate to elevate rights for green OA, nor have to pay or contact the publisher to be granted permission.
- The publisher policy was retrieved from the publishers main website where possible; where such a policy was not available the copyright agreements for a few notable journal titles of the publisher were used instead.
- If the publisher website could not be found the SHERPA/RoMEO database was used for supporting the decision-making since it might contain publisher-provided information which might be useable for decision-making.
- If the journal publisher had an explicit preprint policy, options not mentioned within it are assumed not permitted. Where no such policy existed it was interpreted as green OA of preprint not being formally supported, which is handled separately in the results analysis.
- For accepted manuscripts and final versions everything explicitly not permitted in the publisher policy was interpreted as not permitted.

There is one main weakness with the adopted approach related to the assumption that all journals of a publisher use the same copyright agreement, however, that is commonly the case since the policy documents for publishers are usually found on a common web page linked to from all journals. Where disciplinary differences in copyright restrictions were found they usually related to embargo length, the articles of that publisher were split according to discipline based on the journal they were published in and articles were allocated to their respective embargo lengths.

Another aspect to consider is that some publishers include OA journals in their portfolios among subscription journals. Though it would theoretically be possible to match the Scopus data with the over 9900 gold OA journals currently listed in the DOAJ (Directory of Open Access Journals) it would have the drawback of introducing the need to verify the OA status of the journals retrospectively for the year 2010, a complication which has been dissected in e.g. Laakso and Björk (2012) for a much smaller sample of journals. Since the focus of this study is on green OA, generalizing downwards from the publisher level to the article level, pursuing this additional step was deemed not worth the added investment. This aspect should only have the effect of the results under-estimating the relative share of articles which become available to readers immediately upon publication.

Results

Descriptive statistics

Sample and population sizes are provided in Table 3. The sample is not assumed to be representative of the whole population, Table 3 is provided to convey the scope of the coverage of the analysis. The publishers included in the analysis, the 100 largest publishers by article output in 2010, published a combined 1 150 827 articles in 8 578 journals. This equates to 68.3% of all articles and 49.1% of all journals included in Scopus for the year 2010. After the publisher grouping process described in the methods section 5940 publishers were identified, of which the 100 included in the analysis make up for 1.7%. These figures demonstrate the large variation in size among journal publishers, the fact that 1.7% of publishers output 68.3% of all articles implies that the majority of the publishers outside of the sample have relatively small article output.

	Included in Sample	Full Population	Sample/Population
Articles	1 150 827	1 684 721	68.3%
Journals	8 578	17 470	49.1%
Publishers	100	5 940*	1.7%

*=Approximate value, see article text for further description.

Table 3 Descriptive statistics of sample and full population

Scopus contains a lot of supplemental information about the journals included in the index, of which one section is related to discipline classification. The top level of this classification contains four main scientific disciplines: Social sciences, life sciences, health sciences and physical sciences. Scopus categorizes journals into one or several of these discipline categories. This categorization was utilized in order to get a coarse overview of the discipline distribution of the sampled publications. Discipline data was available for 1 066 079 articles (92.6% of all articles included in the analysis). Since this study emphasizes article-level analysis this approach was used for exploring the discipline distribution as well. Where a journal was categorized into multiple categories the article volume was divided equally between the disciplines. The results of the analysis were the following: Social sciences 91,729 articles (9%), life sciences 202,833 articles (19%), health sciences 282,096 articles (26%), and physical sciences 489,421 articles (46%)

The publisher types were distributed as follows: 42 Commercial publishers, 52 societies or professional associations, 6 University Presses. However, the line between publisher types is not always clear. For example, some societies use commercial publishers to facilitate the publication process. To make interpretation as straightforward and replicable as possible the organization labeled as publisher in Scopus was used for making the classification.

Something which was not accounted for from the outset of the data collection and design of the coding framework was the fact some publishers explicitly limit green OA rights but instead open up all online issues for anyone to read for free after a set embargo period. This type of delayed open access was recently estimated to account for close to 500 high-impact journals publishing over 111 000 articles in 2011, with a 12 month embargo length being the most common (Laakso and Björk 2013). In this study eight such publishers were discovered, all with 12 month embargos, all of which do not permit self-archiving of the published article. Five of these publishers, with a combined article output of 9283 articles, do not permit any self-archiving of the accepted manuscript so are thus excluded from all results calculations. However, since these articles are openly available on the web after 12 months of publication it is important to highlight their contribution towards increasing the share of open access articles despite not being provided through self-archiving.

