sciecominfo
Nordic-Baltic Forum for Scientific Communication

THE SHARE OF OPEN ACCESS IN SWEDEN 2011 – ANALYZING THE OA OUTCOME FROM SWEDISH UNIVERSITIES

Margareta Fathli, Tomas Lundén, Peter Sjögårde

Introduction

During 2013 we performed a study on the state of Open Access (OA) in Sweden. The project was funded by the National Library of Sweden and its wider purpose was to produce a picture of the state of OA in Sweden today, to facilitate well-grounded decisions on how the share of OA can be increased.

The specific purpose of the project was to measure the share of OA at Swedish universities. What share of refereed research articles from Swedish universities was published OA in the year 2011? We studied the total OA uptake, both green and gold, for Sweden as a whole and per university. The year 2011 was chosen, so that embargoed articles would have been opened up. We also examined potential green OA, i.e. to what extent the articles could have been self-archived according to information in SHERPA/RoMEO. A further aim of our study was to develop a method that could be used for continuous measurement of OA in Sweden.

Definitions

The first definition on OA was set in the Budapest Open Access Initiative (BOAI) in 2002. This was when the gold and green roads were outlined (although not named as such). This was followed by the equally influential Berlin Declaration in 2003.[1] Peter Suber has supplied the following summarized definition: "Open access (OA) literature is digital, online, free of charge, and free of most copyright andlicensing restrictions."[2]

Since then a number of varieties of OA have been named, such as *hybrid* and *delayed* for example. In our study we use the following definitions:

Green OA – a copy of an otherwise published article that is deposited and openly available in an institutional or subject based repository. It could be the submitted, accepted or published version of the article.
Gold OA – an article is openly available, immediately upon publication, on a publisher's website. All articles in the journal must be OA.
Hybrid OA – articles that for a fee are made openly available on a publisher's website, but the journal itself maintains the traditional subscription way of publishing.
Delayed OA – a traditional subscription based publishing of an article, that is, after a certain time embargo, openly available on the publisher's website. The embargo time may vary from a couple of months up to a couple of years.
Other OA – all other kinds of making articles freely available on the internet. That could include websites like social media, personal websites etc.

The different forms of publishing can be illustrated as in Figure 3.

Previous studies

Previous research was studied, in particular with focus on research made by Bo-Christer Björk and Mikael Laakso et al, based at Hanken School of Economics.[3] Other important studies include the work made at Université du Quebec à Montréal, led by Stevan Harnad. These studies deal mainly with the global

[1] BOAI: http://www.budapestopenaccessinitiative.org/read . Berlin Declaration: http://openaccess.mpg.de/286432/Berlin-Declaration .

[2] Suber (2012), p. 4.

[3] Research on Open Access Publishing: http://openaccesspublishing.org/ .

uptake of OA, by studying refereed journal articles. Although they use different methods and sometimes different indexes for the material, the overall results show a steady increase in global OA uptake ranging from 19,4 percent for publication year 2007 (Björk group) to 23,1 percent for 2011 (Harnad group). See [5] for an overview of OA studies.

A 2013 study from Science-Metrix produced for the European Commission DG Research & Innovation showed a much larger percent of OA share, both globally and on country level. The global OA share in 2011 was found to be 44 percent, and reaching 50 percent after computing an adjusted OA availability curve.[4] By 50 percent, we would have reached a "tipping point" of openly available refereed journal literature.

The difference between these results may have several explanations and one is the various definitions and classifications (or lack thereof) of OA used in the studies. This involves what is possible to measure, but also what we define as actual OA. For example, is any article, found to be freely available at the time of the study, to be classified as OA, regardless of how and where it is available, and in what version? This was something our project also felt the need to address.

A project that was performed during the same period in Denmark was followed closely, the Danish Open Access Barometer, previously reported on in *ScieCom Info*.[5]

Method

Main study:
The primary source for article data used was the Swedish national publication database SwePub.[6] SwePub includes all publications registered at Swedish universities and covers more of Swedish publications than for example Scopus or Web of Science, especially in the Humanities and Social sciences.[7] It also presents a link to articles that are parallel published in any of the universities' own repository.

Another reason for using SwePub data was to examine how reliable it is as a source for future monitoring of OA in Sweden.

