

Anatomy of Green Open Access

Bo-Christer Björk, Mikael Laakso, and Patrik Welling

Hanken School of Economics, Helsinki Campus, PB 479, 00101 Helsinki, Finland.

E-mail: {bo-christer.bjork, mikael.laakso, patrik.welling}@hanken.fi

Patrik Paetau

Hanken School of Economics, Vaasa Campus, PB 287, 65101 Vaasa, Finland.

E-mail: patrik.paetau@hanken.fi

Open access (OA) is free, unrestricted access to electronic versions of scholarly publications. For peer-reviewed journal articles, there are two main routes to OA: publishing in OA journals (gold OA) or archiving of article copies or manuscripts at other web locations (green OA). This study focuses on summarizing and extending current knowledge about green OA. A synthesis of previous studies indicates that green OA coverage of all published journal articles is approximately 12%, with substantial disciplinary variation. Typically, green OA copies become available after considerable time delays, partly caused by publisher-imposed embargo periods, and partly by author tendencies to archive manuscripts only periodically. Although green OA copies should ideally be archived in proper repositories, a large share is stored on home pages and similar locations, with no assurance of long-term preservation. Often such locations contain exact copies of published articles, which may infringe on the publisher's exclusive rights. The technical foundation for green OA uploading is becoming increasingly solid largely due to the rapid increase in the number of institutional repositories. The number of articles within the scope of OA mandates, which strongly influence the self-archival rate of articles, is nevertheless still low.

Introduction

What Is Open Access and What Are the Benefits?

Open access (OA) is a term used to describe a radical new dissemination model for scientific research publications. OA is gradually replacing the earlier method of selling journal subscriptions and restricting access to paying readers only, a model that matured and established itself during the

era of printed journals. Leading OA advocate Suber (2012) describes it in the following way: “Open Access (OA) literature is digital, online, free of charge, and free of most copyright and licensing restrictions” (p. 4). The most important factor is free access for readers; literature that is merely free without granting liberal re-usage rights is still considered OA.

There are a few commonly used key arguments for why scientific research should be made available OA. One argument is that the increased availability of research results leads to a faster advancement of science, knowledge, and commerce (Willinsky, 2005). Another argument is that because scientific research is predominantly financed by public funds, the results should be considered a public good, which ought to be freely available to the public. An additional argument is that OA, taking into account its effects inside the scientific publishing and dissemination process, would reduce the global costs of the process when compared with the subscription model (Houghton et al., 2009).

Gold and Green OA

There are several variations to the open accessibility of scholarly articles, depending on user rights, the timing of availability, the funding of OA publishing, and whether the reader finds the original or a manuscript copy of the article (Willinsky, 2005). The terms *gold* and *green* OA were coined around 2004, but these two alternative channels for providing open accessibility had arisen much earlier. Harnad et al. (2004, p. 310) define green OA as “publish your article in a non-OA journal but also self-archive it in an OA archive.” Green OA is when such articles, usually in the form of the author manuscripts that preceded the finalized article, are made freely available somewhere on the web. A manuscript is for many uses close enough in content to the final published article, if not for direct citation then at least

Received December 18, 2012; revised February 11, 2013; accepted February 27, 2013

© 2013 ASIS&T • Published online 6 November 2013 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/asi.22963

as a basis for pre-purchase evaluation. The publisher-provided article abstracts seldom succeed in conveying enough information to draw conclusions about the scope and quality of the work. Another key purpose of green OA is to increase the dissemination of the research results by making them available and thus citable to nonsubscribing authors and to those with limited resources to finance their scholarly digest. None of these benefits causes any loss of revenue for the authors themselves (because they receive no royalties), but it increases the potential for readership and citations and opens up the content of journal articles for scrutiny. Green OA was recently argued to be the most cost-effective and affordable means for funders, institutions, and other stakeholders to move toward OA (Houghton & Swan, 2013).

Self-archiving is often used as a synonym for green OA; however, in this study green OA is defined as all freely accessible copies of articles, including different versions of said articles, which exist on other web locations than the original publisher's website. This includes, for example, copies self-archived by authors, copies uploaded to institutional repositories by librarians, and copies stored in subject repositories by the publishers, for example, PubMedCentral (PMC). The Joint Information Systems Committee (JISC) uses the term *mediated deposit* to describe the latter categories (JISC infoNet, 2012). A key difference between gold OA and green OA is that with gold OA, the entire journal content becomes available at a single location on the web, whereas with green OA, copies of a random and limited selection of the articles are scattered around the web. Thus, web search engines are commonly used to establish if a green OA copy of a specific article is available somewhere on the web. Another difference between gold and green OA is that although readers of gold OA articles usually have well-defined reuse and data-mining rights (so-called libre OA, often defined using Creative Commons licenses), readers of green OA copies can usually only read the manuscripts (so-called gratis OA).

Types of Green OA Copies

An article usually completes a number of life-cycle stages before final publication in a journal, and green OA

copies can be made at any of them. In some cases, the manuscript has previously been published as an e-print or working paper. Even if such prior versions seldom are identical to the submitted versions, they regularly show up in web searches for green OA copies if the titles are identical and may provide essentially the same key content to interested readers. Even after acceptance, many publishers still make minor content changes during copy editing and then the manuscript is designed and page numbering is fixed. The key manuscript stages are shown in Table 1 as well as the various terms that have commonly been used to describe them.

Both the SHERPA/RoMEO index of publishers copyright restrictions concerning green OA (SHERPA/RoMEO, 2012) as well as leading OA advocate Suber's OA guidelines (Suber, 2012) use the term *preprint* to refer to the two first stages and *postprint* for the two latter stages (the term *postprint* is rather peculiar because publishing is less dependent on the paper medium and in the past the accepted version was never printed). The copyright agreements that publishers require authors to abide by usually refer to the last three of these stages, and different publishers use slightly different terms. Two useful glossaries for the different stages of article publication are CrossRef (2012) and NISO (2008). For this study, we have opted to refer to the main versions of green OA in the following terms: *submitted manuscript*, *accepted manuscript*, and *published article*.

Taking the perspective of readers, other versions than the published article are usually considered inferior for various reasons. Publishers often require that citations reference the original publication even if the citing author only has access to a green OA copy, and in some disciplines there could be some risk involved in bypassing the published article. Although preprints have a strong presence in some disciplinary cultures, mostly due to the instant availability and wide potential readership they provide, postprints are generally held as the preferable alternative if the article has already been refined into a published journal article. In the social sciences and humanities, references are often made to particular pages in a publication, and in such cases access to the published version, or persistently available green OA copy with identical pagination, is essential.

TABLE 1. Different versions of green OA copies.

Stage	Definition	Terms used
Working paper	A working paper uploaded to an e-print repository	Preprint, Authors, unsubmitted draft
Submitted Manuscript	The version of the manuscript submitted to the journal	Preprint, Author's original draft
Accepted manuscript	The accepted version, after peer review but prior the final copy-editing and layout	Postprint, personal version, accepted author manuscript, final author version
Published article	An exact digital replicate of the published article	Postprint, Version of record, Publisher's version, Published journal article

Note. OA = open access.

Green OA copies, according to the previous broad definition, can be found in different types of locations. The three most popular locations are institutional repositories, subject repositories, and personal/departmental web sites of the authors. Many authors discussing green or self-archived OA have restricted themselves to copies found in repositories, whereas others, this study included, use a broader definition encompassing any location outside the publisher's original site.

The term *institutional repository* (IR) refers to highly structured collections of digital material emanating from scholars employed by a university or a research institution (e.g., the European Organization for Nuclear Research CERN) (Lynch, 2003). In addition to article manuscripts, IRs can contain other types of content, theses in particular, but also teaching material, videos and images, and various data sets. Although authors are the key contributors of content, professional librarians are usually involved in the quality assurance process, checking metadata and permissibility of upload as well as ensuring the long-term preservation of the content.

