View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Helsingin yliopiston digitaalinen arkisto

sciecominfo
Nordic-Baltic Forum for Scientific Communication

THE CURRENT STATE OF OPEN ACCESS TO RESEARCH ARTICLES FROM THE UNIVERSITY OF HELSINKI

Kimmo Koskinen, Arja Lappalainen, Timo Liimatainen, Eija Nevalainen, Arja Niskala and Pekka J. Salminen

Introduction

Open access means the immediate, free and permanent access to the complete text of a scientific publication in internet. One of the first statements in support of open access was the Budapest Open Access Initiative (BOAI). The Berlin declaration on Open Access to Knowledge in the Sciences and Humanities, signed by a number of important research institutes and research councils, has gained much attention.

Open access is important for researchers, institutions and research funders because it facilitates the widest possible visibility and impact for research. There are 44 mandates of international funders (March 2010) requiring open access for the research results they fund, including The National Institutes of Health (USA), The Wellcome Trust (UK), European Research Council, CERN, ICRISAT and some national research funders in Norway, Canada and Australia. The SHERPA/JULIET service has information on funders' policies for open access.

The international scientific community promotes open access to research information in order to ensure the availability of publicly funded research. Research results can be published in open access journals, or results published elsewhere can be subsequently deposited into open digital archives. The majority of international publishers allow the posting of some versions of published articles, sometimes after a delay, so-called embargo, into such repositories. Publishers' copyright policies are listed in the SHERPA/RoMEO service

An increasing number of universities maintain open repositories and require researchers to deposit their research articles as a part of their strategic goal. In March 2010, there are 90 institutional mandates worldwide, e.g. Harvard Faculty of Arts and Sciences, Harvard Law School and MIT. Several universities in Australia, Canada, Great Britain and Germany have already adopted these policies.

The University of Helsinki advocates open access publishing and internet visibility of research. The university has had open access archives for a few years now, but the number of self-archived articles actually deposited to them has remained low (Ilva 2009). As from 1 January 2010, the university requires its researchers to deposit a copy of their scientific peer-reviewed journal articles in HELDA, the open digital archive maintained by the University (University of Helsinki 2010) or in some other open access repositories, or that they publish their articles in an open access journal.

We were interested in the web visibility of the publications of the University of Helsinki prior to the introduction of the mandate. Another purpose of the study was to obtain information for future reference and comparison. The central question in hand was: how many of the research articles can be reached as open access full text? We also studied how many of them were available in electronic form inside the University of Helsinki network. In addition, we made some comparisons between web search engines and metadata harvesters.

Previous studies

There have been plenty of arguments in favour of the open access publishing of research articles and self-archiving their e-versions to a subject repository or depositing them in an institutional repository.
Open access articles have been said to have more visibility in scholarly communication and to get more citations. For example, Norris et al. (2008b) find that open access articles do have a citational advantage. They also point out that the causes to this are not clear.

The direct effect of open access to the amount of citations and impact factors has also been called into question (Craig et al. 2007, Moed 2007). Craig et al. state that the citation differences depend on the quality and the importance of the article. How the article is retrieved is not important. The benefits of self-archiving (open access) are quite uncertain, according to this study.

Björk et al. (2009) studied the annual volume and open access availability of scientific journal publishing. They estimated that after one year 11.3% of the scientific output in 2006 could be found in subjectspecific or institutional repositories or on the home pages of the authors. The overall share of the open access articles was 19.4 % of the annual output. In Björk et al. (2010) it was calculated that the share was 20.4 % of peer reviewed open access articles published in 2008.

Bhat (2009) studied the visibility of publications deposited in open access repositories in computer science and information technology. The visibility of repositories in search engines and data discovery tools ranged from 4% to 92%. The OAI-PMH (Open Archive Initiative Protocol for Metadata Harvesting) compliance enhanced the visibility of the repositories considerably through multiple search engines. Google and MSN retrieved the highest number of documents from the repositories and Gigablast the least.

