

# Open Access to the Scientific Journal Literature – Status and Challenges for the Information Systems Community

by Bo-Christer Björk and Patrik Paetau

## EDITOR'S SUMMARY

Since the 1990s, the aim of open access (OA) has been to enhance scholarly communication by delivering scientific publications on the Internet without fees or restrictions. Adoption of the gold model of OA, electronic access to publishers' scientific journals at no cost, has been slow in the United States and the United Kingdom, even if authors pay an extra OA charge. The more successful route has been green OA, publishing the original or modified versions on authors' personal home pages or in institutional or subject-specific repositories. In a 2008 sample of science disciplines, Bjork found 20.4% were provided OA, with 8.5% on publishers' sites. For information systems articles published in 2009, 21.3% were OA but a mere 0.6% on publishers' sites, the gold model. Of OA articles, 8% were published on domain repositories, 32.7% in institutional repositories, and 59% on other websites. Broad expansion of OA will depend on greater author awareness and wider platform access, and publishers must adopt international bibliographic metadata and standardized author identification.

## KEYWORDS

open access publications	electronic journals	scientific and technical information
scholarly publishing	primary literature	diffusion of innovation
digital repositories	electronic publishing	

Bo-Christer Björk is professor of information systems science at the Hanken School of Economics in Helsinki, Finland. He is the founder of the *Journal of Information Technology in Construction*, one of the earliest open access journals, published since 1996, and has been engaged in national and international activities promoting the open availability of scientific research results on the web, most recently as a board member of the Open Access Scholarly Publishers Association. He has published several peer-reviewed articles concerning OA in *JASIST* and other publications. He can be reached at [bochrister.bjork@hanken.fi](mailto:bochrister.bjork@hanken.fi).

Patrik Paetau is associate professor of information systems science at the Vasa campus of the Hanken School of Economics. His research interests concern development methods for information systems. He can be reached at [patrik.paetau@hanken.fi](mailto:patrik.paetau@hanken.fi).

In the last decade and a half the World Wide Web has revolutionized scientific journal publishing. But despite the fact that all major publishers are today publishing electronic versions in parallel with the paper issues, not all the potential of the web for efficient knowledge transfer has been used. *Open access* (OA) means the free unrestricted access to scientific publications and data in electronic format on the Internet. OA is related to a number of rather similar phenomena that have emerged with the Internet such as *open source*, *open content*, *peer production* and *social media*. A very important feature distinguishing the scientific journal domain of OA from that of other publishing industries such as film, music or books is that scientific knowledge is a public good produced mainly with public funding and that the authors who produce scientific information usually do not get any monetary rewards in the form of sales royalties. Hence from the content providers' (the authors' and reviewers') viewpoint there is no problem with potential piracy; on the contrary as wide a dissemination of the articles as possible is desirable.

The label "open access" was coined at a meeting in Budapest in 2002, but the concept has existed since the earliest days of the World Wide Web. As soon as the web emerged, many scientists rapidly saw its potential for making the sharing of scientific information, in particular peer-reviewed journal articles, more efficient. The first successful OA implementation, the arXiv repository for preprints in high-energy physics and related fields (<http://arxiv.org>), now housing more than 600,000 "papers," was in fact started already in 1991. The earliest peer-reviewed OA journals started to appear via list servers around the same time, and early fully web-based journals include some still-existing publications such as *First Monday* (<http://firstmonday.org>), which has published research concerning the Internet as a phenomenon since 1996.

During the 1990s OA emerged in the form of small-scale voluntary efforts

by individual scholars or groups of scholars, but during the first decade of this millennium several innovative publishers have entered the field. This development has in particular resulted in the creation of OA journals funded by author charges. At the same time universities and research organizations have increasingly established institutional repositories, which often use open source software like D-Space and E-prints and which can be utilized to archive author manuscripts of published articles among other material. Research funders and universities have also started to require that researchers who are funded or employed by them store manuscript copies in repositories, for instance in PubMedCentral. Mandates by public research funders such as the National Institutes of Health (NIH) in the United States have aroused a lot of debate.

In the rest of this article we concentrate on OA to peer-reviewed journal articles. The case for OA is equally strong for conference papers, working papers, theses and research reports as well as datasets, but I will not discuss these formats here.