The earliest successful subject repositories were started by scholars or groups of scholars as voluntary operations. The highly successful arXiv (arXiv, 2012) is by now more than 20 years old and houses more than 800,000 preprints in physics, mathematics, and related fields. In economics, a slightly different model has evolved, with RePec providing an overlay indexing service on top of more than 1,400 archives containing working paper series of individual universities, departments, and so on. PMC, on the other hand, is a highly centralized database of medical publications, maintained by the world's largest funder of medical research, the National Institutes of Health (NIH). Due to the explicit requirement that grantees of NIH must deposit green OA copies of their publications in this database, PMC has become highly influential in setting an example for other research funders to follow.

Especially in the early days of the web, most green OA copies were to be found on the personal web pages of the authors, or the pages of their departments, often linked with the curricula vitae (CVs) or publication lists of the authors in question. At the time, this was the only possibility in most disciplines, but repositories have begun to offer viable alternatives. The major subject repositories and most institutional repositories can be expected to be relatively permanent and able to handle hardware and software upgrades in the future.

When repositories started to emerge there was an effort to standardize their data interfaces in the form of the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) to allow interoperable harvesting of content metadata by third-party web services. In practice, only readers searching for material in the largest subject repositories rely on browsing or searching in the repositories themselves. Instead, academics mostly use general web search engines

or specialized ones, such as Google Scholar, to find articles and to retrieve previously identified articles. The main function of a repository is consequently to secure long-term archiving of manuscripts and articles and facilitate indexing and visibility in search engines.

Previous Research

During the past 15 years, research related to green OA, and in particular to institutional and subject repositories, has ranged from descriptive reporting of individual cases to studies applying rigorous conceptual frameworks. In the following sections, some of the most central studies are briefly reviewed through a categorization into six distinct research areas.

Citation Advantage of OA

There have been dozens of studies focusing on changes in citation rates resulting from articles being openly accessible. Recent reviews of such studies by Swan (2010), Wagner (2010), and the Opcit Project (2012) provide good overviews. Most of the studies have used articles in subscription journals, for which green copies have been made available, to test for any potential citation advantage of OA. It seems almost indisputable that there is some increase in citations, but the degree of influence and other factors at play have been contested. A longer discussion of this topic is, however, outside the scope of this article.

The Prevalence of Green OA

The uptake of green OA has been studied in two ways. In longitudinal studies, the growth in the number of repositories (in particular IRs) and the number of items deposited in them has been the only feasible method. Morrison (2012) documented that the number of repositories registered in the Directory of Open Access Repositories (OpenDOAR, 2012) registry has grown from slightly above 800 in 2006 to more than 2,200 in 2012. One major weakness of this method is that information regarding repository size is usually limited to the total number of items, including other content in addition to green OA copies of journal articles. Another weakness, in view of our definition of green OA, is that it does not lend itself to the study of green OA copies on author home pages and other websites.

An alternative approach is to start with a sample of articles published in peer-reviewed journals (excluding OA journals) and proceeding to check for available green OA copies, an approach usually limited to making snapshots of the situation unless checks are done at multiple points in time. The availability can be checked either automatically by software searching for full-text copies or manually by having someone search for the copies using a web search engine, simulating the circumstances and article retrieval process common for interested readers. The former method is the only feasible one for larger numbers of articles,

whereas the latter enables a more precise classification to be made of the found copies but is limited to smaller article samples due to the time-consuming manual process.

Macro-level studies covering all sciences, for instance Hajjem, Harnad, and Gingras (2005) or Björk et al. (2010), have used article samples from article indexes such as Web of Knowledge (WoK, 2012) or Scopus (Scopus, 2012). For biomedical research, the bibliographical database PubMed has also been used as a source of article metadata (Matsubayashi et al., 2009). Micro-level studies have dealt with individual research disciplines usually by identifying the major journals in that field and then checking the availability of the articles published in them (e.g., Lyons & Booth, 2011).

The choice of metadata source strongly influences green OA prevalence measurements. For instance, Björk et al. (2010) reported a 14% green OA share for WoK-indexed journal articles versus only 5.5% for articles indexed in Scopus but not in WoK, a result perhaps due to the suggested selection bias of authors in choosing their better work for green posting (Moed, 2006; Swan, 2010).

Effects of Mandates on Green OA Uptake

OA mandates are formal requirements issued by either research funders as conditions in the grant contracts or by the employers of the researchers (research institutes or universities), which stipulate that, unless a researcher has published in an OA journal, a green OA copy must be made available. Examples of early and impactful funder mandates are those of the NIH (USA) and the Wellcome Trust (UK), both of which have had their OA requirements in place for a number of years. Funder mandates tend to be discipline-specific (e.g., NIH), whereas institutional mandates usually are multidisciplinary (i.e., University of Minho). A prerequisite is usually that researchers need to follow the copyright rules of the journal in which they have chosen to publish. NIH has, due to its size, been able to exert considerable pressure on the publishers to change their copyright policies, granting special conditions for their grantees. A recent mandate of considerable political importance is the new OA policy of Research Councils UK (RCUK, 2013), which requires that funded researchers publish either in gold OA journals or use the paid OA option in subscription journals, or self-archive copies of articles published in subscription journals. In contrast to many earlier mandates, a researcher can no longer avoid the OA requirement if the publisher does not have a gold or green option, thus excluding such journals as publishing outlets. Mandates come in different variations, for instance Gargouri et al. (2012a) graded the strength of institutional mandates on a scale from 1 to 12, ranging from no requirement to performance evaluation-linked immediate deposit. In the same study, Gargouri et al. (2012a) found a significant correlation between mandate strength and the ratio of deposits made by the institutions, demonstrating that enforcement of strong institutional mandates equals increased green OA deposits.

As for the effectiveness of funder mandates, the NIH has reported a compliance rate of 75% in the form of uploads by the authors of publishers to PMC (Poynder, 2012). The Wellcome Trust has reported a compliance rate of around 55%; however, the major part (85%) of this is achieved via publishing in gold OA journals for which the Trust provides earmarked funding (Finch, 2012). Gargouri et al. (2010) studied the uptake levels (in terms of all published journal articles) for three universities and one research institute with OA mandates and compared those with the uptake levels of a bigger selection of universities without mandates. They found an average uptake of around 60% for institutions with mandates compared to 15% to 20% for other institutions with voluntary upload.

Costs of Setting Up and Maintaining Repositories

Although some early OA advocates claimed green OA to be almost free of cost, there has been a growing realization that setting up and operating institutional repositories requires both human and financial resources, even if Open Source solutions (e.g., DSpace [<http://www.dspace.org>] or EPrints [<http://www.eprints.org/us>]) are mostly used as the information technology (IT) infrastructure for institutional repositories at major universities. In addition to the adaption, installation, and maintenance of the software and servers, library personnel may also be needed to verify the copyrights of uploaded copies, correct references, or provide advice to researchers, among other tasks.

Rough estimates of the costs of uploading and storing green OA copies of articles have been used in scenarios comparing different major strategies for how OA should be achieved. Houghton et al. (2009) estimated the costs for uploading copies of all journal articles in the United Kingdom to repositories at around \$33 USD per article, assuming that it takes 10 minutes of an author's time. One of the most systematic attempts to measure such costs was made in the European Community (EC)-funded Publishing and the Ecology of European Research (PEER) project (PEER, 2011). Average costs for setting up the IT architecture of a full repository were reported to be \$60,000 USD. The personnel cost per article uploaded were within the range \$2 USD to \$53 USD depending on the repository. These results are ambiguous because repositories can contain a wide variety of materials and the cost of setting up the repository and the staff managing it has to be spread over all document types.

