# Proportion of Open Access Peer-Reviewed Papers at the European and World Levels—2004-2011

August 2013

Science-Metrix

Eric Archambault, Didier Amyot, Philippe Deschamps,
Aurore Nicol, Lise Rebout & Guillaume Roberge

# Proportion of Open Access Peer-Reviewed Papers at the European and World Levels—2004-2011

August 2013

Produced for the
**European Commission** DG Research & Innovation
RTD-B6-PP-2011-2: Study to develop a set of indicators to measure open access



by **Science-Metrix Inc.**
**Brussels | Montreal | Washington**
1335 Mont-Royal E., Montréal
Québec, Canada, H2J 1Y6
1.514.495.6505
info@science-metrix.com
www.science-metrix.com

Cover image: iStockphoto

# Executive Summary

This report re-assesses the Open Access (OA) availability of scholarly publications during the 2004 to 2011 period, for 22 fields of knowledge, as well as for the European Research Area countries, Brazil, Canada, Japan, and the US. Using a strategy to increase the number of free articles retrieved (that is, which aims to increasing recall), led to close to a doubling of the proportion of OA estimated by teams lead by Björk[1] and by Harnad[2]. The present report shows that the tipping point for OA (more than 50% of the papers available for free) has been reached in several countries, including Brazil, Switzerland, the Netherlands, the US, as well as in biomedical research, biology, and mathematics and statistics.

### Pilot study

This study comprised an important pilot phase which involved the retrieval of a set of 20,000 randomly selected records corresponding to papers published in 2008 from the Elsevier Scopus database. This set of 20,000 records was then provided to the Harnad team in Montreal for a blind analysis using a protocol they had developed for previous studies. This test suggested that the proportion of total OA was as high as 32%, compared to the 22% Harnad's team had obtained in Thomson Reuters' Web of Science (WoS). Extensive tests performed on a subset of 500 records extracted randomly from the 20,000 records set suggested that 48% of the literature published in 2008 was available for free in December 2012.

We inferred that OA availability had likely passed the tipping point in December 2012 for articles published in 2008 and that the majority of peer-reviewed/scholarly papers published in journals in that year were available for free in one form or another to end-users.

These carefully determined results diverge widely from the measures previously published. Harnad's team measurement of an OA availability rate of only 22% compared to the 48% rate obtained here may be explained by Scopus' broader coverage of the scientific literature compared to WoS and by Google Scholar's imperfect recall. These results also diverge from the measure obtained by Björk's team, who used the Scopus database and suggested that only 20% of the articles published in 2008 were available for free. This discrepancy may be explained by the time required for embargoed articles to appear online and by differences in the methodological approach applied by Björk et al, who sought to measure the share of OA copies available to the average researcher based on Google searches, and excluding papers available in spite of publishers' policies to the contrary. By contrast, our team aimed to measure the share of OA copies available anywhere on the web, regardless of the status of the papers.

The final stage of the pilot study involved drawing a new random sample of 20,000 records from Scopus, and adjusting the sample to include at least 100 records from the smaller fields in terms of number of articles (Philosophy & Theology, Visual & Performing Arts, General Arts, Humanities & Social Sciences—GAHSS, Built Environment & Design). Our research project requires precisely determining the proportion of OA papers by estimating the number of OA

[1] 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). doi: 10.1371/journal.pone.0011273.

[2] Gargouri, Y., Larivière, V. Gingras, Y. and Harnad, S. (2012). Green and Gold Open Access Percentages and Growth, by Field. In Archambault, É, Gingras, Y. and Larivière. V. (2012). Proceedings of 17th International Conference on Science and Technology Indicators, Montréal: Science-Metrix and OST

peer-reviewed papers (the numerator) and dividing this by the number of peer-reviewed articles (the denominator) for 22 fields, and for the total literature. Since there is currently no extensive database of scientific publications, the Ulrich periodical database at the journal level, in conjunction with Scopus at the article level, was used here to provide an estimate of the denominator. Although imperfect, Ulrich remains the most extensive, authoritative and probably the least biased source of data on academic peer-reviewed journals. A sensitivity analysis revealed that the distribution of records in Scopus is only slightly outside the boundaries of three models used to estimate the denominator for 18 out of 22 fields.

#### Large scale study

The subsequent phase used a relatively large-scale measurement of OA availability based on a sample of 320,000 randomly drawn papers for the Scopus database—that is, 40,000 records per publication year between 2004 and 2011.

The same sample of 500 articles used in the pilot study was used for the characterisation of the OA harvester used to measure availability of the 320k sample. A slight variation was observed in the availability of articles in this sample measured in December 2012 (47.6%) and April 2013 (44.8%). It is noteworthy that 249 articles were available for free at one time or another between December 2012 and April 2013, just a hair under 50%. These results suggest that there are important transient aspects that need to be taken into consideration while measuring OA availability. These results also show that the harvesting engine has very good retrieval precision (98%) and fairly good recall (86.6%), resulting in fairly robust measures of OA availability.

At the whole database level, there is an exponential growth of gold OA papers indexed in Scopus. The growth rate is 24% per year (obtained through exponential regression curve fitting), which means that the number of gold papers doubles every 2.9 years.

The availability of gold OA in a random sample of 320,000 papers, not surprisingly, closely follows the population-level statistics. Green and hybrid OA availability grows in the distant past and recedes in the more recent past. This is due, at least in part, to editors having embargo periods on many of the papers in their journals which are sometimes available initially only through subscription and are subsequently being made available for free.

The measurement of overall OA is based on the addition of gold OA and of hybrid and green OA. According to this measure, 38% (the statistical margin of error is ± 0.5 percentage points) of the 2004 peer-reviewed journal articles indexed in Scopus papers are currently available for free. This proportion reached 44% (± 0.5) in 2011. The growth rate is very low, that is, only 1.9% per year. This low growth rate over time likely reflects the translation of the OA availability curve for back years. An adjusted OA availability curve can be computed by applying a conversion factor that accounts for the precision and recall of the instrument (this calibration is based on the analysis of the 500 records sample). This estimation suggests that the tipping point of OA availability was reached in 2011.

Free availability of a majority of articles has been reached in general science & technology, in biomedical research, biology, and mathematics & statistics. The fields where OA availability is most limited are the social sciences and humanities and in the more applied sciences, engineering, and technology. The lowest prevalence of OA availability is in visual and performing arts (13%) and communication & textual studies.

A growth index was computed by dividing the percentage of OA availability in 2008-2011 by that observed in 2004-2007. Overall, between the two periods, there has been an 8% increase in OA availability (slightly more than 3 percentage point). The fields with the fastest growth between the most recent four years and the preceding four years are chemistry, general science & technology, public health and health services, clinical medicine, agriculture, fisheries & forestry, and enabling & strategic technologies.

All the fields derive an OA citation advantage. Interestingly, many of the fields where the OA proportion is low have a sizeable citation advantage, such as in philosophy and theology (54% more cited), general arts humanities and social sciences, communication and textual studies, engineering, and visual & performing arts. What is particularly interesting here is that the citation advantage is derived almost exclusively from the green and hybrid portion, as gold OA is associated with a citation disadvantage on average for all fields except for physics & astronomy.

The statistics on gold journals require careful interpretation. First, many gold journals are younger and smaller, and these factors have an adverse effect on the citation rate and hence on measured ARC values. Authors frequently prefer reading and citing established journals, and it is therefore a challenge to start a journal from scratch, and to have authors submit high quality articles. It takes time to build a reputation and to attract established authors. Importantly though, gold journals might provide an avenue for less mainstream, more revolutionary science. If so, the signature would be a much greater level of variation between the more highly cited papers and the baseline (no citation).

An examination of OA availability was performed for EU28, EFTA (European Free Trade Association), Accession countries, ERA (European Research Area), and for four additional countries, namely, Brazil, Canada, Japan, and the US. For the period 2008-2011 considered as a whole, eight of the EU28 (30%) have reached the tipping point. If the statistical precision and recall of the harvesting instrument are taken into account, 20 out of 27 countries (74%) would have tipped. Calibrating for precision and recall, the proportion of ERA countries having more than 50% of papers in OA is 74%, that is, the same as for the EU28 overall.

In countries outside the ERA, it is noteworthy that the US has passed the tipping point and Canada is approaching. Even more salient is the proportion of 63% observed in Brazil. This is no doubt due to the important contribution of Scielo, which plays a key role in the Southern hemisphere in making scientific knowledge more widely available.

#### State of Open Access scholarly publications

Between 2004 and 2011, the average annual rate of increase of OA availability was relatively limited, with a compound growth rate of 2% per year. In addition to having year-on-year growth, there is an upward translation of the whole availability curve over time. This is due to an increasing number of authors making their manuscripts available for the current year but also for previous years.

There are also transient effects that have to be considered when measuring OA availability, including temporary promotional OA offered by publishers and variations in websites' availability. All in all, more than 50% of the papers could be found for free in November/December 2012 (pilot phase of this study) and in March/April 2013 (1st full measurement stage) but somewhat less so at either time period. This shows that measuring phenomena on the Internet requires particular attention to detail and constant questioning on the meaning of the results.

Green OA appears to be moving slowly, whereas gold and hybrid OA (such as pay-per-article for OA release) appear to be driving in the fast lane. This impression will require further investigation. Efforts should be made to characterise these changes, and to distinguish what percentage of growth comes from green self-archiving and what comes from other forms of hybrid OA.

The fact that the open access tipping point has likely been reached is an important finding for the whole publishing industry. This industry is likely to be undergoing revolutionary change, and at a pace much faster than anticipated, in large part because previous measures of OA availability proved to be misleading. This means that aggressive publishers are likely to gain much in the redesigned landscape, whereas those attached to the old ways are likely to suffer and to lose market share. An important question is whether the switch to a more atomistic, fine-grained market with millions of researchers as buyers will reduce, augment or leave unchanged the negotiating power of publishers.

# Contents



# Tables



# Figures

# 1 Introduction

Since the 1990s, interest in the academic community for Open Access (OA) publications has been increasing steadily, especially following the introduction of the arXiv e-print archives (arXiv.org). Several articles appeared to promote self-archiving in the interest of making scientific knowledge freely available to all. In parallel, an emerging movement aimed to measure and monitor OA availability and impact. Quite early on, proponents of OA used these measures to promote free availability and it is not always easy to distinguish what papers with OA as a subject are attempting to do, i.e. advocate or measure OA. The present paper is all about metrology, not advocacy.

The initial interest in the use of bibliometric methods, focused on accessing the so-called citation advantage of OA as opposed to subscription-based journals (Antelman, 2004; Harnad & Brody, 2004; Craig, 2007). The literature of the time recognised a clear citation advantage to papers available in OA as opposed to papers diffused solely through subscription-based journals. Strong advocacy by authors such as Harnad (2003, 2008, 2012) suggested that benefits would ensue from so-called green OA, that is, research papers self-archived by their authors in various types of repositories. Unsurprisingly, in this context, librarians and information scientists noted that they had a new mission, which meant setting up and curating OA repositories (Proser, 2003; Bailey, 2005; Chan, Kwok, & Yip, 2005; Chan, Devakos & Mircea, 2005 Repanovici, 2012).