The SwePub data (limited to refereed research articles and review articles, a total of 23 905 articles) was analysed for green OA, i.e. links to full text in Swedish university repositories. For the study of gold OA the data was matched by ISSN against DOAJ (Directory of Open Access Journals).[8] For numbers on delayed OA SwePub data was matched against a list of 492 journals compiled by Laakso and Björk. It is to our knowledge the most accurate list to cover delayed OA.[9] Overlaps between the different OA categories were noted and is presented.

The amount of *hybrid* OA is difficult to measure with accuracy (and potentially very labour intensive), since there are no easily available and reliable data.[10] Therefore we were not able to measure this in our main study.

Complementary study:
Since the main study obviously had its limitations (by only capturing green OA in Swedish repositories), the decision was made to supplement it with a manual study of 1.000 randomly selected articles from the total data volume. This was to get some idea of how large a part of articles from Swedish universities is openly available elsewhere on the web.

[4] Archambault et al (2013).
[5] Elbæk (2014).
[6] http://www.swepub.se/ . SwePub harvests records from 35 Swedish HEI local publication databases, using OAI-PMH. SwePub is hosted and maintained at the National Library of Sweden.
[7] At the time of the analysis two universities, Karolinska Institute and Swedish University of Agricultural Sciences, did not deliver records to SwePub. Only Web of Science records were included for these universities.
[8] http://www.doaj.org/
[9] Laakso & Björk (2013). The authors wish to thank Mikael Laakso and Bo-Christer Björk for kindly sharing the data.
[10] See Björk (2012).

The articles were searched for by title in Google and Google Scholar, and the first ten results were examined. The searches were conducted outside campus so that no full text would be accessed through subscription. This method is not without flaws, but it gives a hint of the variety of OA availability globally. The articles we found this way were named "Other OA" and then further categorized.

Potential OA:

We used the Sherpa/RoMEO-database to retrieve the information about what conditions publishers have on parallel publishing. This information is not very precise but nonetheless provides an estimate as to what extent articles are possible to self-archive.[11]

For a fuller description of the method, we refer to the final project report.[12]

Results

Main study:

Slightly more than 10 percent of the articles from Swedish universities published in 2011 were published in OA journals. The number for green OA is close, just below 10 percent. For both gold and green OA the amount rises to 17 percent (subtracting the overlap). With articles from delayed OA journals added, we reach a total of 25 percent.[13] See table 1 and figure 1.

OA-category	Amount	Percentage
Gold	2495	10,4%
Green	2289	9,6%
Delayed	2088	8,7%
Overlap - gold/green	-715	-3,0%
Overlap- green/delayed	-96	-0,4%
Not OA	17844	74,6%
Total	23905	100,0%

Table 1. The amount of journal articles, published in 2011 at Swedish universities, and split in gold, green and delayed OA respectively. The table also shows the overlap between categories.

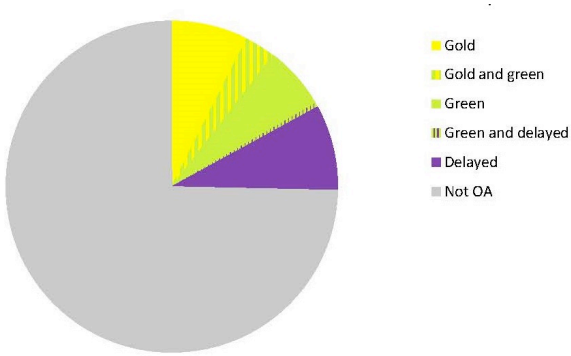


Figure 1. Amount of articles, published in 2011, that are OA.

Complementary study:

Almost 25 percent of the random sample articles (247 out of 1.000) were found openly available on the web, so called "Other OA". They were to a large extent found in subject repositories, institutional repositories outside Sweden or on publisher's websites. But a large number of the articles were found on other websites like the researcher's or department's website or on social media sites, among them ResearchGate.[14] Articles that were found on publisher's websites appeared for the most part to be hybrid OA but could also be delayed OA or "temporary OA".[15] (Table 2.)

OA-category	Amount
Subject repository	68
Institutional repository outside Sweden	53
Publisher's website	84
Other website	143
Overlap	-101
Total	247

Table 2. The share of different kinds of"other OA", based on the manual study.

Taking into account the amount of articles from the random sample that were gold, delayed or green (in Sweden) OA, the share of OA reaches just over 50 percent (Figure 2 and Table 3). This result is in line with the Science-Metrix study by Archambault et al.

[11] http://www.sherpa.ac.uk/romeo/
[12] Fathli et al (2014). Available in Swedish only.
[13] In the full project report numbers are also broken down per university and for subject areas. We also present in which OA and delayed OA journals Swedish authors most frequently published.