Repository Case Studies

There are numerous reports of this particular type, but perhaps the most informative one is by Armbruster (2010), which includes case descriptions of 12 different repositories, many of which are linked with institutional or funder OA policies. Covey (2009) provides an excellent description of the behavior and attitudes of faculty at Carnegie-Mellon University, and Koskinen et al. (2010)

report on the acceptance and usage of the institutional repository at the University of Helsinki.

Davis and Connolly's (2007) study exploring the end-user reasons for accessing the Cornell University repository is interesting because it combines faculty interviews with usage log file analysis of the repository content. The analysis showed that almost all the collections in the repository exhibited either plateau or stair-step patterns in the growth of items, with only 4 of 107 collections exhibiting steady linear growth. This seems to indicate that authors or third parties upload materials to the repository in "batches," either periodically or as one-time efforts, rather than on a continual basis.

Author Attitudes and Behavior

Swan and Brown (2005) found that almost half of the respondents to their survey had self-archived at least one journal article in the past 3 years. Of the authors who had not self-archived, 71% were unaware of this possibility. A total of 81% of authors would willingly comply with an OA mandate from their funder or employer.

In a study commissioned by the Publishing Research Consortium (Morris, 2009), authors in WoK-indexed journals were found to prioritize sending copies of manuscripts and articles directly to colleagues. The posting of green OA copies to their own websites, subject, or institutional repositories was of less importance, in that order of preference. For all methods of dissemination, authors clearly preferred to use the published article, with the accepted manuscript second choice, and the submitted manuscript far less popular. The study found that authors substantially underestimated what publishers allow them to do with the submitted and accepted manuscripts and overestimated what the publishers allow them to do with the published article.

According to a more recent study using a mixed method of web surveys and focus groups (Creaser et al., 2010), more than one half of the respondents had deposited a peer-reviewed journal article in a repository during the past 5 years. The scholars showcased a clear preference to upload to a subject repository when compared with an institutional one, but only 37% knew of a suitable subject repository. A total of 70% of the authors who had uploaded an accepted manuscript reported that they had done so voluntarily.

In a survey of repository usage and scholar attitudes from the universities of New Zealand, Cullen and Chawner (2011) studied the reasons discouraging participants from depositing to an IR. Important barriers for non-depositors, in addition to the institution not having an IR or the author not being aware of one, were that no one had asked them to deposit, that they believed that the copyright policies of the publishers prevented them from doing so, and lack of time and knowledge on how to upload.

There have only been a couple of more ambitious efforts to explain author attitudes and behavior using theories developed in sociology, information systems, and

information science research. Kling and McKim (2000) were interested in the big differences between fields of science in the adoption of electronic communication and in particular the free dissemination of preprints. The authors explain the differences using a social shaping of technology perspective centered around disciplinary constructions of trust and of legitimate communication that develop at different speeds and in different directions depending on the field of science. Kim (2010) studied the motivations and barriers for author self-archiving using the Socio-Technical Interaction Network model proposed by Kling, McKim, and King (2003) and social exchange theory (Kankanhalli, Tan, & Wei, 2005; Molm, 1997) to develop a refined model explaining author behavior in this context. The empirical data were obtained via a survey and interviews with faculty from 17 universities with institutional repositories. Based on the results of the study, Kim (2010) identified the following significant factors influencing self-archiving behavior among authors, listed in descending order of effect size: (a) altruism, (b) perceived self-archiving culture, (c) copyright concerns, (d) technical skills, (e) age, (f) perception of no harmful impact of self-archiving on tenure and promotion, and (g) concerns about additional time and effort required. Age, copyright concerns, and additional time and effort were found to be negatively associated with self-archiving, whereas remaining factors were positively related.

Another model that could be useful in this context is the Unified Theory of Acceptance and Use of Technology (UTAUT), formulated by Venkatesh, Morris, Davis, and Davis (2003). The UTAUT model is a synthesis of a number of earlier competing models explaining technology adoption in voluntary settings and has been used extensively in information systems research. Hedlund (2008) provided an initial connection between UTAUT and OA and also included factors related to green OA. Empirical results of implementing the survey tool have, however, not been published. Dulle and Minishi-Majanja (2011) explored the suitability of the UTAUT model for studying OA adoption among university faculty; however, the study did not separate between gold OA and green OA in the survey questions and handled OA as a singular construct.

Aims and Methods

Aims

This study explores a number of questions regarding the current usage of the green OA alternative, for which satisfactory answers cannot be found in earlier published studies. The research questions are listed below. The answers gained should help academics and academic policymakers better understand the current situation of green OA.

What is the:

- Global share of journal articles available as green OA;
- Variation in green OA uptake between scientific disciplines;

- Location of green OA copies (subject repositories, institutional repositories, author home pages and departmental pages);
- Split among green OA copies over different versions (submitted manuscript, accepted manuscript, published article);
- Time lag from article publication to upload of green OA copy;
- Persistence of green OA copies;
- Availability of appropriate infrastructure and motivation for self-archiving;
- Share of articles in subscription journals for which upload would be copyright-compliant (and time lag distribution for upload embargoes);
- Degree of copyright compliance for uploaded green OA copies.

Methods

The research questions were first reviewed through the literature, and some of them were explored by compiling data and results from earlier studies to establish new insight. However, for questions about which existing studies did not provide sufficient support for drawing general conclusions, new empirical data were collected and analyzed. Detailed method descriptions are given in the subsections where new empirical data are presented. For most questions, only rough estimates (order of magnitude) can be produced because exhaustive large-scale sampling is beyond the resource limitations of this study.

Results

Global Share of Journal Articles Available as Green OA

The following studies were used to compile a longitudinal overview of OA share estimates so far:

- A study of OA journals included in WoK in 2003 and the number of articles published in them (McVeigh, 2004).
- Three studies using automated web searches for open free text versions by a research group led by Harnad (reported in, among others, Hajjem et al. [2005] Gargouri et al. [2012b]). These studies have been designed to test the citation advantage of OA but also provide estimates of the overall OA uptake at three points in time.
- Studies searching for full-text versions using manual search techniques conducted by this research group (Björk, Roos, &

Lauri, 2009; Björk et al., 2010). The samples are smaller than in the robotized studies above but the classification is more precise.

- Full article counts of gold OA journals (Laakso & Björk, 2012), delayed OA journals (Laakso & Björk, 2013), and hybrid journals (Björk, 2012b).

The central results from these studies are presented in Table 2. The figure of 10% for the publication year 2003 was extracted from a diagram in Hajjem et al. (2005), reporting on the OA availability and citation advantage in 2004 for articles published in 1992 through 2003. The last possible years (2006 and 2010) were picked from the tables presented in Gargouri et al. (2012b) to be consistent in measuring OA availability around 1 year after publication.

The overall OA availability has been split into five sub-categories of OA. Green OA is defined according to the definition used in this study. However, the green “self-archived” OA results reported by Gargouri et al. (2012b) were interpreted to also include articles in delayed OA journals, hybrid journals, and articles that publishers have made open for promotional purposes (“Other free articles”), due to the method used of classifying as green OA all copies found on the web except for articles in journals registered in the Directory of Open Access Journals (DOAJ; <http://www.doaj.org/>).

Regardless of methodology, all studies so far point toward a steady growth in overall OA during recent years. On the overall level, the results by Gargouri et al. (2012b) are robust due to the large sample sizes, but one must bear in mind that the study is based on a stratified sample of equal numbers of articles from 14 disciplines without adjusting for the difference in volumes of article production for each discipline, which was done for instance by Björk et al. (2010). The growth in gold OA (immediate full OA journals) is very clear, in particular if the longitudinal growth figures of the most recent study (Laakso & Björk, 2012) are triangulated with the figures of McVeigh (2004) and our own studies using different methods. The low proportion of articles in full OA journals that Gargouri et al. (2012b) found in 2011 is surprising given that the study reportedly had used DOAJ to identify such journals. The figure found for articles published in 2010 (1.2%) was even lower than

TABLE 2. An overview of previous studies reporting OA shares.