A part of the OA literature has discussed how authors, researchers (Pelizzari, 2004; Swan & Brown, 2004; Dubini, Galimberti & Micheli, 2010) and publishers (Morris, 2003; Regazzi, 2004) would react to this new paradigm. Evidently, business and economic models were discussed (Bilder, 2003; Kurek, Geurts & Roosendaal, 2006; Houghton, 2010; Lakshmi Poorna, Mymoon & Hariharan, 2012), but there was also interest in what models academia and libraries would follow (Rowland et al., 2004; Swan et al., 2005; Hu, Zhang & Chen, 2010).

As OA continued to make inroads, a growing number of papers examined the state of development of OA in specific countries (Nyambi & Maynard, 2012; Sawant, 2012; Woutersen-Windhouwer, 2012; Miguel et al., 2013) and in specific fields of research (Abad-Garcí et al., 2010; Gentil-Beccot, Mele, & Brook, 2010; Charles, & Booth, 2011; Henderson, 2013). In this context, it was not surprising to find papers that addressed the general question of OA availability as a proportion of the scientific literature, and the proportion of OA papers available in different fields of science (Björk et al. 2010; Gargouri et al., 2012).

This paper re-assesses OA availability during the 2004-2011 period by carefully tuning harvesting methods in order to increase recall. The current version of the harvesting engine developed by Science-Metrix searches on specific sites including Scielo, PubMed Central and the websites of scientific peer-reviewed journals publishers, uses a locally hosted version of large-scale specialised repositories such as arXiv and CiteSeerX,[3] and systematically harvests metadata from institutional repositories listed in the Registry of Open Access Repositories (ROAR) and the Directory of Open Access Repositories (OpenDOAR).

The approach used here leads to a measurement of OA availability which is close to a doubling of the proportion of OA estimated by Björk et al. and by Gargouri et al. The present paper shows

[3] The authors would like to thank Lee Giles and Douglas Jordan at Penn State for giving access to CiteSeerX data.

that the tipping point for OA (more than 50% of the papers available for free) has been reached in several countries, including Brazil, Switzerland, Netherlands, the US, and in biomedical research, biology, and mathematics and statistics. Data are presented for 22 fields of knowledge, as well as for the European Research Area countries, Brazil, Canada, Japan, and the US. Before entering into the methodological details associated with the measurement of OA, it is important to produce operational definitions of OA, green OA, gold OA, and hybrid OA.

**Types of OA scientific literature:** Peter Suber suggests that '[o]pen-access (OA) literature is digital, online, free of charge, and free of most copyright and licensing restrictions.'[4] A colloquial definition of OA would be 'OA, whether Green or Gold, is about giving people free access to peer-reviewed research journal articles.'[5] The following operational definitions of gold, green and hybrid OA will be used in the present study.

- Gold OA refers to journals that use a funding model that does not charge readers or their institutions for access, and makes all contents available without embargo period.
- Green OA generally refers to authors' self-archiving [of papers accepted in academic journals following a successful peer-review process].
- Hybrid OA is an increasingly important trend in scientific publishing by which authors pay for their papers to be available in OA in an otherwise not OA journal—'[h]ybrid open access journals provide Gold OA only for those individual articles for which their authors (or their author's institution or funder) pay an OA publishing fee.' There are other cases such as the release of subscription-based journal articles after an embargo period, this type of OA articles could also be called delayed OA. There are cases where editors make articles available for free for limited period of time for promotional purpose but then retract them. This is in fact time-limited OA and presents specific measurement problems.

[4] http://www.earlham.edu/~peters/fos/overview.htm.

[5] http://scholarlykitchen.sspnet.org/2011/09/07/oa-rhetoric-economics-and-the-definition-of-research/.

# 2 Methods

This study comprised an important pilot phase (Section 2.1) followed by a phase of relatively large scale measurement (Section 2.2). Key metrology concepts used in this report are presented in Section 2.3.

## 2.1 National and regional policies, incentives and legislation

The pilot phase comprised four stages. A first stage involved the development of a manual retrieval process where we sought to retrieve 20,000 randomly selected papers form the Elsevier Scopus database. The retrieval was made in a manner reminiscent of that used by Björk et al. (2010). Although we later discovered that the sample had some randomness imperfection, it was sufficient for the initial experimental phase. Importantly, this approach was abandoned after three months as it appeared to be prohibitively expensive and extremely slow, and as we noticed that our approach contained a methodological flaw which limited the proportion of papers we retrieved from Google Scholar. Importantly also, Google Scholar routinely blacklisted our manually operated retrieval instrument, thus curtailing our measurement efforts and showing the limits of relying on that source of data for large scale measurement as typically performed in bibliometric studies.

A second stage started with 20,000 records extracted from Scopus for the year 2008 being provided to the Stevan Harnad team in Montreal. This test suggested that the proportion of total OA was as high 32%, compared to the 22% Harnad's team obtained in Thomson Reuters' Web of Science (WoS) (Gargouri et al., 2012). As Björk *et al*. found a score of 20% using Scopus and Google as a search engine, it appeared necessary to perform an in-depth inquiry on a smaller sample to determine whether these new scores were erroneous. After all, this suggested that OA availability was about 50% higher than previous measures suggested.

In the third stage of the pilot study, some 500 of the 20,000 records set were then extracted randomly and extensive tests were performed. The records were all searched manually in Google Scholar, Google, and Microsoft Academics. Records that could be downloaded for free and that came from any of these sources were considered OA, and the carefully verified sample called a 'ground truth.' Importantly, the 'ground truth' can be considered a floor value as none of the search engines used can be considered to have perfect recall, that is, the capacity to retrieve all relevant results.

These tests led to the following observations: Google Scholar and Google have substantial overlap, but each search engine has a somewhat distinct set of positive results (Table I). Microsoft Academics does not add much to the combined results of Google and Google Scholar. Importantly also, the results obtained suggest that the accuracy of the data collection method, and the coverage of the database, are more important than a large sample size (statistical precision).

Extensive testing was done with the subsample of 500 records. The results for the Harnad's team robot are as is and contain a few false positives, so the real positive score is actually lower. The Scholar, Google and Ground Truth results were meticulously validated by hand and the documents downloaded, and as such, they can be considered as being highly accurate. The Ground Truth comprises the combined validated results from Google and Google Scholar in addition to one result from Microsoft Academics. Results from Microsoft Academics are not shown, as only the negative results from Scholar and Google were tested to examine whether this added any substantial results to the previous ones. Please note that these results were obtained in

December 2012. Our tests showed that some of the documents freely available at that time ceased to be free later. This is certainly one difficulty in the measurement of OA, the Internet is very organic and changes constantly.

Table I Availability of OA in a sample of 500 Scopus records, 2008

| Result | UQAM (Gargouri-Harnad) | Scholar | Google | Ground Truth |
|---|---|---|---|---|
| FALSE | 350 | 293 | 290 | 262 |
| TRUE | 150 | 207 | 210 | 238 |
| **Total** | **500** | **500** | **500** | **500** |
| % OA | 30% | 41% | 42% | 48% |

Source: Computed by Science-Metrix

This analysis suggested that 48% of the literature published in 2008 was available for free in December 2012. Despite their high level of sophistication, neither Google nor Google Scholar can be expected to crawl the Web perfectly or to have a search engine so robust that it systematically presents all the relevant records in the first page of results (to which we limited our analysis), and hence cannot be expected to have a 100% recall, especially for academic articles (Arlitsch & O'Brien, 2012). Consequently, we inferred that OA availability had likely passed the tipping point in December 2012 for 2008 articles and that the majority of peer-reviewed/scholarly papers published in journals in that year were available for free in one form or another to end-users.

An important question is why these carefully determined results are diverging so much from the measures previously published, including those published by Harnad's team itself (Gargouri et al., 2012). Our initial tentative explanation was that this difference was likely due to the use of Scopus by our team, as opposed to the Web of Science as Harnad's team had done before. This explanation could account to an increase of 10 percentage points, that is, from 22% availability using WoS to 32% using Scopus with the same harvesting engine used by the Harnad team. Based on his answers and comments made in a conference,[6] Harnad appears to prefer measuring OA availability only in the most highly cited portion of the scientific literature. This can be done using WoS as compared to Scopus which has somewhat more extensive coverage. This is different from the objective of the present team which aims to estimate the proportion of OA availability for all peer- or editorially-reviewed scholarly journals. We feel this objective is important as emerging OA journals will frequently have meagre citation scores, and are thus excluded from the WoS which concentrates on highly cited journals, but that they should be nevertheless be taken into account to warrant a more comprehensive understanding of the evolution of scientific publication. However, it is also obvious that the harvester used by Harnad's team has quite an imperfect recall as it caught only 30% of OA articles using Google Scholar, whereas we retrieved 41% by hand with the same search engine. Combining the use of WoS instead of Scopus, and taking into consideration the imperfect recall of Google Scholar, and the imperfect recall of Harnad's team robot, goes a long way in explaining why Harnad's team measured an OA availability rate of only 22% compared to the 48% rate obtained here.

Another important divergence is with the measure obtained by Björk's team, who used the Scopus database. They suggested that only 20% of the articles published in 2008 were available for free. This is half the figure obtained in our own tests in Google as we were able to retrieve 42%

[6] Science and Technology Indicators (STI) Conference, Montreal 2012.

of the articles for free. At least three aspects may have played a role: 1) it takes a certain time for embargoed articles to make their way to the web, and authors may not always rapidly self-archive or otherwise post their papers on sites such as Research Gate, and as our measures were made three years later (Autumn 2012 versus Autumn 2009), it is very likely that more papers had by then appeared for free; 2) Björk et al aimed to answer "what share of OA copies would the average researcher find" whereas our team aims to answer "what share of OA copies are available somewhere on the web", and hence we did not act as average researchers but as researchers determined to find all articles that Google had indexed, however hard it may be to retrieve them using Google, and other search engines; 3) Björk et al. removed articles which could accidentally be available for free "clearly against the site policy". We made no such attempt as our research question seeks to find what is available for free at a given time.

As discussed later in the paper, there is an important transient aspect to OA availability. Unknown to our team at the time the sample was manually searched, Springer had made several articles available for free during Autumn 2012 but subsequently many of them were withdrawn from free circulation. However, and despite this, six months later the proportion of OA papers found in Google was very similar as new free articles had appeared by then.

The final, fourth stage of the pilot study involved drawing a new random sample of 20,000 records from Scopus. The aim here was twofold; first, to use this sample to calibrate a new OA harvesting engine purposefully developed for this project taking into account what was learned in the previous stages of the pilot study, and second, develop a technique to estimate the overall proportion of OA availability in Scopus using the Ulrich periodical database as a calibration aid.