Publisher types

Comparing the policies between publisher types reveals some interesting differences, the full results are provided as Table 4. Most large publishers have explicit preprint policies, with only 4% of articles output by commercial publishers not being covered by such a policy. For societies and professional associations the respective figure is 15%, while university press publishers have no articles outside explicit preprint policies. When it comes to accepted manuscripts the publisher types are relatively similar in their policies: commercial publishers allow 82% of articles to be self-archived, societies and professional associations 78%, and university presses 96%. When it comes to permitting self-archiving of the published article commercial publishers are the most restrictive with only 1% of the all articles published by this group being granted permission for this. While also restrictive with rights for published articles, society and professional associations allow for 39% to be self-archived at some point in time, and university presses 41% respectively.

Article Version	Policy	All Publisher Types Combined				Commercial Publishers				Societies or Professional Associations				University Presses			
		ART	%	J	P	ART	%	J	P	ART	%	J	P	ART	%	J	P
Preprint	No formal support	75 987	6.6	357	19	36 022	4.2	311	8	39 965	15.1	46	11	0	0	0	0
	Not permitted	125 561	10.9	453	21	54 103	6.3	314	7	70 118	26.5	114	13	1 340	4.0	25	1
	Allowed	949 279	82.5	7 768	60	762 333	89.4	6 710	27	154 414	58.4	614	28	32 532	96.0	444	5
Accepted Manuscript	Immediately	743 613	64.6	4 579	54	545 852	64.0	3 804	19	188 153	71.1	618	33	9 608	28.4	157	2
	6 months	39 009	3.4	246	5	30 708	3.6	233	3	8 301	3.1	13	2	0	0	0	0
	12 months	128 413	11.2	1 631	13	101 923	12.0	1 432	7	6 819	2.6	24	4	19 671	58.1	175	2
	18 months	20 935	1.8	697	2	20 935	2.5	697	2	0	0	0	0	0	0	0	0
	24 months	3 253	0.3	112	1	0	0	0	0	0	0	0	0	3 253	9.6	112	1
	Not permitted	215 604	18.7	1 313	25	153 040	18.0	1 169	11	61 224	23.1	119	13	1 340	4.0	25	1
Publisher Version	Immediately	106 475	9.3	329	16	11 344	1.3	168	2	95 131	36.0	161	14	0	0	0	0
	6 months	9 763	0.8	30	4	0	0	0	0	9 763	3.7	30	4	0	0	0	0
	12 months	15 426	1.3	251	5	1 407	0.2	28	1	0	0	0	0	14 019	41.4	223	4
	18 months	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	24 months	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Not permitted	1 019 163	88.6	7 968	75	839 707	98.5	7 139	39	159 603	60.3	583	34	19 853	58.6	246	2
Total (n)		1 150 827		8578	100	852 458		7335	42	264 497		774	52	33 872		469	6

ART= Articles, J= Journals, P= Publishers

Table 4 Results (Article versions and embargos across publisher types)

Locations

Publisher policies vary in the type of web location they allow posting of a green copy. Figure 3 provides a summary of the publisher policies by permitted locations, measured as the percentage of total articles across the three document types. Personal webpage and institutional repositories were nearly identical in their self-archiving policy profile across all three document types; permission to self-archive preprints and accepted manuscripts is granted by nearly every publisher policy supporting some form of self-archiving and support for self-archiving the published article is within the range of 9% to 11% of all articles. Uploading of accepted manuscripts to subject repositories was generally found to have low support among publishers, with only 32,8% of all articles across all embargo lengths being granted permission to be made available at such locations. Furthermore, most of the publishers allowing subject repository upload do so on the condition that the repository is non-commercial which further narrows down the web services to where the manuscript is allowed upload.

Publishers are generally restrictive with allowing self-archiving on any freely available website, showcased by the range of 1% to 17% of articles permitted on such locations. A detailed breakdown of the location results for each article version including embargo information is provided in Table 5. Though personal homepages and institutional repositories are nearly equal in the amount of articles which could be uploaded to each as Figure 3 highlighted, there is a notable difference in distribution of embargo length for the two locations. The results in Table 5 show that 64.9% of all articles could immediately have had an accepted manuscript uploaded to a personal webpage, while the respective figure for institutional repositories is only 47.7%. However, at 12 months after publication the difference is evened out almost completely.

