

## 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: Björk, B.-C., Laakso, M., Welling, P. and Paetau, P. (2014). Anatomy of green open access. *Journal of the Association for Information Science and Technology*, 65: 237–250.

<http://dx.doi.org/10.1002/asi.22963>

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.

# Anatomy of Green Open Access

## **Bo-Christer Björk**

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

E-mail: [bo-christer.bjork@hanken.fi](mailto:bo-christer.bjork@hanken.fi)

## **Mikael Laakso**

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

E-mail: [mikael.laakso@hanken.fi](mailto:mikael.laakso@hanken.fi)

## **Patrik Welling**

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

E-mail: [patrik.welling@hanken.fi](mailto:patrik.welling@hanken.fi)

## **Patrik Paetau**

Hanken School of Economics, Vaasa Campus, P.O. Box 287, 65101 Vaasa, Finland

E-mail: [patrik.paetau@hanken.fi](mailto:patrik.paetau@hanken.fi)

**Open Access (OA) is the 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 upon current knowledge about green OA. A synthesis of previous studies indicates that the green OA coverage of all published journal articles is approximately 12 %, with substantial disciplinary variation. Typically, green OA copies become available with 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, which is 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. Open Access 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 Peter Suber describes it in the following way; “*Open Access (OA) literature is digital, online, free of charge, and free of most copyright and licensing restrictions.*”(Suber, 2012). The most important factor is free access for readers, literature which 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 since 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 compared to the subscription model (Houghton et al 2009).

### Gold and green OA

There are several variations to the open accessibility of scholarly articles, depending on the user rights, the timing of the availability, the funding of the 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) 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. This is beneficial to readers, as a manuscript in many use cases is close enough to the published article, at least sufficient for pre-purchase evaluation if not for direct citation. 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 the results available and thus citable, to non-subscribing 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 (since 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 enforce movement towards OA (Houghton and 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, as well as copies stored in subject repositories by the publishers, e.g. PubMedCentral (PMC). JISC uses the term “mediated deposit” to describe the latter categories (JISCinfonet 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 while 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 might provide essentially the same key content to interested readers. Even after acceptance, many publishers still make minor changes to manuscript contents during copy editing, and finally the manuscript receives the publisher’s layout 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.

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

Table 1 - Different versions of green OA copies

Both the SHERPA/RoMEO index of publishers copyright restrictions concerning green OA (SHERPA/RoMEO 2012) as well as leading OA advocate Peter Suber’s OA guidelines (Suber 2012) use the term *preprint* to refer to the two first stages and *postprint* to the two latter ones (the term *postprint* is rather peculiar, since publishing is less and less dependent on the paper medium and since the accepted version never even used to be 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 an article are

also 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. Though 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.

## **Locations of green OA copies**

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, while 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 is by now over 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 over 1'400 archives containing working paper series of individual universities, departments etc. PMC on the other hand is a highly centralized database of medical publications, maintained by the world's largest funder of medical research, 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 CVs or publication lists of the authors in question. At the time, this was the only possibility in most disciplines, but with time repositories have started to offer viable alternatives. The major subject repositories and most institutional repositories

can be expected to be relatively permanent storage 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, like Google Scholar, to both 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**

Over the past fifteen 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 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 observing 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 alternative 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 OpenDOAR registry has grown from slightly above 800 in 2006 to over 2200 in 2012. One major weakness of this method is that information regarding repository size is usually limited to the total number of items, including a lot of 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 et al (2005) or Björk et al (2010) have used article samples from article indexes such as Web of Knowledge (WoK) (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 and Booth 2010).

The choice of meta-data 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 (Swan 2010, Moed 2006).

## **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. Currently, the most well known 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) while 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 they have chosen to publish in. 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 2012), 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, 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 15-20 % for other institutions with voluntary upload.

## **Costs of setting up and maintaining repositories**

While 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 or EPrints) are mostly used as the IT-infrastructure of institutional repositories at major universities. In addition to the adaption, installation, and maintenance of the software and servers, library personnel might also be needed to verify the copyrights of uploaded copies, to correct references, 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 UK to repositories at around 33 USD per article, assuming that it takes 10 minutes of the author's time. One of the most systematic attempts to measure such costs was made in the EC funded Publishing and the Ecology of European Research project (PEER) (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 wide range of 2 USD to 53 USD depending on the repository. These results are very ambiguous; in particular since repositories can contain a wide variety of materials, and since the cost of setting up the repository and of staff managing it has to be spread over all these document types.