### Open Access Journals (Gold OA)

The desired end result of the OA movement, free unrestricted access to the full texts of peer-reviewed journal articles, can be achieved in two major ways. First, the scientific journals themselves can start using business models that enable the end product (at least in its electronic format) to be made free for readers. These models can be made possible, on the one hand, by basing the journal publishing on open-source solutions and using volunteers to minimize costs, but also through funding the publication through article charges, advertising, grants and similar sources. Under this model the electronic version of the article is available directly from the publisher's site. In the jargon of the OA community this model is called the "gold route to OA." [1]

There are a number of variations on the gold route. The most straightforward one is where all the content of the journal is available free from day one. These journals are usually indexed in the Directory of Open Access Journals, currently listing some 7000 journals.

For an additional 4000 scientific journals, representing around a quarter

of the total journal portfolio of the 12 biggest scientific journal publishers, the author can open up his article in an otherwise subscription-based journal by paying the publisher an extra OA charge, typically around \$3000 (US). This option is usually referred to as "hybrid OA." So far the uptake by authors of this option has been very low: only around 2% of authors of eligible articles are willing to pay the extra cost [2].

OA journal publishing has been of particular interest to society journals in countries outside the United States and the United Kingdom, which often publish in national languages. In many countries society journals are subsidized by the government in view of their important social function and lack of a commercial subscription market. In other countries such journals receive implicit support via national OA journal portals they can use for no cost such as Scielo (Latin America), J-Stage (Japan) and Hrcak (Croatia).

### Repositories for Manuscript Copies (Green OA)

Authors can also publish their articles in traditional subscription-based journals, but can after the publishing (or even before) upload their manuscripts to internet servers where potential readers without subscription may access them freely. This approach is often called "green OA." [1]

The three major options for such posting are on the homepages of the author or his department, in an institutional repository of the author's employer or in a subject-specific repository. Homepages were the dominant choice in the early days but suffer from a high risk of the material disappearing as authors change jobs, web page structures are reshuffled or other changes occur. Subject-based global repositories have become the primary choice in some disciplines (arXiv for physics, PubMedCentral for biomedicine) but are not available for many fields. The number of institutional repositories is quickly growing and exceeds 1000 [3, 4].

Contrary to common belief among academics a majority of scientific journals explicitly allow green OA in their copyright agreements, but usually not for the exact published versions. Some journals nowadays allow the posting of the author's personal version in a subject-specific repository like PubMedCentral after a delay of, say, six months, in compliance with the requirement of a major research funder like NIH.

### Authors Will Decide If OA Becomes Popular

The authors of manuscripts for peer-reviewed journal articles are in a key position to determine how popular OA becomes. The choice an author faces is three-fold. First, he must decide where to submit his article. If the journal is an OA journal the process ends there. If the journal is not OA but offers hybrid OA, he must next decide whether to pay the charge to “free” the article. If none of these choices has been taken, then for the remaining majority of the articles for which green OA is allowed he must decide whether he is willing to make the extra effort to put a copy of the manuscript in a repository. Slightly different decision-making criteria influence these decisions in which OA is an option.

Due to the rather low uptake of voluntary uploading of manuscripts to repositories, there is an increasing trend for research funders and universities to require green posting from authors receiving funding from/working for them [5, 6]. The best-known funders mandates are those from the NIH and the Wellcome Trust (UK). There are currently more than 100 universities that have also issued mandates of some sort [6]. A recent study has shown that the uptake level for articles indexed by the Institute for Scientific Information (ISI) in the repositories of universities that instituted mandates a few years ago is around 60%, compared to an uptake of around 15% in universities without a mandate [7].

### The Global OA Situation

Our research group studied the OA availability of peer-reviewed journal articles published in 2008 using a sample stratified over research disciplines. The method for checking for OA articles or green copies relied on using a general purpose search engine and on our researchers determining if a full-text OA version was found and on classifying it. The methods and the more detailed results, as well as comparison to earlier and related studies, are described in [8]. Of the articles, 8.5% were freely available in OA journals at the publishers' sites. For an additional 11.9% free manuscript versions could be found using search engines, making the overall OA percentage 20.4%.