This sample was restricted to papers published in 2008, and the results were restricted to original contributions to knowledge (e.g. document type similar to article, notes and reviews). Records where the journal name or the record type contained a conference were excluded as well as those for which the field was unknown. The eligible record set from 2008, comprising somewhat more than 1.36 million records in Scopus, was 'tossed' five times using a pseudo-random method (using the newid() command in SQL Server). Then, a subset of 100,000 records was selected, placed in a subset, and tossed again. These 100,000 records were then imported into Excel, where a straightforward analysis of the distribution of the records by field was performed. This analysis showed that a subsample of 20,000 records would keep few records in three of the smaller fields (Philosophy & Theology, Visual & Performing Arts, and General Arts, Humanities & Social Sciences—GAHSS). For these fields, a random sample of 100 records was selected, and for the Built Environment & Design field, the 101 records that were part of the 100,000 records were all selected. As the objective was to produce a record set of 20,000, a subsequent selection was done for 19,599 records. These were selected by tossing the 100,000 a few more times using the rand() command in Excel, then proceeding to the selection of the required number of records.

An original aspect of the present project is our willingness to assess the proportion of OA availability in all scholarly articles published in peer- and editorially-reviewed journals. There is currently no such extensive database of scientific publications. The Web of Science traditionally aimed to cover only the most highly cited and authoritative journals and while competition from Elsevier's Scopus has had the effect of pushing Thomson Reuters to augment the coverage of the WoS, it is far from being an extensive database of all peer-reviewed journals, as we would have liked for this project. Scopus has a more extensive coverage, and adds journals which typically are somewhat less frequently cited. However, it still leaves out several smaller journals, and it is highly likely that those excluded will frequently be those published in a non-English language. It is remarkable that despite the fact that R&D spending at the world level approaches US $ 2 trillion per year (authors' own estimate using UNESCO statistics on R&D), no one can tell how many scholarly peer-reviewed journals there are currently. This makes the task of measuring the proportion of OA availability in that set a complex exercise as the 'denominator', that is, the total number of peer-reviewed articles published per year, is currently unknown.

Our research project requires precisely determining the proportion of OA papers by estimating the number of OA peer-reviewed papers (the numerator) and dividing this by the number of peer-reviewed articles (the denominator) for 22 fields, and for the total literature. The Ulrich periodical database was used here to provide an estimate of the denominator, that is, the population of peer-reviewed scholarly journals. The strengths and weaknesses of Ulrich data are well known: for example, some journals that should be classified as peer-reviewed are not (and the reverse is also true). A good example of this is the OA journal *Activités*, which mentions that 'Texts that have been submitted to *Activités* (www.activites.org/) will be assessed by two referees (called upon in view of the article). Each will give his or her opinion on the text.'[7] Despite this, and a detailed description that clearly suggests scholarly content and the presence in papers of references to scholarly work, Ulrich has not classified this journal as refereed. Although several journals are likely to be classified 'Academic/Scholarly' in Ulrich and might be considered as contributing to science, this category cannot be included *en masse* as it comprises a substantial amount of material published in universities that has little scientific content. This is the case, for example, with the 'The Hilltop', classified by Ulrich as Academic/Scholarly, and claiming to be the 'The Student Voice of Howard University' (see http://www.thehilltoponline.com/). Consequently, the selection was restricted to Ulrich listed *journals* considered *refereed/peer-reviewed* AND Academic/Scholarly.

Although imperfect, Ulrich remains the most extensive, authoritative and probably the least biased source of data on academic peer-reviewed journals. The reason Ulrich was used is because it was felt that publishers of *article-level* database publishers such as Elsevier (publisher of Scopus) and Thomson Reuters (WoS) were faced with choices having important commercial and profitability impacts. When selecting journals to be included for an article-level database such as Scopus, deciding whether to include a journal has a direct impact on production costs and partly because of this, database publishers tend to have a bias towards larger journals (which help achieve economies of scale) and larger publishers (which help lower transaction costs and economies of scale and several similarly formatted journals are provided).

Whether a journal is small or large in terms of number of articles has substantially fewer consequences when it is included in a journal title database, where journal size can be expected to have comparatively little impact on cost (some differences remain as it is likely easier to find information about the larger journals even when building journal-level databases such as Ulrich). Importantly, one can only assume there are still important economies of scale involved with covering large editors: even for inclusion in Ulrich, it is less costly to input data from one publisher with one thousand journals, and covering one journal each for 2000 publishers. This means that smaller, more marginal journals and editors are more likely to be excluded from Ulrich, and these are perhaps more likely to be non-English journals and editors. Hence, Ulrich is imperfect but it still more likely to be more broadly balanced than Scopus and the WoS.

The technique used to estimate the proportion of each scientific field in the overall total involved the following three steps: 1) journals in Ulrich were matched to those contained in Scopus; 2) journals that intersected were given the field that was already contained in the classification of

[7] http://www.activites.org/resources/activites.eng.book.pdf

Scopus journals used by Science-Metrix. For those that did not intersect, the Ulrich classification was compared with that used in our classification, and a matching table was used to attribute one of 22 fields to each of the journals; and 3) the number of articles per field was counted in the intersecting set, while the number of articles in the Ulrich set with no Scopus counterparts was determined by projecting the average number of articles for the 50% journals in Scopus with the fewest articles per journal. The reason for using the average number of articles for the 50% smaller journals is that experience has revealed that databases such as Scopus and the WoS tend to prioritize the largest journals. For instance, the WoS covers about 12,000 journals, and Scopus about 18,000. Despite a 50% increase in journal coverage, Scopus only has about 20% more articles.

A sensitivity analysis was performed to see the effect of calculating the average for the 75%, 50%, and 25% smallest journals (ranked by decreasing number of articles), and the results were broadly similar. This analysis reveals that the distribution of records in Scopus is inside the boundaries of these models for 18 out of 22 fields, the four fields where Scopus' proportion is outside are green shaded. In relative terms, the largest discrepancies between the Scopus sample are in General Science & Technology, Communication & Textual Studies, Visual & Performing Arts, Philosophy & Theology, and the Social Sciences. Except for the General Science & Technology field, all these are underestimated in Scopus compared with the modelled projection (the 50% model). Importantly though, these are all relatively small fields in terms of peer-reviewed papers and, in the end, the correction of total OA availability based on this approach produced the same overall result as the substantially simpler random sample drawn from Scopus (in the pilot phase, both the random Scopus score and the Ulrich-calibrated score using the purposefully developed OA harvester yielded an availability score of 42%).

Table II Sensitivity analysis of three models compared with Scopus, 2008

| Discipline | 25% smallest | 50% smallest | 75% smallest | Scopus |
|---|---|---|---|---|
| Agriculture, Fisheries & Forestry | 3.75% | 3.84% | 3.90% | 3.74% |
| Biology | 4.81% | 4.96% | 5.15% | 4.29% |
| Biomedical Research | 8.54% | 8.31% | 8.09% | 8.00% |
| Built Environment & Design | 0.49% | 0.50% | 0.50% | 0.55% |
| Chemistry | 8.72% | 8.40% | 8.11% | 8.29% |
| Clinical Medicine | 26.12% | 25.89% | 25.78% | 27.62% |
| Communication & Textual Studies | 0.98% | 1.13% | 1.24% | 0.63% |
| Earth & Environmental Sciences | 3.03% | 3.00% | 3.00% | 3.02% |
| Economics & Business | 2.94% | 3.07% | 3.13% | 2.47% |
| Enabling & Strategic Technologies | 6.54% | 6.39% | 6.34% | 7.62% |
| Engineering | 5.80% | 5.78% | 5.84% | 6.75% |
| General Arts, Humanities & Social Sciences | 0.14% | 0.13% | 0.12% | 0.12% |
| General Science & Technology | 0.80% | 0.81% | 0.83% | 1.45% |
| Historical Studies | 1.03% | 1.13% | 1.16% | 0.91% |
| Information & Communication Technologies | 2.93% | 2.92% | 2.95% | 3.38% |
| Mathematics & Statistics | 3.04% | 3.10% | 3.13% | 2.71% |
| Philosophy & Theology | 0.69% | 0.78% | 0.82% | 0.45% |
| Physics & Astronomy | 9.91% | 9.58% | 9.36% | 9.99% |
| Psychology & Cognitive Sciences | 2.22% | 2.20% | 2.18% | 1.94% |
| Public Health & Health Services | 3.01% | 2.96% | 2.91% | 2.95% |
| Social Sciences | 4.36% | 4.96% | 5.26% | 3.02% |
| Visual & Performing Arts | 0.15% | 0.18% | 0.21% | 0.10% |

Source: Computed by Science-Metrix using Ulrich (Serials Solutions) and the Scopus (Elsevier) databases.

## 2.2 Strategy to measure the proportion of gold, green and hybrid OA in a large sample

Somewhat distinct strategies were used to calculate the occurrence of gold OA and total OA. For gold articles, an estimate of the proportion of papers was made from the random sample by matching the journals that were known to be gold the year a paper was published. These journals were obtained from the Directory of Open Access Journals (DOAJ) and the list of OA journals in PubMed Central. This was done by matching journals' ISSN, E-ISSN and names from Scopus to the relevant records in the sample (the matching had close to 100% precision, but recall may have been imperfect, hence the figures presented here can be considered a floor, rather than a ceiling). Statistics were computed for gold on both the Scopus sample and for the overall population of articles present in the Scopus databases.

Except for the details on the gold journals measurement methods, all the aforementioned methodological details and results were for the pilot phase. The subsequent measurement stage used a relatively large-scale measurement of OA availability based on a sample of 320,000 randomly papers drawn for the Scopus database—that is, 40,000 records per publication year between 2004 and 2011.[8] Additional articles were drawn for smaller fields but these were only used to compute statistics at the field level, the overall proportion relied on the quasi-random selection process of 320,000 papers. Before presenting the results of the study, important metrology concepts are first presented.

## 2.3 Key OA metrology concepts

This paper presents results using two important metrology concepts: (1) *accuracy*, which reflect the quality of the instruments used and the care taken in making measurements; (2) *precision*, which reflect the use of repeated measures, sampling and statistical analysis (see Figure 1)—the later concept will be called *statistical precision* for reasons that will become obvious.

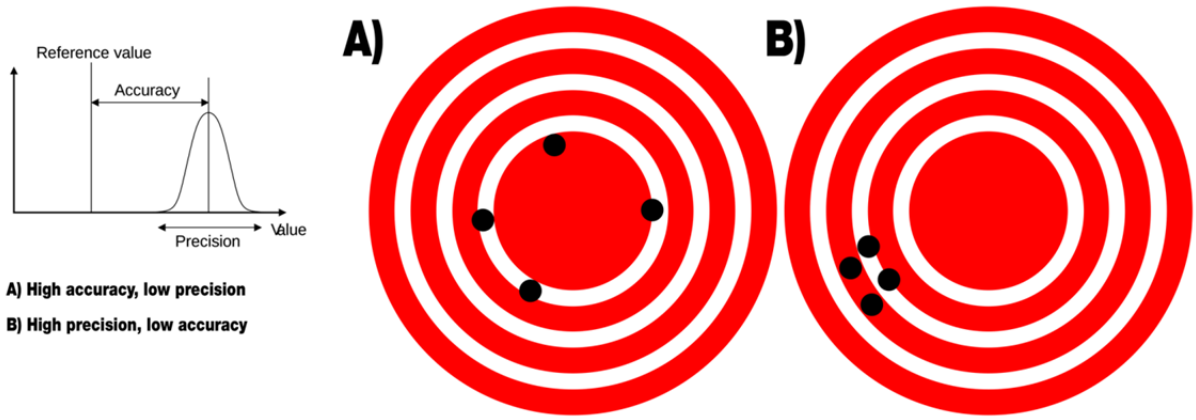


Figure 1 Accuracy and statistical precision
Source: Adapted from http://en.wikipedia.org/wiki/Accuracy_and_precision

[8] With hindsight, the use of a constant number of papers per year (40,000) was not a judicious decision as it does not reflect annual variations in the number of papers. A forthcoming update to this study will sample a large number of papers over the whole period and, in the presence of the law of large numbers, one will be able to assume that this sample will naturally reflect the rate of change observed in the database. This will make it easier to compute accurate statistics and lower error margins.