## **Repository case studies**

There are numerous reports of this particular type, but perhaps the most informative one is the report 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 (2009) study exploring the end-user reasons for accessing the Cornell University repository is interesting since 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 four collections out of 107 exhibiting a 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 three years. Of the authors who had not self-archived 71 % were unaware of this possibility. 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), over half of the respondents had deposited a peer-reviewed journal article to a repository during the past five years. The scholars showcased a clear preference to upload to a subject repository compared to an institutional one, but only 37 % knew of a suitable subject repository. 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 et al (2003) and Social Exchange theory (Molm 2003, Kankanhalli et al 2005) to develop a refined model explaining author behavior in this context. The empirical data was 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, 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 which could be useful in this context is the Unified Theory of Acceptance and Use of Technology (UTAUT), formulated by Venkatesh et al (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, also including factors related to green OA. Empirical results of implementing the survey tool have, however, not been published. Dulle & 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**

The aim of this study is to explore a number of concrete questions about 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 could be explored by compiling data and results from earlier studies to establish new insight. However, for questions where existing studies did not provide sufficient support for drawing general conclusions, new empirical data was collected and analyzed. Detailed method descriptions are given in the sub-sections where new empirical data is presented. For most questions only rough estimates (order of magnitude) can be produced since 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 open access 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 Prof. Stevan Harnad (reported in among others, Hajjem 2005 and Gargouri 2012). 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 et al 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 et al 2012), delayed OA journals (Laakso et al 2013) and hybrid journals (Björk 2012).

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 1992-2003. The last possible years (2006 and 2010) were picked from the tables presented in Gargouri (2012), to be consistent in measuring OA availability around one year after publication.

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, and Lauri 2009	2006	2007	Ulrich's	4.6	3.5	11.3			19.4	
Gargouri et al 2012	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 2012	2010	2011	WoK	1.2	21.9					23.1
Laakso and Björk 2012	2011	2012	Scopus	11.0	5.2	0.7				

Table 2 - An overview of previous studies reporting OA shares, figures reported as % of all articles included in the studied index

The overall OA availability has been split into five subcategories of OA. Green OA is defined according to the definition used in this study. However, the green “self-archived” OA results reported by Gargouri (2012) were interpreted to include also articles in delayed OA journals, hybrid journals and articles which 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 DOAJ.

Regardless of methodology, all studies so far point towards a steady growth in overall OA during recent years. On the overall level the results by Gargouri et al 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 and Björk 2012) are triangulated with the figures of McVeigh and our own studies using different methods. The low proportion of articles in full OA journals that Gargouri et al 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 the average for 2005-2010 (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 (2012). The studies use slightly different discipline categorizations. Björk et al (2010) categorized all articles across a group of 9 main disciplines which were based on aggregating 26 more detailed disciplines from Scopus. Gargouri et al (2012) used a categorization into 14 disciplines originating from the WoK not quite covering all areas and articles. In order to make the studies comparable the 14 disciplines from Gargouri et al (2012) were used to construct a similar breakdown as 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 (2012) 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 (2012) 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.