It is interesting to note the clear differences between the disciplines. Chemistry (13%) had the lowest overall share of OA; earth sciences (33%),

the highest. In medicine, biochemistry and chemistry publishing in OA journals was more common. In all other fields author-posted manuscript copies dominated the picture. The reasons for these differences are likely to be found in variations in the supply of suitable OA journals and repositories, in the funding available for paying publications charges and in the way the publishing cultures have been shaped in each field.

### The OA Situation for Information Systems Scholarship

In the above study the discipline breakdown was very broad, but it is interesting to consider specifically whether researchers in our own field are adopting the new opportunities offered by OA. With this question in mind, we prepared and Googled a sample of information systems (IS) articles from 2009 using the same methods as in our earlier global OA-barometer study and with the same time delay of searching on average one and half years after publication.

The list was stratified in the sense that it consisted of two separate groups. The first was a small group of eight high-impact journals, which a group of leading editors of IS journals have identified as the core journals of the discipline [9]. Checking for the OA availability of manuscript copies of articles from these journals is of particular interest, since their share of references to the whole body of IS journals is very high.

The second much bigger group of journals was compiled using a number of different sources. The basis was the journals ranking pages of the ISworld portal [10], which lists 125 journals, including the eight top journals above. Many of the journals on this list are journals in which IS scholars might publish but which do not have a clear IS focus, dealing with areas such as management, operations research or artificial intelligence, and these journals were dropped from the sample. Two necessary criteria, in addition to subject, were that the journals must be indexed in Scopus and that they had published articles in 2009. These criteria allowed us to extract the article metadata for the sample easily. Additional journals were excluded because they failed to meet these requirements.

We thus ended with a stratified sample of IS articles from 44 journals. For the top eight journals we included all 258 articles in the sample that was

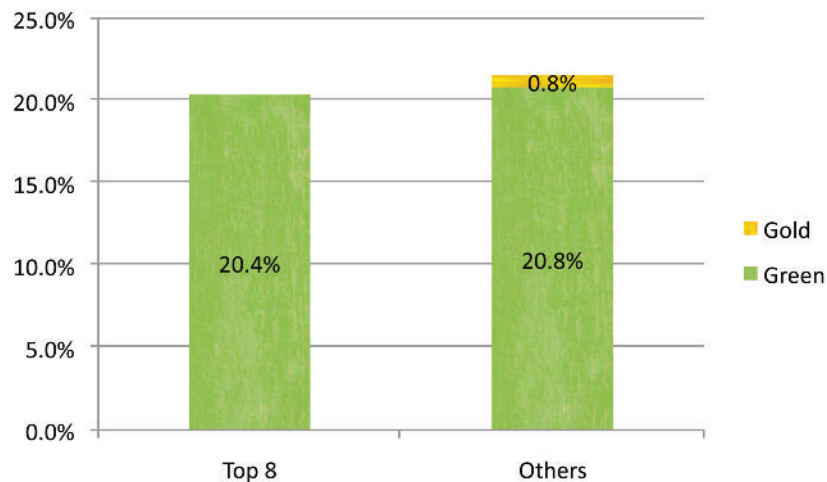
Googled. For the other journals we made a sample of 540 articles from the 1540 published, since the number of articles to Google would otherwise have become too big.

Of all the journals thus identified none is full OA. (The *IBM Systems Journal* would have been on the list but it was merged with another IBM journal beginning in 2009 and ceased to be OA.) We also studied the copyright restrictions of the journals included. On the whole the information shows that for most IS journals authors would be able to post the manuscripts openly on the web.