The article by Norris et al. (2008a) gave us an idea of an empirical study of the open access visibility of research articles published by the researchers of the University of Helsinki. The study showed that 38% of articles in a random sample of 2519 articles in the fields of ecology, economics and sociology had open access versions on the world wide web. Google and Google Scholar found 76% of them. The conclusion was that authors seem to prefer to self-archive their work on personal or departmental web pages that are not reached by metadata harvesters such as OAIster and OpenDOAR. The approach of Norris was chosen as a starting point to study the current state of open access to research articles from the University of Helsinki.

METHODOLOGY

Sample data

We selected peer reviewed articles from the years 2007 to 2008 to be exported from the University of Helsinki publication database JULKI as a sample. JULKI contains information about research project publications and other material published by university staff members. At the time of the data import, JULKI contained a total of 7771 articles from that period. To minimize the effect of certain publishers' embargo rules, articles from year 2009 were excluded from the sample.

A random sample of 407 article references (5.1%) from the original data of 7771 references was chosen for further analysis. The sample included bibliographic information of the research articles from the University of Helsinki. Only peer reviewed journal articles were included in our sample data. This sample gives an overview of the research performed in the University of Helsinki. We randomized the original data and divided it into random groups which were tested separately by six members of the study group.

Searching

Five internet search tools were used to determine the open web accessibility of the sample articles. Two commonly used search engines and two well established open access metadata harvesters were chosen, as well as the University of Helsinki open digital repository HELDA.

The leading web search engine Google was a natural choice, as well as Google Scholar, a search engine for scholarly literature. Two large metadata harvesting services were also chosen: OpenDOAR, an authoritative directory of academic open access repositories, and Scientific Commons, a project of the University of St.Gallen aiming to provide the most comprehensive and freely available access to scientific knowledge on the internet. Also, the University of Helsinki open digital repository HELDA, introduced in 2009, was included in order to get an idea of the usage level of this new service.

The five search tools were used to locate the articles. The searches were performed both inside of the University of Helsinki network and outside of it. These searches were performed at the turn of the years 2009 and 2010. The URLs of the fulltexts found by the searches were saved, as well as some remarks, when necessary. Also, codes by which JULKI describes the departments of the authors were saved to facilitate a later comparison of the open access activity of the faculties of the university.

The searches were executed in the same way with each search tool. A phrase including the whole title of the article was used as a search key. Possible sources of errors, such as transliterations and special characters, were taken into account: for example, the title was divided into parts when necessary. The search results were browsed only using the first result page, and promising links were followed using no more than three clicks. A similar search technique was used by Baldwin (2009) who used Google Scholar to search for online availability of articles.

The search protocol was defined beforehand, but some variation was inevitable in the search practices as six inviduals conducted the searches. For example, the three click rule was sometimes extented into four clicks. We consider that such variations simulate the behaviour of a typical user of search tools and can thereby be accepted.

In this study, we used a simple criterion for open access availability: did the full text version of the article open? Openly available versions of research articles can be found from publisher sites, digital repositories, web pages of research groups, or personal home pages. However, in our study we did not make any further analysis of these different sources of open access.

We computed the amount of articles that were found as full text versions. The results of each five search tools were examined separately. Some statistics concerning the faculties of the authors of articles were also compiled. All these calculations were made both for the searches performed inside the University of Helsinki network and for the searches performed outside of it.

RESULTS AND DISCUSSION

Open availability of research articles

Openly available versions of research articles were found in 49.1% of the sample data. This collective result means that the articles were found by at least one of the search tools studied (Figure 1).

The result differs from the findings of Björk et al. (2009) who found a total open access availability of 19.4% of the worldwide annual scientific output. One possible explanation for this relatively high level of open access in our sample might be the current awareness of open access issues among researchers at the University of Helsinki. Open access has been discussed actively in Finland since 2003, and the university has promoted open access by establishing open digital repositories. Another explanation could be the fact that many research groups include authors from other universities and institutes where open access practices are well established.