[14] http://www.researchgate.net/
[15] By temporary OA we mean an article that is freely available on the publisher's website for a period of time, presumably for marketing purposes, for example a "free sample article". Laakso (2014b) uses the term "promotional OA". No detailed categorization of articles on publisher sites were made, however, due to limitations in time in the project.

OA-category	N	Amount	Percentage
Gold or green	1000	183	18,3%
Gold, green or delayed	1000	275	27,5%
Gold, green, delayed or other	1000	522	52,2%

Table 3. Amount and percentage of journal articles that are OA, based on the manual study.

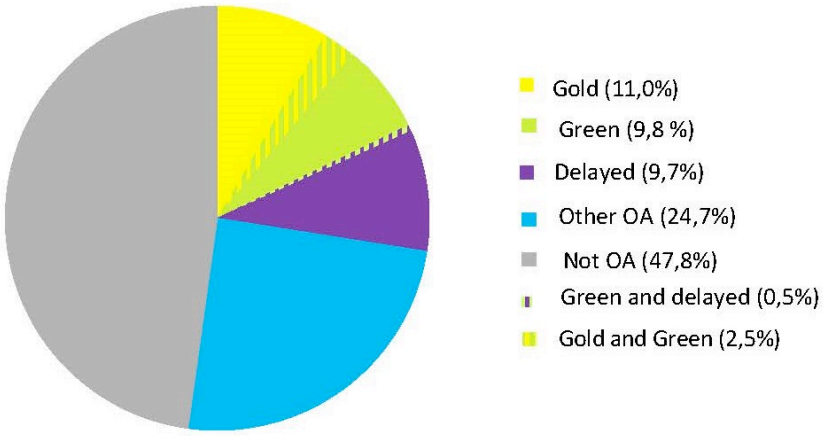


Figure 2. The amount of Open Access articles in the manual study.

Status in SHERPA/RoMEO (PDF or postprint)	N	Green OA	
		Amount	Percentage
Can	11986	1486	12,4%
Cannot	441	4	0,9%
Restricted	7294	366	5,0%
Unclear/unknown	650	105	16,2%
Total	20371	1961	9,6%

Table 4. The share of articles that have been parallel published and their different condition status in Sherpa/Romeo.

Potential OA:

The conditions for parallel publishing may vary, not only between publishers, but also between journals within the same publisher. For a survey on this we used the Sherpa/RoMEO-database. It does not cover 100 percent of the world's publishers and journals, but more than 22.000 journals are registered therein. Of the articles in our survey, the conditions for 85 percent were found in Sherpa/RoMEO.

One severe difficulty for measuring the conditions in Sherpa/RoMEO is that the listed conditions are not machine-readable. They are complex and not standardized. It is however possible to divide the conditions into four categories: 1. those who allow parallel publishing, 2. those who do not allow parallel publishing, 3. those who allow parallel publishing under certain conditions, and 4. those with status unclear. This division is not without defects but it supplies an estimate on the possibilities of parallel publishing.

Another complication is that the publishers differ between article versions: *submitted*, *accepted* or *published* for example.

The result shows that it is possible to parallel publish a much larger amount of articles than actually is being done currently. Only 12,4 % of articles where it is unambiguously allowed have actually been parallel published. Even if we estimate that some articles are parallel published in repositories outside Sweden this figure would only reach about 24 percent (based on our manual study).

Discussion

Our study has showed, not surprisingly, that the ways that research articles are made openly available vary significantly. The same article can be made OA in several ways, and a key problem in measuring OA is to define what we actually mean by it. It may seem surprising that there should still be a need to discuss the definition of OA, but clearly there is, because it affects how we measure it as a phenomenon and what the result is. Largely absent from OA definition discussions are the implications revealed in this and other studies on OA share and growth, that is the locus of availability of and infrastructure surrounding publications taken to be OA.

