Project outline

1. Dissertation advisors endorsing the proposal

Professor Birger Hjørland and associate professor Jeppe Nicolaisen hereby endorse the proposal by Tove Faber Frandsen. The present research project contributes significantly to the field of bibliometrics. The project is highly relevant as it analyzes the consequences of the changing scholarly communication in combination with the increasing need for evaluating research performance. Furthermore, the project is well organized and proceeding according to plan.

2. Introduction

According to Bauer & Bakkalbasi (2005) scholarly communication has changed rapidly over the last decade as scientists now publish their findings in e.g. open access journals (OAJ) and preprint servers. The development in open access (OA) is related to the retrieving of open access documents by search engines. One example is Google Scholar which was introduced in November 2004. These new possibilities may be interesting seen in combination with the known inherent bias of research evaluation using citation analysis. Zhao (2005) finds that the patterns of scholarly communication within a single discipline vary considerably on the Internet and in print journals. He questions the widespread use of the citation databases as the only source in research evaluation and argues that a two-tiered scholarly system is developing. Brody (2004) argues that even with the share of OA available today it is possible to build comprehensive citation databases not limited to proprietary databases only covering a portion of the total literature.

The key issue of this analysis is to investigate if we can supplement and thereby enhance traditional sources for research evaluation by exploiting some of the new possibilities that have emerged with the increased use of open access repositories and archives. A vast amount of information is available and using this we may be able to enhance the evaluation process. These repositories could potentially contain far more material than the citation indexes that only contain journal literature. This could be particularly interesting to investigate in relation to the humanities

and social sciences that traditionally publish a smaller share of their research output in journals. It is well known that especially the publications from the humanities but also to some extent the social sciences are not covered sufficiently to perform acceptable evaluations.

As the means of distributing research results have been expanding over the years to include much more than journal publications the need for analyzing these resources in an evaluation perspective has increased. The above leads to the following main research question: What are the potentials of open access based resources in research evaluation? The main research question can be elaborated in the following preliminary sub-questions:

1. Theoretical framework for OA and research evaluation
2. How do open access bases resources influence research evaluation?
3. Do OA based resources differ from the known services that can be used for research evaluation? Can we treat the citation databases that we normally use for research evaluation and OA based resources as a whole?
4. Do OA change the scholarly communication processes?

The project will contribute to laying the groundwork for measuring research performance in a changing scholarly communication. Furthermore, the project will contribute to the understanding of the changing scholarly communication processes. The project deals with changes in scholarly communication that imply consequences beyond research evaluation and the project is consequently not restricted to research evaluation.

3. Related research

Open access means that scientific publications are made freely available on the Internet, without any access restrictions. According to Moed (2007) the term open access is used with two different definitions. It is used to specify scientific publications published in a journal running under an open access model. However, it is also used to specify scientific publications that are freely available, not considering if they where originally published in a journal running under an OA

model or in a journal managed under other business models but characterised by being deposited in a freely accessible archive such as a personal homepage, institutional repository or subject-based archive (Moed, 2007: 2047).

The open access movement consists of various objectives but essential for the movement is the ambition of distributing peer-reviewed (journal) literature freely available to the public through the internet. The success of open access greatly depends on authors supporting the principle of open access by using an OA model for their publications. Several models for financing open access have been proposed, and Willinsky (2006) identifies ten different financing models. However, simplifying the complexity of financing models open access can be seen as consisting of two main strategies for achieving open access: open access journals and self-archiving.

The green self-archiving strategy comprises of persuading authors to self-archive the articles they publish in traditional toll-access journals in institutional open access archives. The many archives are to be searched collectively provided that they comply with a standard. The gold open access publishing strategy comprises of creating or converting traditional toll-access journals into open access journals. Furthermore, the strategy includes finding funding support for the publication costs and persuading authors to publish in open access journals (Harnad et al., 2004). More colours have been added to the scheme to describe the hybrids (Jenkins et al., 2007).