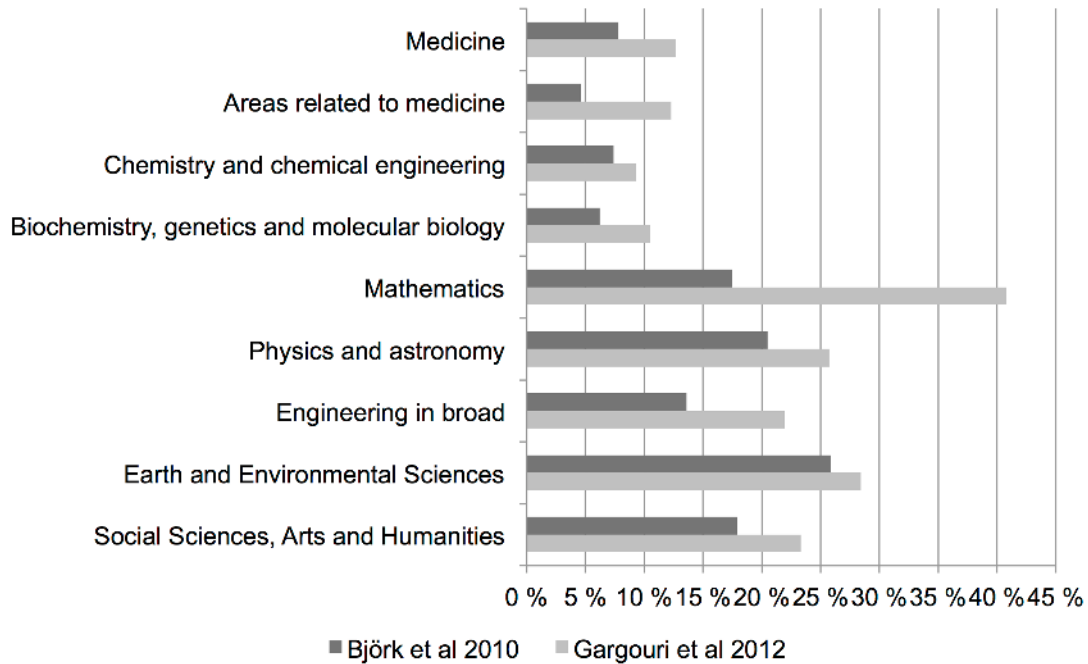


Figure 1 - Green OA uptake between scientific disciplines, comparison of results from two recent studies

In interpreting the results it is important to note three differentiating factors. Firstly, Björk et al 2010 is based on studying article volumes in the broader Scopus index, while Gargouri et al 2012 is based on the more exclusive WoK index. Secondly, Gargouri et al (2012) was conducted two years later allowing more time for green OA copies to be provided. Thirdly, Gargouri et al (2012) implicitly included a number of delayed and hybrid OA journal articles due to the used definition of green OA. 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 (2012) is more than double to the one in Björk et al (2010). At least a partial explanation to this is that the observations were two 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 which is author surveys and interviews. Gadd et al (2003) found that among the 58 % of the 542 respondents to an international author survey 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 interviewed faculty in sixteen US universities that had institutional repositories in October 2006. Of the 70% of respondents that 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 was 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 and Lauri 2009, Björk et al 2010), in specific countries (Hedlund 2010), or in the journals of a particular discipline (Björk and Paetau 2012, Björk 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.

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 (2012)	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%

Table 3 – Green OA location and version distribution across five earlier studies



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 was collected in 2009, with results of earlier survey-based results from Gadd et al (2003) and Kim (2010), where the data was collected in 2002 and 2006 respectively, shows a trend towards 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 publications. The phrasings of the questions make 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 et al 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 are also provided in Table 3.

Unlike the green OA location distribution, the version distribution seems to be more homogeneous across the studies. The outlier is the global study of the Ulrich's Periodicals index, however, that 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, while 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 directly as the articles are accepted or published. Instead many 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.

Most empirical studies of the prevalence of green OA have been snapshots of the availability status at a given date considerable later than the publication dates. 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 actual journal sites. It would be unrealistic to do this on a comprehensive scale, rather it would suffice to do this for a couple of 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.

For the particular case of PMC it was also possible to search for articles which had been published online in a particular month and made available in PMC in another. This enabled the computation of a delay distributions 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.