Statistical precision can be examined with the margin of error (ME). For a proportion ($p$) where the population is finite and known, ($N$) is not systematically much larger than the sample size ($n$), and in which the values are discrete (for example, papers), given a critical score $Z$ (which will be set at 0.95 in the study), ME is calculated as follows:

$$ME = Z\sqrt{\frac{p(1-p)(N-n)}{n(N-1)}} + \frac{0.5}{n}$$

What complicates the use of these definitions is the need to examine accuracy with two more concepts used in information retrieval: recall and precision (hence the need to call the previous concept 'statistical precision'; the second precision-related concept will be referred to as 'retrieval precision'). Recall is the proportion of relevant records that are retrieved, while retrieval precision is the proportion of retrieved records that are relevant. If an instrument retrieves 25 records of which only 20 are relevant, and fails to retrieve 30 additional relevant records, its retrieval precision is 20/25 = 80% while its recall is 25/50 = 50%. Precision is related to Type I errors (false positives), and recall to type II errors (false negatives). Thus, a high recall means that an instrument returned most of the relevant results, while high retrieval precision means that it retrieved more relevant results than irrelevant ones.

# 3 Results

This section examines the quality of the estimates obtained in this study (Section 3.1) followed by a presentation of the proportion of gold OA in Scopus (Section 3.2). The proportion of OA papers published between 2004 and 2011 as measured in April 2013 is presented in Section 3.3 followed by a description of OA availability by field of knowledge production (Section 3.4). The OA citation advantage or disadvantage is briefly examined in Section 3.5 while Section 3.6 provides a brief geographical examination of OA availability.

## 3.1 Quality of the estimates

The results from a perfect classifier would solely comprise a mix of true positives and true negatives. A true positive (*tp*) in the present case is a paper known to be available in OA which is found by the harvesting instrument developed in the current project. A true negative (*tn*) is an article which is not available for free and is not found by the instrument. However, such an instrument rarely is perfect and there are usually false positives (*fp*), that is, articles not available for free but wrongly assigned to that category, and false negatives (*fn*), that is, articles which are available for free but are not found by the instrument. These concepts can be used to characterise how good measurement is with a number of indicators.

Retrieval accuracy provides an overall estimate of the instrument capacity to yield true results as a percentage of records, or more formally, the “closeness of agreement between a measured quantity value and a true quantity value of a measurand”.[9] It is given by:

$$Accuracy = \frac{tp+tn}{tp+tn+fp+fn}$$

Retrieval precision, also called positive predictive value, provides an estimation of how frequently the instrument finds correctly positive results and is given by:

$$Precision = \frac{tp}{tp+fp}$$

Specificity, or true negative rate, is the capacity of the instrument to correctly identify negative results:

$$Specificity = \frac{tn}{tn+fp}$$

Finally, recall, also called true positive rate or sensitivity, is the capacity to correctly identify a large proportion of the positive records:

$$Recall = \frac{tp}{tp+fn}$$

The same sample of 500 articles used in the pilot study was also used for the characterisation of the OA harvester used to measure availability of the 320k sample. Whereas 238 articles were available for free in December 2012 (47.6%), some 224 articles could be found in April 2013 (44.8%). It is also noteworthy that 249 articles were available for free at one time or another between December 2012 and April 2013, just a hair under 50%. Between December 2012 and April 2013, there were 11 new records that appeared in OA, but 25 disappeared, a large part of

[9] BIPM. 2012. *International vocabulary of metrology – Basic and general concepts and associated terms* (VIM). 3rd edition. JCGM 200:2012.

them (13 out of 25) were actually Springer articles which were available for free during the limited time of a promotion. Most of the others were on websites that disappeared, or no longer appeared on the website where they were originally found. These results suggest that there are important transient aspects that need to be taken into consideration while measuring OA availability. Free articles sometimes come and go and it is therefore important to present information not only on the year of publication of articles, but also when the measure was actually made. This also makes replication of results a challenging undertaking.

Table III OA availability in April 2013 of a sample of 500 articles published in 2008

| Type of results | Articles | Quality Tests | Score |
|---|---|---|---|
| True positive (*tp*) | 194 | Retrieval accuracy | 93.2% |
| True negative (*tn*) | 272 | Retrieval precision | 98.0% |
| False positive (*fp*) | 4 | Specificity | 98.6% |
| False negative (*fn*) | 30 | Recall | 86.6% |

These results show that the harvesting engine has very good retrieval precision (98%) and fairly good recall (86.6%). Our design goal is to maximize retrieval precision at the expense of recall, and progressively improve recall while maintaining or improving precision. We therefore consider the harvesting engine's results as consistent with these goals and sufficiently good to present fairly robust measures of OA availability. Please note that even the method used here is a floor of OA availability as neither Google nor Google Scholar, which were used here to determine the 'ground truth', can be expected to have perfect recall themselves.

## 3.2 Gold OA as a proportion of scientific papers

Publications in gold OA journals are examined first at the whole databases level and will subsequently be examined within the sample in greater details. Figure 2 and Figure 3 present current trends on the evolution of gold OA articles availability in Scopus from 1996 to 2011. Figure 2 shows there is an exponential growth of gold OA papers indexed in Scopus. The growth rate is 24% per year (obtained through exponential regression curve fitting), which means that the number of gold papers doubles every 2.9 years. There are currently about 940,000 papers indexed from gold journals in Scopus for the years 1996 to 2011.

The percentage of gold peer-reviewed articles available in Scopus for 1996 was only 0.9% but grew to 11.5% for 2011, the overall compound annual growth rate (CAGR) being 18.7% (Figure 3). Importantly, an availability score for a given year does not mean that such a proportion of articles was initially available at the time. For example, though there are now 0.9% of 1996 papers from Scopus which are available in gold journals, it might have been lower at the time. The reason is that journals sometimes convert to gold status and make their whole back year collection freely available along the way. This creates an upward translation of the availability curve.

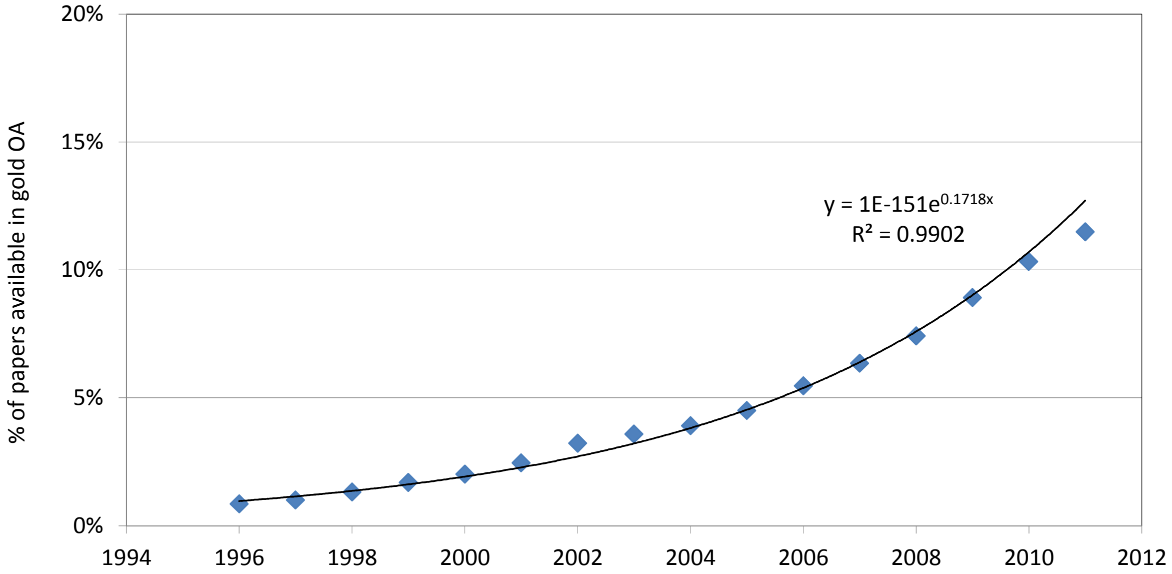


Figure 2 Number of papers from gold journals in Scopus, 1996-2011

Source: Computed by Science-Metrix using DOAJ, PubMedCentral, Scopus.

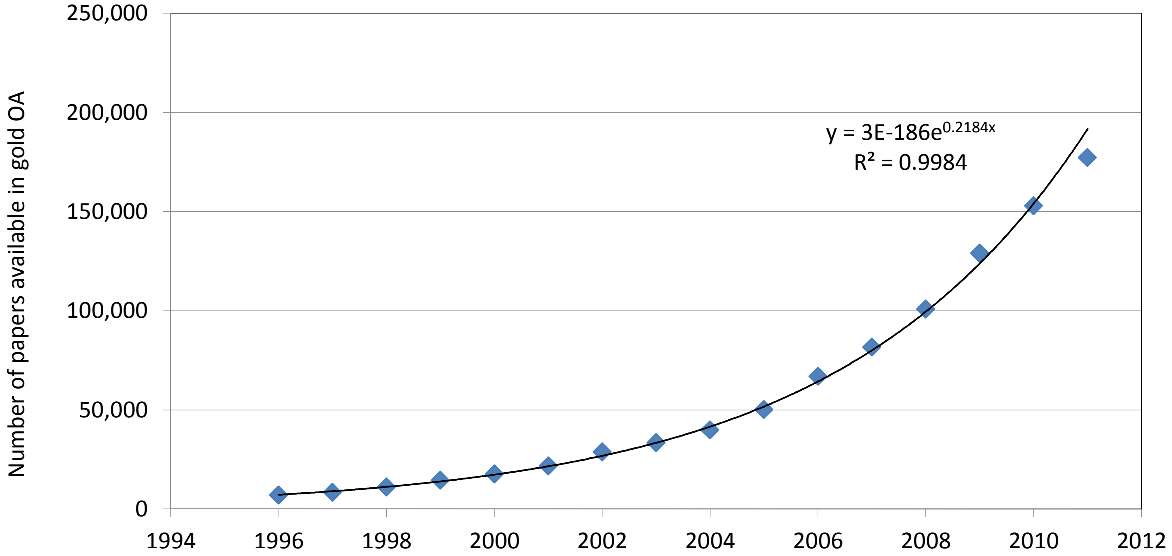


Figure 3 Per cent of papers from gold journals in Scopus, 1996-2011

Source: Computed by Science-Metrix using DOAJ, PubMedCentral, Scopus.