Some of the most significant variables of OA articles are

- Sustainability and integrity of the published fulltext file (incl. long term availability)
- Infrastructure surrounding the availability (searchability, findability etc)
- The point of time that the article is made available
- Version of the article

For the various forms of OA these variables differ (and are in some cases substandard), and this is one reason for the transient nature of OA, as noted by Archambault et al. To try and get a grip on this problem, we use the definition of OA in the Berlin Declaration, and take a closer look on our categories of OA. For it to be an OA contribution, the Declaration states that:

A complete version of the work and all supplemental materials, including a copy of the permission as stated above, in an appropriate standard electronic format is deposited (and thus published) in at least one online repository using suitable technical standards (such as the Open Archive definitions) that is supported and maintained by an academic institution, scholarly society, government agency, or other well-established organization that seeks to enable open access, unrestricted distribution, inter operability, and long-term archiving.[16]

In general terms, we can see that gold OA publishing fits these criteria, as does delayed OA.[17] The same goes for green OA through institutional repositories and well established subject repositories, such as arXiv and PubMed Central.[18] Articles made available through a publisher's website will fit, if they are genuine hybrid OA, but not if they are so called "temporary" OA, for marketing reasons or the like.

As regards articles made available through authors' personal websites and social media sites, we conclude that they do not fit the criteria, either since they are not in a repository and/or not made available through a well-established organization. This would also apply to departmental web pages. These categories are simply lacking in sustainability and/or infrastructural context.[19]

The category of delayed OA has been proposed by Laakso & Björk to qualify as a subset of OA journals in the DOAJ database.[20] These journals are subscription journals, with a free online archive. And it could be argued, that since we accept embargoes (i.e. time delay) for green OA, why not the same for delayed OA via the journals' own web sites? An obvious advantage for delayed over green would be that a delayed OA article gives access to the published version of record, while an embargoed green OA article in most cases gives access to the author version (which has disadvantages when it comes to citability).

So could the OA concept include subscription journals? OA is not a business model in itself, it has been said,[21] and could potentially include several different models. However, including subscription journals within OA does seem like a contradiction in terms. It could be possible to differentiate between article level and journal level, so that the articles are considered OA, but the journal itself is not (this distinction is the case with *hybrid* OA).The discussion in any case points to the highly differentiated, and increasingly complex, forms and models that the dissemination of (open) scientific literature is subject to.

The Science-Metrix study by Archambault et al. states that OA has reached a "tipping point" with a share of 50 percent globally. We suggest that this statement should be taken with caution. A growing number of freely available research articles does not automatically

[16] Berlin Declaration: http://openaccess.mpg.de/Berlin-Declaration .The quote is the second of two definitions. The first states far-reaching re-use rights (also known as *libre* OA). This aspect is not measured in our study, and is a more complex issue. This definition is not taken into account in our argument.

[17] There may be examples of OA journals published by small publishers where sustainability and long term archiving is a problem, but they are typically a minor portion of gold OA publishing. For delayed and green OA, of course, the time delay (or embargo) of availability is an aspect not mentioned in the Declaration, but now mostly accepted for green OA as a (temporary) necessity.

[18] http://arxiv.org/ , http://www.ncbi.nlm.nih.gov/pmc/.

[19] Björk et al (2014) show how articles on personal and institutional web pages are more likely to be missing three years after the initial study.

[20] Laakso & Björk (2013).

[21] Open Access Scholarly Information Sourcebook (OASIS): "Open Access is a means of delivering content to users, not a business model." http://www.openoasis.org/index.php?option=com_content&view=article&id=358&Itemid=263

mean that the shift from traditional subscription based publishing to OA publishing on a global level is imminent. The metaphor of a tipping point suggests a linear process where one form of publishing simply supersedes another. Most forms of OA, it should be noted, are still made available within the framework of traditional publishing (see Figure 3). Only gold OA publishing is a real alternative to traditional publishing.

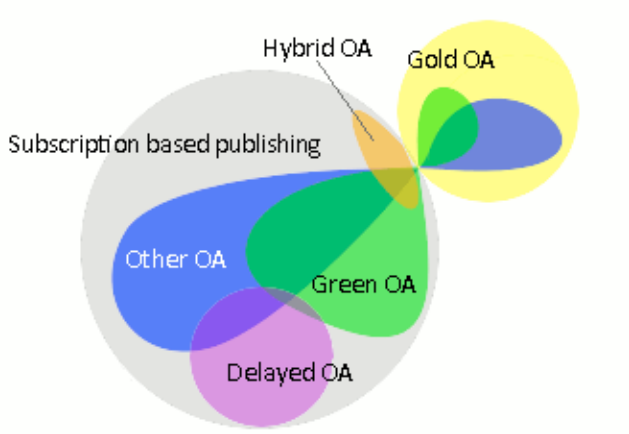


Figure 3. A schematic picture of the OA landscape, showing the overlap of different kinds of publishing.