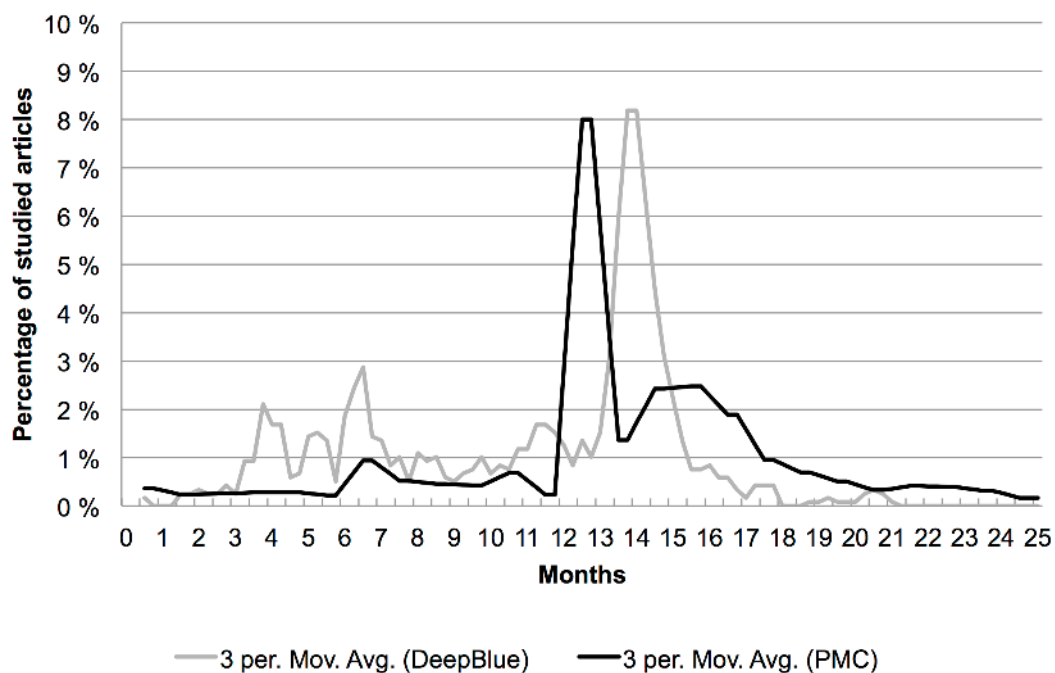


Figure 2 - Distribution curve for delays in upload to PMC and the institutional repository of the University of Michigan (DeepBlue). The delays are calculated from the formal publishing date to the date a full-text version of the article was uploaded to the repository.

In comparing the actual delays with the publisher policies it is evident that PMC delays pretty much adhere to the maximum embargo period of 12 months allowed by the NIH policy, but also that many uploads occur even later than this. The strong peak at exactly the 12-month mark is likely due to the fact that many of the publishers take on responsibility for the PMC upload process as a service to author, making sure that the article is released according to publisher policy. In the Deep Blue case there are many more uploads within a few months compared to PMC, a reflection of what

publishers allow for IRs compared to PMC, but also generally speaking a major delay probably 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, since it might involve changing formats and storage hardware. Unless given special attention, green OA article copies are in risk of disappearing over time. To study the preservation aspect of green OA, the dataset from a previous study (Björk et al 2010) was used, so that the hyperlinks leading to the green article copies were inspected for accuracy three 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, URIs are prone to change due to various technical reasons, such as software migrations on servers, changes in the 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 web sites, 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 above, and partly by for example copyright issues forcing removal or authors changing employment ending the lifespan of the particular website.

Institutional repositories and subject repositories performed better due to them building on clear goals and a systematic approach. Repositories are usually better guarded against technical failures and hyperlink breakage. In most cases, repositories utilize persistent handle systems to circumvent the risk of failure inherent to the URI system. Almost all copies (95%) in subject repositories were still accessible three 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 to exist 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 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

As there are a 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 open access mandates, or less eminent through institutional culture or policies. Technical factors include the technical support and resources allocated to support the task of self-archival. One of these key technical factors influencing author self-archiving behavior is the availability of suitable repositories for document upload.

### Subject repositories

In some scientific discipline there are well-established subject repositories which 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 which 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

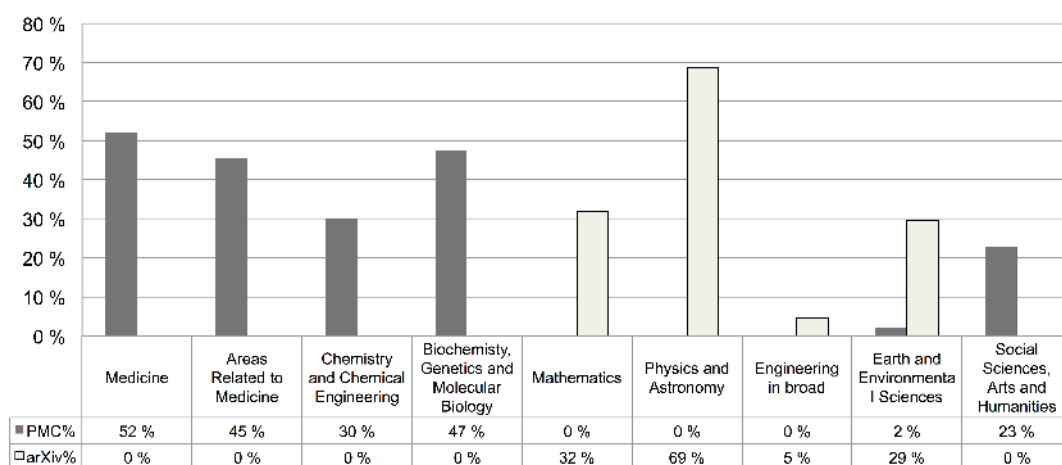


Figure 3 - The share of PMC and arXiv as location for all green OA manuscripts in different disciplines from Björk et al (2010). 239 green OA manuscripts in total.