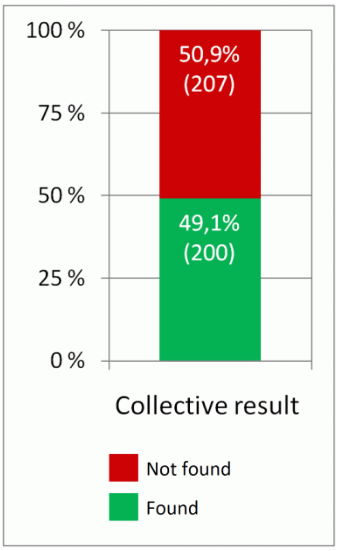


Figure 1: Collective Open Access availability.

Differences between the search tools

The percentage of open access articles found for each search tool was as follows: Google 42.5%, Google Scholar 38.1%, OpenDOAR 14.3% and Scientific Commons 15.7%. (Figure 2). Same searches made in the university network gave the following results: Google 84.3%, Google Scholar 77.4%, OpenDOAR 18.9% and Scientific Commons 18.2%. Also, the HELDA repository was included in the study, but its rate of success (0.7%) was very low. Because HELDA was founded only recently, it does not contain much material as of yet and was therefore excluded from the final figures.

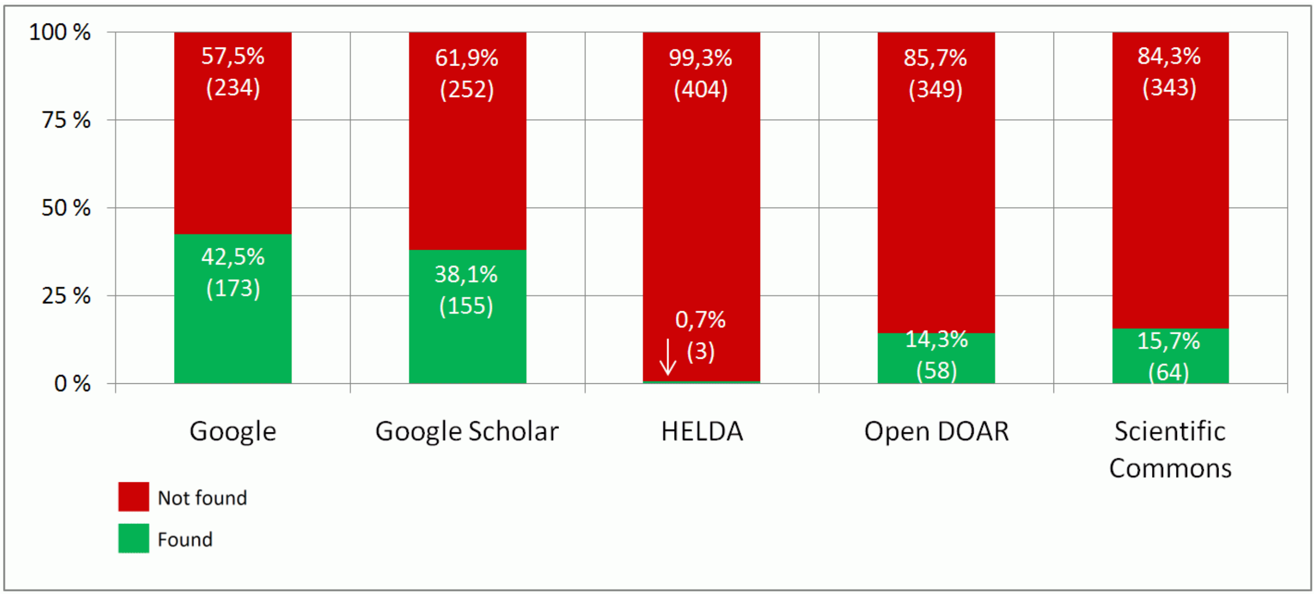


Figure 2: Open Access availability by search tools.