Study	Original Publication Year for Studied Articles	Year Green OA Measurement was Conducted	Index/ Journals Covered	Articles In Full Immediate OA Journals	Articles in Delayed OA Journals	Hybrid OA Articles	Other Free Articles	Green OA Articles	All OA
McVeigh (2004)	2003	2004	WoK	2.9					
Hajjem et al. (2005)	2003	2004	WoK			10.0			
Björk, Roos & Lauri (2009)	2006	2007	Ulrich's	4.6	3.5		11.3		19.4
Gargouri et al. (2012b)	2006	2009	WoK						21.0
Björk et al. (2010)	2008	2009	Scopus	5.3	1.2		2.0	11.9	20.4
Gargouri et al. (2012b)	2010	2011	WoK	1.2			21.9		23.1
Laakso and Björk (2012)	2011	2012	Scopus	11.0	5.2	0.7			

Note. Figures reported as % of all articles included in the studied index. OA = open access; WoK = Web of Knowledge.

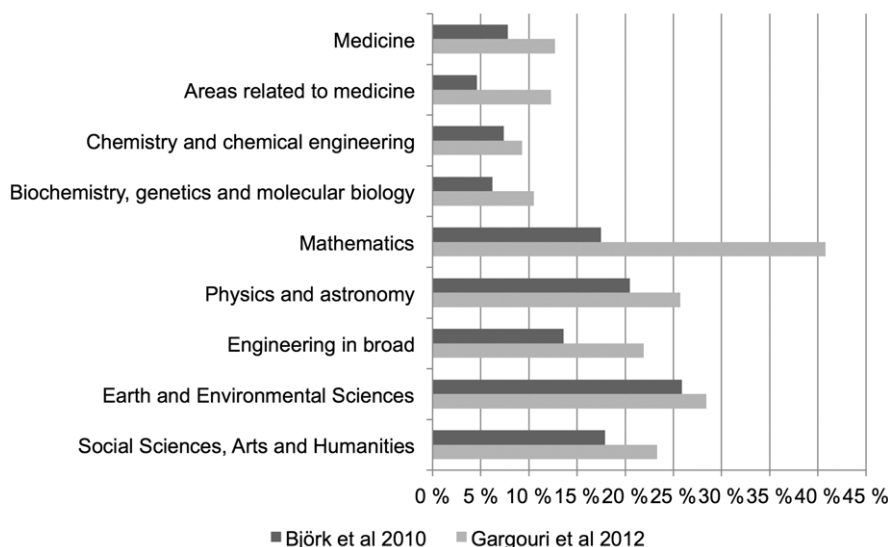


FIG. 1. Green OA uptake between scientific disciplines, comparison of results from two recent studies.

the average for 2005 through 2010 (2.4%). These results can be contrasted with the share of 7.8% DOAJ journal articles of all WoK articles published in 2010 found by Laakso and Björk (2012).

As for the level and development of green OA over time, using our definition of green OA being “all copies found elsewhere than on the publishers’ website,” the picture is not so clear. In combination with the other figures in Table 2 we, nevertheless, suggest that the share of green OA of all the recently published peer reviewed literature is around 12%. It is important to note that this figure means green OA copies of articles not already available in gold, delayed, or hybrid OA journals.

Variation in Green OA Uptake Between Scientific Disciplines

The two most relevant studies providing estimations for the uptake differences between disciplines are Björk et al. (2010) and Gargouri et al. (2012b). The studies use slightly different discipline categorizations. Björk et al. (2010) categorized all articles across a group of nine main disciplines that were based on aggregating 26 more detailed disciplines from Scopus. Gargouri et al. (2012b) used a categorization into 14 disciplines originating from the WoK, not quite covering all areas and articles. To make the studies comparable, the 14 disciplines from Gargouri et al. (2012b) were used to construct a similar breakdown to that used in Björk et al. (2010). For most of the subjects, there was more or less a one-to-one correspondence. The average green OA percentage of the Gargouri et al. (2012b) subjects Earth & Space and Biology were used as a proxy for the subject category for Earth and Environmental Sciences in Björk et al. (2010). The average of the Gargouri et al. (2012b) categories for Psychology, Social Science, Arts, Humanities,

and Professional Fields was used for corresponding to the Social Science, Arts, and Humanities category in Björk et al. (2010). The results are shown in Figure 1.

In interpreting the results it is important to note three differentiating factors. First, Björk et al. (2010) was based on article volumes in the broader Scopus index, whereas Gargouri et al. (2012b) was based on the more exclusive WoK index. Second, Gargouri et al. (2012b) was conducted 2 years later, allowing more time for green OA copies to be provided. Third, Gargouri et al. (2012b) implicitly included a number of delayed and hybrid OA journal articles due to the definition of green OA used. The effect of this is especially strong in increasing the figures in medicine, areas related to medicine and biochemistry, genetics, and molecular biology. Taking these factors into consideration, the differences in popularity of green OA across disciplines follows a similar pattern, with the exception of mathematics, where the figure in Gargouri et al. (2012b) is more than double the one in Björk et al. (2010). At least a partial explanation to this is that the observations were 2 years apart, during which the number of mathematics manuscripts uploaded to arXiv increased rapidly.

Location of Green OA Copies and Split Among Green OA Copies Over Different Versions

The distribution of green OA copies across different types of outlets has been studied through various methods, one type of which is author surveys and interviews. Gadd, Oppenheim, and Probets (2003) found that among the 58% of the 542 respondents to an international author survey who had made papers available, 72% had done so on their own web pages, 37% in a subject repository, 15% in an institutional repository, and 11% on other web sites (the answers were not mutually exclusive). Kim (2010) surveyed and

TABLE 3. Green OA location and version distribution across five earlier studies.

Study Background Information				Green OA Location Distribution			Versions of Green OA Copies Found		
Study	Scope	Data Source	Article Sample	Home pages etc.	Institutional Repositories	Subject Repositories	Submitted Manuscript	Accepted Manuscript	Published Article
Björk (2012a)	Civil eng. journals	13 journals	787	74%	23%	3%	23%	32%	45%
Björk and Paetau (2012)	IS journals	44 journals	798	59%	33%	8%	16%	46%	38%
Hedlund (2010)	5 Nordic countries	Scopus	1260	49%	19%	32%	23%	29%	47%
Björk et al. (2010)	Global	Scopus	1837	33%	24%	43%	15%	46%	38%
Björk, Roos & Lauri (2009)	Global	Ulrichs	300	27%	44%	29%	3%	35%	62%

Note. OA = open access; IS = information science.

interviewed faculty in 16 U.S. universities that had institutional repositories in October 2006. Of the 70% of respondents who had made research material publicly available via the Internet, 66% had used personal web pages, 51% research group web sites, 41% departmental web sites, 28% subject repositories, and 22% institutional repositories.

Another way of studying the location distribution of green OA copies has been via measurements of article volumes and relative shares based on documents made available on the Internet. Data for the breakdown of green OA copies were available from five different data sets obtained from previous studies. The method used in all studies was first to identify a set of articles published in a given year either globally (Björk et al., 2009, 2010), in specific countries (Hedlund, 2010), or in the journals of a particular discipline (Björk, 2012a; Björk & Paetau, 2012), then manually identifying OA copies and finally classifying them through manual inspection. The breakdown according to location is shown in Table 3. The global Scopus-based study (Björk et al., 2010) should be the most reliable global estimate so far.