## 3.3 Gold, and Green and Hybrid OA as a proportion of scientific papers

The results from the harvesting of the random sample of 320,000 papers are presented in Figure 4. Starting from the bottom, one can find the availability of gold OA in this sample and, not surprisingly, the results closely follow the population-level statistics presented in Figure 3. As stated earlier, the green and hybrid OA represent a heterogeneous mix of papers comprising self-archived version of papers published in subscription-based journals, papers made available for free by editors, or because authors paid upfront fees to provide free access to their papers. Papers can also be found on aggregator websites such as CiteSeerX, ResearchGate and others. In fact, the hybrid and green OA category comprises a complex continuum of ways to offer papers for free. As one can see, green and hybrid OA availability grows in the distant past and recedes in the more recent past. This is at least in part due to editors having embargo periods on many of the papers in their journals which are sometimes available initially only through subscription and are subsequently being made available for free.

The estimate of overall OA proportion is based on the sum of gold OA and of hybrid and green OA. According to this measure, there are 38% (the statistical margin of error is ± 0.5 percentage points) of the 2004 peer-reviewed journal articles indexed in Scopus papers which are currently available for free. This proportion reached 44% (± 0.5) for papers published in 2011. The growth rate is very low, that is, only 1.9% per year. This low growth rate over time likely reflects that a core phenomenon to consider is the translation of the OA availability curve for back years. Once again, this translation is due to the fact that when authors and editors decide to embrace OA, they often make both recent and older publications available for free, thus creating an upward translation in OA availability.

An adjusted OA availability curve can be computed by applying a conversion factor that accounts for the precision and recall of the instrument (this calibration is based on the analysis of the 500 records sample). This estimation suggests that the tipping point of OA availability was reached in 2011. Though it is an estimate, one can be relatively confident that this is more a floor estimate than a ceiling as there is currently no gold standard for retrieving every freely available peer-reviewed paper (even by the search engines used to establish the ground truth). Moreover, one may suspect that Scopus (and WoS also) may have a bias towards non-gold journals, and having a perfectly non-biased database as a starting point would likely yield slightly higher scores.

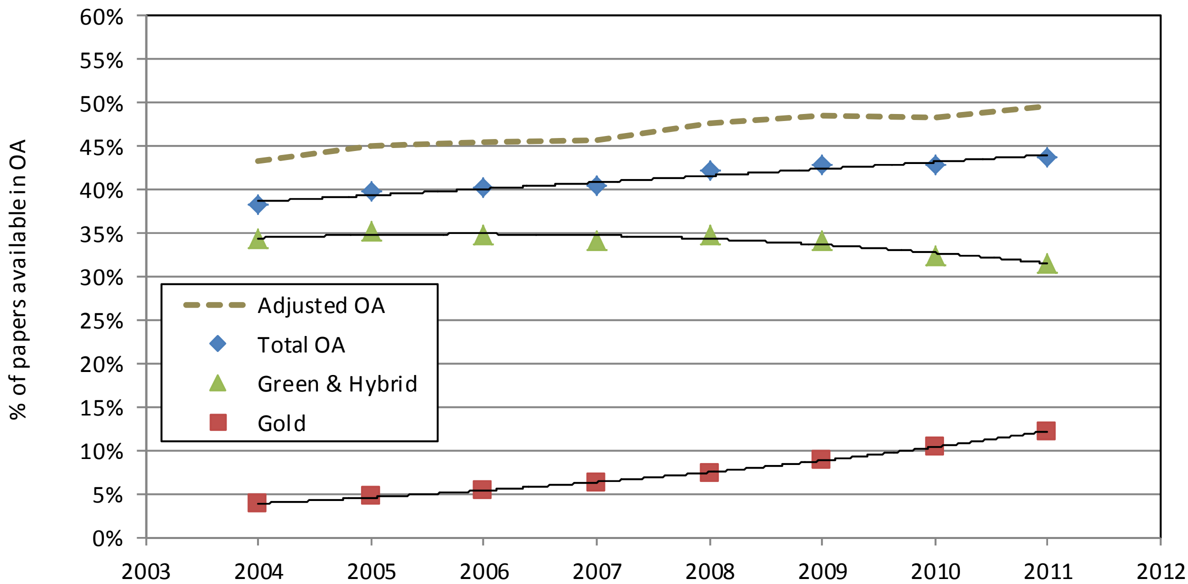


Figure 4 Per cent of freely available peer-reviewed papers, 2004-2011

Source: Computed by Science-Metrix using DOAJ, PubMedCentral, and Scopus.

## 3.4 Availability of OA papers by field

Table IV presents data on the proportion of articles per field which are available in green and hybrid forms, in gold journals, and in OA overall. Considering that these data can be considered as floor values, there is little doubt that free availability of a majority of articles has been reached in general science & technology, in biomedical research, biology, and mathematics & statistics. The general S&T category comprises prestigious journals such as PNAS, Science, Nature, but also gold journals such as PLoS ONE, Scientific Reports, European Journal of Scientific Research, and the Journal of Applied Sciences. The large availability of biomedical papers is owed at least in part to the effort by the US National Institute of Health's OA policy and the presence of the PubMedCentral website which serves as a central platform for the diffusion of scientific evidence in biomedical research and clinical medicine.

Table IV Proportion of OA per field, 4-year non-weighted sampling, 2008-2011

| Field | Papers | Green & Hybrid | | Gold | | OA | | OA Growth | |
|---|---|---|---|---|---|---|---|---|---|
| | | Papers | % | Papers | % | Papers | % | Trend | Index |
| Agriculture, Fisheries & Forestry | 6,142 | 1,655 | 27 ± 1 | 1,033 | 17 ± 0.9 | 2,688 | 44 ± 1 | | 1.13 |
| Biology | 7,031 | 2,749 | 39 ± 1 | 1,224 | 17 ± 0.9 | 3,973 | 57 ± 1 | | 1.08 |
| Biomedical Research | 12,571 | 6,377 | 51 ± 0.9 | 1,346 | 11 ± 0.5 | 7,723 | 61 ± 0.8 | | 1.04 |
| Built Environment & Design | 998 | 237 | 24 ± 3 | 32 | 3 ± 1 | 269 | 27 ± 3 | | 0.99 |
| Chemistry | 13,399 | 2,642 | 20 ± 0.7 | 1,275 | 10 ± 0.5 | 3,917 | 29 ± 0.8 | | 1.39 |
| Clinical Medicine | 42,806 | 15,479 | 36 ± 0.5 | 4,658 | 11 ± 0.3 | 20,137 | 47 ± 0.5 | | 1.17 |
| Communication & Textual Studies | 1,108 | 168 | 15 ± 2 | 63 | 6 ± 1 | 231 | 21 ± 2 | | 1.06 |
| Earth & Environmental Sciences | 4,700 | 1,771 | 38 ± 1 | 339 | 7 ± 0.7 | 2,110 | 45 ± 1 | | 0.95 |
| Economics & Business | 3,950 | 1,583 | 40 ± 2 | 191 | 5 ± 0.7 | 1,774 | 45 ± 2 | | 1.02 |
| Enabling & Strategic Technologies | 12,714 | 2,567 | 20 ± 0.7 | 1,061 | 8 ± 0.5 | 3,628 | 29 ± 0.8 | | 1.11 |
| Engineering | 11,264 | 2,366 | 21 ± 0.8 | 269 | 2 ± 0.3 | 2,635 | 23 ± 0.8 | | 1.03 |
| General Arts, Humanities & Social Sciences* | 8,220 | 2,250 | 27 ± 0.2 | 137 | 2 ± 0.0 | 2,387 | 29 ± 0.2 | | 1.08 |
| General Science & Technology | 2,760 | 720 | 26 ± 2 | 1,036 | 38 ± 2 | 1,756 | 64 ± 2 | | 1.28 |
| Historical Studies | 1,590 | 374 | 24 ± 2 | 124 | 8 ± 1 | 498 | 31 ± 2 | | 0.99 |
| Information & Communication Technologies | 5,620 | 1,993 | 35 ± 1 | 584 | 10 ± 0.8 | 2,577 | 46 ± 1 | | 1.01 |
| Mathematics & Statistics | 4,318 | 2,091 | 48 ± 1 | 310 | 7 ± 0.8 | 2,401 | 56 ± 1 | | 0.95 |
| Philosophy & Theology | 732 | 170 | 23 ± 3 | 38 | 5 ± 2 | 208 | 28 ± 3 | | 1.03 |
| Physics & Astronomy | 15,028 | 5,824 | 39 ± 0.8 | 698 | 5 ± 0.3 | 6,522 | 43 ± 0.8 | | 0.98 |
| Psychology & Cognitive Sciences | 3,027 | 1,204 | 40 ± 2 | 172 | 6 ± 0.8 | 1,376 | 45 ± 2 | | 1.10 |
| Public Health & Health Services | 4,688 | 1,640 | 35 ± 1 | 589 | 13 ± 0.9 | 2,229 | 48 ± 1 | | 1.21 |
| Social Sciences | 5,170 | 1,374 | 27 ± 1 | 489 | 9 ± 0.8 | 1,863 | 36 ± 1 | | 1.00 |
| Visual & Performing Arts* | 6,572 | 609 | 9 ± 0.1 | 231 | 4 ± 0.1 | 840 | 13 ± 0.1 | | 1.00 |
| **Total*** | **160,000** | **53,072** | **33 ± 0.2** | **15,538** | **10 ± 0.1** | **68,610** | **43 ± 0.2** | | 1.08 |

Notes: *In order to decrease the margin of error for the fields of General Arts, Humanities & Social Sciences, and Visual & Performing Arts, the Scopus population of papers were used instead of a sample. However, this does not affect the error margin for the total as this was based on sample data only.

Source: Computed by Science-Metrix using DOAJ, PubMedCentral, and Scopus.

A growth index was computed by dividing the percentage of OA availability in 2008-2011 by that observed in 2004-2007. Overall, between the two periods, there has been an 8% increase in OA availability (slightly more than 3 percentage point). The fields with the fastest growth between the most recent four years and the preceding four years are chemistry, general science & technology, public health and health services, clinical medicine, agriculture, fisheries & forestry, and enabling & strategic technologies. Here, one can suspect once again that the NIH policy is at play (in public health and clinical medicine). Another important observation is that some of the more applied sciences, where OA was not all that prevalent in 2004, appear to be catching up. Also noteworthy is the observation that the cradle of OA availability is now standing still: there is no growth in physics & astronomy. Finally, growth is slightly negative in mathematics & statistics, earth & environmental sciences, built environment, and historical studies.

Gold OA availability is greatest in general S&T (38% of the papers in Scopus) and lowest in general social sciences and humanities (2%) and is also very low in engineering (3%). Other fields with high availability in gold journals include biology as well as agriculture, fisheries & forestry (17%).