Even though we especially searched for scholarly articles, the best results were retrieved by Google, not by Google Scholar. The latter indexes scholarly articles collected from various sites, but the use of its web crawler must be allowed as stated in Google Scholar's publisher policy. Therefore, some cooperation from publishers is required, and this might constitute a threshold for some. Jacsó (2008) points out that more could be expected from Google Scholar's search operations and finding full text because Google Scholar has now obtained permission to index primary documents from some main publishers.

However, when compared to metadata harvesters, Google Scholar does a good job: according to Baldwin (2009), it is because it indexes most of the sources through which online access to full text is available, such as publisher web sites, PubMed Central, institutional repositories and preprint archives. Jacsó (2008) points in his article that only 25% open access PubMed papers are available directly in Google Scholar and that it could cover much more full text repositories.

According to Markland (2006), Google and Google Scholar can especially find articles from repositories when the title is fully known. The result for keyword searches is not as good. Google and Google Scholar differ in giving results even when the same search terms are used. Collectively, Norris et al. (2008a) found 76.8% of the articles. In our study, we used title as a search term and the collective success rate was 47.4%.

Norris et al. (2008a) suppose that the reason of the inability of metadata harvesters to find as many articles as Google and Google Scholar is that most authors prefer to self-archive their work to their personal or departmental web pages. Other explanatory factors might be technical. Repositories still have interoperability problems despite the progress of OAI-PMH (Bhat 2009) and there are problems of illiteracy in the software of Google Scholar (Jacsó 2008).

Metadata harvesters OpenDoar and Scientific Commons differ in their data retrieving practices. Scientific Commons uses the Open Archive Initiative Protocol for Metadata Harvesting (OAI-PMH) to retrieve data from repositories. Repositories which support the OAI-PMH Protocol can add their content to the Scientific Commons manually, according to the Scientific Commons policy. Scientific Commons indexes metadata and full-text documents.

An interesting link between Scientific Commons and OpenDOAR can be found in the background for metadata harvesting procedures: instead of searching the repositories by itself, OpenDOAR uses Google's custom search engine to search articles held in repositories, and Google and Google Scholar search and index data from OAI-PMH compliant repositories (Norris et al. 2008a). The OpenDOAR staff inspects the compliance of every repository before starting harvesting procedures and assigning metadata.

OpenDOAR states on its homepage that it does not include repositories having any kind of access control that would prevent immediate access. Scientific Commons does not say anything about access control restrictions. The lack of interoperability between open access repositories might also give some explanation to these results. OAI-PMH compliance aids the visibility of a repository (Bhat 2009).

It is likely that Google and Google Scholar give better results than OpenDOAR and Scientific Commons because of the differences in their objectives and software. Therefore, we assume that these differences have influenced our results. However, we did not analyze this effect in any detail.