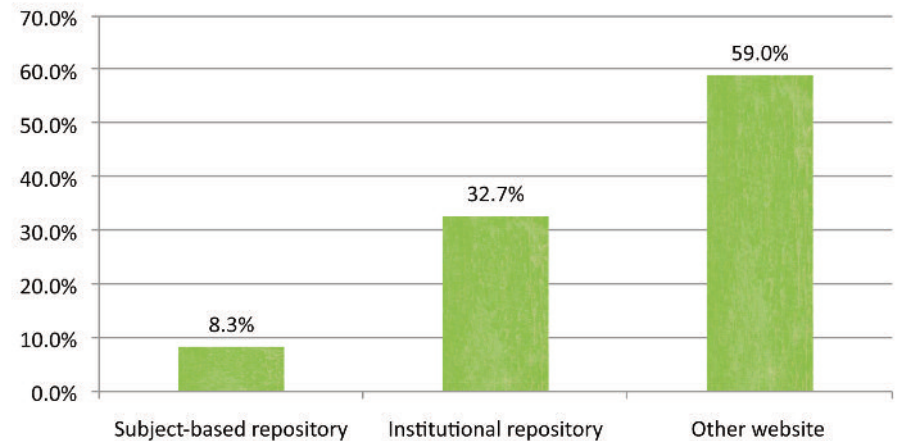
The overall results showed an OA share of 21.3%, split into 0.6% on the journals websites (gold) and 20.7% on personal websites or in repositories (green). Figure 1 shows the corresponding shares separately for the top eight journals and the other journals. Interestingly there was no difference between the high impact journals and the others in contrast with our global study where we found a green percentage of 14% for ISI-indexed journals and 5% for non-ISI (but Scopus) journals.

Thus, overall, the OA share is very close to the global OA share in our

**FIGURE 1.** OA availability of IS articles published in 2009. Left, the top eight journals; right, the remaining 36 journals.



**FIGURE 2.** The split of the green IS copies found according to where they were stored.



earlier study. The results are, however, slightly lower than the result of 27% reported recently by Way in a similar study for the top 20 journals in library and information science [11]. What is very visible is the very low share of gold OA articles, at least among Scopus-indexed journals. There were no purely OA journals in our sample, but we found a few articles that were delayed OA.

The split into types of copies uploaded (38% exact copies, 46% personal versions and 16% preprint versions) is very close to the corresponding global results. Quite a few of the exact copies, usually to be found on the authors' own web pages, were clearly in violation of the copyright rules.

An interesting result (Figure 2) was that only a minority of the copies was found in systematically created subject-based or institutional repositories, which would be the preferred long-term archival solution. In this respect the results differ markedly from the global results, which had 43% in subject-based repositories, 24% in institutional repositories and 33% on other web sites.

There are, of course, several IS journals outside the coverage of Scopus, some of which are in fact openly available. The OA ones can be identified via a manual search in the *Directory of Open Access Journals* ([www.doaj.org](http://www.doaj.org)).

The total number of articles published in 10 such journals was 159. Most of those listed are rather new journals with low article volumes.

If we had been able to also identify all subscription-based IS journals outside the ones indexed in Scopus we could meaningfully have included the above OA journals in the analysis to get a total share of gold OA. This step was, however, not possible for practical purposes, since it would be extremely tedious to identify relevant journals and to obtain samples of articles from them. We can briefly note that if the above OA journals had been directly added to the material (without including non-OA, IS journals outside Scopus) their share of all articles would have been 7.7%.

### Conclusions

**OA as a research topic.** OA is a phenomenon that can be studied through the lenses of several scientific disciplines. It fits naturally into information (library) science, but can also be studied from an e-commerce perspective or from the perspective of information systems, and it has aspects that touch upon science policy and the sociology of science.

One aspect that merits more attention is the publishing process. Scientific publishing is a global, highly integrated information system, which, over the years, has evolved as a result of IT-development by dozens of different stakeholders, rather than as the result of cohesive design [12, 13]. The integrating glue of this system is the bibliographic metadata of a publication. Many of the technical shortcomings of the current system are due to this incremental evolution path. One such shortcoming is, for instance, the lack of an international standardized ID for authors. The system is nowadays used for two purposes, a primary purpose of facilitating communication of research results and a secondary one of facilitating the decision making of universities,

research funders and similar entities that need objective performance measures of researchers, research groups, universities and so forth. The aim of the OA movement is to enhance the communication function of this overall system. Paradoxically the strong reliance of the decision support function on data produced by one stakeholder, ISI's *Journal Citation Report*, contributes to creating a strong inertia in the system by making it difficult for new publishers to enter the market on equal terms and hence stifling innovation.

### What can scholars do to increase the open availability of their articles?

The easiest and quickest route to quickly increase global availability is the green route. There is no lack of suitable repositories for the green approach, and for the majority of journals where we publish copyright is not a barrier.