Article Version	Policy	Any Location		Personal Website		Inst. Repo.		Sub. Repo.		Other Website	
			%		%		%		%		%
Preprint	No formal support	75 987	6.6	75 987	6.6	75 987	6.6	75 987	6.6	75 987	6.6
	Not permitted	125 561	10.9	162 197	14.1	160 962	14.0	293 991	25.5	882 089	76.6
	Allowed	949 279	82.5	912 643	79.3	913 878	79.4	780 849	67.9	192 751	16.7
Accepted Manuscript	Immediately	724 613	64.6	734 812	63.9	545 915	47.4	67 289	5.8	55 299	4.8
	6 months	39 009	3.4	32 457	2.8	39 009	3.4	25 953	2.3	19 401	1.7
	12 months	128 413	11.2	111 813	9.7	310 056	26.9	260 149	22.6	58 917	5.1
	18 months	20 935	1.8	194 92	1.7	20 935	1.8	20 935	1.8	19 492	1.7
	24 months	3 253	0.3	0	0	3 253	0.3	3 253	0.3	0	0
	Not permitted	215 604	18.7	252 253	21.9	231 659	20.1	777 483	67.2	997 718	86.7
Publisher Version	Immediately	106 475	9.3	105 133	9.1	77 166	6.7	4 881	0.4	4 881	0.4
	6 months	9 763	0.8	5 237	0.5	9 763	0.8	3 417	0.3	3 417	0.3
	12 months	15 426	1.3	15 426	1.3	15 426	1.3	1 2491	1.1	0	0
	18 months	0	0	0	0	0	0	0	0	0	0
	24 months	0	0	0	0	0	0	0	0	0	0
	Not permitted	1 019 163	88.6	1 025 031	89.1	1 048 472	91.1	1 140 980	98.2	1 142 529	99.3

Table 5 Main Results (Article versions and embargos across locations)

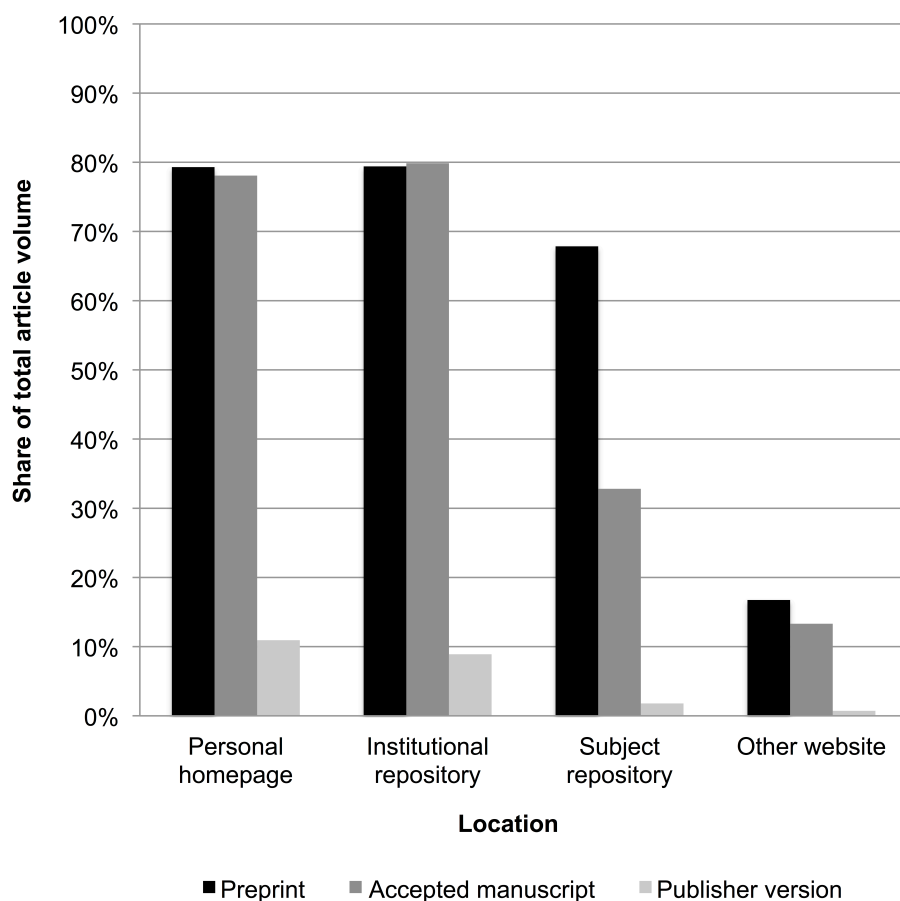


Fig. 3 Support for main green OA location types expressed as percentage of all top 100 publisher articles