The fields where OA availability is most limited are within the social sciences and humanities and in the more applied sciences, engineering, and technology. The lowest prevalence of OA availability is in visual and performing arts (13%) and communication & textual studies. It is also quite low (less than 30% availability) in engineering, built environment & design, enabling & strategic technologies, and chemistry.

The statistics for the 2008–2011 period in Table IV were computed by grouping all the publications for these four years, thus providing a good approximation of the overall trend. However, because sample size remained stable at 40,000 publications per year, this does not reflect the annual variations in the number of papers indexed in Scopus. There is usually an increase in scientific production at the world level in most fields. In Table V, data has been re-weighted to reflect the annual rate of growth of papers in Scopus and in the various fields. By and large, the proportion of OA remains unchanged except for only a few fields, where one or more proportions changed upwards or downwards by one percentage point.

These weighted data show that the absolute number of papers in OA form is rising rapidly as the observed increase in OA proportion noted in Table IV combined by the growth of the number of publications produces a strong overall increase. For example, the growth of the OA proportion in agriculture, fisheries & forestry was 1.13, but the number of papers grew at 1.49 (49% growth in the number of OA papers indexed in Scopus in 2008–2011 compared to the 2004–2007 period). Overall, out of the 5.8 million scientific papers from peer-reviewed journals indexed in Scopus during the 2008–2011 period, 2.5 million were available for free in April 2013. A very large number of papers are available free in clinical medicine, biomedical research, and physics and astronomy. This is partly because of the policy of the US National Institute of Health that mandated the use of the PubMed Central repository for supported research and because of the arXiv e-print archive mirror, which has been largely adopted by researchers in the field of physics.

Table V Number of papers indexed in Scopus available in OA, 2008-2011

| Field | Sample | Scopus | Green & Hybrid | | Gold | | OA | | OA Growth (papers) | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Papers* | Papers | Sco. Papers | % | Sco. Papers | % | Sco. Papers | % | Trend | Index |
| Agriculture, Fisheries & Forestry | 6,142 | 216,797 | 58,363 | 27 ± 2 | 36,644 | 17 ± 2 | 95,007 | 44 ± 2 | | 1.49 |
| Biology | 7,031 | 252,150 | 98,461 | 39 ± 2 | 44,137 | 18 ± 2 | 142,598 | 57 ± 2 | | 1.39 |
| Biomedical Research | 12,571 | 459,911 | 232,867 | 51 ± 2 | 49,745 | 11 ± 1 | 282,613 | 61 ± 2 | | 1.19 |
| Built Environment & Design | 998 | 34,230 | 8,147 | 24 ± 5 | 1,110 | 3 ± 2 | 9,257 | 27 ± 6 | | 1.26 |
| Chemistry | 13,399 | 481,575 | 94,527 | 20 ± 1 | 46,446 | 10 ± 1 | 140,973 | 29 ± 2 | | 1.69 |
| Clinical Medicine | 42,806 | 1,564,310 | 565,524 | 36 ± 0.9 | 171,539 | 11 ± 0.6 | 737,063 | 47 ± 0.9 | | 1.36 |
| Communication & Textual Studies | 1,108 | 42,863 | 6,439 | 15 ± 4 | 2,518 | 6 ± 3 | 8,957 | 21 ± 5 | | 1.58 |
| Earth & Environmental Sciences | 4,700 | 172,130 | 64,813 | 38 ± 3 | 12,391 | 7 ± 1 | 77,204 | 45 ± 3 | | 1.16 |
| Economics & Business | 3,950 | 150,081 | 60,173 | 40 ± 3 | 7,242 | 5 ± 1 | 67,415 | 45 ± 3 | | 1.50 |
| Enabling & Strategic Technologies | 12,714 | 461,531 | 93,056 | 20 ± 1 | 38,645 | 8 ± 1.0 | 131,701 | 29 ± 2 | | 1.60 |
| Engineering | 11,264 | 404,900 | 85,001 | 21 ± 1 | 9,714 | 2 ± 0.6 | 94,716 | 23 ± 2 | | 1.38 |
| General Arts, Humanities & Social Sciences** | 8,220 | 8,220 | 2,250 | 27 | 137 | 2 | 2,387 | 29 | | 1.54 |
| General Science & Technology | 2,760 | 101,464 | 26,239 | 26 ± 3 | 38,662 | 38 ± 3 | 64,901 | 64 ± 4 | | 2.21 |
| Historical Studies | 1,590 | 57,584 | 13,485 | 23 ± 4 | 4,481 | 8 ± 3 | 17,966 | 31 ± 5 | | 1.34 |
| Information & Communication Technologies | 5,620 | 200,487 | 71,230 | 36 ± 3 | 20,781 | 10 ± 2 | 92,011 | 46 ± 3 | | 1.26 |
| Mathematics & Statistics | 4,318 | 157,207 | 75,813 | 48 ± 3 | 11,355 | 7 ± 2 | 87,168 | 55 ± 3 | | 1.23 |
| Philosophy & Theology | 732 | 27,056 | 6,385 | 24 ± 6 | 1,426 | 5 ± 3 | 7,811 | 29 ± 7 | | 1.52 |
| Physics & Astronomy | 15,028 | 552,749 | 213,949 | 39 ± 2 | 25,925 | 5 ± 0.7 | 239,874 | 43 ± 2 | | 1.15 |
| Psychology & Cognitive Sciences | 3,027 | 113,841 | 45,310 | 40 ± 4 | 6,585 | 6 ± 2 | 51,895 | 46 ± 4 | | 1.40 |
| Public Health & Health Services | 4,688 | 173,475 | 60,680 | 35 ± 3 | 21,893 | 13 ± 2 | 82,573 | 48 ± 3 | | 1.59 |
| Social Sciences | 5,170 | 191,674 | 51,011 | 27 ± 2 | 18,105 | 9 ± 2 | 69,116 | 36 ± 3 | | 1.46 |
| Visual & Performing Arts** | 6,572 | 6,572 | 609 | 9 | 231 | 4 | 840 | 13 | | 1.51 |
| **Total**** | **160,000** | **5,830,804** | **1,930,853** | **33 ± 0.5** | **570,947** | **10 ± 0.3** | **2,501,800** | **43 ± 0.5** | | 1.36 |

Notes: * The number of sampled papers is for the whole four-year period (unweighted) whereas the number of papers in Scopus, percentage of OA, and errors use yearly weighted data that reflect growth in Scopus.
**In order to decrease the margin of error for the fields of General Arts, Humanities & Social Sciences, and Visual & Performing Arts, the Scopus population of papers were used instead of a sample. However, this does not affect the error margin for the total as this was based on sample data only.

Source: Computed by Science-Metrix using DOAJ, PubMedCentral, and Scopus.

## 3.5 Citation advantage and disadvantage of OA papers

A question that has animated OA advocates has been the so-called OA citation advantage. Evidence on this question is examined in Table VI using the Average of Relative Citation (ARC), a measure that reflects citation rates and is normalised to account for differences among scientific specialities in the propensity to use references and receive citations. These data present the relative citation rate of OA publications overall, Gold OA and hybrid OA forms relative to all publications in each field. All the fields have been set back to 1.0 to allow the calculation of a citation advantage/disadvantage, this baseline comprises all the papers in a field for the given time period. A score above 1 denotes that papers are more cited than in the field overall, while a score below 1 means that these publications are less frequently cited. For instance, papers in Agriculture, Fisheries & Forestry receive roughly the same level of citation (1.06) in OA overall than they do usually.

All the fields derive an OA citation advantage (rebased ARC > 1). Paradoxically, many of the fields where the OA proportion is low have a sizeable citation advantage such as in philosophy and theology (1.54), general arts humanities and social sciences, communication and textual studies, engineering, and visual & performing arts.

Table VI    Rebased scientific impact (ARC) of OA publications, 2008-2011

| Field | All Publications | Green & Hybrid | Gold | OA |
|---|---|---|---|---|
| Agriculture, Fisheries & Forestry | 1.00 | 1.38 | 0.52 | 1.06 |
| Biology | 1.00 | 1.41 | 0.48 | 1.15 |
| Biomedical Research | 1.00 | 1.25 | 0.76 | 1.17 |
| Built Environment & Design | 1.00 | 1.33 | n.c. | 1.23 |
| Chemistry | 1.00 | 1.38 | 0.36 | 1.09 |
| Clinical Medicine | 1.00 | 1.56 | 0.54 | 1.34 |
| Communication & Textual Studies | 1.00 | 1.66 | 0.88 | 1.46 |
| Earth & Environmental Sciences | 1.00 | 1.30 | 0.82 | 1.22 |
| Economics & Business | 1.00 | 1.32 | 0.22 | 1.20 |
| Enabling & Strategic Technologies | 1.00 | 1.43 | 0.75 | 1.25 |
| Engineering | 1.00 | 1.55 | 0.55 | 1.46 |
| General Arts, Humanities & Social Sciences | 1.00 | 1.53 | 0.10 | 1.46 |
| General Science & Technology | 1.00 | 2.57 | 0.54 | 1.54 |
| Historical Studies | 1.00 | 1.54 | 0.51 | 1.29 |
| Information & Communication Technologies | 1.00 | 1.37 | 0.89 | 1.27 |
| Mathematics & Statistics | 1.00 | 1.22 | 0.71 | 1.16 |
| Philosophy & Theology | 1.00 | 1.56 | n.c. | 1.54 |
| Physics & Astronomy | 1.00 | 1.36 | 1.01 | 1.32 |
| Psychology & Cognitive Sciences | 1.00 | 1.37 | 0.69 | 1.29 |
| Public Health & Health Services | 1.00 | 1.36 | 0.72 | 1.19 |
| Social Sciences | 1.00 | 1.52 | 0.55 | 1.26 |
| Visual & Performing Arts | 1.00 | 1.93 | 0.11 | 1.40 |
| **Total** | **1.00** | **1.41** | **0.60** | **1.24** |

Source:    Computed by Science-Metrix using DOAJ, PubMedCentral, and Scopus.

What is particularly interesting here is that the citation advantage is derived almost exclusively from the green and hybrid portion, as gold OA is associated with a citation disadvantage on average for all fields except for physics & astronomy. In physics & astronomy, information & communication technologies, and in communication & textual studies, there are close to no differences between gold and regular journals in terms of average citation rates. Currently, there is a marked disadvantage for publishing in gold journals in GAHSS, visual & performing arts, economics & business and in chemistry. Interestingly, visual & performing arts has one of the highest advantages derived from the use of green & hybrid OA, yet it is the field with the least prevalent use of OA.

These data reveal that a whole new area in the sociology of science may open up to explain scientists' publication behaviour in the various fields of knowledge production. There are extremely marked communication differences which make inquisitive studies of the kind that Derek de Solla Price initiated in the 1960s very worthwhile again. Considering there are a fairly large number of studies on specific scientific fields currently, there are comparatively little systemic studies of change in contemporary science. This is partly because the sociology and philosophy of science have distanced themselves from, and now largely snob measurement studies such as the present one. The development of OA provides a natural experiment that could revive interest for inquiring into the sociology of science publication, should there be an interest among scientists, science policy practitioners, and scientific journals editors.