OAJs can be seen as the second phase or strategy in the process of achieving open access (Brody, Harnad & Carr, 2005). The open access publishing strategy comprises of creating or converting traditional toll-access journals into open access journals. Furthermore, the strategy includes finding funding support for the publication costs and persuading authors to publish in OAJs. A few years ago, about 4% of scholarly journal titles and 1-2% of articles were directly published as open access (Harnad et al. 2004). Both strategies depend greatly on authors being willing to support open access by making their own work available either through self-archiving or OA publishing.

The coverage of OA is not easily determined. The traditional citation indexes provide information on the indexing policy and the tools to examine it. OA based data sources may have an indexing policy but not necessarily and then it can only be determined through cumbersome empirical

investigations. OA based data sources are often based on authors self-archiving their work. This implies that the data source to a large extent is influenced by the incentives of individuals to make their work more or less visible by choosing to provide OA or not. A believed major benefit of OA for authors is the increased citation to OA publications. However, one should be careful making causal arguments as pointed out by e.g. Craig et al. (2007) stressing that we cannot conclude that OA publication causes more citations as the existing data does not illustrate causation, only association.

The increased number of citations could be caused by other factors than the increased visibility of the work. Other possibilities are self-selection bias caused by authors promoting their best work and early view-effect as online publication date for OA papers is often earlier than the print publication date. Moed (2007) estimates the effect of two factors, "early view" and "quality bias". The two sets of papers, OA and NOA, show no significant difference in citation rates when he controls for the effects of these factors. Davis & Fromerth (2006) find that the OA advantage is not a result of more citations to all the OA papers but by self-selection (authors select their best papers to self-archive). The existing analyses of the self-selection bias or quality bias have mainly focused on comparing citation data for archived and not archived articles.

As pointed out by Gläser (2003) the important issue is to what extent new forms of social order emerges due to the Internet. The use of Internet can be positively related to author productivity (Kaminer, 1998; Barjak, 2006), the Internet has facilitated large-scale collaborations (Finholt, 2002) and new communication regimes in biology based on online databases (Hilgartner, 1995). However, Gläser (2003) argues that the Internet rapidly creates new social phenomena but they are not necessarily sociologically new. The social structure of the scientific communities could remain unchanged although it reforms the work practices. Van Raan (2001: 63) argues that "Plus ça change, plus c'est la même chose". There are examples of the Internet not necessarily changing social phenomena. Lorigo and Pellacini (2007) have shown steady and constant growth in the frequency of long distance scholarly collaborations in a physics community and Mackenzie Owen (2007) finds that OAJs does not transform the research article by incorporating specific digital

properties. It is complex to identify the new forms of social order emerging due to the Internet and thus separating them from new social phenomena that are not sociologically new.

4. Methodological considerations

The methodological considerations are focused on quantitative investigations and reviewing the literature. The sub-questions will be analysed using different datasets. Initially, the characteristics of the open access will be analyzed as a way to explain the patterns discovered in open access repositories concerning e.g. selection bias and citation advantage. Furthermore, some definitions are also needed as the term open access covers varying phenomena. The starting point of analysing the coverage of both citation databases and open access based resources is two disciplines. The study focuses on intra-disciplinary differences within two disciplines (Economics and Psychology) using subject based databases (EconLit and PsychInfo). The point extends to include both the uneven coverage of specialties and research traditions. The data set will be investigated in order to see if we can establish differences in the degree of coverage and whether or not this plays a role when it comes to research evaluation (e.g. if it affects the rankings made on the basis of this material).

The analysis of whether or not open access based resources contribute with valuable information that cannot be discovered using the citation indexes is based on an interaction analysis of a number of journals (both OA and non-OA). Three disciplines are analyzed separately to preserve an overview and control the number of independent variables. This analysis seeks to discover the consequences of coverage on research evaluation. There seems to be lacking studies of differences in content when the type of function served is the same; e.g. publishing in open access (OA) journals and not open access (non-OA) journals. This specific study contributes to the understanding of the consequences of the changing scholarly communication for bibliometric studies as well as the discussion of the open access postulate.