### 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 3290 institutions, which together provide over 80% of the global scientific journal publication output as indexed by Scopus in the period of 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 institutional repositories can be acquired from a number of sources, of which *The Registry of Open Access Repositories (ROAR 2012)*, and the *Directory of Open Access Repositories (OPENDOAR 2012)* can be used to determine whether a particular institution of the 148 has, or lacks, an institutional repository. In ROAR, 1884 institutional or departmental repositories could be found at the time of the study, while OPENDOAR listed 1816 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 institutional repository 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 hints on the mass of authors given the possibility to self-archive in their affiliated institution's repositories.

As a main result of the test, it was concluded that institutional repositories are becoming one of the expected services that university libraries are supposed to offer, and most institutions in fact have a repository in use. Even though the majority of the largest research institutions possess the technical means of supporting open access in the form of author self archiving, this is however no guarantee that the authors in fact 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 open access mandates or strong open access policies. These are usually initiated either by research funders, such as the case of the NIH mandate, or the institutions themselves.

### **OA mandates**

The coverage of open access mandates, as a share of the total global output of documents, could be studied for the same 148 top institutions as above 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. The existence of an open access 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 sub-institutional mandate on the archival of their research output, and these institutions with open access 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 prohibit some or all forms of self-archiving. Publisher 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 a on-going study the 100 largest publishers in terms of the number of articles published annually was identified using data covering over 18,000 journals and 5900 publishers obtained from the Scopus index (Laakso 2013). The copyright policies of each publisher were reviewed, looking specifically if either accepted manuscripts or published versions were allowed to put up by the authors. The Sherpa/Romeo database was used for supporting information but primarily information was accessed directly from the publisher websites where 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.

-	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%
<b>Potential green OA</b>	<b>932 916</b>	<b>81%</b>
Not allowed	217 911	19%
<b>Total articles studied</b>	<b>1 150 827</b>	<b>100%</b>

Table 4 – Results of publisher policy analysis for top 100 publishers by article output in 2010. Data refers to accepted manuscripts or published versions in either institutional or subject repositories.

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 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 in the previous section. The reality looks quite different. Given that there are very few publishers that allow uploading publisher formatted PDFs, the 35-50 % share that such copies constituted of all green OA copies as summarized in Table 4 previously, is surprising. A detailed inspection of such copies showed two dominating 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 themselves 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 are all conflicting factors leading to the current 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 institutional repositories 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 non-systematic storage and breach of copyright makes the 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 prevents more frequent uploading. Even those authors that 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 to gold and delayed OA combined, not just to 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 are Open Access, 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.

## Acknowledgements

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. Available at [http://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=1685855](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1685855)
- arXiv (2012) - Monthly submission rate statistics  
<http://arxiv.org/Stats/hcmonthly.html>
- Björk, B-C. (2012). 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. (2012). The hybrid model for open access publication of scholarly articles: A failed experiment? *J. Am. Soc. Inf. Sci.*, 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., & Guðnason, G. (2010). Open Access to the Scientific Journal Literature: Situation 2009. (E. Scalas, Ed.) *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.
- 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 <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.
- 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 140p. Available at: <http://www.researchinfonet.org/publish/finch/>