The statistics on gold journals require careful interpretation. First, many gold journals are younger and smaller, and these factors have an adverse effect on the citation rate and hence on measured ARC values. Authors frequently prefer reading and citing more established journals, and it is therefore a challenge to start a journal from scratch, and to have authors submit high quality articles. It takes time to build a reputation and to attract established authors. It is also possible that gold journals might provide an avenue for less mainstream, more revolutionary science. If so, the signature would be a much greater level of variation between the more highly cited papers and the baseline (no citation). Also, the ARC is not scale-invariant, and larger journals have an advantage as this measure is not corrected sufficiently for journal size (namely, it is not a scale-independent measure). So it might not always be the gold nature of journals that lowers their 'citedness'; instead several structural aspects might be at play. Even so, the gold journal industry is young, and it is still difficult to separate the wheat from the chaff. In this respect, it might be useful for authors to examine Beall's List of 'potential, possible, or probable predatory scholarly open-access publishers' to lower one's risk of spending money on journals that do not espouse scientific publishing best practices.[10]

## 3.6 Availability of OA papers in European and selected countries

An examination of OA availability was performed for EU28, EFTA (European Free Trade Association), Accession countries, ERA (European Research Area), and four additional countries, namely, Brazil, Canada, Japan, and the US.[11] Please note that these data are not weighted to reflect growth in Scopus. If weighted data were used, some countries would see the proportion of OA go up or down by one percentage point. Countries with a very small sample size (Latvia, Luxembourg, Malta, Liechtenstein, and Macedonia) would show greater variations, but the results presented here for these countries should be interpreted with great care in any case, because of the small sample size and ensuing large error margin.

Fractional counting is used in this table. In fractional counting, if two authors are from separate countries, each country is given half a publication. In contrast, full paper counting would have ascribed one paper to each country. One advantage of fractional counting is that one can add the calculated output of all countries' output in a table and obtain a total of 100%.

---

[10] http://scholarlyoa.com/publishers/.

[11] See Annex I for statistics on OA taking into account yearly growth of the number of publications in the Scopus database.

The EU28, EFTA, and ERA all have roughly the same proportion of OA articles as that observed at the world level (43% for 2008-2011, a figure which is not recalibrated for precision and recall), though there are noticeable differences among countries (Table VII). For the period 2008-2011 considered as a whole, eight of the EU28 (30%) have reached the tipping point, namely (by decreasing percentage of OA): the Netherlands, Portugal, Lithuania, Estonia, Denmark, Malta, Ireland, and Belgium. If the statistical precision and recall of the harvesting instrument are taken into account, 20 out of 27 countries (74%) would have tipped towards a majority of papers published in 2008-2011 being made available for free in April 2013.

Table VII Proportion of OA per country, 4-year non-weighted sampling, 2008-2011

| Group | Country | Papers in Sample | Green & Hybrid | | Gold | | OA | |
|---|---|---|---|---|---|---|---|---|
| | | | Papers | % | Papers | % | Papers | % |
| **EU28** | Austria | 1,349 | 545 | 40 ± 3 | 105 | 8 ± 1 | 650 | 48 ± 3 |
| | Belgium | 2,088 | 939 | 45 ± 2 | 126 | 6 ± 1 | 1,065 | 51 ± 2 |
| | Bulgaria | 293 | 91 | 31 ± 5 | 21 | 7 ± 3 | 112 | 38 ± 6 |
| | Croatia | 403 | 147 | 36 ± 5 | 95 | 24 ± 4 | 242 | 60 ± 5 |
| | Cyprus | 88 | 35 | 40 ± 11 | 3 | 3 ± 4 | 38 | 43 ± 11 |
| | Czech Republic | 1,252 | 411 | 33 ± 3 | 130 | 10 ± 2 | 541 | 43 ± 3 |
| | Denmark | 1,392 | 633 | 45 ± 3 | 103 | 7 ± 1 | 736 | 53 ± 3 |
| | Estonia | 161 | 63 | 39 ± 8 | 23 | 14 ± 6 | 86 | 53 ± 8 |
| | Finland | 1,178 | 488 | 41 ± 3 | 80 | 7 ± 1 | 568 | 48 ± 3 |
| | France | 7,959 | 3,205 | 40 ± 1 | 472 | 6 ± 0.5 | 3,677 | 46 ± 1 |
| | Germany | 10,531 | 4,026 | 38 ± 0.9 | 673 | 6 ± 0.5 | 4,699 | 45 ± 0.9 |
| | Greece | 1,336 | 452 | 34 ± 3 | 116 | 9 ± 2 | 568 | 43 ± 3 |
| | Hungary | 709 | 279 | 39 ± 4 | 49 | 7 ± 2 | 328 | 46 ± 4 |
| | Ireland | 833 | 358 | 43 ± 3 | 75 | 9 ± 2 | 433 | 52 ± 3 |
| | Italy | 6,094 | 2,389 | 39 ± 1 | 421 | 7 ± 0.6 | 2,810 | 46 ± 1 |
| | Latvia | 71 | 23 | 32 ± 11 | 8 | 11 ± 8 | 31 | 44 ± 12 |
| | Lithuania | 228 | 88 | 39 ± 6 | 35 | 15 ± 5 | 123 | 54 ± 7 |
| | Luxembourg | 37 | 13 | 35 ± 17 | 2 | 5 ± 9 | 15 | 41 ± 17 |
| | Malta | 23 | 8 | 35 ± 21 | 4 | 17 ± 17 | 12 | 52 ± 22 |
| | Netherlands | 3,759 | 1,936 | 52 ± 2 | 263 | 7 ± 0.8 | 2,199 | 58 ± 2 |
| | Poland | 2,474 | 707 | 29 ± 2 | 326 | 13 ± 1 | 1,033 | 42 ± 2 |
| | Portugal | 1,047 | 479 | 46 ± 3 | 97 | 9 ± 2 | 576 | 55 ± 3 |
| | Romania | 734 | 289 | 39 ± 4 | 75 | 10 ± 2 | 364 | 50 ± 4 |
| | Slovakia | 368 | 127 | 35 ± 5 | 41 | 11 ± 3 | 168 | 46 ± 5 |
| | Slovenia | 358 | 110 | 31 ± 5 | 50 | 14 ± 4 | 160 | 45 ± 5 |
| | Spain | 5,461 | 2,074 | 38 ± 1 | 604 | 11 ± 0.8 | 2,678 | 49 ± 1 |
| | Sweden | 2,301 | 922 | 40 ± 2 | 181 | 8 ± 1 | 1,103 | 48 ± 2 |
| | United Kingdom | 11,781 | 5,100 | 43 ± 0.9 | 728 | 6 ± 0.4 | 5,828 | 49 ± 0.9 |
| | **Total EU28** | 53,622 | 20,204 | 37 ± 0.4 | 4,192 | 8 ± 0.2 | 24,396 | 45 ± 0.4 |
| **EFTA** | Iceland | 85 | 39 | 46 ± 11 | 3 | 4 ± 4 | 42 | 49 ± 11 |
| | Liechtenstein | 6 | 1 | 17 ± 38 | | | 1 | 17 ± 38 |
| | Norway | 1,159 | 494 | 43 ± 3 | 110 | 9 ± 2 | 604 | 52 ± 3 |
| | Switzerland | 2,642 | 1,214 | 46 ± 2 | 208 | 8 ± 1 | 1,422 | 54 ± 2 |
| | **Total EFTA** | 3,830 | 1,705 | 45 ± 2 | 319 | 8 ± 0.9 | 2,024 | 53 ± 2 |
| **Candidate** | Turkey | 2,873 | 657 | 23 ± 2 | 598 | 21 ± 1 | 1,255 | 44 ± 2 |
| | Macedonia | 39 | 15 | 38 ± 16 | 11 | 28 ± 15 | 26 | 67 ± 16 |
| | **Total Candidate** | 3,303 | 672 | 23 ± 2 | 608 | 21 ± 1 | 1,280 | 44 ± 2 |
| | Israel | 1,376 | 640 | 47 ± 3 | 92 | 7 ± 1 | 732 | 53 ± 3 |
| **Total ERA** | | 59,852 | 22,085 | 37 ± 0.4 | 5,009 | 8 ± 0.2 | 27,094 | 45 ± 0.4 |
| **Others** | United States | 41,740 | 20,894 | 50 ± 0.5 | 2,535 | 6 ± 0.2 | 23,429 | 56 ± 0.5 |
| | Japan | 9,703 | 3,264 | 34 ± 0.9 | 804 | 8 ± 0.5 | 4,068 | 42 ± 1 |
| | Canada | 6,676 | 2,885 | 43 ± 1 | 411 | 6 ± 0.6 | 3,296 | 49 ± 1 |
| | Brazil | 4,224 | 876 | 21 ± 1 | 1,799 | 43 ± 1 | 2,675 | 63 ± 1 |
| **World** | | 160,000 | 53,072 | 33 ± 0.2 | 15,538 | 10 ± 0.1 | 68,610 | 43 ± 0.2 |

Source: Computed by Science-Metrix using DOAJ, PubMedCentral, and Scopus.

Leaving aside smaller countries with only few papers in the sample (less than 100 papers), the countries with the lowest rate of OA adoption are Bulgaria, Poland, Greece, the Czech Republic, and Germany. In EFTA countries, both Norway and Switzerland have clearly tipped towards having a majority of papers in OA, and by counting Israel, there are 14 ERA countries (40%) which had a majority of their 2008-2011 papers being freely available by 2013. Calibrating for retrieval precision and recall, the proportion of ERA countries having more than 50% of papers in OA is 74%, that is, the same as for the EU28.

In countries outside the ERA, it is noteworthy that the US has passed the tipping point and Canada is approaching that point. Even more salient is the proportion of 63% observed in Brazil. This is no doubt due to the important contribution of Scielo, which plays a key role in the Southern hemisphere in making scientific knowledge more widely available.

Publishing in gold journals is much more frequently encountered in the eastern part of Europe. There is a clear pattern that can be observed as the percentage of papers in gold journals is greater in Macedonia, Croatia, Turkey, Malta, Lithuania, Estonia, Slovenia, Poland, Latvia, Slovakia, Spain, the Czech Republic and Romania. One interesting hypothesis is that researchers in these countries may use gold journals because they more frequently allow publishing in languages other than English. Should that be the case, this may also contribute to explaining the lower citation scores received by papers in gold journals as the readership for "vernacular languages", as Eugene Garfield (1998) would put it, is lower and the size of the potential reference pool is consequently also smaller. There is therefore a potentially fertile ground for studying the social and linguistic aspects of science by examining where and why gold open access journals are appearing and who actually makes use of them.