Comparing the results of the five studies suggest that the relative distribution of green OA copy locations varies substantially based on scientific discipline and author affiliation. The two global studies are fairly similar in their distribution with 27% to 33% of green OA copies found on author homepages or other websites, 24% to 44% in institutional repositories, and 29% to 43% in subject repositories. Comparing the results of Björk et al. (2010), for which data were collected in 2009, with results of earlier survey-based results from Gadd et al. (2003) and Kim (2010), for which the data were collected in 2002 and 2006, respectively, shows a trend toward increased use of subject and institutional repositories in comparison to home and departmental web pages.

In their study of the attitudes of academics from New Zealand, Cullen, and Chawner (2011) found that of the respondents who had deposited research publications in the institutional repository of their university, 16% had deposited items prior to peer review, 41% after peer review, and 73% after formal publication. The phrasing of the questions

makes it difficult to know exactly which versions had been uploaded, in particular after formal publication (submitted manuscript or final publication).

In a recent survey limited to physics authors (Nicholas, Rowlands, Watkinson, Brown, & Jamali, 2012), the internal split of versions of journal articles deposited to subject or institutional repositories were as follows: the submitted version (39%), the accepted manuscript (31%), and the published article (30%). The high share of the submitted manuscripts might be explained by the e-print culture in some areas of physics. Somewhat surprising, given the existence of arXiv, the same survey found that institutional repositories were nevertheless slightly more popular (44%) than subject repositories (39%).

The relative distribution of different green OA versions for the five reviewed studies is provided in Table 3.

Unlike the green OA location distribution, the version distribution seems to be more homogeneous across the studies. However, the outlier is the global study of Ulrich's Periodicals index (UlrichsWeb, 2012), which is based on extrapolating results from a small sample of only 300 articles, so it has a wider margin or error in comparison to the other studies based on larger samples. Nevertheless, the main tendency seems to be uploading of accepted manuscripts and published articles in almost equal proportions, whereas submitted manuscripts constitute a smaller proportion of green OA copies.

Time Lag From Article Publication to Green OA

At least two factors limit the upload speed of green OA copies to repositories, in particular institutional ones. The first is the possible embargo period from the publisher. The second factor is the behavior and priorities of authors. Few authors upload copies individually as the articles are accepted or published. Instead, they may upload small batches of manuscripts in connection with the mandatory reporting of metadata of articles published the year before to the current research information systems of their universities. Sometimes articles are also uploaded systematically as larger batch efforts as in the case of the first author of this article.

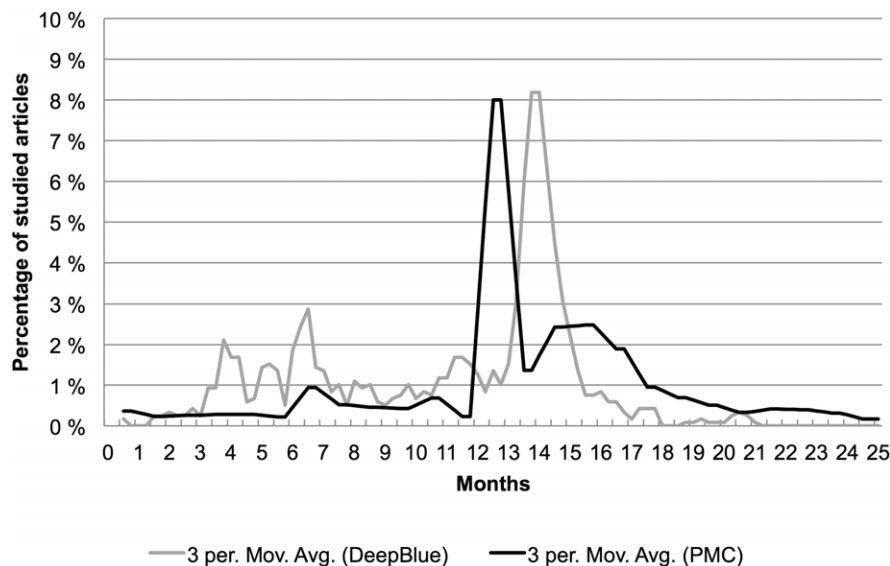


FIG. 2. Distribution curve for delays in upload to PMC and the institutional repository of the University of Michigan (DeepBlue). The delays are calculated from the formal publishing date to the date a full-text version of the article was uploaded to the repository.

Most empirical studies of the prevalence of green OA have been snapshots of the availability status at a given date considerably later than the publication date. It is also often difficult to extract the exact date when a green OA copy has been uploaded to a website or even a repository, and hence studies have generally ignored this aspect.

One way to study the time lag is to use data from selected repositories, provided that the date of deposition of the manuscripts is available in the metadata. The actual publication dates can then be extracted from indexing services or from the journal sites. It would be unrealistic to do this on a comprehensive scale, rather it would suffice to do this for a couple of the bigger universities as a case study, preferably ones without OA mandates. The method was tested in this study by examining the delay distribution for green OA copies in the institutional repository for the University of Michigan, called Deep Blue (<http://deepblue.lib.umich.edu>).

For the particular case of PMC, it was also possible to search for articles that had been published online in a particular month and made available in PMC in another. This enabled the computation of a distribution delay for accepted versions uploaded by the authors. The delay in being made available at PMC for the cohort of articles published online in September 2010 is shown in Figure 2, together with the similar delay curve for DeepBlue.

When comparing the actual delays with the publisher policies, it is evident that PMC delays adhere to the maximum embargo period of 12 months allowed by the NIH policy, but also that many uploads occur even later. The strong peak at exactly the 12-month mark is likely due to the fact that many of the publishers take responsibility for the PMC upload process as a service to authors, ensuring that the articles are released according to pub-

lisher policy. In the case of Deep Blue, there are many more uploads within a few months when compared with PMC, which is a reflection of what publishers allow for IRs compared with PMC, but generally speaking there is probably also a major delay due to the upload patterns of authors.

Persistence of Green OA Copies

The long-term preservation of articles in digital format is somewhat more challenging than for paper publications because it may involve changing formats and storage hardware. Unless given special attention, green OA article copies are at risk of disappearing over time. To study the preservation aspect of green OA, the data set from a previous study (Björk et al., 2010) was used so that the hyperlinks leading to the green article copies were inspected for accuracy 3 years after the initial data collection. As expected, some articles were inaccessible using the original hyperlinks. There were several explanations for hyperlink breakage, some of which were purely technical. As the Internet is in a state of flux, uniform resource identifiers (URIs) are prone to change due to various technical reasons, such as software migrations on servers, changes in the domain name system (DNS) namespace, and renaming of items in server file systems.

The persistence of green OA copies was lowest on arbitrary websites, such as personal or departmental websites, where the items could be found untouched in only 56% of cases. The low percentage can be partly explained by some of the technical issues described previously, and partly by, for example, copyright issues forcing removal or authors changing employment, thus ending the lifespan of the particular website.