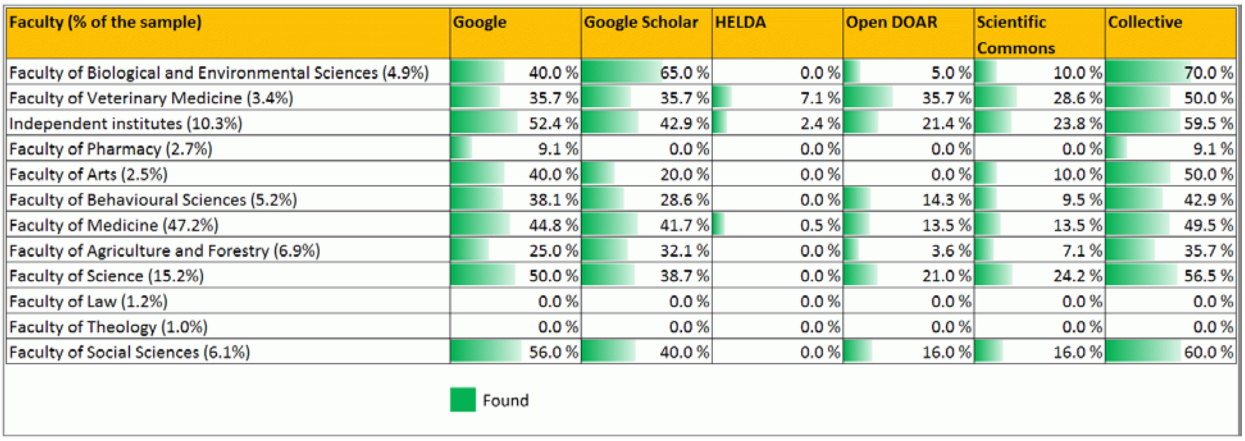
Other observations

Our main focus was to study the open access visibility of articles published by the researchers of the University of Helsinki. As a by-product, the study extended to ponder on the differences between faculties and the effect of the language used in the article. We also found that the availability of full text articles was much higher inside the university network.

Disciplinary differences in publication practices certainly affect the availability of articles. For example, since the mid-1980s, the data on research publications has been collected every 3 to 4 years for the staff in British universities (Meadows 1998). Science, technology and medicine prefer journal articles, whereas the social sciences and humanities favour books. Engineers are in favour of refereed conference articles. For example, some fields of research were not fully represented in our sample because mostly monographs are published in these fields. In the case of our study, it is also possible that some institutions of the University of Helsinki may not have been reporting all their scholarly publications to the JULKI database. We assume that our sample gives a relatively precise approximation of the publishing intensity of peer reviewed articles in different disciplines.

All the faculties of the University of Helsinki were included in our sample data (Figure 3). The Faculty of Medicine was clearly the leader as far as the volume of the publications is concerned. This faculty had authors in 44.2% of the publications in our sample. The Faculty of Science was also well represented with its proportion of 15.2% of the publications.

In our results, most of the sample in the faculty level equated to the average results: about 50% of the articles were found. However, some exceptions are worth mentioning. Only 9.1% of the articles written at the Faculty of Pharmacy were found, whereas 70.0% of the articles of the Faculty of Biosciences were located. These exceptions reflect the different publishing practices between disciplines (compare Björk et al. 2010).



Faculty (% of the sample)	Google	Google Scholar	HELDA	Open DOAR	Scientific Commons	Collective
Faculty of Biological and Environmental Sciences (4.9%)	40.0 %	65.0 %	0.0 %	5.0 %	10.0 %	70.0 %
Faculty of Veterinary Medicine (3.4%)	35.7 %	35.7 %	7.1 %	35.7 %	28.6 %	50.0 %
Independent institutes (10.3%)	52.4 %	42.9 %	2.4 %	21.4 %	23.8 %	59.5 %
Faculty of Pharmacy (2.7%)	9.1 %	0.0 %	0.0 %	0.0 %	0.0 %	9.1 %
Faculty of Arts (2.5%)	40.0 %	20.0 %	0.0 %	0.0 %	10.0 %	50.0 %
Faculty of Behavioural Sciences (5.2%)	38.1 %	28.6 %	0.0 %	14.3 %	9.5 %	42.9 %
Faculty of Medicine (47.2%)	44.8 %	41.7 %	0.5 %	13.5 %	13.5 %	49.5 %
Faculty of Agriculture and Forestry (6.9%)	25.0 %	32.1 %	0.0 %	3.6 %	7.1 %	35.7 %
Faculty of Science (15.2%)	50.0 %	38.7 %	0.0 %	21.0 %	24.2 %	56.5 %
Faculty of Law (1.2%)	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %
Faculty of Theology (1.0%)	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %
Faculty of Social Sciences (6.1%)	56.0 %	40.0 %	0.0 %	16.0 %	16.0 %	60.0 %

🟩 Found

Figure 3: Open Access availability by faculty.
*The sum is more than 100% because of the co-authored articles across faculties. (Figure 3)

50.9% of the sample material was not found by any of the search tools used. We found out that the language of the original article was a significant factor for this result. Of all the articles in English, 48.6% could not be found, whereas 70.5% of the articles in Finnish belonged to this category.