Disciplines

The methods section described the journal-level approach used to allocate publisher article volumes across the four main discipline categories used by Scopus. The following results are based on the result of that discipline categorization of articles, and building upon that data by using the publisher policy coding to calculate the green OA profile for each of the four discipline groups separately. The results of this calculation can be found in Table 6. Despite the caveats of

having to use the publisher–level and very coarse discipline categorization to conduct this analysis, the results are still some of the most comprehensive available to date. For preprints, the social sciences (97.4%) and physical sciences (83.7%) were found to be the most permissive, with the life sciences (81.5%) and health sciences (79.1%) not far behind. For accepted manuscripts the differences across the disciplines were more even, however, while the total share of permitted articles were relatively similar (life sciences 77.1%, health sciences 78%, physical sciences 83.1%, social sciences 85.5%) the embargo distribution was not. To illustrate the differences Figure 4 provides a cumulative article volume embargo diagram for all four disciplines. Physical sciences is notably the only category with minimal embargos, while the three others are much more step-wise in their embargo expiry timelines.

Article Version		Life Sciences	%	Health Sciences	%	Physical Sciences	%	Social Sciences	%
Preprint	No formal support	9 589	4.7	17 059	6.0	4 4214	9.0	982	1.1
	Not permitted	2 8013	13.8	41 987	14.9	35 653	7.3	1 379	1.5
	Allowed	165 232	81.5	223 050	79.1	409 554	83.7	89 369	97.4
Accepted Manuscript	Immediately	120 759	59.5	143 987	51.0	380 486	77.7	41 079	44.8
	6 months	13 495	6.7	15 362	5.4	4 698	1.0	2 468	2.7
	12 months	22 140	10.9	60 696	21.5	21 446	4.4	11 085	12.1
	18 months	0	0.0	47	0.0	128	0.0	20 587	22.4
	24 months	0	0.0	0	0.0	0	0.0	3 253	3.5
	Not permitted	46 440	22.9	62 005	22.0	82 663	16.9	13 257	14.5
Publisher Version	Immediately	9 601	4.7	10 366	3.7	6 8804	14.1	31 73	3.5
	6 months	2 855	1.4	517	0.2	6 346	1.3	46	0.1
	12 months	2454	1.2	4 917	1.7	3 419	0.7	4 261	4.6
	18 months	0	0.0	0	0.0	0	0.0	0	0.0
	24 months	0	0.0	0	0.0	0	0.0	0	0.0
	Not permitted	187 925	92.6	266 297	94.4	410 851	83.9	84 249	91.8
Total		202 833		282 096		489 421		91 729	

Table 6 Main Results (Article versions and embargos across disciplines)