The main barrier for achieving improved availability is, in fact, the lack of *awareness* that only a modest effort on our part could help substantially and that these efforts are in our own best interests. It's really not that difficult. The first author of this article has recently uploaded almost all his journal articles to the repository of his university and also linked them to his publication list. For around 30 articles published from 1989 to 2010 this task required about one week's work, including retrieving and reformatting manuscripts, checking the copyright issues and uploading the metadata and files.

OA publishing is currently growing very fast, on average 30% per year over the last decade [14]. Several established publishers have recently announced new OA journals following in the footsteps of the very successful *PLoS ONE* ([www.plosone.org](http://www.plosone.org)), which is likely to publish more than 14,000 articles in 2011. It will be very interesting to see in the next few years if the effects of OA on scholarly publishing will be as large as *Wikipedia* has been for encyclopedias and Skype for long-distance telephony. ■

*Resources on next page*

### Resources Mentioned in the Article

- [1] Harnad, S., Brody, T., Vallières, F., Carr, L., Hitchcock, S., Gingras, Y., et al. (2004). The access/impact problem and the green and gold roads to open access. *Serials Review*, 30(4), 310-314.
- [2] Dallmeier-Tiessen, S., Goerner, B., Darby, R., Hyppoelae, J., Igo-Kemenes, P., Kahn, D., et al. (July 8, 2010). *Open access publishing - models and attributes: Study of Open Access Publishing*. Max Planck Society Digital Library. Retrieved April 12, 2012, from <http://edoc.mpg.de/478647>
- [3] *Registry of Open Access Repositories (ROAR)*: <http://roar.eprints.org/>
- [4] *Directory of Open Access Repositories (OpenDOAR)*: [www.opendoar.org/](http://www.opendoar.org/)
- [5] *JULIET: Research funders' open access policies*: [www.sherpa.ac.uk/juliet/](http://www.sherpa.ac.uk/juliet/)
- [6] *Registry of Open Access Repository Material Archiving Policies (ROARMAP)*: [www.eprints.org/openaccess/policysignup/](http://www.eprints.org/openaccess/policysignup/)
- [7] Gargouri, Y., Hajjem, C., Larivière, V., Gingras, Y., Carr, L., Brody, T., et al. (2010). Self-selected or mandated, open access increases citation impact for higher quality research. *PLoS ONE* 5(10). Retrieved April 12, 2012, from [www.plosone.org/article/info:doi/10.1371/journal.pone.0013636](http://www.plosone.org/article/info:doi/10.1371/journal.pone.0013636)
- [8] Björk, B-C., Welling P., Laakso, M., Majlender P., Hedlund T. & Gudnasson, G., (2010). Open access to the scientific journal literature: Situation 2009. *PLoS ONE* 5(6). Retrieved April 12, 2012, from <http://dx.doi.org/10.1371/journal.pone.0011273>
- [9] Association for Information Systems. (August 2011). Senior scholars' basket of journals. Retrieved April 12, 2012, from <http://home.aisnet.org/displaycommon.cfm?an=1&subarticlenbr=346>
- [10] Association for Information Systems. (n.d.). MIS journal rankings. Retrieved April 12, 2012, from <http://ais.affiniscape.com/displaycommon.cfm?an=1&subarticlenbr=432>
- [11] Way, D. (July 2010). The open access availability of library and information science literature. *College & Research Libraries*, 71(4), 302-30.
- [12] Hurd, J. (2000). The transformation of scientific communication: A model for 2020. *Journal of the American Society for Information Science*, 51(14), 1279-1283.
- [13] Björk, B-C. (2007, January). A model of scientific communication as a global distributed information system. *Information Research*, 12(2), paper 307. Retrieved April 12, 2012, from <http://informationi.net/ir/12-2/paper307.html>
- [14] Laakso M, Welling P, Bukvova H, Nyman L., Björk B-C, & Hedlund T. (2011). The development of open access journal publishing from 1993 to 2009. *PLoS ONE* 6(6): e20961. Retrieved April 12, 2012, from <http://dx.doi.org/10.1371/journal.pone.0020961>