- Gadd, E., Oppenheim, C., & Proberts, S. (2003). Romeo Studies 3: How Academics Expect to Use Openaccess 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. (R. P. Futrelle, Ed.) *PLoS ONE*, 5(10), e13636. doi:10.1371/journal.pone.0013636.t004
- Gargouri, Y., Larivière, V., Gingras, Y., Carr, L., & Harnad, S. (2012). Green and Gold Open Access Percentages and Growth, by Discipline. In, 17th International Conference on Science and Technology Indicators (STI), Montreal, CA, 05 - 08 Sep 2012. 11pp. Available at: <http://eprints.soton.ac.uk/340294/>
- Gargouri, Y., Larivière, V., Gingras, Y., Brody, T., Carr, L., & Harnad, S. (2012a) Testing the Finch Hypothesis on Green OA Mandate Effectiveness. <http://arxiv.org/vc/arxiv/papers/1210/1210.8174v1.pdf>
- 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). Researcher's Attitudes Towards Open Access and Institutional Repositories: A Methodological Study for Developing a Survey Form Directed to Researchers in Business Schools. Proceedings of the ELPUB 2008 Conference on Electronic Publishing, Toronto, Canada. Available at: [http://elpub.scix.net/cgi-bin/works/Show?015\\_elpub2008](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).
- 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)
- JISCinfonet 2012 - <http://www.jiscinfonet.ac.uk/infokits/repositories/management-framework/self-archiving>
- Kankanhalli, A., Tan, B. C. Y., & Wei, K. K. (2005). Contributing knowledge to electronic knowledge repositories: An empirical investigation. *MIS Quarterly*, 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). <http://www.sciecom.org/ojs/index.php/sciecominfo/article/view/4761>
- 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. *J. Am. Soc. Inf. Sci.*, Accepted for publication
- Laakso, M. (2013). Journal publisher self-archiving policies and the potential for growth in open access. Working paper to be submitted for journal publication
- Lynch, C. A. (2003) Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age. *ARL 226* (February 2003): 1-7. <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 & 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: JMLA*, 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. Available at: <http://ip-science.thomsonreuters.com/m/pdfs/openaccesscitations2.pdf>
- 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). Social exchange and power. In *coercive power in social exchange*. New York: Cambridge University Press. 330p.
- Morris, S. (2009) Journal authors' rights: perception and reality. Publishing Research Consortium Summary Paper 5. Available at: <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), 6.10.2012, <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 Available at: [www.niso.org/publications/rp/RP-8-2008.pdf](http://www.niso.org/publications/rp/RP-8-2008.pdf)
- OpCit Project (2012) Open Citation (OPCIT) Project. <http://opcit.eprints.org/>
- OpenDOAR (2012) - The Directory of Open Access Repositories. <http://www.opendoar.org/>
- PEER Economics Report (2011) Centro ASK, Università Bocconi Available at: [http://www.peerproject.eu/fileadmin/media/reports/PEER\\_Economics\\_Report.pdf](http://www.peerproject.eu/fileadmin/media/reports/PEER_Economics_Report.pdf)
- Poynder, R. (2012) , Open Access Mandates: Ensuring Compliance“Open and Shut” blog, 18.5.2012, <http://poynder.blogspot.fi/2012/05/open-access-mandates-ensuring.html>
- RCUK (2012) Research Councils UK Policy on Access to Research Outputs. Available at: [http://www.rcuk.ac.uk/documents/documents/RCUK%20\\_Policy\\_on\\_Access\\_to\\_Research\\_Outputs.pdf](http://www.rcuk.ac.uk/documents/documents/RCUK%20_Policy_on_Access_to_Research_Outputs.pdf)
- ROAR (2012) ROARMAP: Registry of Open Access Repositories Mandatory Archiving Policies. <http://roarmap.eprints.org/>

- Scimago (2012). SIR World Report 2012 Global Ranking. Available at:  
[http://www.scimagoir.com/pdf/sir\\_2012\\_world\\_report.pdf](http://www.scimagoir.com/pdf/sir_2012_world_report.pdf)
- Scopus (2012) Scopus. <http://www.scopus.com>
- Sherpa/RoMEO (2012) <http://www.sherpa.ac.uk/romeo/>
- Suber, P. (2012). Open Access. MIT Press. 230p.
- Swan, A. (2010). The Open Access citation advantage: Studies and results to date. Key Perspectives Report. 17p. Available at: <http://eprints.ecs.soton.ac.uk/18516>
- Swan, A., & Brown, S. (2005). Open access self-archiving: An author study. Key Perspectives Report. 97p. <http://cogprints.org/4385/>
- UlrichsWeb (2012) UlrichsWeb Serials Solutions.  
<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, 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. <http://apps.webofknowledge.com>
- Willinsky, J. (2005). The Access Principle. MIT Press