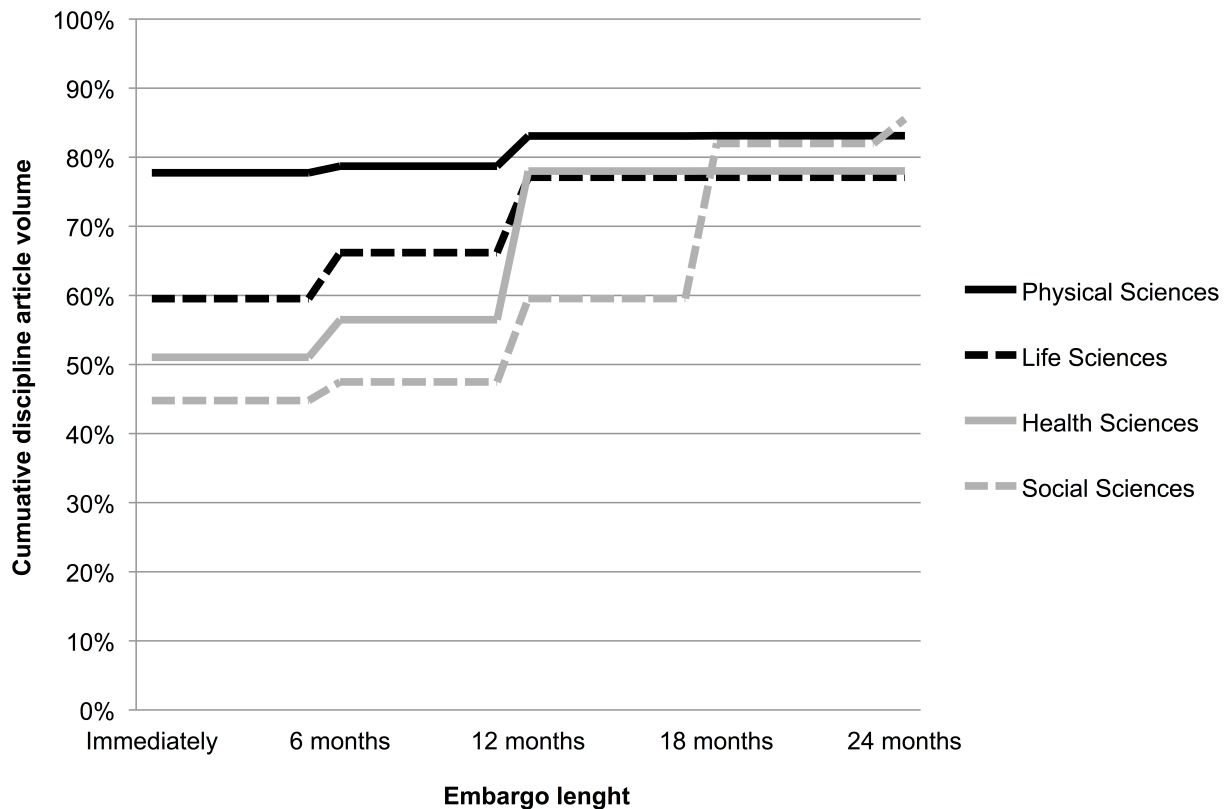


Fig. 4 Disciplinary differences in embargo length for green OA of accepted manuscripts

Cross-analysis for repository upload of accepted manuscript or publisher version

Figure 5 presents the cumulative percentage of articles that could be available as green OA either as accepted or publisher versions in either a subject or institutional repository at various points in time after publication. Since accepted versions should be identical in content to the final article and systematic repository storage the most reliable long-term solution for providing persistent OA this calculation should give a good estimation for what share of subscription journal content could be provided through sustainable green OA and with what kind of embargo distribution.