We assume that this significant difference in availability between articles in English and Finnish has several reasons. First of all, Finnish publishers are usually very small and have not yet embraced open access policies. It is common knowledge that the bigger publishers have already taken their stand on open access. Secondly, though some important international publishers are not yet supporting open access, most of them have already defined their policy. We believe that the researchers of the University of Helsinki prefer their articles to be published in these established journals. This may increase the open access availability. Thirdly, it is known that Google Scholar has started the cooperation with bigger publishers first (Jacsó 2008).

Open access articles are universally available, but articles from commercial publishers are licenced so that they can only be accessed within the subscriber's network. In this study, we made identical searches both inside and outside of the University of Helsinki network. The availability of full text articles was significantly better inside the University network (88.0%) than outside (49.1%) (Figure 4). We assume that the situation is similar in most research universities due to licensing of electronic content. However, 12.0% of the articles produced in the University of Helsinki are not available as full text even inside the university network.

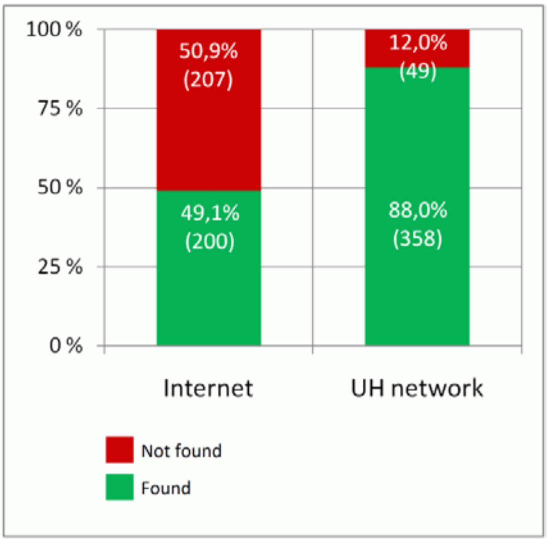


Figure 4: Full text availability outside and inside of the University of Helsinki (UH) network.

Conclusions

Google and Google Scholar dominated the results by a significant margin. However, a small portion of the full text articles were found only when the open access metadata harvesters were used. For example, a total of 16 articles that neither of the Google search engines could not find were accessed easily by OpenDOAR. This means that 27.6% of all the articles found by OpenDOAR were not found by the Google search engines. (Figure 5). As for the results for Scientific

Commons, 16.4% of the articles were inaccessible via the Google search engines.

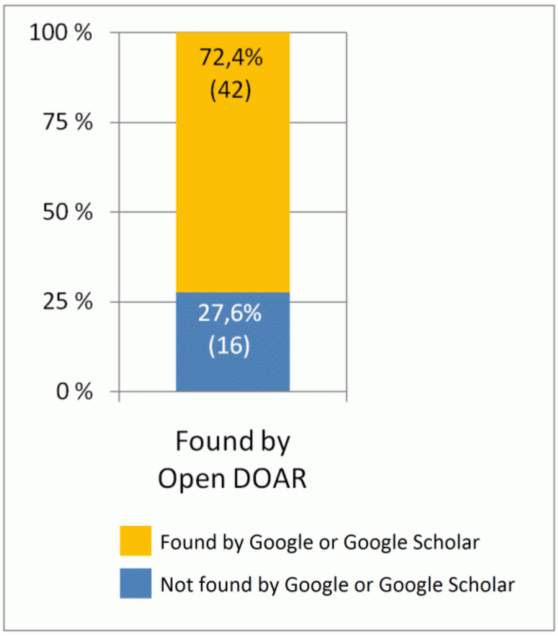


Figure 5: Articles (58) found by OpenDOAR compared with Google and Google Scholar.