The variety of the OA phenomenon is complex and appears in some respects dispersed and fragmented. It is highly desirable that the monitoring and measuring of OA becomes more standardized, with regard to definitions and categories. A standard vocabulary on OA is one way of achieving this. Another, perhaps harder to accomplish, is that a general understanding of what constitutes real OA should be agreed upon.[22]

In our recommendations for future OA monitoring in Sweden we conclude that SwePub should be the prominent tool and index for that purpose. With the ongoing national project to further develop SwePub for statistics and bibliometric analysis, this is already under way.[23] Another recommendation, which is more generally applicable, is that articles made freely available outside of established infrastructure should not be counted as OA. This will exclude articles on personal and departmental websites and on social media sites. This would also be in line with OA mandates by funders and universities, which typically refer only to OA in repositories and/or journals.

[22] See the recent doctoral thesis by Laakso (2014a), especially discussion in section 6.5. Also Laakso (2014b).
[23] The National Library of Sweden has received an assignment by the Swedish government to further develop SwePub to make possible bibliometric analysis and statistics on a national level. The project is ongoing and to be finalized during 2015. http://www.kb.se/aktuellt/nyheter/2014/SwePub-blir-ett-kugghjul-i-forskningens-infrastruktur/.

Table 5. An overview of studies showing the amount of Open Access. The numbers indicate percentage of articles included in the index. Modified after Björk et al (2014).

Study	Publication year for journal articles in the study	Year of survey	Article index	Articles in OA-journals	Articles with delayed OA	Hybrid OA articles	Other freely available articles	Green OA articles in subject repositories	Green OA-articles in institutional repositories	Total OA globally
Björk et al 2009	2006	2007	Ulrich´s	4,6	3,5		3,0	3,3	5,0	19,4
Gargouri et al 2012	2006	2009	WoK	21,0						21,0
Björk et al 2010	2008	2009	Scopus	5,3	1,2	2,0		11,9		20,4
Gargouri et al 2012	2010	2011	WoK	1,2	21,9					23,1
Laakso & Björk 2012	2011	2012	Scopus	11,0	5,2	0,7				
Archambault et al 2013	2011	2013	Scopus	11,5	32,5					44,0
Elsevier 2013	2012 (gold), 2011-2013 (green, hybrid, delayed)	2013	Scopus	9,7	1,0	0,5	11,4			
Study (Swedish OA)	**Publication year for journal articles in the study**	**Year of survey**	**Article index**	**Articles in OA journals**	**Articles with delayed OA**	**Hybrid OA articles**	**Other freely available articles**	**Green OA articles in subject databases**	**Green OA articles in institutional repositories**	**Total OA Sweden**
Hedlund 2010	2008	2009	Scopus	11,4			11,2			22,6
Archambault et al 2013	2008-2011	2013	Scopus	8,0	40,0					48,0

References

Archambault, É., Amyot, D., Deschamps, P., Nicol, A., Rebout, L. Roberge, G. (2013). *Proportion of Open Access Peer-Reviewed Papers at the European and World Levels 2004-2011.* Science-Metrix. European Commission DG Research & Innovation RTD-B6-PP-2011-2: Study to develop a set of indicators to measure open access. http://www.science-metrix.com/pdf/SM_EC_OA_Availability_2004-2011.pdf (Accessed 2 September 2014.)

Björk, B-C. (2012). The hybrid model for open access publication of scholarly articles : A failed experiment? *Journal of the American Society for Information Science and Technology*. 63(8), pp: 1496-1504. doi:10.1002/asi.22709. (Accessed 12 October 2014.)

Björk, B.-C., Laakso, M., Welling, P., Paetau, P. (2014). Anatomy of green open access. *Journal of the Association for Information Science and Technology*, 65(2), pp. 237–250. doi: 10.1002/asi.22963. (Accessed 2 September 2014).

Björk, B-C., Roos, A. & Lauri, M. (2009). Scientific journal publishing: yearly volume and open access availability. *Information Research*, 14(1) paper 391. http://www.informationr.net/ir/14-1/paper391.html (Accessed 3 October 2014.)

Björk, B-C., Welling, P., Laakso, M., Majlender, P., Hedlund, T. & Guðnason, G. (2010). Open Access to the Scientific Journal Literature: Situation 2009. *PLoS ONE*, *5*(6), e11273. doi:10.1371/journal.pone.0011273 (Accessed 3 October 2014).