# 4 Discussion

If one makes a straight reading of the data presented here, between 2004 and 2011, the average annual rate of increase of OA availability was relatively limited, with a compound growth rate of 2% per year. Importantly though, examining the growth curve is misleading. Though the data does not support this hypothesis, there seems to be an important translation movement in OA availability. This means that in addition to having year-on-year growth, there is an upward movement of the whole availability curve over time. This is due to an increasing number of authors making their manuscript available for the current year but also for previous years. The same is true for journal publishers who may decide to opt for an OA model and make back files freely available as well. Finally, when one looks at the availability curve, it can be seen that the availability falls somewhat for the most recent few years. This is likely due to the presence of embargo for some of the subscription based journals which make all articles freely available to all after a few years—a practice which may be called "delayed OA".

There are also transient effects that have to be considered when measuring OA availability. Many articles that were available for free in December 2012 were no longer available for free in April 2013. This is in part due to a promotion Springer was running in late 2012 by making several subscription-journal papers available for free, and later making them available for a fee again. Some articles disappeared from websites, and some websites were not responding when visited in April thus reducing further OA availability. Conversely, many new papers then became available for free. All in all, more than 50% of the papers could be found for free in November/December 2012 (pilot phase of this study) and in March/April 2013 (1st full measurement stage) but somewhat less so at either time period. This shows that measuring phenomena on the Internet requires particular attention to detail and constant questioning on the meaning of the results—one has to ask whether these results are permanent or transient.

Other authors have presented results suggesting that OA availability was only half as high as measured here, a reminder that it is important to be reflective and self-critical in all bibliometric exercises. Whereas other authors have measured what appeared in specific databases, or what specific search engines were able to do, the more important question is that of overall free availability of peer-reviewed, academic-level journal papers available. Measuring how well Google Scholar fares at identifying a part of this literature is certainly an interesting exercise in itself, but it does not inform on the state of affairs of the scientific publishing industry, and what researchers who are reasonably tenacious will find for free.

Suggesting that the tipping point has been reached in open access certainly has important implications for academia, for university librarians, and perhaps even more so for the scientific, technical and medical publishing industry.

Much has been said about the cost of publishing in gold and hybrid OA being disadvantageous relative to self-archiving, but one has to place this in perspective. The cost of academic papers in the US is over $100,000—which is calculated by dividing the higher education expenditures on R&D (HERD) by the number of papers published by academia. Adding to or included in this amount, a $2,000 OA publication fee only accounts for a few percentage points of a typical research project budget, especially in the natural and health sciences.

If the toll plaza moves from the end of the scientific publication process to the beginning, one category of workers is likely to be highly affected: the university and research centre librarian. Librarians have been highly affected already by the shift from paper to digital media, and losing the financial responsibility of spending the large sums paid in journal subscriptions will certainly create another large dent in their traditional sphere of responsibilities. If the toll plaza is moved before the publication process and subscriptions are obsolete, it means that researchers will gain control over financial matters associated with scientific publishing, and that librarians will lose control over sizeable part of their budgets. The market power will shift tremendously from the tens of thousands of buyers that publisher sales staff have nurtured in the past to the millions of researchers that will now make the atomistic decision of how best to spend their publication budget.

The impression gained from carrying out this study and developing our OA harvester is that the toll plaza is being moved to the beginning of the publishing process, away from the back-end of the process, and thus away from librarians and closer to researchers. Despite what several authors thought, and argued for, green OA only appears to be moving slowly, whereas gold and hybrid OA (such as pay-per-article for OA release) appear to be driving in the fast lane. This impression will require further verifications. If confirmed, it would mean that the industry is increasingly placing the "toll before" as opposed to placing the "toll after". Efforts need to be made to characterise these changes, and to distinguish what percentage of growth comes from green self-archiving and what comes from other forms of hybrid OA.

Using a wider lens, the fact that the tipping point has likely been reached in open access is an important finding for the whole publishing industry. This industry is likely to be undergoing revolutionary change, and at a pace much faster than anticipated, in large part because previous measures of OA availability proved to be misleading—having missed more than half of what was available for a number of reasons. This means that aggressive OA publishers are likely to gain much in the redesigned landscape, whereas those attached to the old ways are more likely to suffer and to lose market share. An important economic question is whether the switch to a more atomistic, fine-grained market with millions of researchers as buyers will reduce, augment or leave unchanged the negotiating power of publishers.

# References


Abad-García, M. F., Melero, R., Abadal, E., & González-Teruel, A. (2010). Self-archiving of biomedical papers in open access repositories. *Autoarchivo de artículos biomédicos en repositorios de acceso abierto, 50*(7), 431-440. doi: 10.1371/journal.pbio.0040157.

Antelman, K. (2004). Do open-access articles have a greater citation impact? *College & Research Libraries*, 65(5), 372–382.

Arlitsch, K., & O'Brien, P. S. (2012). Invisible institutional repositories: Addressing the low indexing ratios of IRs in Google Scholar. *Library Hi Tech*, *30*(1), 60-81. doi: 10.1108/07378831211213210.

Bailey Jr, C. W. (2005). The role of reference librarians in institutional repositories. *Reference Services Review, 33*(3), 259-267.

Bilder, G. (2003). Ingenta's economic and technical models for providing institutional OA archives. *Information Services and Use, 23*(2-3), 111-112.

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). doi: 10.1371/journal.pone.0011273.

Carbone, P. (2007). Consortium negotiations with publishers - Past and Future. *LIBER Quarterly*, 17(2).

Chan, D. L. H., Kwok, C. S. Y., & Yip, S. K. F. (2005). Changing roles of reference librarians: The case of the HKUST Institutional Repository. *Reference Services Review, 33*(3), 268-282. doi: 10.1108/00907320510611302

Chan, L., Devakos, R., & Mircea, G. (2005). *Workshop 2: Implementing and filling institutional repositories introduction*, Leuven-Heverlee.

Charles, L., & 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

Craig, I. D., Plume, A. M., McVeigh, M. E., Pringle, J., & Amin, M. (2007). Do open access articles have greater citation impact? A critical review of the literature. *Journal of Informetrics, 1*(3), 239-248. doi: 10.1016/j.joi.2007.04.001.

Dubini, P., Galimberti, P., & Micheli, M. R. (2010). Authors publication strategies in scholarly publishing. In *ELPUB 2010 International Conference on Electronic Publishing*, Helsinki (Iceland), 16-18 June 2010.

Garfield, E. (1998). The Use of Journal Impact Factors and Citations Analysis for Evaluation of Science. Presented at Cell Separation, Hematology and Journal Citation Analysis Mini Symposium in tribute to Arne Bøyum Rikshospitalet, Oslo, April 17, 1998. Retrieved from http://www.garfield.library.upenn.edu/papers/eval_of_science_oslo.html.

Gargouri, Y., Larivière, V. Gingras, Y. and Harnad, S. (2012). Green and Gold Open Access Percentages and Growth, by Field. In Archambault, É, Gingras, Y. and Larivière. V. (2012). *Proceedings of 17th International Conference on Science and Technology Indicators*, Montréal: Science-Metrix and OST.

Gentil-Beccot, A., Mele, S., & Brooks, T. C. (2010). Citing and reading behaviours in high-energy physics. *Scientometrics, 84*(2), 345-355. doi: 10.1007/s11192-009-0111-1

Harnad, S. (2003). The research-impact cycle. *Information Services and Use, 23*(2-3), 139-142.

Harnad, S. (2008). Waking OA's "slumbering giant": The university's mandate to mandate open access. *New Review of Information Networking, 14*(1), 51-68. doi: 10.1080/13614570903001322.

Harnad, S. (2012). Open access: A green light for archiving. *Nature, 487*(7407), 302. doi: 10.1038/487302b

Harnad, S., & Brody, T. (2004). Comparing the impact of open access (OA) vs. non-OA articles in the same journals. *D-Lib Magazine*, 10(6).

Henderson, I. (2013). Open-Access and Institutional Repositories in Fire Literature. *Fire Technology, 49*(1), 155-161. doi: 10.1007/s10694-010-0198-1

Houghton, J. W. (2010). Economic implications of alternative publishing models: Self-archiving and repositories. *LIBER Quarterly, 19*(3-4), 275-292.

Hu, C., Zhang, Y., & Chen, G. (2010). Exploring a New Model for Preprint Server: A Case Study of CSPO. *Journal of Academic Librarianship, 36*(3), 257-262. doi: 10.1016/j.acalib.2010.03.010

Kurek, K., Geurts, P. A. Th M., & Roosendaal, H. E. (2006). The split between availability and selection: Business models for scientific information, and the scientific process? *Information Services and Use, 26*(4), 271-282.

Lakshmi Poorna, R., Mymoon, M., & Hariharan, A. (2012). A study of select open access journals and their business models listed in DOAJ in the fields of civil and structural engineering. *Journal of Structural Engineering (India), 39*(4), 458-468. doi: 10.1371/journal.pone.0020961.

Larivière, V., Archambault, É., Gingras, Y. & Vignola-Gagné, É. (2006). The Place of Serials in Referencing Practices: Comparing Natural Sciences and Engineering with Social Sciences and Humanities, *Journal of the American Society for Information Science and Technology*, 57(8), 997-1004. doi: 10.1002/asi.20349.

Miguel, S., Bongiovani, P. C., Gómez, N. D., & Bueno-de-la-Fuente, G. (2013). Prospect for Development of Open Access in Argentina. *Journal of Academic Librarianship, 39*(1), 1-2. doi: 10.1016/j.acalib.2012.10.002

Morris, S. (2003). Open Publishing: How publishers are reacting. *Information Services and Use, 23*(2-3), 99-101.

Nyambi, E., & Maynard, S. (2012). An investigation of institutional repositories in state universities in Zimbabwe. *Information Development, 28*(1), 55-67. doi: 10.1177/0266666911425264

Pelizzari, E. (2004). Academic authors and open archives: A survey in the social science field. *Libri, 54*(2), 113-122.

Prosser, D. (2003). Institutional repositories and Open access: The future of scholarly communication. *Information Services and Use, 23*(2-3), 167-170.

Regazzi, John. (2004). The Shifting Sands of Open Access Publishing, a Publisher's View. Serials Review, 30(4), 275-280. doi: http://dx.doi.org/10.1016/j.serrev.2004.09.010

Repanovici, A. (2012). Professional profile of digital repository manager. *Library Hi Tech News, 29*(10), 13-20. doi: 10.1108/07419051211294473

Rowland, F. et al. (2004). Delivery, management and access model for e-prints and open access journals. *Serials Review, 30*(4), 298-303. doi: 10.1016/j.serrev.2004.09.006

Sawant, S. (2012). Past and Present Scenario of Open Access Movement in India. *Journal of Academic Librarianship*. doi: 10.1016/j.acalib.2012.11.007

Swan, A. et al. (2005). Developing a model for e-prints and open access journal content in UK further and higher education. *Learned Publishing, 18*(1), 25-40. doi: 10.1087/0953151052801479

Swan, A., & Brown, S. (2004). Authors and open access publishing. *Learned Publishing, 17*(3), 219-224. doi: 10.1087/095315104323159649

Woutersen-Windhouwer, S. (2012). The Future of Open Access Publishing in the Netherlands: Constant Dripping Wears Away the Stone. *Journal of Academic Librarianship*. doi: 10.1016/j.acalib.2012.11.015