Although Google and Google Scholar are currently the best tools for finding online full texts of scientific open access articles, our results suggest that the open access metadata harvesters should not be disregarded. Similar results were observed by Jamali and Asadi (2010), so the importance of services like OpenDOAR and Scientific Commons must not be understated. There are open access repositories that Google or Google Scholar cannot reach, and therefore it is advisable to use open access directories besides Google in order to obtain comprehensive search results.

We predict that the University of Helsinki open access mandate, recommending that the researches of the university publish their articles as open access starting from year 2010, will affect the publishing and depositing practices, but further studies are needed in the near future. A possible point in time for a follow-up study could be after a few years, when HELDA (University of Helsinki DigitalArchive) has come of age.

Currently, a half of the research articles of the University of Helsinki are openly available. Depending on the perspective, a glass of wine can be either half full or half empty at the same time.

References

Baldwin, V. A. 2009. Using Google Scholar to Search for Online Availability of a Cited Article in Engineering Disciplines. Issues in Science and Technology Librarianship, (56). http://www.istl.org/09-winter/article1.html

Bhat, M. H. 2009.Interoperability of open access repositories in computer science and IT – an evaluation. Library Hi Tech, 28 (1). http://www.emeraldinsight.com/10.1108/07378831011026724

Björk, B.-C., Roos, A. and Lauri, M. 2009. Scientific journal publishing: yearly volume and open access availability. Information Research, 14 (1). http://informationr.net/ir/14-1/paper391.html

Björk, B.-C. et al. 2010. Open Access to the Scientific Journal Literature: Situation 2009. PLoS ONE 5(6). http://www.plosone.org/article/info:doi/10.1371/journal.pone.0011273

Iain D. Craig et al. 2007. Do open access articles have greater citation impact?: A critical review of the literature. Journal of Informetrics, 1 (3). http://dx.doi.org/10.1016/j.joi.2007.04.001

Ilva, J. 2009. Building a Repository Infrastructure for Finland. ScieCom Info, 5 (3). http://www.sciecom.org/ojs/index.php/sciecominfo/article/viewFile/1763/1392

Jacsó, P. 2008. Google Scholar revisited (Savvy searching). Online Information Review, 32 (1). http://www.emeraldinsight.com/10.1108/14684520810866010

Jamali, H. R., Asadi, S. 2010. Google and the scholar: the role of Google in scientists' information-seeking behaviour. Online Information Review, 34 (2).http://www.emeraldinsight.com/10.1108/14684521011036990

Markland, M. 2006. Institutional repositories in the UK: What can the Google user find there? Journal of Librarianship and Information Science, 38 (4). http://lis.sagepub.com/cgi/reprint/38/4/221

Meadows, A. J. 1998. Communicating research. San Diego, Academic Press. ISBN: 0124874150.

Moed, H.F. (2007). The effect of "Open Access" upon citation impact: An analysis of ArXiv's Condensed Matter Section. Journal of the American Society for Information Science and Technology, 58(13).http://arxiv.org/pdf/cs/0611060v1

Norris, M., Oppenheim, C. and Rowland, F. 2008a. Finding open access articles using Google, Google Scholar, OAIster and OpenDOAR. Online Information Review 32 (6). http://hdl.handle.net/2134/4084

Norris, M., Oppenheim, C. and Rowland, F. 2008b. The citation advantage of open-access articles. Journal of the American Society for Information Science and Technology, 59(12). http://onlinelibrary.wiley.com/doi/10.1002/asi.20898/full

About the authors

The authors, Kimmo Koskinen, Arja Lappalainen, Timo Liimatainen, Eija Nevalainen, Arja Niskala and Pekka J. Salminen work in different tasks in the Helsinki University Library, Finland, and together have deepened their knowledge of open access in a self-directing study group. It started working in 2007 as one of the groups of the mentorship project. In the study group the participants have been able to wonder, discuss enthusiastically and learn more about openaccess developments and their applications at the University of Helsinki.