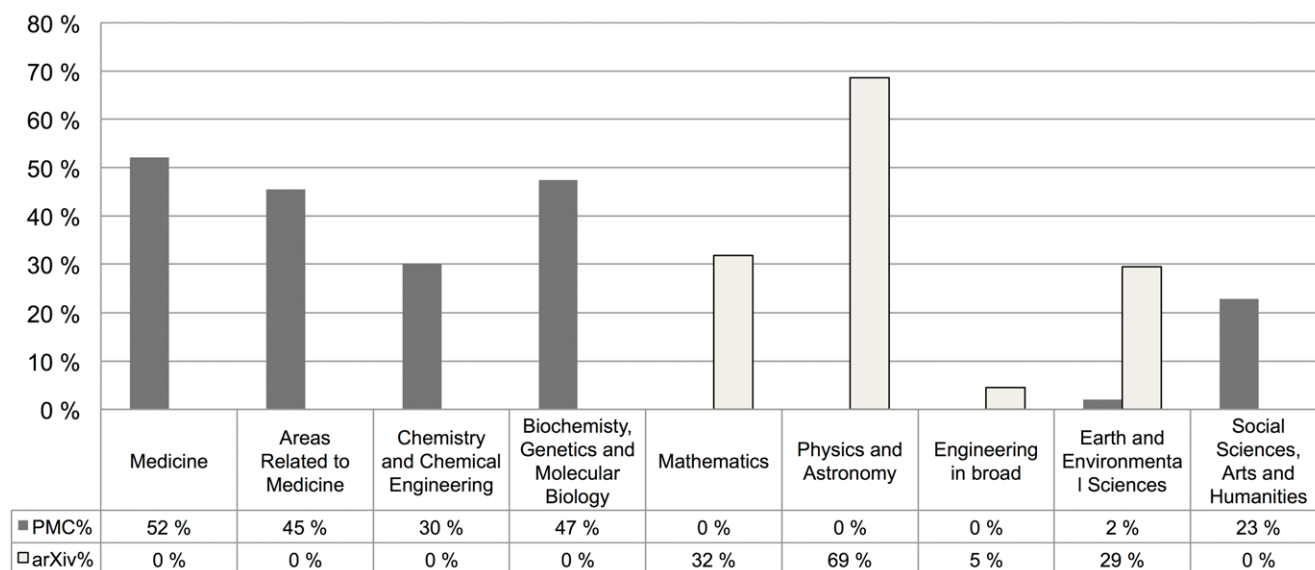


FIG. 3. The share of PMC and arXiv as location for all green OA manuscripts in different disciplines from Björk et al. (2010). There were 239 green OA manuscripts in total.

IRs and subject repositories performed better due to being built with clear goals and a systematic approach. Repositories are usually better guarded against technical failures and hyperlink breakage. In most cases, repositories use persistent handle systems to circumvent the risk of failure inherent to the URI system. Almost all copies (95%) in subject repositories were still accessible 3 years after the initial encounter, the more esteemed ones, such as PMC and arXiv, outperforming others. Most of the items in institutional repositories (80%) were still found to be intact and accessible.

The test was continued by trying to find out if another green copy of the disappeared articles could be located elsewhere on the web. The follow-up test revealed that in several cases where the item could no longer be found at its original location, another copy was stored elsewhere on a new site. This was particularly true with arbitrary websites. In some cases where repository copies had become unavailable, the article could still be found in the archive, however with a new URI. Reasons for this may be technical, such as renaming of the file, or changing the type of uploaded version, for example exact copy to preprint. A replacement green OA copy was found for approximately one half of the cases where the original green copy was inaccessible.

The disappearance of a share of green OA copies decreases the value of green OA as a systematic solution to the access dilemma. To fully harness the potential of author self-archiving, or green OA, the articles should be archived in proper repositories to maximize visibility and ensure that the items will also stay accessible in the future.

Availability of Appropriate Infrastructure and Motivation for Self-Archiving

Because of the multitude of factors contributing to the frequency of articles becoming self-archived by their

authors, these can be roughly divided into technical and motivational factors. Motivational factors are strongly influenced by obligation, either explicit in the form of OA mandates, or less directly through institutional culture or policies. Technical factors include the technical support and resources allocated to support the task of self-archiving. One of the key technical factors influencing author self-archiving behavior is the availability of suitable repositories for document upload.

Subject repositories. In some scientific disciplines there are well-established subject repositories that provide the natural first choice for authors wishing to upload a green OA copy. In particular, in biomedicine (PMC), and in physics and mathematics (arXiv), such repositories have become the norm. Using data from our study of OA prevalence (Björk et al., 2010), it was possible to estimate the share of the green OA copies we had that were in either PMC or arXiv by using the stored hyperlinks. Together PMC and arXiv contributed 38% of all green copies found and 94% of all copies in subject repositories. PMC dominated in the life sciences and arXiv in Physics and Mathematics. The results for nine disciplines are shown in Figure 3.

Institutional repositories. The number of institutional repositories has grown rapidly, but the interesting question is what proportion of authors has support for self-archiving provided by their universities. To study this, SCImago institutional ranking was used as a basis (Scimago, 2012). The ranking includes 3,290 institutions, which together provide more than 80% of the global scientific journal publication output as indexed by Scopus during the period 2006 to 2010. Of these institutions, the 148 top performers measured in counts of articles were chosen for the study, with a total output of 4,240,150 scientific documents during the period,

which is approximately 42% of all items indexed by Scopus in the same timeframe.

Additional information on IRs can be acquired from a number of sources, of which The Registry of Open Access Repositories (ROAR, 2012) and the OpenDOAR (2012) can be used to determine whether a particular institution of the 148 has, or lacks, an IR. In ROAR, 1,884 institutional or departmental repositories could be found at the time of the study, whereas OpenDOAR listed 1,816 repositories with similar criteria. In cases where neither ROAR nor OpenDOAR could provide any information on the institution, information was searched for through web search engines, and by visiting the institutional websites. Of the 148 most productive institutions, 82% had at least one IR for the purpose of collecting, preserving, and disseminating the intellectual output of the institution. The count of scientific documents for those institutions amounts to 3,620,234 (85%), which also gives some indication of the mass of authors given the possibility to self-archive in their affiliated institutions' repositories.

As a main result of the test, it was concluded that IRs are becoming one of the expected services that university libraries are supposed to offer, and most institutions have a repository in use. Although the majority of the largest research institutions possess the technical means of supporting OA in the form of author self-archiving, this is no guarantee that the authors comply and archive their work. For author self-archiving to become a widely adopted practice among researchers, several motivating factors have to be taken into account. The most effective on the institutional level are OA mandates or strong OA policies. These are usually initiated either by research funders, such as the case of the NIH mandate, or by the institutions themselves.

OA mandates. The coverage of OA mandates, as a share of the total global output of documents, could be studied for the same 148 top institutions mentioned previously by providing additional information on which institutions have mandated the use of their repositories for archiving scholarly literature. This information can be retrieved from the Registry of Open Access Repositories Mandatory Archiving Policies (ROARMAP) database (ROAR, 2012). The existence of an OA mandate was studied for each of the 148 top institutions measured by scientific output. The results show that approximately 15% of the 148 institutions either have a full institutional or a subinstitutional mandate on the archiving of their research output, and these institutions with OA mandates cover roughly 20% of the total output of the 148 top institutions.

Share of Articles in Subscription Journals For Which Upload Would Be Copyright Compliant

If researchers strictly follow the publication agreements they have signed, there is an upper limit to the extent of green OA. A minority of journals and publishers strictly prohibits some or all forms of self-archiving. Publisher

TABLE 4. Results of publisher policy analysis for top 100 publishers by article output in 2010.

	Article count	% of studied articles
Immediately upon publication	709,773	62
6 months	47,023	4
12 months	151,932	13
18 months	20,935	2
24 months	3,253	0
Potential green OA	932,916	81
Not allowed	217,911	19
Total articles studied	1,150,827	100

Note. Data refer to accepted manuscripts or published versions in either institutional or subject repositories. OA = open access.

copyright policies have been quantitatively analyzed in several studies in order to establish to what degree journal articles could potentially be provided as copyright-compliant green OA, if all authors would exercise their granted rights. Most earlier studies have been limited to the journal level, looking only at the number of journals allowing or prohibiting uploading of article manuscripts (e.g., Miguel et al., 2011), or have not discussed embargoes set by publishers for delaying self-archiving in the analysis (e.g., Morris, 2009).