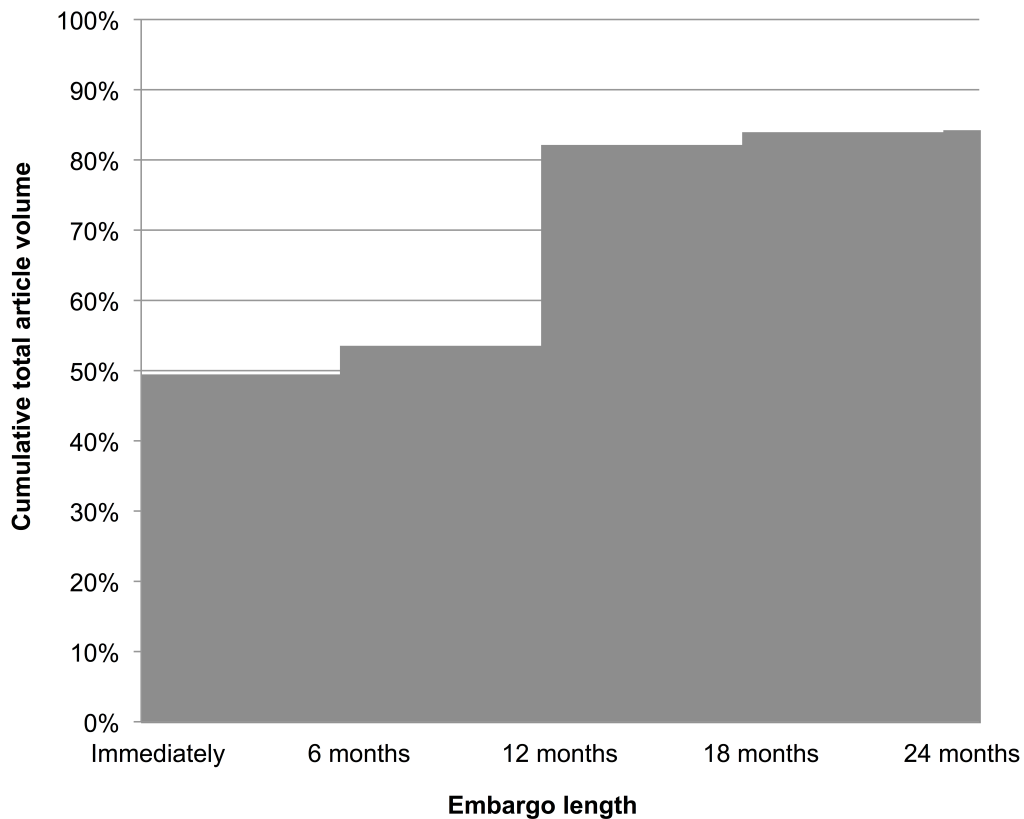


Fig. 5 Percentage of articles output by the top 100 largest journal publishers in 2010 that in theory could have been made available as green OA as accepted manuscripts or publisher versions in either institutional or subject repositories

As seen in Figure 5, nearly half of all articles (548,718) are permitted upload immediately upon publication with the share rising to a cumulative 80.4% of all articles after 12 months of publication (924,725). Only 2.1% (24,188) of the articles would be available after a longer embargo. Though repository self-archiving is restricted by embargos to a larger extent than author website self-archiving which is rarely embargoed, the dominant length of repository embargos is 12 months after which the difference in article shares allowed in either personal homepages or repositories is very similar.

Discussion and conclusions

The results of the study fall in line and extend upon the indications provided from earlier studies conducted on other bibliographic databases in combination with the SHERPA/RoMEO policy database. Publishers are relatively liberal in permitting distribution of accepted manuscripts (81% of all articles permitted) while distribution of the publisher version is considerably more restricted (11% of all articles permitted). Green OA has a lot of room to grow in terms of uptake, showcased by the discrepancy between the results of this study and recent studies measuring realized self-archiving which suggests realized green OA to be in the range of 12% of annual articles (Björk et al 2014).

The main contribution of this study is arguably the new insight gained into the green OA policies of journal publishers. There is also a methodological contribution providing a framework for compiling the raw data on publisher policies using openly available sources, and a coding scheme for dissecting the publisher policies for author-retained rights. The natural next step in this line of research would be to extend the coverage of publishers included in the analysis beyond just the 100 largest ones. The problem is the diminishing returns in terms of articles covered by each publisher, making it a very time-consuming to substantially increase the article coverage. A study based on a systematic random sample of smaller publishers would be one way to address this issue. Precision could also be added by performing the coding on a journal-level, where possible, rather than going by the often broader publisher policy. However, policies coded and analyzed today are likely outdated two years into the future considering the pace at which the scholarly publishing sector is currently evolving. This reduces the incentive to invest time and effort into an endeavor which only has limited temporal utility and cannot not be built upon further without re-doing most of the manual work involved.

In an age where automated indexing and crawling of web content leaves few stones unturned it is interesting to see substantial differences between the locations publishers permit for green OA. Where personal webpages and institutional repositories were found to be nearly universally permitted locations for green OA, with largely just a difference in embargo length, considerably fewer publishers supported subject repositories or other websites. While repository storage has been found to be more permanent and resilient to the dynamic nature of the web compared to files stored on websites (Björk et al 2014), social media (Tenopir et al 2013) and free full-text harvesting services like Google Scholar have arguably made location a less relevant factor when it comes to reach and visibility. Among web search engine results hyperlinks to green OA versions of articles are aggregated and placed side by side with the link to the publishers paid version for non-subscribers.