The reasons for differences in citing behaviour are investigated by a bibliometric author analysis. The existing literature can provide numerous author analyses in relation to OA and several of

them indicate various author reservations. The existing author analyses are limited to experiments with heavily subsidised author charges, low response rates, limited samples. Furthermore, they are all based on either surveys or interviews. There seems to be lacking investigations on a larger scale looking more into the actual publishing behaviour of authors.

A bibliometric study of the citing behaviour in open access publications compared to traditional toll-access journal publications cast light on whether open access based resources differ from the known services that can be used for research evaluation.

Finally, the citation impact of a selection of working papers is measured to detect an open access advantage. Instead of focusing on comparing articles in traditional toll access journals (OA and non-OA) this study focuses on analyzing the development over time and is thus a supplement to the existing analyzes of a potential OA advantage.

5. Schedule of completion

Fall 2006: Reading. Teaching.

Spring 2007: Laying the theoretical framework for the project. Doing coursework and completing teaching obligations.

Fall 2007: Collecting data, completing coursework.

Spring 2008: Collecting data, analysing the data. Visiting student at Virtual Knowledge Studio, Amsterdam, the Netherlands.

Fall 2008: Analysing the data; writing.

Spring 2009: Completing the thesis by April 31.

6. Literature

Barjak, F. (2006). The role of the Internet in informal scholarly communication. Journal of the American Society for Information Science and Technology, 57(10), 1350-1367.

Bauer, K. & Bakkalbasi, N. (2005). An Examination of Citation Counts in a New Scholarly Communication Environment. D-Lib Magazine, 11(9).

Brody, T. (2004). Citation Analysis in the Open Access World. Interactive Media International. http://eprints.ecs.soton.ac.uk/10000/ [Downloaded 01-01-2008].

Brody, T., Harnad, S. and Carr, L. (2005) Earlier Web Usage Statistics as Predictors of Later Citation Impact. Journal of the American Association for Information Science and Technology (JASIST), 57(8), 1060-1072.

Craig, I.; Plume, A.; McVeigh, M.; Pringle, J. & Amin, M. (2007) Do Open Access Articles Have Greater Citation Impact? A critical review of the literature. Journal of Informetrics, 1(3), 239-248.

Davis, P.M. & Fromerth, M. J. (2007). Does the arXiv lead to higher citations and reduced publisher downloads for mathematics articles? Scientometrics, 71(2), 203-215.

Gläser, J. (2003). What Internet use does and does not change in scientific communities. Science Studies 16(1), 38-51.

Harnad, S.; Brody, T.; Vallieres, F.; Carr, L.; Hitchcock, S.; Gingras, Y.; Oppenheim, C.; Stamerjoanns, H. & Hilf, E.R. (2004) The green and the gold roads to Open Access. Nature (web focus). http://www.nature.com/nature/focus/accessdebate/21.html.

Hilgartner, S. (1995). Biomolecular databases: new communication regimes for biology? Science Communication, 17(2), 240-63.

Jenkins, C.; Probets, S. G.; Oppenheim, C. & Hubbard, B. (2007). 'ROMEO Studies 8: Self-Archiving: The Logic Behind the Colour Coding Used in the Copyright Knowledge Bank. Program, 41(2), 124-133.

Kaminer, N. and Braunstein, Y. M. (1998). Bibliometric Analysis of the Impact of Internet Use on Scholarly Productivity. Journal of the American Society for Information Science, 49(8), 720-730.

Lorigo, L. & Pellacini, F. (2007). Frequency and structure of long distance scholarly collaborations in a physics community. Journal of the American Society for Information Science and Technology, 58(10), p 1497-1502.

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), 2047-2054.

Van Raan, A.F.J. (2001). Bibliometrics and internet: Some observations and expectations. Scientometrics 50(1), 59-63.

Willinsky, J. (2006). The Access Principle – the case for open access to research and scholarship. The MIT Press 2006.

Zhao, D. (2005). Challenges of scholarly publications on the web to the evaluation of science: a comparison of author visibility on the web and in print journals. Information Processing and Management, 41(6), 1403-1418.