In an ongoing study, the 100 largest publishers in terms of the number of articles published annually was identified using data covering more than 18,000 journals and 5,900 publishers obtained from the Scopus index (Laakso, 2013). The copyright policies of each publisher were reviewed, looking specifically at whether either accepted manuscripts or published versions were allowed to be put up by the authors. The Sherpa/Romeo database was used for supporting information but information was primarily accessed directly from the publisher websites when available.

Together, the top 100 publishers were responsible for 68% of all articles indexed in Scopus during 2010. Table 4 shows the distribution of what the analyzed publishers allow for accepted manuscripts or final versions in institutional or subject repositories, calculated over the number of articles output.

The results can, with some hesitation, be generalized to all articles and suggest that for four out of five articles green uploading is allowed, and for two out of three cases it is allowed immediately.

Degree of Copyright Compliance for Uploaded Green OA Copies

In much of the literature on green OA there is an implicit assumption that authors largely abide by the restrictions and embargoes stipulated by the publishers, as discussed. The reality looks quite different. Given that there are very few publishers that allow uploading publisher-formatted PDFs, the 35% to 50% share that such copies constituted of all green OA copies, as summarized in Table 4, is surprising. A detailed inspection of such copies showed two common

types as the origin, PDFs downloaded via the authors' institutional subscription, and author proofs received by the authors just prior to publishing. In strictly curated subject repositories like PMC and many IRs the rules can be assumed to be enforced and followed. But in reality authors seem to care little for such restrictions, in particular for copies uploaded to personal home pages or departmental pages.

Covey (2009) comments on this state of affairs as follows: Publisher policy appears to influence neither the decision to self-archive nor the article version that is self-archived. Also, in a study of author self-archiving behavior in the social sciences, Antelman (2006) found no relationship between publisher policy and self-archiving behavior.

Discussion

The presented dissection of the anatomy of green OA shows a complex structure shaped by publisher restrictions, university and funder policies, an evolving repository infrastructure, and the individual behavior of academic authors. Despite the benefits to the authors in terms of increased dissemination and citations, factors such as peer pressure and culture, academic reward systems, availability of suitable repositories, and lack of awareness and time all inhibit uptake and structure of green OA.

The overall uptake of green OA, using our definition, "all freely accessible copies of articles, including different versions of said articles, which exist on other web locations than the original publishers website," we estimate to be around 12%, based on a synthesis of a number of previous studies.

There are considerable differences in the uptake of green OA between disciplines, influenced by factors such as the uneven existence of preprint cultures, subject repositories, high-quality OA journals, and mechanisms for funding article processing charges. Our results suggest that IRs have increasingly become available as a viable option for green OA upload. The effect of mandates on uploading behavior seems undisputable, but so far such mandates cover only a small proportion of articles.

A closer, article-level analysis of publisher restrictions concerning green uploads shows that the top 100 publishers measured by output volume are surprisingly liberal and would in 62% of cases allow upload of accepted version manuscripts in institutional or subject repositories immediately upon publication, and a whole 79% within a year of publication. In analyzing actual uploaded green OA copies, it turns out that many authors can be assumed to knowingly break copyright rules by posting the actual published articles, in particular on home pages or departmental pages. This is cause for some concern for the sustainability of access to such copies as both the nonsystematic storage and breach of copyright make long-term availability uncertain.

The real barrier to green OA is author behavior. Many authors are unaware of what they can do and lack of time

and other priorities also prevent more frequent uploading. Even authors who use the opportunity do so only periodically, and there are strong indications that roughly half of green OA copies in repositories are uploaded a year or more after publishing. This means that the impact of green OA should be compared with gold and delayed OA combined, not only gold OA.

Only time will tell how much green OA will contribute to the overall open availability of the scientific journal literature. Much depends on the actions of research funders, who increasingly require that the results of funded research be OA, mostly in a "color-neutral" way. Above all, the overall OA share, as well as the relative contributions to this of gold and green, will be determined by the choices of the individual authors.

Acknowledgments

We deliberately chose to submit this article to a subscription journal allowing the immediate uploading of a green accepted manuscript version to the institutional repository of our university.

References

- Antelman, K. (2006). Self-archiving practice and the influence of publisher policies in the social sciences. *Learned Publishing*, 19(2), 85–95.
- Armbruster, C. (2010). Implementing open access: policy case studies. Retrieved from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1685855
- arXiv. (2012). Monthly submission rate statistics. Retrieved from http://arxiv.org/show_monthly_submissions
- Björk, B.-C. (2012a). A comparison of subscription and open access journals in construction management and related fields. *Australasian Journal of Construction Economics and Building*, 12(2), 42.
- Björk, B.-C. (2012b). The hybrid model for open access publication of scholarly articles: A failed experiment? *Journal of the American Society for Information Science and Technology*, 63, 1496–1504. doi: 10.1002/asi.22709
- Björk, B.-C., & Paetau, P. (2012). Open access to the scientific journal literature: Status and challenges for the information systems community. *Bulletin of the American Society for Information Science and Technology*, 38(5), 39–44.
- 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., & Gudnason, G. (2010). Open Access to the scientific journal literature: Situation 2009. *PLoS ONE*, 5(6), e11273. doi: 10.1371/journal.pone.0011273.t003
- 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. doi: 10.1353/pla.0.0042
- 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, 145–161. doi: 10.1080/13614533.2010.518851
- CrossRef. (2012). CrossRef glossary. Retrieved from <http://crossref.org/02publishers/glossary.html>
- Cullen, R., & Chawner, B. (2011). Institutional repositories, open access, and scholarly communication: a study of conflicting paradigms. *The Journal of Academic Librarianship*, 37(6), 460–470. doi: 10.1016/j.acalib.2011.07.002