While this study has further confirmed the large potential increase in the share of green OA articles, there is no telling how publishers would react if the rate of realized green OA would substantially increase. It is unlikely that radical increases will happen purely based on more widespread voluntary uploading, however, green OA mandates enforced by institutions and research funders might be such an accelerant. Such a scenario might be met with publishers moving of the goalposts; if seen as a threat to the subscription-based business model the publishing terms might change towards being more restrictive, like introduction of longer embargos, stronger enforcement of copyright compliant self-archiving, and not permitting uploading to subject repositories. Indications of pre-emptive protection against systematic green OA posting can be found in several of the reviewed publisher policies, most notably Elsevier which otherwise permits accepted manuscripts to be uploaded to institutional repositories explicitly prohibits such uploading if authors' affiliated institution has OA mandate (Elsevier 2013b). Another signal of publishers taking a protective stand against potential growing systematic green OA posting can be seen in the CHORUS (ClearingHouse for the Open Research of the United States)(Publishers.org 2013) proposal which the Association of American Publishers have put together as a publisher-friendly answer to work around a potential full OA mandate for all Federally funded research in the United States (Whitehouse.gov 2013). Since many large publishers provide direct OA publishing of individual articles as an author-pays service, either through newly founded OA journals or hybrid OA, there are financial interests among such publishers in limiting green OA posting. If authors are obliged by their research funders to provide OA to their research but face prohibitive green OA restrictions from journal publishers the only viable option for authors obliged to provide timely OA copy of their work, a process which recent UK research policy recommendations might initiate as a side-effect (Finch, 2012).

The manual data collection process exposed the wide breadth of definitions, descriptions, and restrictions that publishers use to instruct authors about their retained green OA rights. Some publishers have the key information summarized in a single table or a few bullet points which makes it very accessible, while others sprinkle author permissions and restrictions across much longer bodies of text covering everything from manuscript formatting to technical requirements for figures. Authors would likely benefit from a higher degree of standardization in how the information is presented, making comparisons across journals straightforward. This kind of standardization development would also be beneficial for building new web services aggregating publisher policy information, much like SHERPA/RoMEO is today but always up to date with more detailed information available in a structured format (e.g. embargos, locations). Acknowledging the volatile nature of the policies, a request for this type of web service has also been expressed in a recent update by the UK working group on OA research policy recommendations (Finch 2013).

However, regardless of what potential future scenarios might entail there is something to be said for the importance of increasing author awareness of retained rights. At least until something more technically advanced comes around, the SHERPA/RoMEO database with its simplified policy classification scheme and open application programming interface is indispensable for making green OA provision a conscious decision for authors: simplified publisher policy information can be displayed as part of the process of entering a new publication into an institutional publication

database. The threshold for making a green OA copy available voluntarily is in such cases low, but what remains to be aligned is author attitude.

A study similar to this one could likely not have been performed ten years ago due to the lack of explicit green OA policies tailored for the dissemination possibilities available today. Even though policies now exist for most large publishers, the contents of them are not set in stone. Publishers are exposed to the pressures of various stakeholders and the competitive marketplace, making this study a snapshot of the status at the point in time of data collection. Based on the results from this study it is impossible to say whether the trend is moving towards expansion or restriction of author-retained rights for green OA, however, regardless of the direction it can be assumed that policies will become even more detailed and specific as the possibilities for systematic digital dissemination advance and commercial interests intersect.