Elbæk, M.K. (2014). Danish Open Access Barometer : mapping Open Access to Danish research and creation of an online prototype for automated open access monitoring. *ScieCom Info* Vol 10 (1). http://journals.lub.lu.se/index.php/sciecominfo/article/view/10238 (Accessed 3 October 2014.)

Elsevier (2013). *International Comparative Performance of the UK Research Base – 2013 : A report prepared by Elsevier for the UK's Department of Business, Innovation and Skills (BIS).* https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/263729/bis-13-1297-international-comparative-performance-of-the-UK-research-base-2013.pdf (Accessed 3 October 2014.)

Fathli, M., Lundén, T., Sjögårde, P. (2014). *Open Access-publicering vid svenska lärosäten : en kartläggning av året 2011.* Kungl biblioteket. http://www.kb.se/Dokument/Om/projekt/open_access/2014/OA-publicering%20vid%20svenska%20l%C3%A4ros%C3%A4ten%20-%20slutrapport.pdf (Accessed 3 October 2014.)

Gargouri, Y., Lariviere, V., Gingras, Y., Carr, L. & Harnad, S. (2012). Green and Gold Open Access percentages and growth, by discipline. *In, 17th International Conference on Science and Technology Indicators (STI), Montréal, CA, 05 - 08 Sep 2012.* http://eprints.soton.ac.uk/340294/ (Accessed 3 October 2014.)

Hedlund, T. (2010). Open Access availability of articles by Nordic authors. *ScieCom Info*, Vol 6(4). http://journals.lub.lu.se/index.php/sciecominfo/article/view/4760 (Accessed 3 October 2014.)

Laakso, M. (2014a*). Measuring Open Access : Studies of Web-enabled Innovation in Scientific Journal Publishing.* Diss. Hanken School of Economics. http://hdl.handle.net/10138/45238 (Accessed 9 October 2014.)

Laakso, M. (2014b). The past, present & future of Open Access. *ScieCom Info* Vol 10(1). http://journals.lub.lu.se/index.php/sciecominfo/article/view/10236 (Accessed 9 October 2014).

Laakso, M. & Björk, B-C. (2012). Anatomy of open access publishing: a study of longitudinal development and internal structure. *BMC Medicine* 10:124. doi:10.1186/1741-7015-10-124 (Accessed 3 October 2014).

Laakso, M. & Björk, B-C. (2013). Delayed open access : an overlooked high-impact category of openly available scientific literature. *Journal of the American Society for Information Science and Technology*, 64(7), 1323–1329. doi:10.1002/asi.22856. (Accessed 4 September 2014.)

Suber, P. (2012). *Open Access*. MIT Press. (Digital version: bit.ly/oa-book.)

Websites

ArXiv: http://arxiv.org/ (Accessed 12 October 2014.)

Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities: http://openaccess.mpg.de/Berlin-Declaration (Accessed 4 September 2014.)

Budapest Open Access Initiative: http://www.budapestopenaccessinitiative.org/read (Accessed 3 October 2014.)

Directory of Open Access Journals (DOAJ): http://doaj.org/ (Accessed 4 September 2014.)

Open Access Scholarly Information Sourcebook (OASIS): http://www.openoasis.org/ (Accessed 4 September 2014.)

PubMed Central: http://www.ncbi.nlm.nih.gov/pmc/ (Accessed 12 October 2014.)

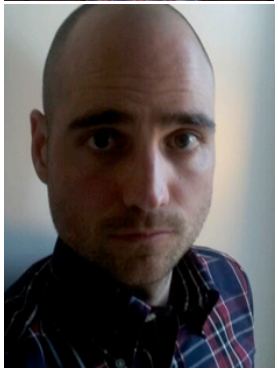
Research on Open Access Publishing (Hanken School of Economics): http://openaccesspublishing.org/ (Accessed 4 September 2014.)

ResearchGate: http://www.researchgate.net/ (Accessed 3 October 2014.)

Sherpa/RoMEO: http://www.sherpa.ac.uk/romeo/ (Accessed 3 October 2014.)

Margareta Fathli Librarian, with focus on Open Access, at KTH Royal Institute of Technology, School of Education and Communication in Engineering Science, Unit for Publication Infrastructure.

Tomas Lundén
Librarian, with focus on scholarly communication, at SLU University Library, Swedish University of Agricultural Sciences. At the time of the project employed by Gothenburg University Library.



Peter Sjögårde
Bibliometric analyst at KTH Royal Institute of Technology, School of Education and Communication in Engineering Science, Unit for Publication Infrastructure.