- Davis, P.M., & Connolly, M.J.L. (2007). Institutional repositories: Evaluating the reasons for non-use of Cornell University's installation of DSpace. *D-Lib Magazine*, 13(3/4).
- Dulle, F.W., & Minishi-Majanja, M.K. (2011). The suitability of the Unified Theory of Acceptance and Use of Technology (UTAUT) model in open access adoption studies. *Information Development*, 27(1), 32–45. doi: 10.1177/0266666910385375
- 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—the Finch Group. Finch Report. Retrieved from <http://www.researchinfonet.org/publish/finch/>
- Gadd, E., Oppenheim, C., & Proberts, S. (2003). Romeo Studies 3: How academics expect to use open access research papers. *Journal of Librarianship and Information Science*, 35(3), 171–187. doi: 10.1177/0961000603353005
- Gargouri, Y., Hajjem, C., Larivière, V., Gingras, Y., Carr, L., Brody, T., & Harnad, S. (2010). Self-selected or mandated, open access increases citation impact for higher quality research. *PLoS ONE*, 5(10), e13636. doi: 10.1371/journal.pone.0013636.t004
- Gargouri, Y., Larivière, V., Gingras, Y., Brody, T., Carr, L., & Harnad, S. (2012a). Testing the Finch hypothesis on green OA mandate effectiveness. Retrieved from <http://arxiv.org/vc/arxiv/papers/1210/1210.8174v1.pdf>
- Gargouri, Y., Larivière, V., Gingras, Y., Carr, L., & Harnad, S. (2012b, September). Green and gold open access percentages and growth, by discipline. Paper presented at 17th International Conference on Science and Technology Indicators (STI), Montreal, Canada. Retrieved from <http://eprints.soton.ac.uk/340294>
- Hajjem, C., Harnad, S., & Gingras, Y. (2005). Ten-year cross-disciplinary comparison of the growth of open access and how it increases research citation impact. *IEEE Data Engineering Bulletin*, 28(4), 39–47.
- Harnad, S., Brody, T., Vallieres, F., Carr, L., Hitchcock, S., Gingras, Y., Oppenheim, C., et al. (2004). The access/impact problem and the green and gold roads to open access. *Serials Review*, 30(4), 310–314.
- Hedlund, T. (2008). Business school researchers' attitudes towards open access and institutional repositories: A study on user acceptance and user behavior. In L. Chan & S. Mornati (Eds.), *Open scholarship: Authority, community, and sustainability in the age of Web 2.0*. Proceedings of the 12th International Conference on Electronic Publishing (ELPUB2008) (pp. 15–22). *ELectronic PUBlishing (Elpub) Digital Library*. Retrieved from http://elpub.scix.net/cgi-bin/works/Show?015_elpub2008
- Hedlund, T. (2010). Open access availability of articles by Nordic authors. *ScieCom Info*, 6(4). Retrieved from <http://www.sciecom.org/ojs/index.php/sciecominfo/article/view/4760/4321>
- Houghton, J., Rasmussen, B., Sheehan, P., Oppenheim, C., Morris, A., Creaser, C., Greenwood, H., et al. (2009). Economic implications of alternative scholarly publishing models: Exploring the costs and benefits. A report to the Joint Information Systems Committee (JISC), 256.
- Houghton, J., & Swan, A. (2013). Planting the green seeds for a golden harvest: Comments and clarification on “going for gold.” *D-Lib Magazine*, 19(1/2). Retrieved from <http://www.dlib.org/dlib/january13/houghton/01houghton.html>
- JISC infoNet. (2012). Digital repositories. Retrieved from <http://www.jiscinfonet.ac.uk/infokits/digital-repositories/>
- Kankanhalli, A., Tan, B.C.Y., & Wei, K.K. (2005). Contributing knowledge to electronic knowledge repositories: An empirical investigation. *MIS Quarterly*, 29(1), 113–143.
- Kim, J. (2010). Faculty self-archiving: Motivations and barriers. *Journal of the American Society for Information Science and Technology*, 61(9), 1909–1922. doi: 10.1002/asi.21336
- Kling, R., & McKim, G. (2000). Not just a matter of time: Field differences and the shaping of electronic media in supporting scientific communication. *Journal of the American Society for Information Science*, 51(14), 1306–1320.
- Kling, R., McKim, G., & King, A. (2003). A bit more to it: Scholarly communication forums as socio-technical interaction networks. *Journal of the American Society for Information Science and Technology*, 54(1), 47–67.
- Koskinen, K., Lappalainen, A., Liimatainen, T., Niskala, A., Salminen, P. J., & Nevalainen, E. (2010). The current state of open access to research articles from the University of Helsinki. *ScieCom Info*, 6(4). Retrieved from <http://www.sciecom.org/ojs/index.php/sciecominfo/article/view/4761>
- Laakso, M. (2013). Journal publisher self-archiving policies and the potential for growth in open access. Working paper to be submitted for journal publication.
- 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 and Technology*. In press.
- Lynch, C.A. (2003). Institutional repositories: Essential infrastructure for scholarship in the digital age. *ARL*, 226, 1–7. Retrieved from <http://www.arl.org/resources/pubs/br/br226/br226ir.shtml>
- Lyons, C., & Booth, H.A. (2011). An overview of open access in the fields of business and management. *Journal of Business and Finance Librarianship*, 16(2), 108–124. doi: 10.1080/08963568.2011.554786
- Matsubayashi, M., Kurata, K., Sakai, Y., Morioka, T., Kato, S., Mine, S., & Ueda, S. (2009). Status of open access in the biomedical field in 2005. *Journal of the Medical Library Association*, 97(1), 4–11. doi: 10.3163/1536-5050.97.1.002
- McVeigh, M. (2004). Open access journals in the ISI citation databases: Analysis of impact factors and citation patterns: A citation study from Thomson Scientific. Retrieved from <http://ip-science.thomsonreuters.com/m/pdfs/openaccesscitations2.pdf>
- 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
- Moed, H.F. (2006). The effect of “Open Access” upon citation impact: An analysis of ArXiv's Condensed Matter Section. *Journal of the American Society for Information Science and Technology*, 58(13), 2047–2054.
- Molm, L.D. (1997). *Coercive power in social exchange*. New York: Cambridge University Press.
- Morris, S. (2009). Journal authors' rights: Perception and reality. Publishing Research Consortium Summary Paper 5. Retrieved from <http://www.publishingresearch.net/documents/JournalAuthorsRights.pdf>
- Morrison, H. (2012). Thank you, open access movement! September 30, 2012 Dramatic Growth of Open Access. *The Imaginary Journal of Poetic Economics Blog*, 2012 October 6. Retrieved from <http://poeticeconomics.blogspot.fi/2012/10/thank-you-open-access-movement.html>
- Nicholas, D., Rowlands, I., Watkinson, A., Brown, D., & Jamali, H.R. (2012). Digital repositories ten years on: what do scientific researchers think of them and how do they use them? *Learned Publishing*, 25(3), 195–206. doi: 10.1087/20120306
- NISO. (2008). NISO-RP-8-2008. Journal Article Versions (JAV): Recommendations of the NISO/ALPSP JAV Technical Working Group April 2008. A Recommended practice of the National Information Standards Organization in partnership with the Association of Learned and Professional Society Publishers. Retrieved from <http://www.niso.org/publications/rp/RP-8-2008.pdf>
- OpCit Project. (2012). The Open Citation (OPCIT) Project. Retrieved from <http://opcit.eprints.org>
- OpenDOAR. (2012). The Directory of Open Access Repositories. Retrieved from <http://www.opendoar.org>
- PEER. (2011). PEER Economics Report. Centro ASK, Università Bocconi. Retrieved from http://www.peerproject.eu/fileadmin/media/reports/PEER_Economics_Report.pdf
- Poynder, R. (2012). Open access mandates: Ensuring compliance, Open and Shut Blog. 2012 May 18. Retrieved from <http://poynder.blogspot.fi/2012/05/open-access-mandates-ensuring.html>

- ROAR. (2012). ROARMAP: Registry of Open Access Repositories Mandatory Archiving Policies. Retrieved from <http://roarmap.eprints.org>
- RCUK. (2013). RCUK Policy on Open Access and Supporting Guidance. Retrieved from <http://www.rcuk.ac.uk/documents/documents/RCUKOpenAccessPolicyandRevisedguidance.pdf>
- Scimago. (2012). SIR World Report 2012 Global Ranking. Retrieved from http://www.scimagoir.com/pdf/sir_2012_world_report.pdf
- Scopus. (2012). Scopus. Retrieved from <http://www.scopus.com>
- SHERPA/RoMEO. (2012). SHERPA/RoMEO Publisher copyright policies and self-archiving. Retrieved from <http://www.sherpa.ac.uk/romeo>
- Suber, P. (2012). Open Access. MIT Press.
- Swan, A. (2010). The Open Access citation advantage: Studies and results to date. Key Perspectives Report. Retrieved from <http://eprints.ecs.soton.ac.uk/18516>
- Swan, A., & Brown, S. (2005). Open access self-archiving: An author study. Key Perspectives Report. Cornwall, UK: Key Perspectives Limited. Retrieved from <http://cogprints.org/4385>
- UlrichsWeb. (2012). UlrichsWeb Global Serials Directory. Retrieved from <http://ulrichsweb.serialssolutions.com>
- Venkatesh, V., Morris, M., Davis, G., & Davis, F. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly*, 27(3), 425–478.
- Wagner, A.B. (2010). Open access citation advantage: An annotated bibliography. *Issues in Science and Technology Librarianship*. doi: 10.5062/F4Q81B0W
- WoK. (2012). Thomson Reuters Web of Knowledge. Retrieved from <http://apps.webofknowledge.com>.
- Willinsky, J. (2005). The access principle. Cambridge, MA: MIT Press.