References

- Antelman, K. (2006). Self-archiving practice and the influence of publisher policies in the social sciences. *Learned Publishing*, 19(2), 85–95.
- Björk, B.-C., Laakso, M., Welling, P., & Paetau, P. (2014). Anatomy of Green Open Access. *Journal of the American Society for Information Science and Technology*. In Press. doi: 10.1002/asi.22963
- Covey, D. T. (2009). Self-archiving journal articles: A case study of faculty practice and missed opportunity. *Portal: Libraries and the Academy*, 9(2), 223–251.
- Creaser, C., Fry, J., Greenwood, H., Oppenheim, C., Proberts, S., Spezi, V., & White, S. (2010). Authors' Awareness and Attitudes Toward Open Access Repositories. *New Review of Academic Librarianship*, 16(sup1), 145–161. doi:10.1080/13614533.2010.518851
- DOAJ (2013) Directory of Open Access Journals. Available at: <http://www.doaj.org>
- Elsevier. (2013a). What does it Cover? Content coverage guide for SciVerse/Scopus. Retrieved from <http://www.info.sciverse.com/scopus/scopus-in-detail/facts>
- Elsevier. (2013b). Article posting policy. <http://www.elsevier.com/about/open-access/open-access-policies/article-posting-policy#accepted-author-manuscript>
- Finch, J. (2012). Accessibility, sustainability, excellence: how to expand access to research publications .140p. Available at: <http://www.researchinfonet.org/publish/finch/>
- Finch, J. (2013). Accessibility, Sustainability, Excellence: How to Expand Access to Research Publications - A Review of Progress in Implementing the Recommendations of the Finch Report. 74p. Available at: <http://www.researchinfonet.org/wp-content/uploads/2013/02/Final-version.pdf>
- Gadd, E., Oppenheim, C., & Proberts, S. (2003). RoMEO studies 4: an analysis of journal publishers' copyright agreements. *Learned Publishing*, 16(4), 293–308.
- Hansen, D. (2012). Understanding and Making Use of Academic Authors' Open Access Rights. *Journal of Librarianship and Scholarly Communication*, 1(2), eP1050. doi:10.7710/2162-3309.1050
- Harnad, S., Brody, T., Vallieres, F., Carr, L., Hitchcock, S., Gingras, Y., et al. (2004). The access/impact problem and the green and gold roads to open access. *Serials Review*, 30(4), 310–314.
- Laakso, M., & Björk, B.-C. (2012). Anatomy of open access publishing: a study of longitudinal development and internal structure. *BMC Medicine*, 10, 124. doi:10.1186/1741-7015-10-124
- 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*, 64(7), 1323–1329. doi:10.1002/asi.22856
- Larsen, P. O., & Ins, M. (2010). The rate of growth in scientific publication and the decline in coverage provided by Science Citation Index. *Scientometrics*, 84(3), 575–603. doi:10.1007/s11192-010-0202-z
- Miguel, S., Chinchilla-Rodríguez, Z., & de Moya-Anegón, F. (2011). Open access and Scopus: A new approach to scientific visibility from the standpoint of access. *Journal of the American Society for Information Science and Technology*, 62(6), 1130–1145. doi:10.1002/asi.21532
- Morris, S. (2009). Journal Authors' Rights: perception and reality (PRC Summary Paper 5), <http://www.publishingresearch.net/documents/JournalAuthorsRights.pdf>
- Publishers.org (2013) Understanding CHORUS <http://www.publishers.org/press/107/>
- Rowlands, I., & Nicholas, D. (2005). Scholarly communication in the digital environment: The 2005 survey of journal author behaviour and attitudes. *Aslib Proceedings*, 57(6), 481–497. doi:10.1108/00012530510634226
- Scopus (2013) The Scopus bibliographic database. Available at: <http://www.scopus.com/> (subscription required)
- SCImago. (2007). SJR – SCImago Journal & Country Rank. Retrieved from <http://www.scimagojr.com>
- SHERPA/RoMEO (2013) SHERPA/RoMEO: Publisher copyright policies & self-archiving. Available at: <http://www.sherpa.ac.uk/>

SHERPA/RoMEO statistics (2013) SHERPA/RoMEO statistics. Available at:
<http://www.sherpa.ac.uk/romeo/statistics.php?la=en&fidnum=|&mode=simple>

Solomon, D. J., Laakso, M., & Björk, B.-C. (2013). A longitudinal comparison of citation rates and growth among open access journals. *Journal of Informetrics*, 7(3), 642–650. doi: 10.1016/j.joi.2013.03.008

Spezi, V., Fry, J., Creaser, C., Proberts, S., & White, S. (2013). Researchers' green open access practice: a cross-disciplinary analysis. *Journal of Documentation*, 69(3), 334–359.

Swan, A., & Brown, S. (2005). Open access self-archiving: An author study (p. 97). Key Perspectives Report. Available at: http://www.jisc.ac.uk/uploaded_documents/Open%20Access%20Self%20Archiving-an%20author%20study.pdf

Tenopir, C., Volentine, R., & King, D. W. (2013). Social media and scholarly reading. *Online Information Review*, 37(2), 193–216. doi:10.1108/OIR-04-2012-0062

UlrichsWeb (2013) UlrichsWeb Serials Solutions. Available at: <http://ulrichsweb.serialssolutions.com>

Web of Knowledge (2013). Thomson Reuters Web of Knowledge. Available at: <http://apps.webofknowledge.com> (Subscription required)

Whitehouse.gov (2013) Memorandum for the heads of executive departments and agencies
http://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf