Accessibility and Self Archiving of Conference Articles: A Study on a Selection of Swedish Institutional Repositories

Peter LINDE[a], Jörgen ERIKSSON[b], Lars KULLMAN[c 1], Margareta FATHLI[d], Klemens KARLSSON[d], Marjatta SIKSTRÖM[e], Ylva SKÖLD[f], Ingela TÅNG[e]

[a]*Blekinge Institute of Technology*
[b]*Lund University*
[c]*Chalmers University of Technology*
[d]*KTH Royal Institute of Technology*
[e]*Stockholm University*
[f]*Luleå University of Technology*

Abstract. The main purpose of this project has been to examine the accessibility of refereed conference articles and the OA- and publishing policies of conferences in order to in this way elucidate different aspects concerning self-archiving in Swedish institutional repositories. For this purpose, the project participants have examined a number of conferences and references to conference articles via their institutional repositories during a specific time period and described these from the perspective of a common scheme. The study has showed that the local institutional repositories fill an important role to make conference publications visible. We have found that ca. 50 % of the conference papers can be published as postprints in our institutional repositories. We have noticed that ca. 15% or the studied conference articles are not available at all. It is, therefore, of great importance to use local institutional repositories as a publishing channel, not only for primary published material such as dissertations and reports, but also as a source for finding these conference articles "without a home". Between 20-25 % of the examined articles were found in some type of OA archive; ca. half of these were found in one of the project participants' own institutional repositories. This indicates that the publishing database of respective higher education institution is an important factor for open accessibility. Ca. 10% of the conferences in the study had an explicit OA policy or expressed such a policy by openly making conference articles accessible on their conference sites. A big problem when it comes to self-archiving of conference articles is the lack of information about OA policy. The landscape of conference publishing is complex and the self-archiving of documents from conferences is very time-consuming. Above all, we would wish a policy resource for conferences similar to the SHERPA/RoMEO. At present, however, there is no other alternative than scrutinizing the conferences' copyright information to the authors and from this attempt to draw conclusions about possible self-archiving.

---

[1] Corresponding Author: Lars Kullman, Chalmers University of Technology, The Library; E-mail: lars.kullman@chalmers.se

To facilitate the future handling and classification of conference articles in Swedish institutional repositories a number of recommendations are suggested.

Introduction

The following study was conducted from February 2011 till January 2012 at Blekinge Institute of Technology (BTH), Chalmers University of Technology (Chalmers), KTH Royal Institute of Technology (KTH), Lund University (LU), Stockholm University (SU) and Luleå University of Technology (LTU). The project was funded by the programme OpenAccess.se run by the National Library of Sweden [1].

A more comprehensive report of the findings in this article, including more detailed data and figures will be published and available in Swedish.

1. Background

In the past few years, numerous studies have been conducted on the activities of self-archiving - the green road to Open Access. However, the focus in these studies has mainly been on journal articles. Since institutions and research funders frequently requires deposit in public-access repositories, also of quality reviewed conference publications, this also increases the pressure on researchers and repository managers to handle all of the issues when it comes to self-archive conference articles. The situation is particularly important to the universities of technology where, not seldom, conference articles constitute a major part of the articles produced.

Conference articles are not published in the same homogenous way as journal articles traditionally are with their well established publishing channels with clear rules on copyright and licenses and with policy tools such as SHERPA/RoMEO. The channels and routines when it comes to conference articles are more difficult to grasp and explicit information about Open Access and where and how the articles are published can sometimes be impossible to find.

A study of institutional repositories in Hong Kong from 2009 observes that only about 3% of the conference articles published by Hong Kong researchers since 10 years back are openly accessible in full text via the institutional repositories of the universities [2]. The numbers indicate the difficulty to self-archive conference articles. One of the reasons for the low numbers in the Hong Kong study may be that the world of science, traditionally, has looked upon conference articles as a second-rate publication. This is not the way forward in the modern information landscape with its great diversity of information products of varying degrees of formality and visibility.

Moed & Visser have, among others, in a research report to The Netherlands Organization for Scientific Research shown how important conference articles can be in terms of citations generated both to proceedings and to journals. Within some research areas, as for example computer science, conference proceedings rather than journals constitute the main channel of scientific publishing [3]. Despite the importance of conference articles in research communication and despite the generally low

occurrence of full-texts in the institutional repositories, studies around this topic are lacking.

2. Purpose and Problems

The purpose of this study is to examine the self-archiving activity, Open Access and publishing policies of quality reviewed conferences in order to clarify different aspects and general problems concerning self-archiving of conference articles. Are conference articles more difficult to find in full text compared to journal articles? Are Open Access policies for conference articles more difficult to find and thus making self-archiving more laborious? Our purpose is to find answers to questions like these and make our findings serve as a background to suggested recommendations for a "best practice" regarding conference articles.

We have been interested in the following questions:

- What does the relationship organizer/publisher look like; who are the publishers and in what type of publications are the conference contributions published?
- What is the frequency of self -archiving?
- What does the accessibility look like?
- Can conference articles be found to a somewhat great extent in full text in the institutional repositories of the participating institutions?

3. Concepts and Definitions

We have used the concept "self-archiving" throughout the report. Self-archiving here refers to a publication which has first been published in a scientific journal or proceedings and then made freely accessible in an open digital archive. Self-archiving is the dominating terminology used within the Open Access movement and in the international literature.

Since the publication databases at the participating institutions also serve as bibliographical databases we have used a wide definition of the term "institutional repository" as a database not only containing full text records.

By conference articles we are referring to full-length articles published in proceedings or other media, and thus not conference contributions in the length of an abstract. Quality reviewed means that the person who has entered the article in respective institutional repository has stated that it is "peer-reviewed".

4. Method

The project participants have studied a selection of conferences and conference articles taken from their institutional repositories (together the project participants have about

25,000 bibliographical references to conference articles in their institutional repositories).

The project constitutes of two main parts:

1. An accessibility study conducted on a selection of conference articles based on bibliographic information retrieved from the project participants' repositories. Through what channels can the articles be freely accessible in full text?
2. A conference study where we examine the Open Access policies, publishing channels, etc. of a selection of conferences.

4.1. Selection

4.1.1. The accessibility study

300 references to conference articles from the institutional repository of each higher education institution participating in the study were selected. To give the selection a spreading over time, the material was distributed to 100 references per year for the three publication years 2005, 2007 and 2009. The quantity of 100 per year was determined given the great variation that existed in the total number of references in the institutional repositories of the higher education institutions. As, for example, BTH has fully 100 references to refereed conference articles in its publishing database per year, the limit was set at 100 per year. This also meant that there were great variations in the share of conference articles that the selection covers in relation to the total amount of references to conference articles at the respective institutions. For BTH, for example, the selection covers ca. 75% of the references for the respective years while the selection of LU covers ca. 10% of all refereed conference articles.

The starting year was chosen to be 2005 as, for example, Chalmers did not until 2004 have the requirement that all references be reported in their publishing database, something that was considered as established in 2005. The final year 2009 was chosen as we assumed that most records for this year by now (2011) should have been entered in the institutional repositories.

To some degree the selection was made on subjective grounds as the references had been collected in different ways. The principal goal was to generate 300 references from the respective institution. The selection, thus, lacks a scientific basis and can therefore not form the basis for any scientific conclusions. But this was not, however, the purpose of the study. This means that the numbers and results that are reported cannot, strictly, be given validity for anything else than the selection that we chose to study. However, we believe that the numbers and results point to tendencies and connections that are well worth noting, and that these gathered experiences also can form the basis for certain assumptions regarding conference articles in relation to open access.

The main stress has been on information science, computer science, science of engineering and applied science, etc., as these are the disciplines that use conference articles as their main channel. However, the project also comprises social science and the humanities.

Each project participant retrieved 100 conference articles from the publishing years 2005, 2007 and 2009, that is, a total of 300 articles per institution (a smaller number of articles have, from SU, been transferred from the year 2004 to 2005 as there was not a sufficient quantity from the year of 2005). The selection was made according to the following:

BTH: The following number of references to refereed conference articles was found for the years in question: 2005=125; 2007=138; 2009=142. The 100 first articles in alphabetical title order were retrieved for each year and used in the study.

KTH: The following number of references of refereed conference articles was found for the years in question: 2005=1005; 2007=994; 2009=1127. Each 3rd − 4th article sorted on the author's name was taken out and used in the study.

Chalmers: The following number of references of refereed conference articles was found for the years in question: 2005=647; 2007=695; 2009=715. For each year, all references were sorted according to conference title and then a reference was picked out from the most commonly occurring conferences.

Lund: The following number of references of refereed conference articles was found for the years in question: 2005=981; 2007=919; 2009=754. Each 3rd − 4the article sorted on the author's name was taken out and used in the study. The ambition was here also to have each 3rd or 4th selected reference belong to a Hum./Soc. Sci. conference.

SU: The following number of references of refereed conference articles was found for the years in question: 2004-2005=119; 2007=153; 2009=243. For the examined years, 300 articles were selected randomly. All of them were noted as Hum./Soc.Sci. articles (Note that many of these belonged to the subject computer- and systems sciences which, in turn, belongs to the Faculty of Social Sciences at SU).

Luleå: The following number of references of refereed conference articles was found for the years in question: 2005=325; 2007=304; 2009=354. Among these, 63 articles in the Hum./Soc. Sci. field were retrieved randomly. In addition to these, the first 100 articles in alphabetical title order for every year were used.

In all, 1787 conference articles have been examined, of which ca. 1/4 have been classified as Social Science/Humanities (446 articles) and the rest as Natural Science/Technology (1341 articles).

4.1.2. The conference study

Each project participant selected 50 conferences from her/his institutional repository according to the following:

BTH: The 50 most frequently occurring conferences.

KTH: Search on conferences in 2010. The first 50 in the hit list were selected.

Chalmers: The 50 most frequently occurring conferences.

Lund: The same principle as for the accessibility study was followed.

SU: The 39 most frequently occurring conferences from the selection of conference articles, and 13 conferences that were frequently occurring in the field of Natural Science.

Luleå: Conferences were selected randomly but with the thought of getting as big a spreading as possible. Every 5th conference that was selected had to be of the type Hum./Soc. Sci. and there was an attempt to avoid too many closely related conferences within the same subject, for example, IEEE, ACM etc.

After the selection of both conference articles and conferences, the lists have been checked and possible duplicates have been removed and, if possible, been replaced by new titles. Finally, 297 conferences were examined according to the description below.

5. Procedures

5.1. The accessibility study

All the articles were examined according to the following matrix created in a spreadsheet with column headings according to the list below:

- Title of article: Title of the whole article. This can then be copied to conduct a search in Google.
- Name of conference.
- Year: 2005, 2007 and 2009.
- Institution.
- OA Home page: Is the article openly accessible in full text on any type of Home page? Home page is here defined as personal web pages, for example web pages at web hosts or at a university webserver or at any other resources that offers personalized webpages for use by individuals or professionals.
- OA archive. Does the article exist in some kind of OA archive? Examples of these are institutional repositories at universities and higher education institutions, open conference archives, subject archives of different types etc.
- Subscription: Is the article locked behind subscription filters or other cost barriers? If Yes, we have tried to note where.
- Accessibility for conference participants: Is the article only accessible for conference participants, distributed on a USB, CD etc.? This information is noted if it is accessible on a conference site or similar.
- Print: Does the article exist in print? Here our departure point has been if the article has formed part of any proceedings with an ISBN or other identification and if this is made apparent from the search on the Web.
- Error 404. Do links to the article give Error 404?
- Published later? Has the article been published later in the same or an extended form compared to the publishing in possible proceedings, in any journal/book series? If Yes, we have tried to indicate where.
- Subject: Hum/Soc. Sci. or Tech/Nat. Sci.

The accessibility was examined by copying the whole or parts of the title of the article and conducting a search in Google. We followed links on maximum three Google pages. If no result was obtained, we searched on the conference name and lastly we searched on the organizer, if this information was available in the data. Karlsruhe Virtual Catalogue[4] or the catalogue of Library of Congress[5].

5.2. The conference study

The articles were examined according to the following matrix:

- The higher education institution in the project – noted by an acronym.
- Subject, either Hum/Soc. Sci. or Nat. Sci./Techn.
- AUS ranking according to the ERA Conf. List [6]. Used when we want to see if there are differences in OA between highly and lowly ranked conferences;.
- Acronym – the acronym of the conference.
- Conference name.
- ISSN if available.
- Publishing channels. Marked if the conference publishes on its own Web pages or uses a third part for publishing.
- Agreements Yes/No. If the conference on its Web pages provides information about applicable copyright agreements for authors this is noted here. The information usually is found on the conference webpages or via references to publisher information.
- OA policy. Does the conference give information on its Web pages about its OA policy or show openly in any other way that the conference articles are published in a freely accessible form?
- URL. If you do no find the URL to the Web page of the conference, the conference page for last year's conference is used.
- Local observations. Serial numbers can be entered here, etc.

6. Results

6.1. Where do you find the full text?

Of the 1787 conference articles that were examined, we found 823 of them openly accessible in full text on some Web page. That means just over 46 % of the whole material which is about twice the number that Björk et al. found in their study of the open access to refereed journal articles in full text 2009[7].

We have looked especially thoroughly at the distribution between different types of OA archives and Web pages to understand what sources are important in this context. The articles found in full text were distributed according to the sources in Figure 1.
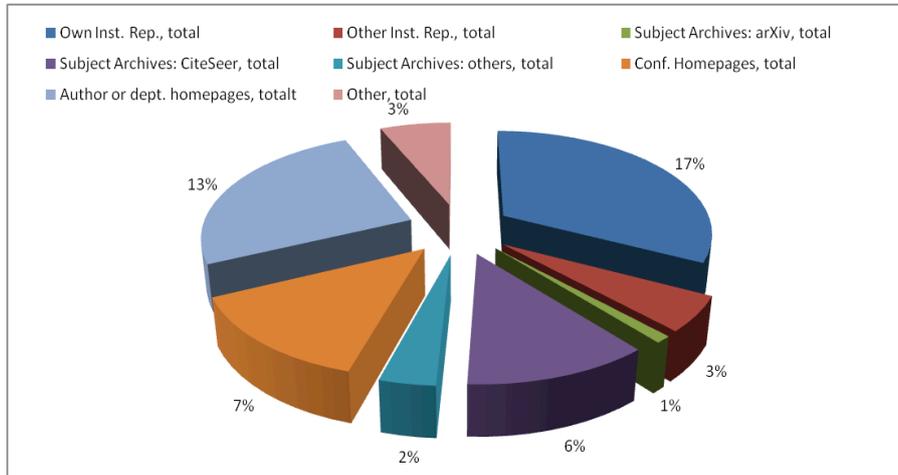
Figure 1. Distribution of sources for found full texts.

The numbers show that the institutional repositories of respective higher education institution is the most important source of the full texts followed by the author's/departments' web pages and subject archives of the type CiteSeer etc. The Chinese study, which we referred to in the beginning of this report, showed that only 3 % of the conference articles were found openly accessible via the institutional repositories of the universities. In our material (which is smaller and builds on a later time interval) the number is significantly higher – fully 17 %!

The differences between the contents in the project participants' institutional repositories differ greatly. The scope of the archives spans from under 2 % of found full texts in the respective repository to over 50%.

The increase of full-text material in the participating institutions repositories when comparing publishing year 2005 to publishing year 2009 is ca. 50%, from 77 full texts 2005 to 114 in 2009.

We cannot see any major difference between Tech./Nat. Sci. or Hum./Soc. Sci. conference articles regarding publishing in OA archives.

6.2. The grey material

In our material, we saw rather quickly that a relatively large share could not be retrieved at all, i.e., these texts were not represented in any channel. Therefore, we wanted to look a little closer at precisely that which we in the library world often categorize as "grey material," or such material that exists as reference but that is hard to retrieve in full text or in print. We studied 5 categories of this material in particular according to the table below:

Table 1. Full text papers that was available via at least one channel or none at all.

Category	Quantity	%
Not found anywhere	264	15%
Only print	153	9%
Only OA repositories	128	7%
Only OA web page	63	4%
Only subscription	22	1%
Only published later	22	1%
Only error 404	9	1%

It is interesting to note that 264 of the conference articles (15 %) are not found in any of the examined channels. Only small parts of the material are represented in only one channel, except "Only print," which lies somewhat higher than the other ones. Since as much as 15 % of all articles do not seem to be accessible in any way in full text, there is reason to believe that publishing channels are not always available for this type of publications. Another conclusion could be that many small conferences simply do not have the possibility to publish the articles on the Internet or in print. Worth noting is also the fact that ca. 7% of the articles are only found in OA archives.

6.3. The conference study

Five project participants out of six registered an AUS ranking. Out of 250, 37 conferences were ranked. Three of them had an OA policy that is somewhat lower than the average in the study, 8% compared to 12%. One should note that the majority of the conferences on this list are in the field of Tech./Nat. Sci. and therefore the result corresponds rather well to what we found for non-ranked conferences generally in the field of technology (9 %). The conclusion, thus, is that in regard to Open Access policy, we see no differences when it comes to highly or lowly ranked conferences.

About two out of three conferences had information about publication agreements on their web pages. When looking at the differences subject-wise, one can observe that humanities conferences, compared to technology conferences, more often seem to lack information about publication agreements. Approximately half of the humanities conferences had information about publication agreements.

About 10 % per cent of the 297 conferences had an OA policy or expressed an OA policy by making conferences easily accessible on their conference sites.

When we looked at the conference articles that are found in channels that expressly allow so-called postprint publishing, the result is, generally, that about half of the conferences make use of channels that allow self-archiving. An extra high number is demonstrated by higher education institutions with many conferences using IEEE-explore and ACM's services to publish their conference articles.

7. Discussion and conclusions

Self-archiving of conference articles is often a more time-consuming occupation than similar work with journal articles as information about the open access policy of conferences is often lacking (cf. SHERPA/RoMEO for journal articles). Furthermore, conference publishing seems to be more unstable and heterogeneous than journal publishing. Self-archiving of conference articles has, thus, often been characterized by insecurity regarding both mode of procedure and use.

After the work in our project we may, however, observe that:

- Almost 46 % of the examined articles are openly accessible in full text.
- There has been an increase of the number of full-text conference articles with ca. 50 % in the project participants' institutional repositories when comparing publishing year 2005 to publishing year 2009.
- Ca. 20-25 % of the examined articles exist in full text in some type of OA archive.
- Ca. 17% of the examined articles exist in full text in one of the project participants' own publishing database.
- Ca. 15 % of the examined articles were not accessible in any channel at all. Furthermore, the humanities/social sciences have a bigger share of documents not accessible in any channel at all (21 %).
- It is possible to Self-archive at least half of the examined articles as postprint versions of the full text.
- A minority of the conferences, ca. 10% of the ones that we examined, has an OA policy or openly shows that they publish the conference articles freely accessible.

To contact the contact persons of the conferences can be an alternative to sort out OA policy matters. In our study, a dozen mails with requests were sent and only 2 were answered but these were then followed up by further correspondence and questions. The result was too poor and the method too time-consuming to be used further.

On most conference sites you find agreement policies for authors, either under the heading "Call for Papers" or under "Instructions for Authors". If you read these agreements, you can sometimes conclude if the contributions can be self-archived. But this is not always the case. It is clearly significantly more time-consuming to investigate the possibilities of self-archiving for conference articles than what it is for journal articles.

We did not find any common denominator for the conferences that do have an OA policy or that openly publish their contributions. We did not see any difference when we, via the AUS ranking, compared big recognized conferences to the rest of the conferences. There was, however, a tendency showing that Hum./Soc. Sci. conferences more often had an OA policy than Nat. Sci./Tech. conferences. Particularly in Social Science conferences, Open Access channels have, for a long time, been established where working papers/preprints have been distributed and shared between researchers. This practice is also seen at Hum./Soc. Sci. conferences. Here you more often run into conferences that in some form are published in Open Access.

The study has shown that the institutional repositories fill an important function in spreading knowledge about conference publications. We have found that no less than ca. 50 % of the conference articles can be published as postprints in our institutional repositories. It is, therefore, of great importance to continue the work of making use of the local institutional repositories as a publishing channel. One should not only use them as a tool for self-archiving and primary published material, such as dissertations and reports, but also as a source for, above all, conference articles "without a home," which we now know constitute a large part of our studied material (ca. 15 %).

At the same time as we, during the course of the project, have realized how complicated the landscape of conference publishing is, we have also become convinced of the great possibilities that are given here for self-archiving of different types of conference documents. A tool similar to SHERPA/RoMEO would of course be of much help. But conferences start up and are terminated, and they are hived off and find new forms constantly. To keep alive a register of this type would be enormously demanding in terms of resources. However, SHERPA/RoMEO can be used in the cases where conference documents are published by big organizations or publishing firms as IEEE, ACM, Springer, Elsevier etc. Here you often find the OA policy information at the level of the publisher/organization in the case when the web pages of the separate conferences do not provide any guidance.

8. Recommendations

To strengthen the work at the higher education institutions for self-archiving of conference articles through increased resources and understanding of the complexity of this type of research documents. Self-archiving of similar documents is significantly more time-consuming, but, however, definitely as possible as of journal articles. To be aware of the fact that a considerable part of conference articles are not published in a any channel and that the local institutional repository possibly, may offer a possibility of making the document visible.

In many cases information on the conference web sites lack instructions as to whether the author has the right to self-archive their text, must transfer their copyrights etc. Therefore we recommend efforts to make researchers aware of the importance of copyright control. Generally, as long as they have not signed any agreement they retain their copyright and the conference article can be self archived. If the copyright is transferred the agreement must be saved for later reference. With some conferences it is sufficient to submit the article via an electronic system. By doing so, in most cases, you transfer the copyright to the publisher.

Preprint publishing. It sometimes happens that an author wishes to self-archive her/his conference contribution before the conference takes place. In the instructions given to the authors, either on the Web pages of the conferences or on Web pages of the publishers that publish the conference contributions, it is often difficult to find any information concerning self-archiving before the taking place of the conference. It is often stated, however, that for admission to the conference in question, the contribution cannot be published anywhere else. There is, therefore, reason to be generally reserved about allowing self-archiving of conference articles in local institutional repositories before the conference has been carried out. A recommendation to the authors should, therefore, be to wait with the self-archiving of the conference contribution until after

the conference. This does not constitute a problem from the perspective of parallel publishing.

As conference contributions are published in so many different ways it is difficult to find a common denominator regarding cover pages. On the other hand, this is also why it is important to facilitate the process of self-archiving and avoid a complicated handling of many different Cover pages. The lowest common denominator can, however, be said to be the name, number, acronym, place and point of time of the conference, as well as information about the organizer and possible publishing firm and a unique identifier, of the type DOI. This information can be found in a structured form as metadata in all our different institutional repositories and it should be possible to automatically transfer this information to a Cover page. Basically, this information complies with that which most publishing firms request for self-archiving namely a correct reference which clearly states that this is a self-archived version of a conference contribution that has originally been published somewhere else, to which the information on the Cover page also refers.

References

[1]   OpenAcces.se. Available from: http://www.kb.se/OpenAccess/
[2]   D.L.H. Chan, An Integrative View of the Institutional Repositories in Hong Kong: Strategies and Challenges, *Serials Review* 35 (2009),119-124.
[3]   H.F. Moed, M.S. Visser, Developing Bibliometric Indicators of Research Performance in Computer Science: An Exploratory Study, *Centre for Science and Technology Studies. Leiden University, the Netherlands,* 2007, 117
[4]   Karlsruhe Virtual Catalogue. Available from: http://www.ubka.uni-karlsruhe.de/kvk_en.html
[5]   Library of Congress Catalogue. Available from: http://catalog.loc.gov/
[6]   The ERA Conference Ranking Exercise. Available from: http://core.edu.au/index.php/categories/conference%20rankings/1
[7]   B-C. Björk et al. Open Access to the Scientific Journal Literature: Situation 2009. *PLoS ONE* 5(6) (2010) Available from: doi:10.1371/journal.pone.0011273